



**2023 Keystone Technical Report: Algebra I, Biology, and Literature**

Provided by Data Recognition Corporation

This document has been formatted to be ADA compliant.

# TABLE OF CONTENTS

---

- Preface: An Overview of the Assessments . . . . . 7**
  - The Keystone Exams from 2008 to Present . . . . . 7
  - Assessment Activities Occurring from 2010 to Present . . . . . 7
- Chapter One: Background of the Keystone Exams . . . . . 15**
  - Assessment History in Pennsylvania . . . . . 15
  - The Keystone Exams . . . . . 15
- Chapter Two: Test Development Overview of the Keystone Exams. . . . . 19**
  - Keystone Blueprint/Assessment Anchors and Eligible Content. . . . . 19
  - High-Level Test Design Considerations . . . . . 21
  - Online Testing Design Considerations . . . . . 21
  - Algebra I . . . . . 22
  - Biology . . . . . 24
  - Literature. . . . . 25
  - Literature Passages . . . . . 26
- Chapter Three: Item and Test Development Processes . . . . . 28**
  - General Keystone Test Development Processes . . . . . 28
  - General Test Definition . . . . . 28
  - Algebra I Test Definitions . . . . . 29
  - Biology Test Definitions . . . . . 30
  - Literature Test Definitions . . . . . 32
  - Item Development Considerations . . . . . 34
  - Item and Test Development Cycle . . . . . 36
  - General Item and Test Development Process . . . . . 39
- Chapter Four: Universal Design Procedures Applied to the Keystone Exams Test Development Process . . 43**
  - Universal Design . . . . . 43
  - Elements of Universally Designed Assessments . . . . . 43
  - Guidelines for Universally Designed Items . . . . . 44
  - Item Development. . . . . 45
  - Item Format . . . . . 46
  - Assessment Accommodations . . . . . 47
- Chapter Five: Embedded Field Test. . . . . 48**
  - Field Test Overview . . . . . 48
  - Classical Item Analysis . . . . . 49
  - Criteria for Identifying Items . . . . . 51
  - Review of Items with Data . . . . . 55

<b>Chapter Six: Operational Forms Construction for 2023 Administrations</b> .....	<b>56</b>
Final Selection of Items and Keystone Forms Construction .....	56
Special Forms Used with the Operational 2023 Keystone Exams .....	57
<b>Chapter Seven: Test Administration Procedures</b> .....	<b>59</b>
Sections, Sessions, Timing, and Layout of the Keystone Exams .....	59
Sections and Sessions .....	59
Timing .....	60
Layout .....	61
Shipping, Packaging, and Delivery of Materials .....	62
<b>Chapter Eight: Processing and Scoring</b> .....	<b>65</b>
Receipt of Materials .....	65
Scanning of Materials .....	66
Materials Storage .....	68
Online Testing .....	68
Scoring Multiple-Choice Items .....	70
Rangefinding .....	70
Rater Recruitment/Qualifications .....	70
Leadership Recruitment/Qualifications .....	71
Training .....	71
Handscoring Process .....	73
Handscoring Validity Process .....	73
Quality Control .....	74
<b>Chapter Nine: Description of Data Sources</b> .....	<b>79</b>
Student Filtering Criteria .....	79
Key Verification Data .....	80
Pre-Equating Verification .....	80
Final Data .....	80
Spiraling of Forms .....	80
<b>Chapter Ten: Summary Demographic and Accommodation Data for Spring 2023 Keystone Exams</b> .....	<b>83</b>
Assessed Students .....	83
Reasons for Student Non-Assessment .....	84
Demographic Characteristics of Students Receiving Test Scores .....	86
Test Accommodations Provided .....	95
Glossary of Accommodation Terms .....	110
<b>Chapter Eleven: Classical Item Statistics</b> .....	<b>113</b>
Item-Level Statistics .....	113
Item Difficulty .....	113
Item Discrimination .....	114

Scatter Plots of Item Discrimination and Difficulty . . . . .	115
Observations and Interpretations . . . . .	120
<b>Chapter Twelve: Rasch Item Calibration. . . . .</b>	<b>121</b>
Description of the Rasch Model . . . . .	121
Checking Rasch Assumptions . . . . .	122
Rasch Item Statistics . . . . .	126
<b>Chapter Thirteen: Standard Setting. . . . .</b>	<b>133</b>
Standard Setting and Performance Level Descriptors . . . . .	133
Development Overview for the Performance Level Descriptors. . . . .	133
Performance Level Descriptors Meeting 1 . . . . .	134
Performance Level Descriptors Meeting 2 . . . . .	136
Standard Setting . . . . .	138
<b>Chapter Fourteen: Scaling . . . . .</b>	<b>151</b>
Raw Scores to Rasch Ability Estimates . . . . .	151
Rasch Ability Estimates to Scaled Scores . . . . .	152
Raw-to-Scaled-Score Tables . . . . .	153
<b>Chapter Fifteen: Equating. . . . .</b>	<b>154</b>
Pre- vs. Post-Equating . . . . .	154
Equating Design for Keystone Exams . . . . .	155
Scale Linking . . . . .	155
Pre-Equating Verification . . . . .	156
Performance Level Classification . . . . .	159
Scale Stability and Maintenance . . . . .	160
Test Characteristic Curves and Logit Plots . . . . .	160
<b>Chapter Sixteen: Scores and Score Reports . . . . .</b>	<b>168</b>
Scoring . . . . .	168
Description of Total-Test Scores. . . . .	168
Highest Total Test Scaled Score to Date . . . . .	169
Description of Module Scores . . . . .	170
Appropriate Score Use . . . . .	171
Cautions for Score Use. . . . .	171
Report Development. . . . .	172
Reports . . . . .	172
<b>Chapter Seventeen: Operational Test Statistics. . . . .</b>	<b>177</b>
Performance Level Statistics . . . . .	177
Scaled Scores. . . . .	178
Raw Scores. . . . .	178

<b>Chapter Eighteen: Reliability</b> .....	<b>188</b>
Reliability Indices .....	189
Coefficient Alpha .....	189
Further Interpretations .....	190
Standard Error of Measurement .....	193
Results and Observations .....	194
Rasch Conditional Standard Errors of Measurement .....	195
Results and Observations .....	196
Reliability of Performance Level Classification Decisions .....	201
Rater Agreement .....	202
<b>Chapter Nineteen: Validity</b> .....	<b>206</b>
Purposes and Intended Uses of the Keystone Exams .....	206
Evidence Based on Test Content .....	206
Evidence Based on Response Process .....	208
Evidence Based on Internal Structure .....	208
Evidence Based on Relationships with Other Variables .....	212
Evidence Based on Consequences of Tests .....	213
Evidence Related to the Use of Rasch Model .....	214
Validity Evidence Summary .....	214
<b>Chapter Twenty: Special Study on Test Administration Mode</b> .....	<b>215</b>
Mode DIF Summary of Operational Items .....	216
Raw-To-Scaled-Score Comparison .....	216
Mode DIF Summary of Field-Test Items .....	222
<b>Appendix A: Understanding Depth of Knowledge and Cognitive Complexity</b> .....	<b>223</b>
Algebra I Depth of Knowledge .....	224
Biology Depth of Knowledge .....	226
Literature Depth of Knowledge .....	229
<b>Appendix B: General Scoring Guidelines</b> .....	<b>231</b>
Algebra I .....	231
Biology .....	232
Literature .....	233
<b>Appendix C: Item and Test Development Process for the Keystone Exams</b> .....	<b>234</b>
<b>Appendix D: Item and data Review Card examples</b> .....	<b>237</b>
Item Review Card Example .....	237
Data Review Card Example .....	238
<b>Appendix E: Item Rating Sheet</b> .....	<b>240</b>
<b>Appendix F: Tally Sheets</b> .....	<b>244</b>

<b>Appendix G: Keystone Exams Module Layout Plans</b> . . . . .	<b>262</b>
<b>Appendix H: Mean Raw Scores by Form</b> . . . . .	<b>263</b>
Algebra I: Spring . . . . .	264
Biology: Spring . . . . .	265
Literature: Spring . . . . .	266
<b>Appendix I: Demographic and Accommodation Data</b> . . . . .	<b>267</b>
Winter . . . . .	267
Summer . . . . .	289
<b>Appendix J: Item Statistics</b> . . . . .	<b>313</b>
Multiple-Choice Items . . . . .	314
Constructed-Response Items . . . . .	350
<b>Appendix K: Raw-To-Scale Score Conversion Tables</b> . . . . .	<b>357</b>
Winter . . . . .	358
Spring . . . . .	364
Summer . . . . .	370
<b>Appendix L: Pre-Equating Verification Results</b> . . . . .	<b>376</b>
Winter . . . . .	376
Spring . . . . .	392
Summer . . . . .	408
<b>Appendix M: Reliabilities</b> . . . . .	<b>424</b>
<b>References</b> . . . . .	<b>434</b>

# **PREFACE: AN OVERVIEW OF THE ASSESSMENTS**

---

## **THE KEYSTONE EXAMS FROM 2008 TO PRESENT**

### **COMPREHENSIVE GRADUATION COMPETENCY ASSESSMENT PROGRAM**

In 2008, the Commonwealth of Pennsylvania initiated a comprehensive graduation competency assessment program. The goals of this program include the following:

- To provide a system that is aligned, focused, standards-based, accurate, universally applicable, and publicly accessible
- To develop, produce, distribute, administer (both online and in paper-and-pencil), collect, score, analyze, track, and report results of graduation competency assessments for ten high school-level content areas: Algebra I, Algebra II, Biology, Chemistry, Civics and Government, English Composition, Geometry, Literature, U.S. History, and World History, with each area or course comprised of modules containing unique content
- To provide graduation competency testing opportunities for students three times each school year—spring, summer, and fall—with students permitted to retake modules until proficiency is achieved on each module
- To report graduation competency results under accelerated timelines
- To ensure validity and reliability of the assessment systems through technically sound test development and psychometric practices, detailed statistical analyses and research studies, and well-documented processes and quality procedures

The Keystone Exams, as the graduation competency assessments are named, are just one component of Pennsylvania's system of high school graduation requirements. Keystone Exams are designed to help school districts guide students toward meeting state standards—standards aligned with expectations for success in college and the workplace. In order to receive a diploma, students are also required to meet local district credit and attendance requirements and to complete a culminating project, along with any additional district requirements.

For graduating classes, students are to demonstrate successful completion of secondary-level course work in Algebra I, Biology, and Literature, in which the Keystone Exam served as the final course exam. Students' Keystone Exam scores count for at least one-third of the final course grades.

Based upon Chapter 4 regulations, each Keystone Exam is designed in modules that reflect distinct, related academic content common to the traditional progression of course work. Students who do not score Proficient or above on a Keystone Exam module may choose to complete a project-based assessment for that module based upon other specific requirements.

## **ASSESSMENT ACTIVITIES OCCURRING FROM 2010 TO PRESENT**

The first assessment activities took place in the 2010–2011 school year. Prior to November 2010, there were no Keystone Exams assessment events. The table below outlines the field tests and operational exams administered during the 2010–11 school year.

Following the development of Assessment Anchors and Eligible Content, exams were developed for the initial field test in 2010 and were subsequently administered as operational exams in 2011. Additional exams, which were based on the Assessment Anchors and Eligible Content developed in 2009 and 2010, were developed for the initial field test in 2011. Detailed information about the operational exam activities that occurred during the 2010–2011 school year is in the *Keystone Exams Spring 2011 Algebra I, Biology, and Literature Technical Report*.

## Field Test and Operational Exams during the 2015–16 School Year

Exam	Assessment Activity	Date
Algebra I	Initial Stand-Alone Field Test	Fall 2010 (November)
Algebra I	Inaugural Operational Exam Administration	Spring 2011 (May)
Algebra II	Initial Stand-Alone Field Test	Spring 2011 (May)
Biology	Initial Stand-Alone Field Test	Fall 2010 (November)
Biology	Inaugural Operational Exam Administration	Spring 2011 (May)
English Composition	Initial Stand-Alone Field Test	Spring 2011 (May)
Geometry	Initial Stand-Alone Field Test	Spring 2011 (May)
Literature	Initial Stand-Alone Field Test	Fall 2010 (November)
Literature	Inaugural Operational Exam Administration	Spring 2011 (May)

Following a one-year program hiatus in 2012, the field-test items embedded in the Spring 2011 operational forms were used to construct the forms for the next four administrations (spring, summer, winter, and possible breach). The table below outlines exams administered during the 2012–13 school year. Detailed information about the operational exam activities that occurred during the 2012–2013 school year is in the *Keystone Exams Spring 2013 Algebra I, Biology, and Literature Technical Report*.

## Operational Exams during the 2012–13 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2012/2013 (December–January)
Algebra I	Operational Exam Administration	Spring 2013 (May)
Algebra I	Operational Retest Exam Administration	Summer 2013 (August)
Biology	Operational Retest Exam Administration	Winter 2012/2013 (December–January)
Biology	Operational Exam Administration	Spring 2013 (May)
Biology	Operational Retest Exam Administration	Summer 2013 (August)
Literature	Operational Retest Exam Administration	Winter 2012/2013 (December–January)
Literature	Operational Exam Administration	Spring 2013 (May)
Literature	Operational Retest Exam Administration	Summer 2013 (August)

Some of the field-test items embedded in the Spring 2013 operational forms were used to construct the forms for the next four administrations (spring, summer, winter, and possible breach). The core items on the 2013–2014 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2013–14 school year. Detailed information about the operational exam activities that occurred during the 2013–2014 school year is in the *Keystone Exams Spring 2014 Algebra I, Biology, and Literature Technical Report*.



### Operational Exams during the 2013–14 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2013/2014 (December–January)
Algebra I	Operational Exam Administration	Spring 2014 (May)
Algebra I	Operational Retest Exam Administration	Summer 2014 (August)
Biology	Operational Retest Exam Administration	Winter 2013/2014 (December–January)
Biology	Operational Exam Administration	Spring 2014 (May)
Biology	Operational Retest Exam Administration	Summer 2014 (August)
Literature	Operational Retest Exam Administration	Winter 2013/2014 (December–January)
Literature	Operational Exam Administration	Spring 2014 (May)
Literature	Operational Retest Exam Administration	Summer 2014 (August)

Some of the field-test items embedded in the Spring 2014 operational forms were used to construct the forms for the next year’s administrations (spring, summer, winter). The core items on the 2014–2015 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2014–15 school year.

### Operational Exams during the 2014–15 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2014/2015 (December–January)
Algebra I	Operational Exam Administration	Spring 2015 (May)
Algebra I	Operational Retest Exam Administration	Summer 2015 (August)
Biology	Operational Retest Exam Administration	Winter 2014/2015 (December–January)
Biology	Operational Exam Administration	Spring 2015 (May)
Biology	Operational Retest Exam Administration	Summer 2015 (August)
Literature	Operational Retest Exam Administration	Winter 2014/2015 (December–January)
Literature	Operational Exam Administration	Spring 2015 (May)
Literature	Operational Retest Exam Administration	Summer 2015 (August)

Some of the field-test items embedded in the Spring 2015 operational forms were used to construct the forms for the next year’s administrations (spring, summer, winter). The core items on the 2015–2016 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2015–16 school year.

### Operational Exams during the 2015–16 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2015/2016 (December–January)
Algebra I	Operational Exam Administration	Spring 2016 (May)
Algebra I	Operational Retest Exam Administration	Summer 2016 (August)
Biology	Operational Retest Exam Administration	Winter 2015/2016 (December–January)
Biology	Operational Exam Administration	Spring 2016 (May)
Biology	Operational Retest Exam Administration	Summer 2016 (August)
Literature	Operational Retest Exam Administration	Winter 2015/2016 (December–January)
Literature	Operational Exam Administration	Spring 2016 (May)
Literature	Operational Retest Exam Administration	Summer 2016 (August)

Some of the field-test items embedded in the Spring 2016 operational forms were used to construct the forms for the next year’s administrations (spring, summer, winter). The core items on the 2016–2017 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2016-17 school year.

### Operational Exams during the 2016–17 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2016/2017 (December–January)
Algebra I	Operational Exam Administration	Spring 2017 (May)
Algebra I	Operational Retest Exam Administration	Summer 2017 (July–August)
Biology	Operational Retest Exam Administration	Winter 2016/2017 (December–January)
Biology	Operational Exam Administration	Spring 2017 (May)
Biology	Operational Retest Exam Administration	Summer 2017 (July–August)
Literature	Operational Retest Exam Administration	Winter 2016/2017 (December–January)
Literature	Operational Exam Administration	Spring 2017 (May)
Literature	Operational Retest Exam Administration	Summer 2017 (July–August)

Some of the field-test items embedded in the spring 2017 operational forms were used to construct the forms for the next year’s administrations (spring, summer, winter). The core items on the 2017–2018 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2017–2018 school year.

### Operational Exams during the 2017–18 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2017/2018 (December–January)
Algebra I	Operational Exam Administration	Spring 2018 (May)
Algebra I	Operational Retest Exam Administration	Summer 2018 (July–August)
Biology	Operational Retest Exam Administration	Winter 2017/2018 (December–January)
Biology	Operational Exam Administration	Spring 2018 (May)
Biology	Operational Retest Exam Administration	Summer 2018 (July–August)
Literature	Operational Retest Exam Administration	Winter 2017/2018 (December–January)
Literature	Operational Exam Administration	Spring 2018 (May)
Literature	Operational Retest Exam Administration	Summer 2018 (July–August)

Some of the field-test items embedded in the spring 2018 operational forms were used to construct the forms for the next year’s administrations (spring, summer, winter). The core items on the 2018–2019 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2018–2019 school year.

### Operational Exams during the 2018–19 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2018/2019 (December–January)
Algebra I	Operational Exam Administration	Spring 2019 (May)
Algebra I	Operational Retest Exam Administration	Summer 2019 (July–August)
Biology	Operational Retest Exam Administration	Winter 2018/2019 (December–January)
Biology	Operational Exam Administration	Spring 2019 (May)
Biology	Operational Retest Exam Administration	Summer 2019 (July–August)
Literature	Operational Retest Exam Administration	Winter 2018/2019 (December–January)
Literature	Operational Exam Administration	Spring 2019 (May)
Literature	Operational Retest Exam Administration	Summer 2019 (July–August)

Some of the field-test items embedded in the spring 2019 operational forms were used to construct the forms for the next year’s administrations (summer, winter). The core items on the 2019–2020 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2019–2020 school year.

The spring administration of the Keystone exams was cancelled due to Coronavirus (COVID-19) mitigation efforts. On March 27, 2020, the U.S. Department of Education (USDE) approved Pennsylvania’s request to waive federal assessment requirements for the 2019–20 school year, along with accountability and certain reporting requirements based on data derived from the 2019–20 school year.

The waiver’s coverage of assessment requirements applies to the cohort of 2019–2020 test takers who were scheduled to take one or more Keystone Exams in the spring of 2020 (Spring 2019–2020 Cohort). Accordingly, the Federal government is not requiring any student enrolled in a Keystone Exam trigger course (Algebra I, Biology, English Literature) during the spring of the 2019–2020 school year, regardless of their current grade level or expected graduation date, to take the associated Keystone Exam(s) once schools reopen and federal assessment requirements resume. Because Keystone Exams scores are “banked” for accountability purposes and applied in grade 11, this waiver will affect students enrolled in grades other than grades 11 and 12, including as low as grade 6.

Students who took the Keystone Exams during the 2019–2020 school year prior to the pandemic (i.e., summer 2019 or winter 2020) may use those results to satisfy their individual Act 158 requirements and for any local purposes. However, because 2019–2020 school year cohort results will not be comparable with prior or future years results, these results will not be factored into future accountability determinations such as cyclical Comprehensive Support and Improvement (CSI) and Additional Targeted Support and Improvement determinations (A-TSI) (scores of students who took the Keystone Exams in the summer of 2019 or the winter of 2019–2020 as a re-tester will be attributed for the purposes of federal accountability and reporting.)

For this reason, these partial results were not reported through Future Ready in fall 2020.

### Operational Exams during the 2019–20 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2019/2020 (December–January)
Algebra I	Operational Exam Administration	Cancelled due to Coronavirus (COVID-19) mitigation efforts.
Algebra I	Operational Retest Exam Administration	Autumn 2020 (September–October)  *previously a summer administration
Biology	Operational Retest Exam Administration	Winter 2019/2020 (December–January)
Biology	Operational Exam Administration	Cancelled due to Coronavirus (COVID-19) mitigation efforts.
Biology	Operational Retest Exam Administration	Autumn 2020 (September–October)  *previously a summer administration
Literature	Operational Retest Exam Administration	Winter 2019/2020 (December–January)
Literature	Operational Exam Administration	Cancelled due to Coronavirus (COVID-19) mitigation efforts.
Literature	Operational Retest Exam Administration	Autumn 2020 (September–October)  *previously a summer administration

The core items on the 2019–2020 forms also consisted of items that appeared on the core forms of the administrations two years prior. More information on these core-to-core overlap items can be found in Chapter Three. The table below outlines exams administered during the 2020–2021 school year.

### Operational Exams during the 2020–21 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2020/2021 (December–January)
Algebra I	Operational Exam Administration	Spring 2021 (May–September)
Biology	Operational Retest Exam Administration	Winter 2020/2021 (December–January)
Biology	Operational Exam Administration	Spring 2021 (May–September)
Literature	Operational Retest Exam Administration	Winter 2020/2021 (December–January)
Literature	Operational Exam Administration	Spring 2021 (May–September)

The core items on the 2020–2021 forms also consisted of items that appeared on the core forms of prior administrations. More information on these core-to-core overlap items can be found in Chapter Three.

### Operational Exams during the 2021–22 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2021/2022 (December–January)
Algebra I	Operational Exam Administration	Spring 2022 (May)
Algebra I	Operational Retest Exam Administration	Summer 2022 (July–August)
Biology	Operational Retest Exam Administration	Winter 2021/2022 (December–January)
Biology	Operational Exam Administration	Spring 2022 (May)
Biology	Operational Retest Exam Administration	Summer 2022 (July–August)
Literature	Operational Retest Exam Administration	Winter 2021/2022 (December–January)
Literature	Operational Exam Administration	Spring 2022 (May)
Literature	Operational Retest Exam Administration	Summer 2022 (July–August)

The core items on the 2021–2022 forms also consisted of items that appeared on the core forms of prior administrations. More information on these core-to-core overlap items can be found in Chapter Three.

## Operational Exams during the 2022–23 School Year

Exam	Assessment Activity	Date
Algebra I	Operational Retest Exam Administration	Winter 2022/2023 (December–January)
Algebra I	Operational Exam Administration	Spring 2023 (May)
Algebra I	Operational Retest Exam Administration	Summer 2023 (July)
Biology	Operational Retest Exam Administration	Winter 2022/2023 (December–January)
Biology	Operational Exam Administration	Spring 2023 (May)
Biology	Operational Retest Exam Administration	Summer 2023 (July)
Literature	Operational Retest Exam Administration	Winter 2022/2023 (December–January)
Literature	Operational Exam Administration	Spring 2023 (May)
Literature	Operational Retest Exam Administration	Summer 2023 (July)

The core items on the 2022–2023 forms also consisted of items that appeared on the core forms of prior administrations. More information on these core-to-core overlap items can be found in Chapter Three.

# CHAPTER ONE: BACKGROUND OF THE KEYSTONE EXAMS

---

This brief overview of the Pennsylvania Keystone Exams summarizes the history of the program’s development process, intent and purpose, and recent changes.

## ASSESSMENT HISTORY IN PENNSYLVANIA

Pennsylvania’s involvement in statewide assessment actually began in the 1969–1970 school year with a purely school-based assessment known as *Educational Quality Assessment (EQA)*, which continued through the 1987–1988 school year. A state-mandated student competency testing program called *Testing for Essential Learning and Literacy Skills (TELLS)* also operated from the school years of 1984–1985 through 1990–1991. Also in 1990, the state initiated an on-demand writing assessment.

The Pennsylvania System of School Assessment (PSSA) program was instituted in 1992 as a school evaluation model with reporting at the school level only. The PSSA initially measured performance in the content areas of mathematics and reading at grades 5, 8, and 11, and in writing at grades 6 and 9. Starting in 1994, as part of Chapter 5 regulations, the PSSA added student-level reports. In 1999, as part of Chapter 4 regulations, the State Board of Education adopted the Pennsylvania Academic Standards for mathematics and for reading, writing, speaking, and listening. Proficiency levels for Advanced, Proficient, Basic, and Below Basic were defined in 2000. In 2001 and 2004, the reading and mathematics assessments underwent various content enhancements to improve alignment to the 1999 Academic Standards. Grade 11 was added to the writing assessment in 2001. Then, in 2004–2005, the PSSA Assessment Anchors and Eligible Content were developed to clarify content structure and improve articulation between assessment and instruction. In addition, in 2005, the grade 6 and 9 writing assessments were moved to grades 5 and 8. By 2006, the operational mathematics and reading assessments incorporated grades 3 through 8 and 11. In 2007, the PSSA and the PSSA Assessment Anchors and Eligible Content underwent additional content enhancements. In 2008, science was added to the PSSA as an operational assessment. Starting with the 2013 field test, PSSA began a multiyear transition to a new set of standards called the Pennsylvania Core Standards. Detailed information about the operational exam activities that occurred during the 2013–2014 school year is in the *2014 PSSA Technical Report*.

## THE KEYSTONE EXAMS

In 2008, the Commonwealth of Pennsylvania initiated a comprehensive graduation competency assessment program. As a key piece of this program, the Keystone Exams are designed to assess proficiency in various subject areas, including Algebra I, Algebra II, Biology, Chemistry, Civics and Government, English Composition, Geometry, Literature, U.S. History, and World History. The Keystone Exams are just one component of Pennsylvania’s high school graduation requirements. Students must also earn state-specified credits, fulfill the state’s service learning and attendance requirements, and complete any additional local school system requirements to receive a Pennsylvania high school diploma.

The stated goals of the Keystone program are to

- provide for a system that is aligned, focused, standards-based, accurate, universally applicable, and publicly accessible.
- develop, produce, distribute, administer (both online and in paper-and-pencil), collect, score, analyze, track, and report results of graduation competency assessments for ten high school-level content areas: Algebra I, Algebra II, Biology, Chemistry, Civics and Government, English Composition, Geometry, Literature, U.S. History, and World History, with each area or course composed of modules containing unique content.
- provide graduation competency testing opportunities for students three times each school year—spring, summer, and fall—with students permitted to retake modules until proficiency is achieved in each module.
- report graduation competency results under accelerated timelines.
- ensure validity and reliability of the assessment systems through technically sound test development and psychometric practices, detailed statistical analyses and research studies, and well-documented processes and quality procedures.

## **GRADUATION REQUIREMENTS AND THE KEYSTONE EXAMS**

Based upon Chapter 4 regulations, each Keystone Exam is designed in modules that reflect distinct, related academic content common to the traditional progression of coursework. Students who do not score Proficient or above on a Keystone Exam module may choose to complete a project-based assessment for that module based on the requirements detailed below.

If a student is unable to meet the requirements in § 4.24(b)(1)(iv)(A) (relating to high school graduation requirements) after two attempts on a Keystone Exam, the student may supplement a Keystone Exam score with satisfactory completion of a project-based assessment. Points earned through satisfactory performance on one or more project modules related to the Keystone Exam module or modules that the student did not pass shall be added to the student's highest Keystone Exam score.

A student may qualify to participate in one or more project-based assessments if the student has met all of the following conditions:

1. The student has taken the course.
2. The student was unsuccessful in achieving a score of Proficient on the Keystone Exam after at least two attempts.
3. The student has met the district's attendance requirements for the course.
4. The student has participated in a satisfactory manner in supplemental instructional services under § 4.24(i).

## **KEYSTONE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT**

In 2009, the state initiated development of test designs and test blueprints for the Keystone Exams based on Pennsylvania Keystone Course Standards. Committees of Pennsylvania educators met in 2009, 2010, and 2011 to write, review, and approve Assessment Anchors and Eligible Content statements and sample exam items. To provide initial focus, each test blueprint committee was presented with materials specific to the exam in question, including a basic blueprint structure, the Pennsylvania State Standards, and draft Eligible Content statements based on the standards. The results from the initial committee work were evaluated by national, state, and local subject experts, and following revisions, they were ultimately validated by another committee of Pennsylvania educators. Following committee approval, the Keystone Assessment Anchors and Eligible Content statements for literacy, mathematics, and science were approved by the State Board of Education in September 2010.

- Mathematics
  - The first committee meetings took place in April 2009, where initial drafts of the test blueprints were developed for Algebra I, Algebra II, and Geometry.
  - A follow-up committee meeting for the three mathematics exams was held in August 2009.
- Literacy
  - The first committee meetings took place in April 2009, where initial drafts of the test blueprints were developed for English Composition and Literature.
  - A follow-up committee meeting for the two literacy exams was held in November 2009.
- Science
  - The first committee meetings took place in October 2009, where the initial draft of the test blueprint was developed for Biology.
  - A follow-up committee meeting for Biology was held in January 2010.
  - In addition, in January 2010, the initial draft of the test blueprint was developed for Chemistry.
  - Chemistry was part of a follow-up committee meeting held in late January 2010.



- Social Studies
  - The first committee meetings took place in November 2010, where initial drafts of the test blueprints were developed for Civics and Government, U.S. History, and World History.
  - A follow-up committee meeting for the Civics and Government exam was held in October 2011.
  - A follow-up committee meeting for U.S. History and World History remains unscheduled pending further decisions about the future of these Keystone exams.

## WAVE IMPLEMENTATION OF THE EXAMS

The implementation plan for the Keystone Exams envisioned the ten Keystone Exams becoming operational through a series of waves. The initial wave included Algebra I, Biology, and Literature. These first three exams were field-tested in fall 2010 and reached operational status with the spring 2011 administration. The second wave included Algebra II, English Composition, and Geometry; these were field-tested in spring 2011. English Composition is scheduled to reach operational status at a future date. Civics and Government is projected to reach operational status following English Composition. The implementation of the five remaining courses, Algebra II, Geometry, Chemistry, U.S. History, and World History, is currently unscheduled. The Pennsylvania Department of Education continues to evaluate the implementation schedule. Table 1–1 reflects the implementation plans as of September 2021.

**Table 1–1. Keystone Exams Wave Implementation Plan**

Wave	Exam(s)	Initial Field Test	First Operational
1	Algebra I, Biology, Literature	Fall 2010	Spring 2011
2	English Composition	Spring 2011	TBD
2	Algebra II, Geometry	Spring 2011	Not Scheduled
3	Civics and Government	TBD	Not Scheduled
4	Chemistry, U.S. History, World History	TBD	Not Scheduled

## MODE OF DELIVERY FOR THE EXAMS

One key feature of the Keystone Exams is the dual mode of delivery of the testing materials that is available to districts. In addition to the traditional paper-and-pencil format, the Keystone Exams are also available in a computer-based online format using test-delivery software.

While exam materials are still available in the traditional format (two pieces of exam materials—a test book and a separate answer book [or, in the case of English Composition, a single test/answer book]), districts are given the option to administer the exams using computer-based online testing software instead of the paper-and-pencil format.

For more information about how the online exams were developed in concert with the traditional paper-and-pencil format, see Chapter Three.

## MULTIPLE TESTING OPPORTUNITIES

Another key feature of the Keystone Exams is the multiple testing opportunities provided to students. Main administrations in both spring and winter provide options for students completing course work at various times of the year and accommodate both traditional and block scheduling. In addition, a summer retest opportunity is also available. More information about the spring, winter, and summer administrations can be found in Chapter Seven.

## **PERFORMANCE LEVELS FOR THE KEYSTONE EXAMS**

The State Board approved a set of criteria defining Advanced, Proficient, Basic, and Below Basic levels of performance for the Keystone Exams. More information about these Performance Level Descriptors (PLDs) is found in Chapter Thirteen.

## **OPERATIONAL TEST DESIGN INFORMATION**

The test definition of each of the operational Keystone Exams, including information about exam-specific test designs, test blueprints, test layouts, item types, and other exam elements, is detailed in Chapter Three.

# CHAPTER TWO: TEST DEVELOPMENT OVERVIEW OF THE KEYSTONE EXAMS

---

## KEYSTONE BLUEPRINT/ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

The Keystone Test Blueprints—known as the Keystone Exams Assessment Anchors and Eligible Content—are based on Pennsylvania Keystone Course Standards and the Pennsylvania Core Standards. Prior to the development of the Assessment Anchors, multiple groups of Pennsylvania educators convened to create a set of standards for each of the Keystone Exams. Derived from a review of existing standards, these Enhanced Standards (Course Standards) focus on what students need to know and be able to do in order to be ready for college and career.

Although the Keystone Course Standards indicate what students should know and be able to do, Assessment Anchors are designed to indicate the parts of the Keystone Course Standards (Instructional Standards) that will be assessed on the Keystone Exams. Based on recommendations from Pennsylvania educators, the Assessment Anchors were designed as a tool to improve the articulation of curricular, instructional, and assessment practices. The Assessment Anchors clarify what is expected and focus the content of the standards into what is assessable on a large-scale exam. The Assessment Anchor documents also serve to communicate Eligible Content—the range of knowledge and skills from which the Keystone Exams are designed.

The Keystone Exams Assessment Anchors and Eligible Content have been designed to hold together, or anchor, the state assessment system and curricular/instructional practices in schools by following these design parameters:

- **Clear:** The Assessment Anchors are easy to read and user friendly; they clearly detail which standards are assessed on the Keystone Exams.
- **Focused:** The Assessment Anchors identify a core set of standards that can be reasonably assessed on a large-scale assessment; this keeps educators from having to guess which standards are critical.
- **Rigorous:** The Assessment Anchors support the rigor of the state standards by assessing higher order and reasoning skills.
- **Manageable:** The Assessment Anchors define the standards in a way that can be easily incorporated into a course to prepare students for success.

The Assessment Anchors and Eligible Content are organized into cohesive blueprints, each structured with a common labeling system. This framework is organized by increasing levels of detail: first, Module (Reporting Category); second, Assessment Anchor; third, Anchor Descriptor; and fourth, Eligible Content statement. The common format of this outline is followed across the Keystone Exams.

A description of each level in the labeling system for the Keystone Exams is as follows:

- **Module:** The Assessment Anchors are organized into two thematic modules for each of the Keystone Exams, and these modules serve as the Reporting Categories for the Keystone Exams. The module title appears at the top of each page in the Assessment Anchor document. The module level is also important because the Keystone Exams are built using a module format, with each of the Keystone Exams divided into two equally sized test modules. Each module is made up of two or more Assessment Anchors.
- **Assessment Anchor:** The Assessment Anchor appears in the shaded bar across the top of each Assessment Anchor table in the Assessment Anchor document. The Assessment Anchors represent categories of subject matter that anchor the content of the Keystone Exams. Each Assessment Anchor is part of a module and has one or more Anchor Descriptors unified under it.
- **Anchor Descriptor:** Below each Assessment Anchor in the Assessment Anchor document is a specific Anchor Descriptor. The Anchor Descriptor level details the scope of content covered by the Assessment Anchor. Each Anchor Descriptor is part of an Assessment Anchor and has one or more Eligible Content unified under it.

- **Eligible Content:** The column to the right of the Anchor Descriptor in the Assessment Anchor document contains the Eligible Content statements. The Eligible Content is the most specific description of the content that is assessed in the Keystone Exams. This level is considered the assessment limit. It helps educators identify the range of content covered on the Keystone Exams.
- **Enhanced Standard:** In the column to the right of each Eligible Content statement is a code representing one or more Enhanced Standards that correlate to the Eligible Content statement. Some Eligible Content statements include annotations that clarify the scope of an Eligible Content.
- **Notes:** There are three types of notes included in the Assessment Anchor document.
  - “e.g.” (“for example”)—sample approach, but not a limit to the Eligible Content
  - “i.e.” (“that is”)—specific limit to the Eligible Content
  - “Note”—content exclusions or definable range of the Eligible Content

The Assessment Anchor’s coding is read like an outline. The coding includes the Subject (Exam), Reporting Category/Module, Assessment Anchor, Anchor Descriptor, and Eligible Content. Each exam has two modules. Each module has two or more Assessment Anchors. Each of the Assessment Anchors has one or more Anchor Descriptors, and each Anchor Descriptor has at least one Eligible Content (generally more than one). The Assessment Anchors form the basis of the test design for the exams undergoing test development. In turn, this hierarchy is the basis for organizing the total module and exam scores (based on the core [common] portions).

**Table 2–1. Sample Keystone Assessment Anchor Coding**

Sample Code	Subject (Exam)	Reporting Category (Module)	Assessment Anchor (AA)	Anchor Descriptor (AD)	Eligible Content (EC)
A1.1.1.2.1	<b>A1</b> Algebra I	<b>1</b> Operations and Linear Equations & Inequalities	<b>1</b> Linear Equations	<b>2</b> Write, solve, and/or graph linear equations using various methods.	<b>1</b> Write, solve, and/or apply a linear equation (including problem situations).
BIO.A.2.1.1	<b>BIO</b> Biology	<b>A</b> Cells and Cell Processes	<b>2</b> The Chemical Basis for Life	<b>1</b> Describe how the unique properties of water support life on Earth	<b>1</b> Describe the unique properties of water and how these properties support life on Earth (e.g., freezing point, high specific heat, cohesion).
L.F.2.4.1	<b>L</b> Literature	<b>F</b> Fiction	<b>2</b> Analyzing and Interpreting Literature—Fiction	<b>4</b> Use appropriate strategies to interpret and analyze the universal significance of literary fiction.	<b>1</b> Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.

The complete set of Assessment Anchors and Eligible Content aligned to the Pennsylvania Academic Standards can be referenced at PDE’s website: [www.education.pa.gov](http://www.education.pa.gov).

## HIGH-LEVEL TEST DESIGN CONSIDERATIONS

The Keystone Exams employ two types of test items (questions): multiple-choice and constructed-response. These item types assess different levels of knowledge and provide different information about achievement. Psychometrically, multiple-choice items are very useful and efficient tools for collecting information about a student's academic achievement. Constructed-response performance tasks generally generate fewer scorable points than multiple-choice items generate in the same amount of testing time; however, they provide tasks that are more realistic and sample eligible content that best lends itself to this item type. Furthermore, well-constructed scoring guides have made it possible to include constructed-response tasks in large-scale assessments, and trained scorers apply the scoring guides to efficiently score large numbers of student responses in a highly reliable way. The design of the Keystone Exams attempts to achieve a reasonable balance between the two item types.

**Table 2–2. Keystone Exams High-Level Design Considerations**

Exam	MC as Percentage of Core Points	CR as Percentage of Core Points	Number of Points per MC	Number of Points per CR	Number of Modules	Number of Assessment Anchors	Number of Eligible Content
Algebra I	60	40	1	4	2	6	33
Biology	73	27	1	3	2	8	38
Literature	65	35	1	3	2	4	56

### DEPTH OF KNOWLEDGE

The goal of each Keystone Exam is for each item to be of sufficient rigor, or Webb's Depth of Knowledge (DOK) Level 3. Webb's DOK was created by Norman Webb of the Wisconsin Center for Education Research. Webb's definition of depth of knowledge is the degree or complexity of knowledge that the content curriculum standards and expectations require. Therefore, when reviewing items for depth of knowledge, the item is reviewed to determine whether it is as demanding cognitively as what the actual content curriculum standard expects. In the case of the Pennsylvania Keystone items, the item meets the criterion if the DOK of the item is in alignment with the DOK of the Assessment Anchor as defined by the Eligible Content. Webb's DOK includes four levels, from the lowest (basic recall) level to the highest (extended thinking) level.

In some specific cases, DOK level 2 was allowed when the cognitive intent of an Eligible Content was level 2. For more information on DOK, see Chapter Three and Appendix A.

### ONLINE TESTING DESIGN CONSIDERATIONS

The Keystone Exams were designed from the beginning to provide a dual mode of test delivery, using traditional paper-and-pencil forms and using computer-based online forms. The computer-based online testing environment (called INSIGHT) is designed to provide a testing experience that mirrors the elements of traditional paper-and-pencil-based test delivery. This includes not only standard ancillary testing materials available in or with the printed forms, like formula sheets, periodic tables, scoring guidelines, and response spaces, but also analogs of the mechanical elements of response generation not necessarily associated with a computer-screen interface. These elements include line guides, rulers, screen highlighters, magnifiers, equation-building software, online calculators and graphing tools, and keyboard shortcuts.

Consideration of other components of online testing—like item layout, passage layout, font, screen resolution, navigation tools, and other interface mechanisms—all played a role in the overall design constraints, with some considerations having a more meaningful impact on specific exams. For more information on how the online test design impacted the overall test design considerations, see the sections below under each exam.

Online testing also provides an opportunity to utilize software to generate scores for student responses. In cases where responses to questions invoke numerical strings or equations, online responses can be scored through the use of lookup tables. Lookup tables are automated scoring rubrics that contain common correct and incorrect responses. When a response does not match a record in the lookup table, a human scorer is used to adjudicate the score. Operational autoscoring was only used for the Algebra I Exam; see below for more information on its use in Algebra I. For more information on scoring, see Chapter Eight.

## **ALGEBRA I**

The Keystone Algebra I Exam has two reporting categories: Module 1, Operations and Linear Equations & Inequalities, and Module 2, Linear Functions and Data Organizations. Both modules include three Assessment Anchors. Module 1 has 18 Eligible Content, and Module 2 has 15 Eligible Content. Each module corresponds to specific content aligned to statements and specifications included in the course-specific Assessment Anchor documents. The Algebra I content included in the Keystone Algebra I multiple-choice items aligns with the Assessment Anchors and Eligible Content statements. The process skills, directives, and action statements also specifically align with the Assessment Anchors and Eligible Content statements. The content included in Algebra I constructed-response items aligns with content included in the Eligible Content statements. The process skills, directives, and action statements included in the performance demands of the Algebra I constructed-response items align with specifications included in the Assessment Anchor statements, the Anchor Descriptor statements, and/or the Eligible Content statements. In other words, the verbs or action statements used in the constructed-response items or stems can come from the Eligible Content, Anchor Descriptor, or Assessment Anchor statements.

### **ALGEBRA I ONLINE CONSIDERATIONS**

Students taking the computer-based online delivery of the Algebra I exam are provided with online versions of several common tools typically available to a student taking a traditional paper-and-pencil exam. Each student has access to the following online tools: a standard four-function calculator, a scientific calculator, a graphing tool (similar, but not identical to, a graphing calculator), a ruler (available in metric and English units), a highlighter, a line guide, a magnifier, a sticky note generator, and a cross-off tool. In addition, an equation builder—which allows students to generate complex equations not normally possible with a standard keyboard—is also made available with all constructed-response items. Also, if the constructed-response item requests that the student draw, label, or otherwise change a graph, special graph-drawing tools are provided for on-screen graph generation. The Algebra I general scoring guideline and formula sheets are also available to students.

Layout of both the multiple-choice and constructed-response items is optimized for minimal screen manipulation (minimal scrolling required to see graphics or text that extend beyond the visible working space on the computer screen), and exam items are scrutinized carefully in both print and online versions for continuity and accuracy.

### **ALGEBRA I MULTIPLE-CHOICE ITEMS**

Sixty percent of the possible points on the Algebra I Exam are derived from multiple-choice items. This item type is especially efficient for measuring a broad range of content. Each multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent incorrect concepts, incorrect logic, incorrect application of an algorithm, or computational errors.

Algebra I multiple-choice items are intended to take about one and a half minutes of response time per item. They are used to assess a variety of skill levels, including problem solving. Algebra I items involving application emphasize the requirement to carry out some mathematical process to find an answer rather than simply recalling information from memory.

## ALGEBRA I CONSTRUCTED-RESPONSE ITEMS

Constructed-response items (tasks) require that students read a problem description and develop an appropriate solution. Algebra I constructed-response items are designed to take about ten minutes of response time per item. Most of the constructed-response items have several components in the overall task that may enable students to enter or begin the problem at different places. In some items, each successive component is designed to assess progressively more difficult skills or higher knowledge levels. Certain components may ask students to explain their reasoning for applying particular operations or for arriving at certain conclusions. The types of tasks utilized do not necessarily require computations. Students may also be asked to perform such tasks as constructing a graph, shading some portion of a figure, or listing object combinations that meet specified criteria.

Constructed-response tasks are especially useful for measuring students' problem-solving skills in Algebra. They offer the opportunity to present real-life situations that necessitate that the students solve problems using mathematics abilities learned in the classroom. Students must read the task carefully, identify the necessary information, devise a method of solution, perform the calculations, enter the solution directly in the answer document, and, when required, offer an explanation. This provides insight into the students' mathematical knowledge, abilities, and reasoning processes.

The constructed-response Algebra items are scored on a 0–4 point scale using an item-specific scoring guideline. The item-specific scoring guideline outlines the requirements for each score point. Item-specific scoring guidelines are based on the Algebra I General Description of Scoring Guidelines. The general guidelines describe a hierarchy of responses, which represent the five score levels, see Appendix B.

The Algebra I Keystone Exam includes two types of constructed-response items: Scaffolded Constructed-Response Items (SCR) and Extended Constructed-Response Items (ECR). Both types are scored on the same 0–4 point scale using the same Algebra I General Description of Scoring Guidelines as the base. SCR items are constructed to generally elicit four distinct responses (a response may contain more than one answer blank), and each response has the potential to earn a discrete number of score points (generally just one [1] score point per response). In turn, the four distinct responses are generally organized into four sections, with each labeled as a "Part" within an SCR. The next table shows a generic (nonauthentic) illustration of the application of this concept.

**Table 2–3. Generic Example [Nonauthentic] Showing Concept of Four Distinct Responses**

Stem	Part A	Part B	Part C	Part D
Presents a numerical distribution	In the answer spaces, write the list of numbers from least to greatest.	Write the mean in an answer blank.	Write the median in an answer blank.	Write the mode in an answer blank.
4 points	1 distinct point even though students enter more than one number	1 distinct point with one distinct entry	1 distinct point with one distinct entry	1 distinct point with one distinct entry

- SCR items do not require narrative, explanation, or "show all your work" responses.
- Most SCR item responses lend themselves to automatic scoring; however, **not all items can be automatically scored exclusively with the use of lookup tables.** The full application of Assessment Anchors and Eligible Content sometimes requires item construction that is incompatible with lookup tables.

In familiar and probably the most descriptive terms, Algebra I ECR items—in form, format, and scoring provisions—adhere to the philosophy of PSSA OE item format. Like SCR items, development is based on the item qualities that best measure the skills and concepts with which the item aligns.

- ECR items intentionally elicit narrative, explanation of reasoning, “explain why . . .”, and/or “show your work” responses.
- In contrast to SCR items, in which DOK level 3 cognitive engagement is inferred from student responses, ECR items (through explanations and recorded work) often provide direct evidence of DOK level 3 engagement. This aspect of ECR items is intentionally included during development. Following initial development, the ECR item will be approved by PDE as accepted by the review committee, or PDE and DRC will collaborate in amending the item.

## **BIOLOGY**

The Keystone Biology Exam has two reporting categories: Module 1[A], Cells and Cell Processes; and Module 2[B], Continuity and Unity of Life. Both modules have four Assessment Anchors. Module A has 16 Eligible Content, and Module B has 22 Eligible Content. Each module corresponds to specific content aligned to statements and specifications included in the course-specific assessment anchor documents. The Biology content included in the Keystone Biology multiple-choice items aligns with the Assessment Anchors and Eligible Content statements. The process skills, directives, and action statements also specifically align with the Assessment Anchors and Eligible Content statements. The content included in Biology constructed-response items aligns with content included in the Eligible Content statements. The process skills, directives, and action statements included in the performance demands of the Biology constructed-response items align with specifications included in the Assessment Anchor statements, the Anchor Descriptor statements, and/or the Eligible Content statements. In other words, the verbs or action statements used in the constructed-response items or stems can come from the Eligible Content, Anchor Descriptor, or Assessment Anchor statements.

### **BIOLOGY ONLINE CONSIDERATIONS**

Students taking the computer-based online delivery of the Biology Exam are provided with online versions of several common tools typically available to a student taking a traditional paper-and-pencil exam. Each student has access to the following online tools: a highlighter, a line guide, a magnifier, a sticky note generator, and a cross-off tool. The Biology general scoring guideline and a periodic table are also provided to students.

Layout of both the multiple-choice and constructed-response items is optimized for minimal screen manipulation (minimal scrolling to see graphics or text that extend beyond the visible working space on the computer screen), and exam items are scrutinized carefully in both print and online versions for continuity and accuracy.

### **BIOLOGY MULTIPLE-CHOICE ITEMS**

Seventy-three percent of the possible points on the Biology Exam are derived from multiple-choice items. Multiple-choice items are especially efficient for measuring a broad range of content. Each multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent incorrect concepts, incorrect logic, or incorrect application of a biological principle.

Biology multiple-choice items are intended to take about one and a quarter minutes of response time per item. They are used to assess a variety of skill levels, including the application of Biology content. Biology items involving application emphasize the requirement to utilize science content to find an answer rather than simply recalling information from memory.



## **BIOLOGY CONSTRUCTED-RESPONSE ITEMS**

Constructed-response items (tasks) require students to read a description of a Biology problem and to develop an appropriate solution. Biology constructed-response items are designed to take about eight minutes of response time per item. Constructed-response tasks are especially useful for measuring students' skills in biology. These tasks may present real-life situations that require students to solve problems using science abilities learned in the classroom. Students must read a task carefully, identify the necessary information, devise a method of solution, enter the solution directly into the answer document, and when required, offer an explanation. This provides insight into students' science knowledge, abilities, and reasoning processes.

The constructed-response science items are scored on a 0–3 point scale with an item-specific scoring guideline, and each task is carefully constructed with a scoring guideline reflecting the task requirements. The general guidelines describe a hierarchy of responses, which represents the four score levels. Each item-specific scoring guideline outlines the requirements at each score point, and each item-specific scoring guideline is based on the Biology General Description of Scoring Guidelines, see Appendix B.

## **LITERATURE**

The Keystone Literature Exam has two reporting categories: Module 1, Fiction; and Module 2, Nonfiction. Both modules have two Assessment Anchors. Module 1 has 25 Eligible Content, and Module 2 has 33 Eligible Content. The Literature Exam employs two types of test items, multiple-choice and constructed-response, and the content included aligns with content included in the Eligible Content statements. The items are designed to measure students' comprehension of the content contained in the literature passages. Each module corresponds to specific content aligned to statements and specifications included in the course-specific Assessment Anchor documents. The Literature content included in the Keystone Literature multiple-choice items aligns with the Assessment Anchors and Eligible Content statements. The process skills, directives, and action statements also specifically align with the Assessment Anchors and Eligible Content statements. The content included in Literature constructed-response items aligns with content included in the Eligible Content statements. The process skills, directives, and action statements included in the performance demands of the Literature constructed-response items align with specifications included in the Assessment Anchor statements, the Anchor Descriptor statements, and/or the Eligible Content statements. In other words, the verbs or action statements used in the constructed-response items or stems can come from the Eligible Content, Anchor Descriptor, or Assessment Anchor statements.

### **LITERATURE ONLINE CONSIDERATIONS**

Students taking the computer-based online delivery of the Literature Exam are provided with online versions of several common tools typically available to a student taking a traditional paper-and-pencil exam. Each student has access to the following online tools: a highlighter, a line guide, a magnifier, a sticky note generator, and a cross-off tool. The Literature general scoring guideline is also provided to students.

Layout of passages, multiple-choice items, and constructed-response items is optimized for minimal screen manipulation (minimal scrolling to see text and graphics that extend beyond the visible working space on the computer screen), and exam items are scrutinized carefully in both print and online versions for continuity and accuracy. In addition, the amount of space devoted to the passage compared to the amount of space devoted to the exam questions was also optimized.

### **LITERATURE MULTIPLE-CHOICE ITEMS**

Sixty-five percent of the possible points on the Literature Exam are derived from multiple-choice items. Literature multiple-choice items are intended to take about one minute of response time per item. They are designed to measure how well students comprehend the overall meaning of a passage or make basic inferences about it. At times, asking students to choose a preferred answer is the best way to determine whether they have gleaned certain information from a story. Such information may include central idea, setting, or main events and their sequence.

Each Literature multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent some kind of misinterpretation, predisposition, unsound reasoning, or casual reading.

## **LITERATURE CONSTRUCTED-RESPONSE ITEMS**

Constructed-response items (tasks) are designed to address comprehension of text in ways that multiple-choice items cannot. Literature constructed-response items are designed to take about five minutes of response time per item. A short written response allows students to prepare an answer and summarize using supporting details or examples derived from the text.

The Literature constructed-response items are scored on a 0–3 point scale using an item-specific scoring guideline. Each task is text-dependent and is carefully constructed with the scoring guideline reflecting the task requirements. All item-specific scoring guidelines are based on the Literature General Description of Scoring Guidelines. The general guidelines describe a hierarchy of responses, which represent the four score levels, see Appendix B.

## **LITERATURE PASSAGES**

One of the key requirements of the Keystone Literature Exam is that students should be able to read and comprehend both literature and informational texts of sufficient text complexity and quality as required by the Assessment Anchors and Eligible Content. For example, the Literature Keystone Assessment Anchors and Eligible Content require students to engage with appropriately complex literary fiction, literary nonfiction, and informational works. Passage genres include, but are not limited to, the following: stories; excerpts from novels, biographies, and autobiographies; letters; dramas; poems; myths from diverse cultures and different time periods; texts in history/social studies, science, and other disciplines; seminal U.S. documents; the classics of American, British, and world literature; and current articles and editorials.

## **TEXT COMPLEXITY**

Text complexity involves three components: matching reader to text and task, qualitative evaluation of the text, and quantitative evaluation of the text.

### MATCHING READER TO TEXT AND TASK

A number of factors are taken into consideration when deciding whether a passage will be placed in the pool for possible use on the Keystone Literature Exam. The factors include, but are not limited to, the following:

- Are the conceptual load, vocabulary, syntactic patterns, sentence length, and clarity appropriate for the grade level?
- Does the passage stand the test of time as an example of literary fiction, literary nonfiction, and/or informational text, and is it judged by the committee of Pennsylvania educators as having sufficient quality?
- Is the passage “rich” enough to generate a variety of items?
- Do the passages represent a range of reading levels appropriate to the grade level?
- Do the passages lend themselves well to measuring the Keystone Assessment Anchors and Eligible Content, including text structures and elements?
- Are the passages free of issues of bias, fairness, and/or sensitivity?
- Does the pool of passages represent diversity in the areas of gender, culture, ethnicity, urban/rural status, socioeconomic status, physical differences, and age?

### QUALITATIVE EVALUATION OF THE TEXT

Evaluating the text complexity of a passage is essentially a judgmental process by individuals familiar with the classroom context and what is linguistically appropriate at a given grade level. All Keystone passages to be included in the pool of passages for possible use on the Keystone Literature Exam are reviewed and approved by PDE and the Pennsylvania Reading Content Committee (a committee of Pennsylvania educators). The passages are reviewed by Pennsylvania educators to judge whether each passage meets the criteria outlined above. All potential passages are also reviewed by the Pennsylvania Bias, Fairness, and Sensitivity Committee.

## QUANTITATIVE EVALUATION OF THE TEXT

Each readability program uses different methods to determine the readability for a particular passage (e.g., syllables, sentence length, number of words, vocabulary lists). Each readability formula is designed for a particular grade range of materials. When using the various readability formulas, a wide range of readability levels may be identified for a particular passage. Some readability formulas are better suited to a particular grade level. If a particular formula being used is outside of the intended range, then the results may be unreliable.

Readability of the Keystone Literature Exam passages has been determined using several of the most widely accepted readability formulas. These formulas are not used in a rigid way, but rather more informally to provide for several “snapshots” of a passage. The readability formulas used for the passages that appear on the Keystone Literature Exam are the Dale-Chall Formula, the Flesch Grade Level Formula, and the Fry Graph.

# CHAPTER THREE: ITEM AND TEST DEVELOPMENT PROCESSES

## GENERAL KEYSTONE TEST DEVELOPMENT PROCESSES

The 2023 Keystone Exams continued to use the core-to-core biennial overlap. Approximately 30% to 50% of the operational points in each module overlap with items used operationally 2 years prior. The 2023 Keystone Exam cores were made up of items that had appeared on the Spring 2021, Summer 2021, and/or Winter 2021/2022 cores. The remainder of the operational 2023 exams were made up of items that were field-tested on the Spring 2022 Keystone Exams embedded field-test administration. Table 3–1 is a graphic representation of the basic process flow and overlap of the development cycles.

**Table 3–1. General Development and Usage Cycle of the Algebra I, Biology, and Literature Keystone Exams**

Admin Year	Events Occurring in Year 1	Events Occurring in Year 2	Events Occurring in Year 3	Events Occurring in Year 4	Events Occurring in Year 5
Year 1	Winter Admin; New Item Dev. for Spring Year 2	Spring Admin (Oper. & Embedded FT); Data Review of previous year's Spring FT; Summer Admin	Biennial Core-to-Core Overlap (a portion of the year 1 core included as a portion of the year 3 core)		
Year 2		Winter Admin; New Item Dev. for Spring Year 3	Spring Admin (Oper. & Embedded FT); Data Review of previous year's Spring FT; Summer Admin	Biennial Core-to-Core Overlap (a portion of the year 2 core included as a portion of the year 4 core)	
Year 3			Winter Admin; New Item Dev. for Spring Year 4	Spring Admin (Oper. & Embedded FT); Data Review of previous year's Spring FT; Summer Admin	Biennial Core-to-Core Overlap (a portion of the year 4 core included as a portion of the year 5 core)

## GENERAL TEST DEFINITION

The plan for the Keystone Exam was developed through the collaborative efforts of the Pennsylvania Department of Education (PDE) and Data Recognition Corporation (DRC). The exams are presented online or in two printed testing materials, a test book and a separate answer book. The test book contains multiple-choice (MC) items. The answer book contains scannable pages for multiple-choice responses, constructed-response (CR) items with response spaces, and demographic data collection areas. All MC items are worth 1 point. Algebra I CR items receive a maximum of 4 points (on a scale of 0–4), and all Biology and Literature CR items receive a maximum of 3 points (on a scale of 0–3).

## CORE-TO-CORE OVERLAP ITEMS

The operational items consist of a set of core items taken by all students. Starting in 2014 these core items included core-to-core overlapping items, which are items that also appeared on the core form of the administration two years before. The overlap connects the spring and summer administrations of year  $x$  and the winter administration of year  $x+1$ , with the year  $x+2$  spring and summer and year  $x+3$  winter administrations. The first biennial core-to-core overlap from the spring 2011 and winter 2011–2012 core was scheduled to begin with the spring 2013 administration. However, when the program was placed on hiatus during the 2011–2012 school year, the overlap was moved to the spring 2014 administration.

## ALGEBRA I TEST DEFINITIONS

The Spring 2023 Algebra I Keystone Exam was composed of 24 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. The following two tables display the design for Algebra I for forms 1 through 24. The column entries for these tables denote the following:

- Number of unique core MC items
- Number of unique core CR items
- Number of embedded MC field-test items
- Number of embedded CR field-test items
- Total number of MC and CR items in the form

**Table 3–2. Algebra I Test Plan (Spring 2023) per Operational Form**

Module	Core per Form MC Items	Core per Form CR Items	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	18	3	5	1	23	4
2	18	3	5	1	23	4
Total	36	6	10	2	46	8

**Table 3–3. Algebra I Test Plan (Spring 2023) per 24 Operational Forms**

Module	Core per 24 Forms MC Items	Core per 24 Forms CR Items	Field Test per 24 Forms MC Items	Field Test per 24 Forms CR Items	Total per 24 Forms Core & FT MC Items	Total per 24 Forms Core & FT CR Items
1	18	3	120	24	138	27
2	18	3	120	24	138	27
Total	36	6	120	48	276	54

The operational (core) portions of the Spring 2023, Summer 2023, and the Winter 2023/2024 administrations came from the same sources. Therefore 30% to 50% of the 2023/2024 Winter, Spring and Summer cores overlap with the Spring 2021, Summer 2021, and Winter 21/22 cores. The remaining core items that appeared on the 2023/2024 forms were field-tested on prior administrations. Although each spring administration includes embedded field-test items, the summer, winter, and breach forms do not include any embedded field-test items due to the lower  $n$ -counts for these administrations. However, summer, winter, and breach forms still include the same number of items that appear in the spring administration. Instead of field-test items, the slots in these exams are filled by placeholder (PH) items. Table 3–4 displays the design for the Algebra I Summer, Winter, and Breach operational forms.

**Table 3–4. Algebra I Test Plan (2023 Summer, Winter, and Breach) per Operational Form**

Module	Core per Form MC Items	Core per Form CR Items	Placeholders per Form MC Items	Placeholders per Form CR Items	Total per Form Core & PH MC Items	Total per Form Core & PH CR Items	Number of Forms Master Core	Number of Forms Scrambles
1	18	3	5	1	23	4	1	3
2	18	3	5	1	23	4	1	3
Total	36	6	10	2	46	8	1	3

Since an individual student’s score is based solely on the operational (or core) items, the total number of operational points is 60 for Algebra I. The total score is obtained by combining the points from the core MC (1 point each) and core CR (up to 4 points each) portions of the exam as follows:

**Table 3–5. Algebra I Core Points**

Category	Module 1 MC Items	Module 1 CR Items	Module 2 MC Items	Module 2 CR Items	Total MC Items	Total CR Items
Total Points	30 (50%)		30 (50%)		60 (100%)	
Core Items	18	3	18	3	36	6
Core Points	18	12	18	12	36	24

The Algebra I Exam results will be reported in two categories based on the two modules of the Algebra I Exam. The code letters for these Assessment Anchor categories are

1. Operations and Linear Equations & Inequalities
2. Linear Functions and Data Organization

The distribution of Algebra I items into these two categories is shown in the following table.

**Table 3–6. Algebra I Module and Anchor Distribution**

Algebra I Module	Raw Points	Module Weight	Number of Anchors	Number of Eligible Content
1	30	50%	3	18
2	30	50%	3	15

The reporting categories are further subdivided for specificity and Eligible Content (limits). Each subdivision is coded by adding an additional character to the framework of the labeling system. These subdivisions are called Assessment Anchors and Eligible Content. More information about Assessment Anchors and Eligible Content is in Chapter Two.

For more information concerning the process used to convert the Algebra I operational test plan into forms (i.e., form construction), see Chapter Six.

For more information concerning the test sessions, timing, and layout for the Algebra I operational exam, see Chapter Seven.

## **BIOLOGY TEST DEFINITIONS**

The Spring 2023 Biology Keystone Exam was composed of 6 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. The following two tables display the design for Biology for forms 1 through 6. The column entries for these tables denote the following:

- Number of unique core MC items
- Number of unique core CR items
- Number of embedded MC field-test items
- Number of embedded CR field-test items
- Total number of MC and CR items in the form

**Table 3–7. Biology Test Plan (Spring 2023) per Operational Form**

Module	Core per Form MC Items	Core per Form CR Items	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	24	3	8	1	32	4
2	24	3	8	1	32	4
Total	48	6	16	2	64	8

**Table 3–8. Biology Test Plan (Spring 2023) per 6 Operational Forms**

Module	Core per 6 Forms MC Items	Core per 6 Forms CR Items	Field Test per 6 Forms MC Items	Field Test per 6 Forms CR Items	Total per 6 Forms Core & FT MC Items	Total per 6 Forms Core & FT CR Items
1	24	3	48	6	72	9
2	24	3	48	6	72	9
Total	48	6	96	12	144	18

The operational (core) portions of the Spring 2023, Summer 2023, and the Winter 2023/2024 administrations came from the same sources. Therefore 30% to 50% of the 2023/2024 Winter, Spring and Summer cores overlap with the Spring 2021, Summer 2021, and Winter 21/22 cores. The remaining core items that appeared on the 2023/2024 forms were field-tested on prior administrations. Although each spring administration includes embedded field-test items, the summer, winter, and breach forms do not include any embedded field-test items due to the lower *n*-counts for these administrations. However, summer, winter, and breach forms still include the same number of items that appear in the spring administration. Instead of field-test items, the slots in these exams are filled by placeholder (PH) items. Table 3–9 displays the design for the Biology Summer, Winter, and Breach operational forms.

**Table 3–9. Biology Test Plan (2023 Summer, Winter, and Breach) per Operational Form**

Module	Core per Form MC Items	Core per Form CR Items	Placeholders per Form MC Items	Placeholders per Form CR Items	Total per Form Core & PH MC Items	Total per Form Core & PH CR Items	Number of Forms Master Core	Number of Forms Scrambles
1	24	3	8	1	32	4	1	3
2	24	3	8	1	32	4	1	3
Total	48	6	16	2	64	8	1	3

Since an individual student’s score is based solely on the operational (or core) items, the total number of operational points is 66 for Biology. The total score is obtained by combining the points from the core MC (1 point each) and core CR (up to 3 points each) portions of the exam as follows:

**Table 3–10. Biology Core Points**

Category	Module 1 MC Items	Module 1 CR Items	Module 2 MC Items	Module 2 CR Items	Total MC Items	Total CR Items
Total Points	33 (50%)		33 (50%)		66 (100%)	
Core Items	24	3	24	3	48	6
Core Points	24	9	24	9	48	18

The Biology Exam results will be reported in two categories based on the two modules of the Biology Exam.

1. Cells and Cell Processes
2. Continuity and Unity of Life

The distribution of Biology items into these two categories is shown in the following table.

**Table 3–11. Biology Module and Anchor Distribution**

Biology Module	Raw Points	Module Weight	Number of Anchors	Number of Eligible Content
1	33	50%	4	16
2	33	50%	4	22

The reporting categories are further subdivided for specificity and Eligible Content (limits). Each subdivision is coded by adding an additional character to the framework of the labeling system. These subdivisions are called Assessment Anchors and Eligible Content. More information about Assessment Anchors and Eligible Content is in Chapter Two.

For more information concerning the process used to convert the Biology operational test plan into forms (i.e., form construction), see Chapter Six.

For more information concerning the test sessions, timing, and layout for the Biology operational exam, see Chapter Seven.

## LITERATURE TEST DEFINITIONS

The Spring 2023 Literature Keystone Exam was composed of 24 forms. All of the forms contained operational core items identical for all students and sets of generally unique items. The following two tables display the design for Literature for forms 1 through 24. The column entries for these tables denote the following:

- Number of unique core passages
- Number of unique core MC items
- Number of unique core CR items
- Number of embedded field-test passages
- Number of embedded MC field-test items
- Number of embedded CR field-test items
- Total number of passages, MC items, and CR items in the form

**Table 3–12. Literature Test Plan (Spring 2023) per Operational Form**

Module	Core per Form Passages	Core per Form MC Items	Core per Form CR Items	Field Test per Form Passages	Field Test per Form MC Items	Field Test per Form CR Items	Total per Form Passages	Total per Form Core & FT MC Items	Total per Form Core & FT CR Items
1	2	17	*3	1	6	1	3	23	4
2	2	17	*3	1	6	1	3	23	4
Total	4	34	6	2	12	2	6	46	8

\*For each module, one core passage has two CRs and one core passage has one CR.



**Table 3–13. Literature Test Plan (Spring 2023) per 24 Operational Forms**

Module	Core per 24 Forms Passages	Core per 24 Forms MC Items	Core per 24 Forms CR Items	Field Test per 24 Forms Passages	Field Test per 24 Forms MC Items	Field Test per 24 Forms CR Items	Total per 24 Forms Passages	Total per 24 Forms Core & FT MC Items	Total per 24 Forms Core & FT CR Items
1	2	17	*3	12	144	24	26	161	27
2	2	17	*3	12	144	24	26	161	27
Total	4	34	6	24	288	48	52	322	54

\*For each module, one core passage has two CRs and one core passage has one CR.

The operational (core) portions of the Spring 2023, Summer 2023, and the Winter 2023/2024 administrations came from the same sources. Therefore 30% to 50% of the 2023/2024 Winter, Spring and Summer cores overlap with the Spring 2021, Summer 2021, and Winter 21/22 cores. The remaining core items that appeared on the 2023/2024 forms were field-tested on prior administrations. Although each spring administration includes embedded field-test items, the summer, winter, and breach forms do not include any embedded field-test items due to the lower *n*-counts for these administrations. However, summer, winter, and breach forms still include the same number of items that appear in the spring administration. Instead of field-test items, the slots in these exams are filled by placeholder (PH) items. Table 3–14 displays the design for the Literature Summer, Winter, and Breach operational forms.

**Table 3–14. Literature Test Plan (2023 Summer, Winter, and Breach) per Operational Form**

Module	Core per Form MC Items	Core per Form CR Items	Placeholders per Form MC Items	Placeholders per Form CR Items	Total per Form Core & PH MC Items	Total per Form Core & PH CR Items	Number of Forms Master Core	Number of Forms Scrambles
1	2	17	*3	1	6	1	1	3
2	2	17	*3	1	6	1	1	3
Total	4	34	6	2	12	2	1	3

\*For each module, one core passage has two CRs and one core passage has one CR.

Since an individual student’s score is based solely on the operational (or core) items, the total number of operational points is 52 for Literature. The total score is obtained by combining the points from the core MC (1 point each) and core CR (up to 3 points each) portions of the exam as follows:

**Table 3–15. Literature Core Points**

Category	Module 1 Passages	Module 1 MC Items	Module 1 CR Items	Module 2 Passages	Module 2 MC Items	Module 2 CR Items	Total Passages	Total MC Items	Total CR Items
Total Points	26 (50%)			26 (50%)			52 (50%)		
Core Items	2	17	3	2	17	3	4	34	6
Core Points	NA	17	9	NA	17	9	NA	34	18

The Literature Exam results will be reported in two broad categories based on the two modules of the Literature Exam.

1. Fiction Literature
2. Nonfiction Literature

The distribution of Literature items into these two categories is shown in the following table.

**Table 3–16. Literature Module and Anchor Distribution**

Literature Module	Raw Points	Module Weight	Number of Anchors	Number of Eligible Content
1	26	50%	2	25
2	26	50%	2	31

The reporting categories are further subdivided for specificity and Eligible Content (limits). Each subdivision is coded by adding an additional character to the framework of the labeling system. These subdivisions are called Assessment Anchors and Eligible Content. More information about Assessment Anchors and Eligible Content is in Chapter Two.

For more information concerning the process used to convert the Literature operational test plan into forms (i.e., form construction), see Chapter Six.

For more information concerning the test sessions, timing, and layout for the Literature operational exam, see Chapter Seven.

## ITEM DEVELOPMENT CONSIDERATIONS

Alignment to the Keystone Assessment Anchors and Eligible Content, course-level appropriateness (as specified by PDE), depth of knowledge (DOK), item/task level of complexity, estimated difficulty level, relevancy of context, rationale for distractors, style, accuracy, and correct terminology were major considerations in the item development process. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014) and *Universal Design* (Thompson, Johnstone, & Thurlow, 2002) guided the development process. In addition, *Fairness in Testing: Training Manual for Issues of Bias, Fairness, and Sensitivity* (DRC, 2010) was used for developing items. All items were reviewed for fairness by bias and sensitivity committees and for content by Pennsylvania educators and field specialists.

## BIAS, FAIRNESS, AND SENSITIVITY OVERVIEW

At every stage of the item and test development process, DRC employs procedures that are designed to ensure that items and tests meet Standard 7.4 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

Standard 7.4: Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.

To meet Standard 7.4, DRC uses a series of internal quality steps. DRC provides specific training for test developers, item writers, and reviewers on how to write, review, revise, and edit items related to issues of bias, fairness, and sensitivity (as well as based on technical quality). Training also includes an awareness of and sensitivity to issues of cultural diversity. In addition to providing *internal* training in reviewing items in order to eliminate potential bias, DRC also provides *external* training to the review panels of minority experts, teachers, and other stakeholders.

DRC’s guidelines for bias, fairness, and sensitivity include instruction concerning how to eliminate language, symbols, words, phrases, and content that might be considered offensive by members of racial, ethnic, gender, or other groups. Areas of bias that are specifically targeted include, but are not limited to, stereotyping, gender, region/geography, ethnic/cultural, socioeconomic/class, religion, experience, and biases against a particular age group (ageism) or persons with disabilities. DRC catalogues topics that should be avoided and maintains balance in gender and ethnic emphasis within the pool of available items and passages.

See the sections below in this chapter for more information about the Bias, Fairness, and Sensitivity Review meetings conducted for the Keystone Exams.

## **UNIVERSAL DESIGN OVERVIEW**

The principles of universal design were incorporated throughout the item development process to allow participation of the widest possible range of students in the Keystone Exams. The following checklist was used as a guideline:

- Items measure what they are intended to measure.
- Items respect the diversity of the assessment population.
- Items have a clear format for text.
- Stimuli and items have clear pictures and graphics.
- Items have concise and readable text.
- Items allow changes to other formats, such as Braille, without changing meaning or difficulty.
- The arrangement of the items on the test has an overall appearance that is clean and well organized.

A more extensive description of the application of the principles of universal design is provided in Chapter Four.

## **DEPTH-OF-KNOWLEDGE OVERVIEW**

An important element in statewide graduation exams is the alignment between the overall assessment system and the state's standards. A methodology developed by Norman Webb (1999, 2006) offers a comprehensive model that can be applied to a wide variety of contexts. With regard to the alignment between standards statements and the assessment instruments, Webb's criteria include five categories, one of which deals with content. Within the content category is a useful set of levels for evaluating DOK. According to Webb (1999), "depth-of-knowledge consistency between standards and assessments indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards" (Webb, 1999, pp. 7–8). The four levels of cognitive complexity (i.e., DOK) are as follows:

- Level 1: Recall
- Level 2: Application of Skill/Concept
- Level 3: Strategic Thinking
- Level 4: Extended Thinking

DOK levels were incorporated into the item writing and review process, and items were coded with respect to the level they represented. The DOK level for MC and CR items are at Level 3, Level 2, or Level 1 depending on the cognitive intent of an Eligible Content. DOK Level 4 items are not included on the Keystone Exams. For more information on DOK (and a comparison of DOK to Bloom's Taxonomy), see Appendix A.

## **PASSAGE READABILITY OVERVIEW**

Evaluating the readability of a passage is essentially a judgment by individuals familiar with the classroom context and what is linguistically appropriate (PDE recommends that the Literature Keystone Exam be administered at grade 10). Although various readability indices were computed and reviewed, it is recognized that such methods measure different aspects of readability and are often fraught with particular interpretive liabilities. Thus, the commonly available readability formulas were not used in a rigid way, but more informally to provide for several snapshots of a passage that senior test development staff considered, along with experience-based judgments in guiding the passage-selection process. In addition, passages were reviewed by committees of Pennsylvania educators who evaluated each passage for readability and grade-level appropriateness. For more information on Literature passages, see Chapter Two and the literature passage-selection process described below.

## TEST ITEM READABILITY OVERVIEW

Careful attention was given to the readability of the items to make certain that the assessment focus of the item did not shift based on the difficulty of reading the item. Subject/course areas such as Algebra I or Biology contain many content-specific vocabulary terms. As a result, readability formulas were not used. However, wherever it was practicable and reasonable, every effort was made to keep the vocabulary at or one level below the course level for non-Literature exams. There was a conscious effort made to ensure that each question was evaluating a student's ability to build toward mastery of the course standards rather than evaluating the student's reading ability. Resources used to verify the vocabulary level were the *EDL Core Vocabularies* and the *Children's Writer's Word Book*.

In addition, every test question is brought before several different committees composed of Pennsylvania educators who are course-level/grade-level experts in the content field in question. They review each question from the perspective of the students they teach, determine the validity of the vocabulary used, and work to minimize the level of reading required.

Vocabulary was also addressed at the Bias, Fairness, and Sensitivity Review, although the focus was on how certain words or phrases may represent possible sources of bias or issues of fairness or sensitivity. See the sections that follow in this chapter for more information about the Bias, Fairness, and Sensitivity Review meetings conducted for the Keystone Exams.

## ITEM AND TEST DEVELOPMENT CYCLE

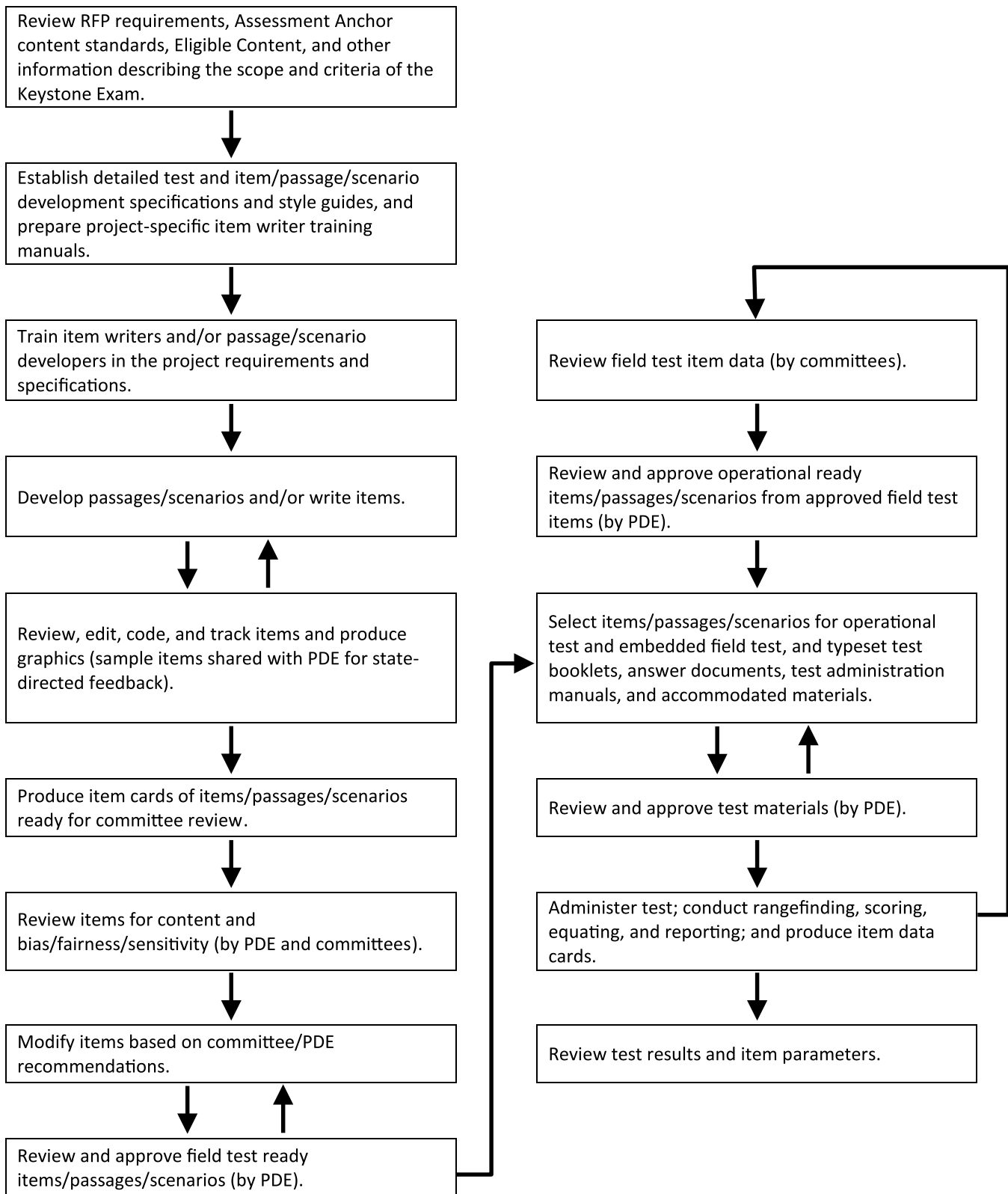
The item development process for items followed a logical cycle and time line, which are outlined in the table and figure on the next pages. On the front end of the schedule, tasks were generally completed with the goal of presenting field-test candidate items to committees of Pennsylvania educators. On the back end of the schedule, all tasks led to the field-test data review and operational test construction. This process represents a typical life cycle for an embedded Keystone field-test event, not a stand-alone field-test event or an accelerated development cycle.

The process flowchart, also shown below, illustrates the interrelationship among the steps in the primary cycle that occurs in a normal process of development (i.e., when the items for field testing are primarily from new development, as opposed to being selected from an existing item bank). In addition, a detailed process table describing the item and test development processes also appears in Appendix C.

**Table 3–17. General Item and Test Development Life Cycle for Spring Keystone Administrations**

Cycle	Steps in Development Life Cycle	Time Line	Approximate Window	
Primary	Development Planning	Summer/Fall	Month 1–4	Jul–Oct
Primary	Literature Passage Selection	Summer/Fall	Month 1–6	Jul–Dec
Primary	Item Writer Training	Fall	Month 5	Nov
Primary	Initial Item Authoring	Fall/Winter	Month 5–9	Nov–Mar
Primary	Internal Reviews and PDE Reviews	Fall to Spring	Month 6–12	Dec–Jun
Primary	Bias, Fairness, and Sensitivity Review	Summer	Month 13	Jul
Primary	New Item Content Committee Review (PA Educators)	Summer	Month 13	Jul
Primary	Post-Review Resolution and Cleanup	Summer	Month 13–14	Jul–Aug
Primary	Build Field-Test Forms	Summer/Fall	Month 15–16	Sep–Oct
Primary	Internal Form Reviews and PDE Reviews	Summer/Fall	Month 15–18	Sep–Dec
Primary	Final Form and Printer Proof Approvals	Fall/Winter	Month 18–19	Dec–Jan
Primary	Ancillary and Accommodated Form Development	Fall/Winter	Month 18–20	Dec–Feb
Primary	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 19–22	Jan–Apr
Primary	Field-Test Administration	Spring	Month 23	May
Primary	Material/Data Processing, Rangefinding, and Scoring	Spring/Summer	Month 23–26	May–Aug
Primary	Field-Test Item Data Review (PA Educators)	Summer	Month 27	Sept
Primary	Select Operational Items	Summer/Fall	Month 27–28	Sep–Oct
Primary	Build Operational Forms	Fall	Month 28–29	Oct–Nov
Primary	Internal Form Reviews and PDE Reviews	Fall	Month 29–30	Nov–Dec
Primary	Final Form and Printer Proof Approvals	Fall/Winter	Month 30–31	Dec–Jan
Primary	Ancillary and Accommodated Form Development	Fall/Winter	Month 31–33	Jan–Mar
Primary	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 31–33	Jan–Mar
Primary	Operational Test Administration	Spring	Month 35	May
Primary	Material/Data Processing and Scoring	Spring/Summer	Month 35–36	May–Jun
Primary	Score Reporting	Summer	Month 35–39	May–Sep
Secondary	Select Biennial Core-to-Core Overlap Items (Operational Items)	Summer/Fall	Month 51–52	Sep–Oct
Secondary	Build Operational Forms	Fall/Winter	Month 52–53	Oct–Nov
Secondary	Internal Form Reviews and PDE Reviews	Winter	Month 53–54	Nov–Dec
Secondary	Final Form and Printer Proof Approvals	Winter	Month 54–55	Dec–Jan
Secondary	Ancillary and Accommodated Form Development	Winter/Spring	Month 55–57	Jan–Mar
Secondary	Form Printing, Spiraling, Packaging, and Shipping	Winter/Spring	Month 56–58	Feb–Apr
Secondary	Second Operational Test Administration	Spring	Month 59	May
Secondary	Material/Data Processing and Scoring	Spring/Summer	Month 59–60	May–Jun
Secondary	Score Reporting	Summer	Month 59–63	May–Sep
Tertiary	Release Core-to-Core Overlap Items in Samplers	Fall	Month 63	Sep

**Figure 3–1. DRC Item and Test Development Primary Cycle**



## GENERAL ITEM AND TEST DEVELOPMENT PROCESS

This section describes the processes which lead up to an operational exam. These processes were used to develop the entire pool of items that appeared in the field-test and operational administrations.

### ITEM DEVELOPMENT PLANNING MEETING

Prior to the start of any item development work, DRC’s test development staff meets with PDE’s assessment office to discuss the test development plans for the next administration, including the test blueprint, the field-test plan (including development counts), procedures, timelines, etc. With a complete development cycle lasting about three years (from item authoring through field test, data review, and operational usage), the initial planning begins well in advance of the anticipated administration.

### ITEM WRITER TRAINING

Item writers were selected and trained for the subject areas of Algebra I, Biology, and Literature. Qualified writers—either hired independently by the testing vendor, DRC, or through subcontractors like Victory—were college graduates with teaching experience and a demonstrated base of knowledge in the content area. Many of these writers were content assessment specialists and curriculum specialists. The writers were trained individually and had previous experience in writing MC and CR items. Prior to developing items for the Keystone Exams, the cadre of item writers was trained with regard to the following areas:

- Keystone Assessment Anchors and Eligible Content
- Webb’s levels of cognitive complexity, DOK
- Subject-specific general scoring guidelines
- Specific and general guidelines for item writing
- Bias, fairness, and sensitivity guidelines
- Principles of universal design
- Item quality technical style guidelines
- Reference information
- Sample items

### LITERATURE PASSAGE SELECTION

The task of searching for passages was conducted by DRC professionals with classroom experience in reading/ language arts. These professionals also underwent specialized training (provided by DRC) in the characteristics of acceptable passages. Guidelines for passage selection included appropriate length, text structure, density, and vocabulary. A judgment was also made about whether the reading level required by a particular passage was at the independent level—that is, where the average student should be able to read 90 percent of words in the text independently. Passage finders were given the task of searching for a specified number of passages for each genre. Generally, they looked for at least twice as many passages as were needed. Passages acquired were either authentic (permissioned), in that they were culled from published materials, or commissioned by experienced authors. See Chapter Two for more information on the types of passages used on the Literature Keystone Exam.

For permissioned passages, approval to reprint was secured from the publishers. Passages underwent an internal review by several test development content editors to judge their merit with regard to the following criteria:

- Passages have interest value for students.
- Passages are appropriate in terms of vocabulary and language characteristics.
- Passages are free of bias, fairness, and sensitivity issues.
- Passages represent different cultures.
- Passages are from a variety of sources.
- Passages are able to stand the test of time.

- Passages are sufficiently rich to generate a variety of MC and CR items.
- Passages are complete with all necessary permissions documentation.
- Passages avoid dated subject matter unless a relevant historical context is provided.
- Passages should not require students to have extensive background knowledge in a certain discipline or area to understand a text.

Once through the internal review process, the passages deemed potentially acceptable were reviewed by the Reading Content Committee and the Bias, Fairness, and Sensitivity Committee for final approval.

### **ITEM AUTHORIZING AND TRACKING**

Initially, items are generated with software-prepared Keystone Exams Item Cards, which allow for preliminary sorting and reviewing. Although very similar, the Keystone Exams Item Card for Multiple-Choice Items differs from the Keystone Exams Item Card for Constructed-Response Items in that the former has a location at the bottom of the card for comments regarding the distractors. Blank examples of these two cards are shown in Appendix D. In both instances, a column against the right margin includes codes to identify the subject area, grade, content categories, passage information (in the case of reading), item type, DOK (cognitive complexity), estimated difficulty, answer key (for MC items), and calculator use (for mathematics items).

All items undergoing field testing were entered into the DRC Item Development and Educational Assessment System (IDEAS), which is a comprehensive, secure, online item banking system. It accommodates item writing, item viewing and reviewing, and item tracking and versioning. IDEAS manages the transition of an item from its developmental stage to its approval for use in a test form (for both print and online delivery). The system supports item history records that include item usage within a form, item-level notes, content categories and subcategories, item statistics from both classical and Rasch item analyses, and classifications derived from analyses of differential item functioning (DIF). A sample IDEAS Item Card is presented in Appendix D.

### **INTERNAL REVIEWS AND PDE REVIEWS**

To ensure that the items produced were sufficient in number and adequately distributed across subcategories and levels of difficulty, item writers were informed of the required quantities of items. As items were written, an item authoring card was completed. It contained information about the item, such as subject, content category, and subcategories. Based on the item writer's classroom teaching experience, his/her knowledge of the content area curriculum, and the cognitive demands required by the item, estimates were recorded for level of cognitive complexity and difficulty level. Items were written to provide for a range of difficulty and for cognitive complexity focused on DOK Level 3.

As part of the item construction process, each item was reviewed by content specialists and editors at DRC. Content specialists and editors evaluated each item to make sure that it measured the intended Eligible Content and/or Assessment Anchor. They also assessed each item to make certain that it was appropriate for the intended grade and that it provided and cued only one correct answer (MC items only). In addition, the difficulty level, DOK, graphics, language demand, and distractors were also evaluated. Other elements considered in this process included, but were not limited to, universal design, bias, source of challenge, grammar/punctuation, and Keystone style.

Following this internal process, items were reviewed by content specialists at PDE, who then consulted with DRC about any general issues or concerns (e.g., style, format, interpretation of Assessment Anchors and Eligible Content) and about edits to specific items. Following PDE's review, the items were prepared for the content review meetings conducted with Pennsylvania educators.



## ITEM CONTENT REVIEWS

Prior to the 2023 field testing, all newly developed test items were submitted to content committees for review. The content committees consisted of Pennsylvania educators from school districts throughout the Commonwealth of Pennsylvania, some with postsecondary university affiliations. The primary responsibility of the content committee was to evaluate items with regard to quality and content classification, including grade-level (course) appropriateness, estimated difficulty, DOK, and source of challenge. They also suggested revisions and made recommendations for reclassification of items. In some cases when an item was deleted, the committee suggested a replacement item and/or reviewed a suggested replacement item provided by the facilitators. The committee also reviewed the items for adherence to the principles of universal design, including language demand and issues of bias, fairness, and sensitivity.

With source of challenge, items were identified where the cognitive demand was focused on an unintended content, concept, or skill (Webb, 2002). Source of challenge may be a contributing factor if the reason that an answer could be given results from a cultural bias, an inappropriate reading level, a flawed graphic in an item, or if an item requires specialized, noncontent-related knowledge to answer. Source of challenge could result in a student who has mastered the intended content or skill answering the item incorrectly or a student who has not mastered the intended content or skill answering the item correctly. Committee members were asked to note any items with a source of challenge and to suggest revisions to remove the source of challenge.

DRC content assessment specialists facilitated the reviews and were assisted by representatives of PDE. Committee members, grouped by exam, worked through and reviewed the items for quality and content, as well as for the following categories:

- Assessment Anchor alignment
- Content limits
- Grade-level (course-level) appropriateness
- Difficulty level
- DOK
- Appropriate source of challenge
- Correct answer
- Quality of distractors
- Graphics in regards to appropriateness
- Appropriate language demand
- Freedom from bias

The members then came to consensus and assigned a status to each item: Approved, Accepted with Revision, or Rejected. All comments were recorded, and a master rating sheet was completed. Committee facilitators recorded the committee consensus on the Item Review Rating Sheet. A sample form and rating criteria may be found in Appendix E.

Security was addressed by adhering to a strict set of procedures. Items in binders were distributed for committee review by number and signed for by each member on a daily basis. All attendees, with the exception of PDE staff, were required to sign a confidentiality agreement. All materials not in use at any time were stored in a locked room. Secure materials that did not need to be retained after the meetings were deposited in secure barrels, the contents of which were shredded.

## **BIAS, FAIRNESS, AND SENSITIVITY REVIEWS**

Prior to field testing, all newly developed test items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. The committee's primary responsibility was to evaluate items with regard to bias, fairness, and sensitivity issues. They also made recommendations for changes or deletion of items in order to remove the potential for issues of bias, fairness, and/or sensitivity. Included in the review were proposed reading passages. An expert, multiethnic committee composed of men and women was trained by a DRC test development lead to review items for bias, fairness, and sensitivity issues. Training materials included a manual developed by DRC (DRC, 2017). Members of the committee also had expertise with special needs students and English Learners (EL). PDE staff members were also trained and participated in the review. All items were read by a cross-section of committee members. Each member noted bias, fairness, and/or sensitivity comments on tracking sheets and on the item, if needed, for clarification. Committee members individually categorized any concerns as related to ageism, disability, ethnicity/culture, gender, region, religion, socioeconomics, or stereotypes. These categories were the framework through which recommendations for modification or rejection of items occurred during the subsequent committee consensus process. The committee discussed each of the issues as a group and came to consensus as to which decisions should represent the view of the committee. All consensus comments were then compiled, and the suggested actions on these items were recorded and submitted to PDE. This review followed the same security procedures as outlined above.

# CHAPTER FOUR: UNIVERSAL DESIGN PROCEDURES APPLIED TO THE KEYSTONE EXAMS TEST DEVELOPMENT PROCESS

---

## UNIVERSAL DESIGN

Universally designed assessments allow participation of the widest possible range of students and contribute to valid inferences about participating students. Principles of Universal Design are based on the premise that each child in school is a part of the population to be tested and that testing results should not be affected by disability, gender, race, or English language ability (Thompson, Johnstone, & Thurlow, 2002). At every stage of the item and test development process, procedures were employed to ensure that items and subsequent tests (in both print and online delivery methods) were designed and developed using the elements of universally designed assessments established by the National Center on Educational Outcomes (NCEO).

Federal legislation addresses the need for universally designed assessments. The No Child Left Behind Act (Elementary and Secondary Education Act) requires that each state must “provide for the participation in [statewide] assessments of all students” [Section 1111(b)(3)(C)(ix)(I)]. Both Title I and IDEA regulations call for universally designed assessments that are accessible and valid for all students, including English Learners and students with disabilities. The benefits of universally designed assessments not only apply to these groups of students, but to all individuals with wide-ranging characteristics.

DRC’s test development team was trained in the elements of Universal Design as they relate to developing large-scale statewide assessments. Team leaders were trained directly by NCEO, and other team members were subsequently trained by team leaders. Committees involved in content review included some members who were familiar with the unique needs of students with disabilities and English Learners. Likewise, some members of the Bias, Fairness, and Sensitivity Committee were conversant with these issues. What follows are the Universal Design guidelines that were followed during all stages of the item development process for the Keystone Exams.

## ELEMENTS OF UNIVERSALLY DESIGNED ASSESSMENTS

After a review of research relevant to the assessment development process and the Principles of Universal Design (Connell et al., 1997), NCEO has produced seven elements of Universal Design as they apply to assessments (Thompson, Johnstone, & Thurlow, 2002). These elements served to guide item development for the Keystone Exams.

- **Inclusive Assessment Population**

The target population includes students attending Commonwealth schools who participate in one or more of the graduation competency exams.

- **Precisely Defined Constructs**

An important function of well-designed assessments is that the assessments actually measure what they are intended to measure. The Keystone Exams Assessment Anchor Content Standards (Assessment Anchors) provided clear descriptions of the constructs to be measured on each of the exams. Universally designed assessments must remove all non-construct-oriented cognitive, sensory, emotional, and physical barriers.

- **Accessible, Nonbiased Items**

DRC conducted both internal and external reviews of items and test specifications to ensure that they did not create barriers due to lack of sensitivity to disability, culture, or other subgroups. Items and test specifications were developed by a team who understood the varied characteristics of items that might create difficulties for any group of students. Accessibility is incorporated as a primary dimension of test specifications, so accessibility was woven into the fabric of the test rather than being added after the fact.

- **Amenable to Accommodations**

Even though items on universally designed assessments are accessible for most students, there are some students who continue to need accommodations. This essential element of a universally designed assessment requires that the exam is compatible with accommodations and a variety of widely used adaptive equipment and assistive technology (see also the section on Assessment Accommodations later in this chapter).

- **Simple, Clear, and Intuitive Instructions and Procedures**

Assessment instructions should be easy to understand regardless of a student’s experience, knowledge, language skills, or current concentration level. Questions that are posed using complex language can invalidate the test if students cannot understand how they are expected to respond to a question. To meet this guideline, directions and questions were prepared in simple, clear, and understandable language that underwent multiple reviews.

- **Maximum Readability and Comprehensibility**

A variety of guidelines exist to ensure the maximum readability and comprehensibility of a test. These features go beyond what is measured by readability formulas. Readability and comprehensibility are affected by many factors, including student background, sentence difficulty, and text organization. All of these features were considered as item text was developed.

Plain language is a concept now being highlighted in research on assessments. Plain language has been defined as language that is straightforward and concise. The following strategies for editing text to produce plain language were used during the editing process of the new Keystone Exam items:

- Reduction of excessive length
- Use of common words
- Avoidance of ambiguous words
- Avoidance of irregularly spelled words
- Avoidance of proper names
- Avoidance of inconsistent naming and graphic conventions
- Avoidance of unclear signals about directing attention

- **Maximum Legibility**

Legibility is the physical appearance of text, the way that the shapes of letters and numbers enable people to read text easily. Bias can result when tests contain physical features that interfere with a student’s focus on or understanding of the constructs that test items are assessing. A style guide was developed and utilized that included dimensions of style consistent with Universal Design.

## **GUIDELINES FOR UNIVERSALLY DESIGNED ITEMS**

All test items written and reviewed adhered closely to the following guidelines for Universal Design. Item writers and reviewers used a checklist during the item development process to ensure that each aspect was followed. For more information on the checklist, see the Universal Design Overview section in Chapter Three of this report.

1. **Items measure what they are intended to measure.** Item writing training included making certain that writers and reviewers had a clear understanding of Pennsylvania’s Academic Standards and the Keystone Assessment Anchors. During all phases of test development, items were presented with content-standard information to ensure that each item reflected the intended Assessment Anchor. Careful consideration of the content standards was important in determining which skills involved in responding to an item were extraneous and which were relevant. With certain types of items, an additional skill was necessary, such as the Algebra I test, which requires the student to read.
2. **Items respect the diversity of the assessment population.** To develop items that avoid content that might unfairly advantage or disadvantage any student subgroup, item writers, test developers, and reviewers were trained to write and review items to avoid issues of bias, fairness, and sensitivity. Training also included an awareness of and sensitivity to issues of cultural and regional diversity.

3. **Items have a clear format for text.** Decisions about how items were presented to students must allow for maximum readability for all students. Appropriate fonts and point sizes were employed with minimal use of italics, which are far less legible and are read considerably more slowly than standard typeface. Captions, keys, and legends were at least a 12-point size, while footnotes and sentence numbers used a 10-point font.<sup>1</sup> Legibility was enhanced by sufficient spacing between letters, words, and lines. Blank space was used around paragraphs and between columns and staggered right margins.
4. **Stimuli and items have clear pictures and graphics.** When pictures and graphics were used, they were designed to provide essential information in a clear and uncluttered manner. Illustrations were placed directly next to the information to which they referred, and labels were used when possible. Sufficient contrast between the background and text, with minimal use of shading, increased readability for students with visual impairments. Color was not used to convey important information.
5. **Items have concise and readable text.** Linguistic demands of stimuli and items can interfere with a student's ability to demonstrate knowledge of the construct being assessed. During item writing and review, the following guidelines were used:
  - Simple, clear, commonly used words were used whenever possible.
  - Extraneous text was omitted.
  - Vocabulary and sentence complexity were appropriate for the grade level being assessed.
  - Technical terms and abbreviations were used only if they were related to the content being measured.
  - Definitions and examples were clear and understandable.
  - Idioms were avoided unless idiomatic speech was being assessed.
  - Questions to be answered were clearly identifiable.
6. **Items allow changes to format without changing meaning or difficulty.** A Braille version was available for each operational exam. Attention was given to using items that allow for Braille. Specific accommodations were permitted, such as signing to a student, the use of oral presentation under specified conditions, and the use of various assistive technologies. A Spanish version of the Algebra I and Biology exams was available for use by English Learners who would benefit from this accommodation and who were in U.S. schools for less than three years.
7. **The test has an overall appearance that is clean and organized.** Information was organized in a left-right, top-bottom format. Images, pictures, and text that may not be necessary (e.g., sidebars, overlays, callout boxes, shading, visual crowding caused by excess information) and that could be potentially distracting to students were avoided. Also avoided were purely decorative features that did not serve a purpose.

## ITEM DEVELOPMENT

DRC worked closely with the Pennsylvania Department of Education to ensure that the Keystone Exams complied with nationally recognized principles of Universal Design. The implementation of accommodations on large-scale statewide assessments for students with disabilities was supported in the development of the Keystone Exams. In addition to the principles of Universal Design as described in the Pennsylvania Technical Report, DRC applied to each exam the standards for test accessibility as described in *Tests Access: Making Tests Accessible for Students with Visual Impairments—A Guide for Test Publishers, Test Developers, and State Assessment Personnel* (Allman, 2004).

---

<sup>1</sup> While font size follows specific requirements during online setup of an exam, the screen resolution used at the local level can impact whether the effective font size is visible to the student.

To this end, DRC embraces the following precepts:

- Test directions are worded to allow for alternate responses to constructed-response items.
- During item and bias reviews, committee members are made aware of the Principles of Universal Design and of issues that may adversely affect students with disabilities. The goal is to make certain that the Keystone Exams are bias free for all students. With the goal of ensuring that the Keystone Exams are accessible to the widest range of diverse student populations, PDE instructs DRC to limit item types that are difficult to format in Braille and that may become distorted when published in large print. DRC is instructed to limit the following on the Keystone Exams.
  - Algebra I: Complicated tessellations; charts or graphs that extend beyond one page
  - Literature: Graphics and illustrations that are not germane to the content presented
  - All exams: Unnecessary boxes and framing of text, unless enclosing the text provides necessary context for the student; use of italics (limited to only when it is absolutely necessary, such as with variables)

## ITEM FORMAT

For all Keystone Exams (both online and print), DRC formats the items to maximize accessibility for all students by using text that is in an easily readable size and font style. DRC limits shading, graphics, charts, and the number of items per page so that there is sufficient white space on each page. Whenever possible, DRC ensures that graphics, pictures, diagrams, charts, and tables are positioned on the page with the associated test items. DRC uses high contrast for text and background when possible to convey pertinent information. Tests are published on dull-finish paper to avoid the glare encountered on glossy paper. DRC pays close attention to the binding of the exam books to ensure that they lie flat for two-page viewing and ease of reading and handling.

DRC ensures consistency across Keystone Exams by following these Principles of Universal Design:

- High contrast and clarity is used to convey detailed information.
- Typically, shading is avoided; when necessary for content purposes, 10-percent screens are used as the standard.
- Overlaid print on diagrams, charts, and graphs is avoided.
- Charts, graphs, diagrams, and tables are clearly labeled with titles and with short descriptions when applicable.
- Only relevant information is included in diagrams, pictures, and graphics.
- Symbols used in keys and legends are meaningful and provide reasonable representations of the topics they depict.
- Pictures that require physical measurement are true to size.

## ASSESSMENT ACCOMMODATIONS

While universally designed assessments provide for participation of the widest range of students, many students require accommodations in order to participate in the regular assessment. Clearly, the intent of providing accommodations for students is to make certain that students are not unfairly disadvantaged during testing and that the accommodations used during instruction, if appropriate, are made available as students take the test. The literature related to assessment accommodations is still evolving and often focuses on state policies regulating accommodations rather than on providing empirical data that supports the reliability and validity of the use of accommodations. On a yearly basis, the Pennsylvania Department of Education examines accommodations policies and current research to ensure that valid, acceptable accommodations are available for students. Accommodations manuals for Pennsylvania assessments titled *Accommodations Guidelines* and *Accommodations Guidelines for English Learners* were developed for use with the Keystone Exams. The PDE guideline manuals can be accessed by going to [www.education.pa.gov](http://www.education.pa.gov).

In addition, Spanish-language versions, translated from the original English versions were made available for both the Algebra I Exam and the Biology Exam. The Spanish-translation editions of the exams are discussed in Chapter Six.

# CHAPTER FIVE: EMBEDDED FIELD TEST

## FIELD TEST OVERVIEW

Every Keystone spring administration, field-test items are embedded on operational forms. The main purposes of field testing items prior to future operational use are (a) to calculate item statistics, (b) to determine if items meet the criteria with respect to statistical properties for future operational use, and (c) to obtain item parameters for pre-equating purposes. In comparison to stand-alone field tests, embedded field tests (EFT) allow for more accurate item statistics and item parameters by alleviating concerns of whether students may perceive differences between field-test and operational items. The EFT approach allows item parameters to be used for future pre-equating purposes and is based on the assumption that students should be equally motivated to take the operational and field-test items because they are not aware of which items are field-test items. To minimize item context and item position effects (i.e., lack of motivation and fatigue), field-test items were interspersed within the operational sections. With this design, students have a smaller chance of knowing the field-test item positions.

Keystone test forms contained common operational items that were identical on all forms along with embedded field-test items. In most instances, the field-test items were unique to a given form; however, there were instances in which an embedded field-test item appeared on more than one form. More information on the field-test designs for all subjects can be found in the content-specific portions of Chapter Three. In general, the field-test items for each year represent a portion of the previous year’s operational forms (Spring, Summer, and Winter). For example, items from the field-test in Spring 2018 were selected for the Spring 2019, Summer 2019, and Winter 2019–2020 operational forms. This chapter presents information about the 2023 EFT, including classical item analyses, Differential Item Functioning (DIF) analyses, identification of items for data review, and outcomes from data review.

The number of field-test positions on spring forms varies by subject. Table 5–1 displays the number of field-test positions by subject and module. It is important to note that the number of item positions and the number of field-test items on a spring administration may differ slightly.

**Table 5–1. Number of FT Item Positions by Item Type (MC, CR), Content Area and Module**

Content Area	Module	MC Item Positions per Form	CR Item Positions per Form	MC Item Positions (Total)*	CR Item Positions (Total)*
Algebra I	1	5	1	120	24
Algebra I	2	5	1	120	24
Algebra I	Total	10	2	240	48
Biology	1	8	1	48	6
Biology	2	8	1	48	6
Biology	Total	16	2	96	12
Literature	1	6	1	144	24
Literature	2	6	1	144	24
Literature	Total	12	2	288	48

\*Total FT Item Positions across 6 Biology forms, and 24 Literature and Algebra I forms.

*Note.* Each FT passage is repeated across 2 forms with different sets of FT items to yield the highest volume of eligible items for subsequent forms selection.



## CLASSICAL ITEM ANALYSIS

Classical item analyses of field-test items are conducted in order to assess the items' quality and to identify items for data review. Specifically, item difficulty (also referred to as  $p$ -values, which represent the proportion of students correctly answering the item) and item-total correlations (the relationship between whether a student answers an item correctly and that student's total test score) are estimated for the item, for each option for MC items, and for each score point for multi-point items.

### ITEM DIFFICULTY

At the most general level, an item's difficulty is indicated by its mean score in some specified group (e.g., content area).

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

In the mean score formula above, the individual item scores ( $x_i$ ) are summed and then divided by the total number of students ( $n$ ). For MC items, student scores are represented by 0's and 1's (0 = wrong, 1 = right). With 0/1 scoring, the equation above also represents the number of students correctly answering the item divided by the total number of students. So this is also the *proportion correct* for the item, or as it is better known, the  $p$ -value. In theory,  $p$ -values can range from 0.00<sup>1</sup> to 1.00 on the proportion-correct scale. For example, if an item has a  $p$ -value of 0.89, it means 89 percent of students answered the item correctly. Additionally, this value might suggest that the item is relatively easy or that the students who attempted the item are relatively high achievers. In other words, item difficulty and student ability are somewhat confounded.

For CR items, mean scores can range from the minimum possible score (usually zero) to the maximum possible score (e.g., four points in the case of Algebra I CR items, three points in the case of Biology and Literature CR items). Sometimes a *pseudo p*-value is provided for a CR item by dividing the mean item score by the maximum possible item score.

The minimum and maximum extremes of the difficulty scale are virtually never seen in applied practice. However, understanding what those values are helps illustrate that relatively lower values correspond to more difficult items and that relatively higher values correspond to easier items. (Because of this, some assert that this index would be better referred to as the item's *easiness*.)

Item difficulty is an important consideration for the Keystone Exams because of the various student achievement levels (Below Basic, Basic, Proficient, and Advanced). Items that are either very hard or very easy provide little information about student differences in achievement. However, an item answered correctly by a high percentage of students would suggest that the knowledge or skill the item measures has been mastered by most students. Conversely, an item answered correctly by a low percentage of students would suggest that few students have mastered the knowledge or skill the item measures. So on criteria-referenced tests like the Keystone Exams, a test development goal is to include a wide range of item difficulties.

### ITEM DISCRIMINATION

At the most general level, item discrimination<sup>2</sup> indicates an item's ability to differentiate between high and low achievers. It is expected that students with high ability (i.e., those who perform well on the Keystone Exams overall) would be more likely to answer any given item correctly, while students with low ability (i.e., those who perform poorly on the Keystone Exams overall) would be more likely to answer the same item incorrectly. For the Keystone Exams, Pearson's product-moment correlation coefficient between item scores and test scores is used to indicate discrimination. As commonly practiced, Data Recognition Corporation (DRC) removes the item score from the total score so that the resulting correlations will not be spuriously high. The correlation coefficient can range from negative 1.0 to positive 1.0. If higher-scoring students tend to answer the item correctly while lower-scoring students do not answer the item correctly, the correlation between the item score and the total test score

---

<sup>1</sup> For MC items with four response options, pure random guessing would lead to an expected  $p$ -value of 0.25.

<sup>2</sup> As noted earlier, the discrimination index for dichotomous MC items is typically referred to as the *point-biserial correlation coefficient*. For CR items, the term *item-test correlation* is sometimes used.

will be both positive and noticeably large in its magnitude (i.e., well above zero), indicating that the item is good at discriminating between higher-scoring and lower-scoring students.

Item-total correlation for each option is another indicator of an item’s ability to differentiate between high and low achievers. It is expected that students with high ability would be less likely to choose a distractor, while students with low ability would be more likely to choose a distractor. In other words, the item-total correlations for the distractors are expected to be negative.

In summary, the correlation will be positive in value when the mean test score of the students answering the item correctly is higher than the mean test score of the students answering the item incorrectly.<sup>3</sup> In other words, students who did well on the total test tended to do well on the item as well. However, an interaction can exist between item discrimination and item difficulty. Items answered correctly (or incorrectly) by a large proportion of examinees (i.e., items with extreme  $p$ -values) can have reduced power to discriminate and thus can have lower correlations.

## CLASSICAL ITEM ANALYSIS RESULTS

Table 5–2 provides the summary statistics for the difficulty and discrimination for the 2023 field-test items with respect to each content area and item type (see Chapter Eleven for summary statistics for operational items). There is a range of  $p$ -values across all content areas, where mean  $p$ -values were between 0.49 to 0.64 for MC items and 0.29 to 0.51 for CR items. The mean item-total correlations were between 0.36 to 0.39 for MC items and 0.60 to 0.74 for CR items.

**Table 5–2. Summary Statistics of Difficulty and Discrimination by Content Area and Item Type**

Content Area	Item Type	N	Mean $p$ -value	Min $p$ -val.	Median $p$ -value	Max $p$ -value	Mean I-T Corr.	Min I-T Corr.	Median I-T Corr.	Max I-T Corr.
Algebra I	CR	48	0.29	0.09	0.28	0.68	0.72	0.58	0.72	0.82
Algebra I	MC	240	0.50	0.12	0.49	0.86	0.39	0.05	0.41	0.61
Biology	CR	12	0.41	0.14	0.35	0.71	0.60	0.37	0.62	0.72
Biology	MC	96	0.49	0.21	0.47	0.86	0.36	-0.04	0.39	0.56
Literature	CR	48	0.51	0.42	0.51	0.60	0.74	0.70	0.74	0.78
Literature	MC	288	0.64	0.13	0.64	0.96	0.37	-0.08	0.39	0.56

Note. I-T Corr. is the item-test score correlation.

## DIFFERENTIAL ITEM FUNCTIONING

Differential item functioning (DIF) occurs when examinees who have the same ability level but belong to different groups do not have the same probability of answering an item correctly. When the probability differs, it is important for content experts to review items for any potential *item bias*. It is important to note that, as a statistical concept, DIF is different from item bias. DIF detects a difference in performance after controlling for student ability, whereas bias is a content issue that can arise in situations where something other than the intended construct of measurement affects the probability of a correct response for a particular group. For example, bias is present when an item presents negative group stereotypes that draw the attention of the examinee, uses non-construct relevant language that is more familiar to one subpopulation than to another, or is presented in a non-construct relevant format that disadvantages certain learning styles. While the source of item bias can be plain to trained judges, DIF may have no clear cause. In such cases, something other than bias, including construct relevant content, may be explaining the differential performance on the item. Flagging items with DIF provides the opportunity for reviewers to assess and correct potential bias, but DIF does not necessarily mean that bias is present.

## DIF DETECTION PROCEDURES

For MC items, the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959) for detecting DIF is a commonly used technique. It does not depend on the application or the fit of any specific measurement model. However, it does have significant philosophical overlap with the Rasch model since it uses a test’s total score for the analysis.

<sup>3</sup> It is legitimate to view the point-biserial correlation as a standardized mean. A positive value indicates that students who chose that response had a higher mean score than the average score; a negative value indicates that students who chose that response had a lower mean score than the average score.

The MH procedure as implemented by DRC contrasts a focal group with a reference group. While it makes no practical difference in the analysis which group is defined as the focal group, the group most likely to be disadvantaged by a biased measurement is typically defined as the focal group. In these analyses, the focal group was female for gender-based DIF, black and Hispanic for ethnicity-based DIF, and the computer-based-test (CBT) group for the test administration mode-based DIF; reference groups were male, white, and the paper-and-pencil test (PPT) group for the gender, ethnicity, and mode-based DIFs, respectively. The MH statistic for each item is computed from a contingency table. The table has two groups (focal and reference) and two outcomes (right or wrong). The ability groups are defined by the total score distribution for the total examinee populations.

The basic MH statistic is a single degree of freedom chi-square that compares the observed number in each cell to the expected number. The expected counts are computed to ensure that the analysis is not confounded with differences in the achievement level of the two groups.

For CR items, a comparable statistic is computed based on the standardized mean difference (SMD) (Dorans, Schmitt, & Bleistein, 1992), computed as the differences in mean scores for the focal and reference groups if both groups had the same score distribution.

To assist the review committees in interpreting the analyses, all items are assigned a severity code (A, B, or C) based on the magnitude of the MH statistic and a direction (minus or plus) based on the direction of the MH statistic. Items classified as A+ or A- have little or no statistical indication of DIF. Items classified as B+ or B- have some indication of DIF but may be judged to be acceptable for future use. Items classified as C+ or C- have strong evidence of DIF and should be reviewed and possibly rejected from the eligible item pool. A plus sign indicates that the item favors the focal group, and a minus sign indicates that the item favors the reference group.

## LIMITATIONS OF STATISTICAL DETECTION

No statistical procedure should be used as a substitute for rigorous, hands-on reviews by content and bias specialists. The statistical results can help organize reviews so the effort is concentrated on the most problematic cases. Further, no items should be automatically rejected simply because a statistical method flagged them or automatically accepted because the statistical method did not flag them.

Statistical detection of DIF is not an exact science. There have been a variety of methods proposed for detecting DIF, but no single statistic can be considered either necessary or sufficient. Different methods are more or less successful and detect DIF at different rates. No analysis can guarantee that a test is free of bias, but thoughtful item development and field-test analysis can prevent potentially biased items from unfairly impacting student scores.

A fundamental shortcoming of all statistical methods used in DIF evaluation is that they are all intrinsic to the test being evaluated. If a test is unbiased overall but contains one or two DIF items, any method can identify DIF. However, because all current methods use total test performance as the measure on which to control for group abilities, a test with all DIF items will not be able to separate DIF effects from true differences in achievement on the test.

## CRITERIA FOR IDENTIFYING ITEMS

As previously discussed, all field-test items were analyzed statistically using conventional item analysis methods. For MC items, classical item statistics included the corrected point-biserial correlation (Pt. Bis.) for the correct and incorrect responses (distractors), the percentage correct ( $p$ -value), and the percentage incorrect. For constructed-response (CR) items, the statistical indices included the item-test correlation, the point-biserial correlation for each score point, the percentage of responses in each score point, and the percentage of non-scorable responses.

In general, more capable students are expected to respond correctly to easy items and less capable students are expected to respond incorrectly to difficult items. If either of these situations does not occur, the item in question will be reviewed by DRC test development staff and committees of Pennsylvania educators to determine the nature of the problem and the characteristics of the students affected. The primary ways of detecting such conditions are through the point-biserial correlation coefficient for dichotomously scored MC items and the item-total correlation for polytomously scored CR items. In each case the statistic will be positive if the total test mean score is higher for the students who respond correctly to MC items (or attain a higher CR item score) and negative when the reverse is true.

The following set of criteria was used to identify items for additional review.

For an MC item to be flagged, the item needed to meet any of the following criteria:

1. Percentage Correct ( $p$ -value) less than 0.3 or greater than 0.9
2. Item-total correlation for the correct response less than 0.25
3. Point-biserial correlation for any incorrect response greater than 0.0
4. Percentage responding to any incorrect responses greater than the percentage correct
5. Gender, ethnic, or mode DIF code of C+ or C-

For a CR item to be flagged, the item needed to meet any of the following criteria:

1. Percentage Correct ( $p$ -value) less than 0.3 or greater than 0.9
2. Item-total correlation for the correct response less than 0.25
3. Score proportion less than 0.05
4. Gender, ethnic, or mode DIF code of C+ or C-

The intent of the above criteria is to flag items for review. For this purpose, the preference is to over-identify rather than under-identify items to ensure they are appropriate for future use. Any of the criteria would cause the item to be reviewed by content experts, but there are many reasons the experts may want to keep an item in spite of the statistics.

## RESULTS AND OBSERVATIONS

Each form has a set of operational items along with embedded FT items. Test forms are spiraled to ensure that a representative sample of students answer each set of FT items (see Chapter Nine for additional information about spiraling). For MC items, all data is used for classical item analyses and DIF analyses. In contrast, a random sample of responses are selected to be scored for FT analyses for CR items.

This section focuses on reporting the number ( $N$ ) and percentage (%) of items flagged by different criteria (see Tables 5–3 to 5–5). For the DIF analysis, the number and percentage of items were provided not only for the C- and C+ bias codes, which were used as the criteria to flag items, but also for the bias codes A-, A+, B-, and B+.

**Table 5–3. Summary of Items Flagged by the Classical Item Statistics**

Item Type	Flagging Criterion	Alg. I Total N	Alg. I N	Alg. I %	Bio. Total N	Bio. N	Bio. %	Lit. Total N	Lit. N	Lit. %
MC	1 (Low $P$ -value)	240	19	7.9	96	6	6.3	288	11	3.8
MC	2 (Low Point-Biserial Correlation)	240	25	10.4	96	15	15.6	288	45	15.6
MC	3 (Non-negative distractor correlation)	240	31	12.9	96	13	13.5	288	35	12.2
MC	4 (High distractor proportion)	240	28	11.7	96	11	11.5	288	18	6.3
CR	1 (Low $P$ -value)	48	26	54.2	12	2	16.7	48	0	0.0
CR	2 (Low Point-Biserial Correlation)	48	0	0.0	12	0	0.0	48	0	0.0
CR	3 (Low score-point proportion)	48	25	52.1	12	2	16.7	48	2	4.2

*Note.* Refer to Criteria for Identifying Items.

**Table 5–4. DIF Summary – MC Items**

Reference Group	Focal Group	Bias Code	Alg. I (Total N=240) N	Alg. I (Total N=240) %	Biology (Total N=96) N	Biology (Total N=96) %	Literature (Total N=288) N	Literature (Total N=288) %
Male	Female	A-	105	43.8	42	43.8	168	58.3
Male	Female	A+	123	51.3	53	55.2	91	31.6
Male	Female	B-	9	3.8	1	1.0	20	6.9
Male	Female	B+	1	0.4	0	0.0	1	0.3
Male	Female	C-	2	0.8	0	0.0	8	2.8
Male	Female	C+	0	0.0	0	0.0	0	0.0
White	Black	A-	151	62.9	51	53.1	178	61.8
White	Black	A+	79	32.9	45	46.9	90	31.3
White	Black	B-	10	4.2	0	0.0	15	5.2
White	Black	B+	0	0.0	0	0.0	1	0.3
White	Black	C-	0	0.0	0	0.0	4	1.4
White	Black	C+	0	0.0	0	0.0	0	0.0
White	Hispanic	A-	128	53.3	51	53.1	186	64.6
White	Hispanic	A+	109	45.4	45	46.9	79	27.4
White	Hispanic	B-	3	1.3	0	0.0	18	6.3
White	Hispanic	B+	0	0.0	0	0.0	0	0.0
White	Hispanic	C-	0	0.0	0	0.0	5	1.7
White	Hispanic	C+	0	0.0	0	0.0	0	0.0
PPT	CBT	A-	85	35.4	36	37.5	102	35.4
PPT	CBT	A+	155	64.6	60	62.5	177	61.5
PPT	CBT	B-	0	0.0	0	0.0	2	0.7
PPT	CBT	B+	0	0.0	0	0.0	7	2.4
PPT	CBT	C-	0	0.0	0	0.0	0	0.0
PPT	CBT	C+	0	0.0	0	0.0	0	0.0

**Table 5–5. DIF Summary – CR Items**

Reference Group	Focal Group	Bias Code	Alg. I	Alg. I	Biology	Biology	Literature	Literature
			(Total N=48) N	(Total N=48) %	(Total N=12) N	(Total N=12) %	(Total N=48) N	(Total N=48) %
Male	Female	A-	16	33.3	3	25.0	0	0.0
Male	Female	A+	32	66.7	9	75.0	37	77.1
Male	Female	B-	0	0.0	0	0.0	0	0.0
Male	Female	B+	0	0.0	0	0.0	11	22.9
Male	Female	C-	0	0.0	0	0.0	0	0.0
Male	Female	C+	0	0.0	0	0.0	0	0.0
White	Black	A-	39	81.3	7	58.3	33	68.8
White	Black	A+	9	18.8	1	8.3	15	31.3
White	Black	B-	0	0.0	3	25.0	0	0.0
White	Black	B+	0	0.0	0	0.0	0	0.0
White	Black	C-	0	0.0	1	8.3	0	0.0
White	Black	C+	0	0.0	0	0.0	0	0.0
White	Hispanic	A-	40	83.3	10	83.3	32	66.7
White	Hispanic	A+	7	14.6	2	16.7	16	33.3
White	Hispanic	B-	1	2.1	0	0.0	0	0.0
White	Hispanic	B+	0	0.0	0	0.0	0	0.0
White	Hispanic	C-	0	0.0	0	0.0	0	0.0
White	Hispanic	C+	0	0.0	0	0.0	0	0.0
PPT	CBT	A-	33	68.8	6	50.0	23	47.9
PPT	CBT	A+	15	31.3	6	50.0	25	52.1
PPT	CBT	B-	0	0.0	0	0.0	0	0.0
PPT	CBT	B+	0	0.0	0	0.0	0	0.0
PPT	CBT	C-	0	0.0	0	0.0	0	0.0
PPT	CBT	C+	0	0.0	0	0.0	0	0.0

## REVIEW OF ITEMS WITH DATA

In the preceding section, it was stated that test development content-area specialists used certain statistics from classical item analyses and DIF analyses of the 2022 field test to identify items for review by Pennsylvania educators. Items not identified for this review had good statistical characteristics and, consequently, were entered into the eligible pool for future item selection. DRC content-area test development specialists and DRC psychometric specialists identified the remaining items for further review by a committee of Pennsylvania educators. The intent was to capture all items that needed a closer review; thus, the criteria employed tended to over-identify rather than under-identify items.

The review of the relevant items was conducted by more than 30 Pennsylvania educators (teachers and PDE staff) broken out into exam-based committees. The review took place on September 7 and 8, 2022. In these sessions, committee members were first trained by a representative from DRC’s psychometrics staff with regard to the statistical indices used in item evaluation. This training was followed by a discussion with examples concerning reasons an item might be retained regardless of the statistics. The committee review process involved a brief exploration of possible reasons for the statistical profile of an item (e.g., possible bias, grade appropriateness, instructional issues) and a decision regarding acceptance. DRC content-area test development specialists facilitated the review of the items. Each committee reviewed the pool of flagged EFT items and made recommendations on each item. The results of the committee reviews are shown in the table below.

**Table 5–6. Data Review Committee Results**

Exam	Module	Total No. of FT Items	No. Reviewed Items	% of Reviewed Items	No. of Items Rejected by Committee	% of Items Rejected by Committee	No. of Items Classified as Rejected*	% of Items Classified as Rejected*
Algebra I	1	144	53	36.8%	0	0.0%	1	0.7%
Algebra I	2	144	59	41.0%	0	0.0%	0	0.0%
Biology	1	180	57	31.7%	14	7.8%	17	9.4%
Biology	2	180	46	25.6%	6	3.3%	10	5.6%
Literature	1	168	42	25.0%	4	2.4%	4	2.4%
Literature	2	168	35	20.8%	3	1.8%	3	1.8%
	Total	984	292	29.7%	27	2.7%	35	3.6%

\* Items Classified as “Rejected” from Spring 2023 EFT (all sources: Data Review Committee, PDE, and DRC)

## CHAPTER SIX: OPERATIONAL FORMS CONSTRUCTION FOR 2023 ADMINISTRATIONS

---

### FINAL SELECTION OF ITEMS AND KEYSTONE FORMS CONSTRUCTION

Approximately 50% of the items that made up the Spring 2023, Summer 2023, and Winter 2023/2024 operational forms emerged from the Spring 2022 embedded field test. The remaining operational (core) items were part of the biennial core-to-core overlap. For more information about the core-to-core overlapping items, please see Chapter Three. Prior to being placed on the operational tests, these items had undergone multiple reviews, including the following:

- Reviews by Data Recognition Corporation (DRC) content-area test development specialists and curriculum specialists to ensure that all items were properly aligned with content standards
- Formal bias, fairness, and sensitivity review by the Bias, Fairness, and Sensitivity Committee, which consisted of a multiethnic group of men and women having expertise with special-needs students and English Learners (EL)
- Formal review by the content committees consisting of Pennsylvania educators, including teachers as well as district personnel
- Pennsylvania Department of Education (PDE) review
- Item data review by members of the PDE subject-area teacher committees

The item and bias reviews are detailed in Chapter Three. The results of the data review are summarized in Chapter Five.

The end product of the above process was an item status designation for each field-test item. All items having an item status code of Accepted/Operational Ready were candidates to be selected for the 2022 Keystone Exams. To have an item status code of Accepted/Operational Ready meant that the item met the following criteria:

- Appropriately aligned with its designated Keystone Assessment Anchor Content Standard (Assessment Anchor) and subclassifications
- Acceptable in terms of bias/fairness/sensitivity issues, including differential item functioning (for gender and ethnicity)
- Acceptable in terms of psychometric standards, including a special review of flagged items

Next, all relevant information regarding the acceptable items, including associated graphics, was entered into the item banking system known as IDEAS (Item Development and Education Assessment System). From IDEAS and other database sources, Microsoft Excel files were created for each exam. These files contained all relevant content codes and statistical characteristics. IDEAS also created an item card displaying each acceptable item, any associated graphic, and all relevant exam codes and item statistics for use by the subject-area test development specialists and psychometric services staff.

DRC test development specialists reviewed the test design blueprint, including the number of items per strand for each content-area test. Psychometricians provided content-area test development specialists with an overview of the psychometric guidelines for forms construction.

Senior DRC content-area test development specialists reviewed all items in the operational pool to make an initial selection (pull) for common (core) positions according to test blueprint requirements and psychometric guidelines. Changes to items were not encouraged since alterations could affect how an item would perform in subsequent testing.



For these common items, this meant that the combination of multiple-choice (MC) and constructed-response (CR) items would yield the appropriate range of points while tapping an appropriate variety of the Assessment Anchors and related Eligible Content within each Reporting Category (module). Items selected in the first round were examined with regard to how well they fit together as a set. Of particular concern were the following:

- One item providing cues as to the correct answer to another item
- Context redundancy (e.g., mathematics items with a sports context)
- Presence of clang (distractors not unique from one another)
- Diversity of names and artwork for gender and ethnicity

A core-building software tool known as PerForm was used in concert with performance data and metadata from IDEAS to aid in the organization and communication of the pulled data. PerForm automatically tabulates the statistical characteristics of the proposed core, updating instantly whenever item swaps were performed. Using PerForm, the first round of items was then evaluated for statistical features such as an acceptable point-biserial correlation and whether correct answers were distributed equally—that is, whether approximately 25 percent of correct answers appeared in each of the four possible positions (A, B, C, or D). Selected items that were deemed psychometrically less advantageous in contrast to the overall psychometric characteristics of the core resulted in a search by the senior reviewer for suitable replacements. At this point, the second round of items was analyzed. If necessary, this iterative process between content-based selections and statistical properties continued in an effort to reach the best possible balance.

Once the recommendations were finalized for the core items, they were submitted to PDE for review. Department staff provided feedback, which could be in the form of approval or recommendations for replacement of certain items. Any item replacement was accomplished by the collective effort of the test development specialists, psychometricians, and PDE staff until final PDE approval was given. See Appendix F for the Keystone Exams Tally Sheets.

Following final approval by PDE, test development specialists developed print and online forms based on the approved core and approved embedded field-test items. Both modes of delivery were built using IDEAS. Highly skilled test development specialists and editors used specialized checklists to verify accuracy of layout and formatting in both modes of delivery. Following final approval to print, the documents were prepared for the printing presses.

## **SPECIAL FORMS USED WITH THE OPERATIONAL 2023 KEYSTONE EXAMS**

### **SPANISH TRANSLATION**

Starting with the operational exams in spring 2011, school personnel had the option of allowing Spanish-speaking students who had been enrolled in schools in the United States for less than three years to respond to a Spanish version of the Keystone Exams for Algebra I and Biology. The original translation of the items and the *Directions for Administration Manual* was initiated by Tri-Lin Integrated Services Inc. and completed/verified by Exact Communications. These companies use translators with varying cultural and regional backgrounds to create the Spanish versions. The translations were then reviewed and verified by DRC's internal Spanish group. As part of the internal review, a Spanish style guide is maintained to document Spanish word choice from administration-to-administration and across exams within an administration.

Following PDE's approval of the translation, the translated text was typeset into print delivery forms. The test book is constructed with a side-by-side format with the English text and Spanish-translated text on facing pages. The Spanish-translated text is on the left-hand side followed by the original English text on the right-hand (facing) side. Each CR item covered either two or four pages in the answer book, depending on the length of the original English-language item. In the case of four-page open-ended items, the first set of facing pages of an item was presented in Spanish. The second set of facing pages of an item was presented in the original English.

Those students using this accommodated version are permitted to write their answers on either the English language pages or on the translated Spanish language pages. Their answers can be written in English, Spanish, or a combination of both Spanish and English because all pages are evaluated and scored, and the highest possible scores from those combinations are recorded for the students.

On a yearly basis, the PDE examines accommodations policies and current research to ensure that valid, acceptable accommodations are available for students. Accommodations manuals for Pennsylvania assessments titled *Accommodations Guidelines* and *Accommodations Guidelines for English Learners (ELs)* were developed for use with the Keystone Exams. The PDE guideline manuals can be accessed by going to [www.education.pa.gov](http://www.education.pa.gov). For more information about the general on-screen testing aids available to students taking the online mode of delivery, see Chapter Two.

## **AUDIO**

For students requiring an auditory presentation accommodation, a text-to-speech synthesizer is available to students taking the Algebra I and/or Biology Exams using the online mode of test delivery. For each operational exam, one form was selected for the creation of the audio version. Special scripts are crafted, writing out each item, distractor, graphic, and directions to utilize the rich, synthesized voice features while accounting for specific nuances of the intended sounds. The resulting audio information is provided to students receiving the accommodation. Since additional software is required to generate the vocalization from the scripted text and since headphones are required to minimize disruptions within a computer lab setting, local school personnel generally must preplan to use the audio version in order to ensure that the student has a properly equipped computer and a proper setting.

## **BRILLE, LARGE PRINT, AND VIDEO SIGN LANGUAGE**

Students were able to respond to test materials that were available in Braille, large print, or Video Sign Language. At each grade level assessed, one form was selected for the creation of these accommodations.

The large print edition is a replication of the standard print form; 8.5×11 standard form is enlarged to an 11×17 page format to achieve a font size of approximately 18-point. A side-by-side verification is completed between the standard print and large print forms to ensure that the integrity of all formatting and graphics is maintained on the large print forms.

For Braille production, the final selected form is delivered to American Printing House for the Blind (APH) via APH's secure website. APH ensures that all tests are translated correctly and accurately by using a translator and a validator. After all Braille booklets are printed, APH conducts a quality assurance step to ensure all items are bound in order and directions are included. All Braille booklets are shipped from APH to DRC via UPS.

DRC applies a security barcode to each large print and Braille booklet for purposes of shipping, distributing, and collecting the materials. This security barcode is used with DRC's Operations Materials Management System (Ops MMS).

School personnel were directed to transcribe all student answers (SR and CR) into scannable answer documents exactly as the student responded. No alterations or corrections of student work were permitted, and the transcribed answer document had to have the same form designation as the Braille and large print version.

DRC utilizes Wolfgang Productions for the production of Sign Language Videos. The items are passed to Wolfgang Productions via a secure ftp site. Two to three different interpreters are used to interpret and validate the translations during video recording. After the interpretations are recorded and returned to DRC via a secure ftp site, DRC loads these videos in the online test engine. When school personnel assign the specific sign language accommodation, the student will be able to play each video next to the item.

## CHAPTER SEVEN: TEST ADMINISTRATION PROCEDURES

### SECTIONS, SESSIONS, TIMING, AND LAYOUT OF THE KEYSTONE EXAMS

The design for most Keystone Exams utilizes separate test books and answer books. An answer book is used to respond to the multiple-choice (MC), evidence-based selected-response (EBSR), and constructed-response (CR) items and to collect demographic information. The MC items and all stimulus text are placed within the test book. One exam uses a single consumable book. When a single scannable answer book is utilized, the contents of the answer book and the test book are combined into one integrated book. The table below identifies the exam material format for each 2023 Keystone Exam.

**Table 7–1. Book Type by Exam**

Exam	Test Book	Answer Book	Single Consumable Book
Algebra I	✓	✓	
Biology	✓	✓	
Literature	✓	✓	

Generally, a separate test book and answer book are used to separate the MC items and the CR items. For passage-based exams, like Literature, the separate exam materials allow the students to reference stimulus materials at the same time that a response to a CR item is composed. In addition, since all student responses must be scanned for scoring and storage purposes, a separate answer book limits the volume of data that must be stored.

### SECTIONS AND SESSIONS

Each operational Keystone Exam is organized around two equally sized test modules; the focus of each is on two or more specific, thematically linked Assessment Anchors and Eligible Content. The content in each module remains separate, and items measuring the Eligible Content in a module appear only in that module. The module design is identical in the print (paper-and-pencil) and online modes of delivery.

Each exam section is administered in an exam session. Local districts must schedule the two modules as two separate exam sessions (morning and afternoon or two separate days), and an individual module must be completed in one exam session.

Each test session is to be completed within a prescribed testing window. The testing windows below reflect both online and paper-based administrations in the 2022–2023 school year. The testing windows also include all make-up testing. Schools were able to choose one of the two testing windows (“waves”) for the winter administration. Two windows were provided to accommodate different semester end dates for schools with block scheduling.

**Table 7–2. Winter 2022/2023 Operational Keystone Exam Testing Windows**

Exams	Wave 1 Dates	Wave 2 Dates
Algebra I, Biology, Literature	December 5–16, 2022	January 4–18, 2023

**Table 7–3. Spring 2023 Operational Keystone Exam Testing Window**

Exams	Dates
Algebra I, Biology, Literature	May 15–26, 2023

**Table 7–4. Summer 2023 Operational Keystone Exam Testing Window**

Exams	Dates
Algebra I, Biology, Literature	July 24–28, 2023

## TIMING

In general, the estimated testing times allow 1–2 minutes per MC item on the Keystone Exams, depending on the exam. The CR items are estimated to take approximately 5–10 minutes per item, also depending on the exam. Each stimulus passage on the Literature exam is estimated to take about 10 minutes to read. There was no difference in the timing for online and print forms of delivery.

Test administrators were instructed that each section (module) in a form should be scheduled as a separate exam session. Exam modules were not to have been scheduled back-to-back in the morning (or in the afternoon). Instead, the exam modules were to be divided across two days or divided across the morning and afternoon of the same day.

Since not all students are expected to finish the exam sections at the same time, test administrators are advised to use the flexibility of the time limits to the students' advantage. For example, test administrators manage the testing time so that students do not feel rushed while they are taking any assessment section, and no student is penalized because he or she works slowly. It is also stressed to test administrators that a student should not be given an opportunity to waste time. Students are told to close their exam materials when they have finished the section of the exam in which they have been working. Students who finish early are allowed to sit quietly or read for pleasure until all students have finished. Students with special requirements and/or abilities (i.e., physical, visual, auditory, or learning disabilities as defined by their IEP or service contracts) and students who just work slowly may require extended time. Special assessment situations are arranged for these students. When all students in a testing session indicate that they have finished an exam section, test administrators end the section.

Scheduled extended time is provided by a test administrator, and students are allowed to request extended time if they indicate that they have not completed the task. Such requests are granted if the test administrator finds the request to be educationally valid. Test administrators are advised that not permitting ample time for students to complete the assessment might impact the students' and school's performances.

As a general guideline, however, when all students indicate that they have finished a section, that section is closed. Students requiring time beyond the majority of the student population are allowed to continue immediately following the regularly scheduled session in another setting. When such accommodations are made, school personnel ensure that students are monitored at all times to prevent sharing of information. Students are not permitted to continue a section of the assessment after a significant lapse of time from the original session.

**Table 7–5. Keystone Testing Load and Duration by Exam**

<b>Exam</b>	<b>Total No. of MC Items per Form per Administration</b>	<b>Total No. of CR Items per Form per Administration</b>	<b>Total Estimated Testing Time per Form (in minutes)</b>	<b>Total Estimated Administration Time per Form (in minutes)</b>
Algebra I	46	8	150	170–180
Biology	64	8	144	164–174
Literature	46	8	146	166–176

**Table 7–6. Keystone Testing Load and Duration by Type per Unit (in minutes)**

Exam	Administration Tasks	Stimulus Passages	MC Points per Minute [PPM]*	CR [PPM]*	Estimated Overall PPM**
Algebra I	24	–	1.5 [0.670]	10 [0.400]	0.400
Biology	24	–	1.25 [0.800]	8 [0.375]	0.458
Literature	24	–	1 [1.000]	5 [0.600]	0.356

\*Based on rates per item type

\*\*Based on total testing time

Prior to beginning the exam, students were asked to verify that they understood the *Code of Conduct for Test Takers* by marking the circle in the exam. Additionally, an Attention statement was added to the beginning of the exams to notify students of the penalties incurred if exam materials are copied.

## LAYOUT

The layout of the operational Keystone Exams follows a general sequence regardless of the exam. Each exam is divided into thematically linked sets of content called modules. Within each module, there are core (common) items and field-test items. Both core and field-test items are represented through MC and CR items.

Stimulus material (like passages), text for MC items, answer options, and any stimulus materials associated with MC items or answer options appear in the test book. Answer bubbles, text for CR items, and associated response spaces appear in the answer book.

Within a non-passage-based module (like Algebra I and Biology), the sequencing of items follows this pattern:

- 1st: Approximately half of the MC items
- 2nd: Half of the CR items
- 3rd: Remaining half of the MC items
- 4th: Remaining CR items

Within a passage-based module (like Literature), the sequencing of items follows this pattern:

- 1st: Stimulus Passage X
- 2nd: MC items associated with Passage X
- 3rd: CR items associated with Passage X
- 4th: Stimulus Passage Y
- 5th: MC items associated with Passage Y
- 6th: CR items associated with Passage Y
- 7th: Stimulus Passage Z
- 8th: MC items associated with Passage Z
- 9th: CR items associated with Passage Z

Regardless of sequencing pattern, the field-test items appear in the relative middle of each module, and item sequencing is self-contained within a module.

For more information about the test layout of the operational Keystone Exams, see Appendix G.

## SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS

Participating sites receive the *Handbook for Assessment Coordinators*, the *Directions for Administration Manuals*, the administrative materials (e.g., Return Shipping labels, District/School labels, Do Not Score labels, Student Precode labels) and secure materials (e.g., consumable test/answer books) for each tested subject in Algebra I, Biology, and Literature Keystone Exams. All materials arrive at least two weeks prior to the start of the testing window.

DRC ensured that all exam materials were assembled correctly prior to shipping. DRC operations staff used the automated Operations Materials Management System (Ops MMS) to assign secure materials to a school at the time of ship out. This system used barcode technology to provide an automated quality check between items requested for a site and items shipped to a site. A shipment box manifest was produced for and placed in each box shipped. DRC operations staff double-checked all box contents with the box manifest prior to sealing the box for shipping to ensure accurate delivery of materials. DRC operations staff performed lot acceptance sampling on both shipments. Districts and schools were selected at random and examined for correct and complete packaging and labeling. This sampling represented a minimum of 10 percent of all shipping sites.

DRC's materials management system, along with the systems of shippers, allowed DRC to track materials from DRC's warehouse facility to receipt at the district, school, or testing site. All DRC shipping facilities, materials processing facilities, and storage facilities are secure. Access is restricted by security code. Non-DRC personnel are escorted by a DRC employee at all times. Only DRC inventory control personnel have access to stored secure materials. DRC employees are trained in and made aware of the high level of security that is required.

DRC used United Parcel Service (UPS) to deliver the secure materials to the testing sites.

## ONLINE TESTING

Online administration is managed through the DRC INSIGHT Portal that provides tiered, secure access to all required administrative functions. Within the DRC INSIGHT Portal, users manage student information and create test sessions.

Student information from the Pennsylvania Information Management System (PIMS) is imported into the DRC INSIGHT Portal Test Setup application via file transfer. If a record was not transferred via the PIMS file, LEAs also have the opportunity to upload a student(s) directly into the DRC INSIGHT Portal so the student can be included in a test session.

Once the student data is loaded into Test Setup, users organize students into test sessions. Test sessions can be created by class, grade, or school. Through Test Setup, users can also update student accommodation information, print test tickets, and monitor student testing status.

The student login ticket contains unique login credentials used by the student to access the testing software. For a selected test session, users can download and print a PDF document containing instructions, a roster of student tickets being printed, and the actual test tickets. Student test tickets are considered secure materials and LEAs are required to keep printed tickets in a predetermined, locked, secure storage area.

The web-based test engine, DRC INSIGHT Online Learning System, is downloaded onto computers that students will access during the assessment. Test items and forms can only be accessed using a valid test ticket. During testing, responses are sent to a DRC server each time the student navigates away from an item or clicks the *Next* button to submit an answer. The system is configured to allow students to review answers before submitting their test.

## TEST SECURITY MEASURES

Test security is essential to obtaining reliable and valid scores for accountability purposes. Test Security Certifications were required to be signed by each building Principal, School Assessment Coordinator, District Assessment Coordinator, Test Administrator, and Proctor after the assessment is administered. All signed Certifications were returned to the Chief School Administrator who must retain the Certifications for three years. The purpose of the Certifications was to serve as a tool to document that the individuals responsible for administering the assessments both understood and acknowledged the importance of test security and accountability. Additional details can be found in the *Handbook for Assessment Coordinators*. A screen shot of the Test Administrator Certificate is provided in Figure 7–1.

**Figure 7–1. Test Administrator and Proctor Test Security Certification**

### **Keystone Exam Test Security Certification Form**

(Test Administrator and Proctor)

District: \_\_\_\_\_

School: \_\_\_\_\_

AUN: \_\_\_\_\_

Maintaining the security and integrity of all assessment materials, preventing any dishonest or fraudulent behavior in the administration and handling of the assessment, and promoting a fair and equitable testing environment are essential in order to obtain reliable and valid student scores. In that regard, I certify the following:

Prior to the administration of the assessment, I completed the Pennsylvania State Test Administration Training, and I understand that the assessment materials are secure, confidential, and proprietary documents owned by the Pennsylvania Department of Education.

I have not reviewed, discussed, disseminated, described, or otherwise revealed the contents of the assessment to anyone. I have not removed any assessment materials from the school building unless I was specifically authorized to administer the assessment to a student on homebound instruction. I have not kept, copied, reproduced, released, or used any assessment, assessment question, specific assessment content, or examinee response to any item or any section of the secure assessment in any manner that is inconsistent with the instructions provided by or through the Pennsylvania Department of Education. I have not provided any examinee with an answer to an assessment question or in any way influenced an examinee's response to any assessment question. I have not in any manner altered or caused the alteration of any examinee response, assessment booklet, or papers used by examinees.

I understand that any breach in assessment security could result in the invalidation of assessment results, professional discipline, and/or criminal prosecution.

I understand that false statements herein are made subject to the penalties of 18 Pa.C.S. § 4904.

\_\_\_\_\_  
Administrator/Proctor Name

\_\_\_\_\_  
Administrator/Proctor Signature

\_\_\_\_\_  
Date of Signature

## **SAMPLE MANUALS**

Copies of the *Handbook for Assessment Coordinators* and the *Directions for Administration Manuals* can be found on the PDE website at [www.education.pa.gov](http://www.education.pa.gov).

## **TESTING WINDOW ASSESSMENT ACCOMMODATIONS**

PDE develops an *Accommodations Guidelines* handbook for use with the Keystone Exams. This manual can be found on the PDE website at [www.education.pa.gov](http://www.education.pa.gov). Additional information regarding assessment accommodations can be found in Chapter Four and Six of this report.



## CHAPTER EIGHT: PROCESSING AND SCORING

---

### RECEIPT OF MATERIALS

Receipt of Pennsylvania Keystone Exams' test materials began five days after the start of the test window. DRC's Operations Materials Management System (Ops MMS) was utilized to receive assessment materials securely, accurately, and efficiently. This system features innovative automation and advanced barcode scanners. Captured data was organized into reports, which provided timely information with respect to suspected missing material.

The first step in Ops MMS was Box Receipt. When a shipment arrived at DRC, the boxes were removed from the carrier's truck and passed under a barcode reader, which read the barcode printed on the return label and identified the district and school. The number of boxes was immediately compared to what was picked up at the district. The data collected in this process was stored in the Ops MMS database. After the barcode data was captured, the boxes were placed on a pallet and assigned a corresponding pallet number.

Once the Box Receipt process was completed, the Materials Separation phase began. Warehouse personnel opened the boxes and sorted materials by grade, subject, and status (used and/or unused booklets) into scanning boxes. Every booklets' security barcode and precode barcode were hand scanned to link each document to the original box. As the booklets were sorted, Ops MMS guided the floor operator to the box in which to place the document. Ops MMS kept count and record of the materials placed in each box. This count remained correlated to the box as an essential quality-control step throughout the secure booklet processing and provided a target number for all steps of the check-in process. Once a box was closed, an MMS Processing Label was placed on that box.

Once labeled, the sorted and counted boxes proceeded to Quality Assurance, where a secure booklet check-in operator used a hand scanner to scan the MMS Processing Label. This procedure identified the material type and quantity parameters for what Ops MMS should expect within a box. The box contents were then loaded into the stream feeder.

The documents were fed past oscillating scanners that captured both the security code and precode from the booklets. A human operator monitored an Ops MMS screen that displayed scan errors, an ordered accounting of what was successfully scanned, and the document count for each box. The system ensured that all material within the box matched the information obtained from the original hand-scanning process.

When all materials were scanned and the correct document count was confirmed, the box was sealed and placed on a pallet. If the correct document count was not confirmed, or if the operator encountered difficulties with material scanning, the box and its contents were delivered to an exception-handling station for resolution.

This check-in process occurred immediately upon receipt of materials; therefore, DRC provided feedback to districts and schools regarding any missing materials based on actual receipt versus expected receipt. Sites that had 100 percent of their materials missing after the date they were due to DRC were contacted, and any issues were resolved.

Throughout the process of secure booklet check-in, DRC project management ran a daily Missing Materials Report. Every site that was missing any number of booklets was contacted by DRC. Results of these correspondences were recorded for inclusion in the final Missing Materials Report if the missing booklets were not returned by the testing site. DRC produced the Missing Materials Report for PDE upon completion of secure booklet check-in. The report listed all schools in each participating district, along with security barcodes for any booklets not returned to DRC.

After scannable materials (used answer booklets) were processed through booklet check-in, the materials became available to the DRC Document Processing log-in staff for document log-in. The booklets were logged in using the following process:

- A DRC scannable barcode batch header was scanned, and a batch number was assigned to each box of booklets.
- The DRC box label barcode was scanned into the system to link the box and booklets to the newly created batch and to create a Batch Control Sheet.
- The DRC box label barcode number and the number of booklets in the box were printed on the Batch Control Sheet for document-tracking purposes. All booklets linked to the box barcode were assigned to the batch number and tracked through all processing steps. As booklets were processed, DRC staff dated and initialed the Batch Control Sheet to indicate that proper processing and controls were observed.

Before the booklets were scanned, all batches went through a quality inspection to ensure batch integrity and correct document placement.

After a quality check-in at the DRC Document Processing log-in area, the spines were cut off the scannable documents, and the pages were sent to DRC's Imaging and Scoring System.

## **SCANNING OF MATERIALS**

Customized scanning programs for all scannable documents were prepared to read the books and to format the scanned information electronically. Before materials arrived, all image-scanning programs went through a quality review process that included scanning of mock data from production books to ensure proper data collection.

DRC's image scanners were calibrated using a standard deck of scannable pages with 16 known levels of gray. On a predefined page location, the average pixel darkness was compared to the standard calibration to determine the level of gray. Marks with an average darkness level of 4 or above on a scale of 16 (0 through F) were determined to be valid responses, per industry standards. If multiple marks were read for a single item and the difference between the grayscale reads was greater than four levels, the lighter mark was discarded. If the multiple marks had fewer than four levels of grayscale difference, the response was flagged and forwarded to an editor for resolution.

DRC's image scanners read selected-response, demographic, and identification information. The image scanners also used barcode readers to read preprinted barcodes from a label on the book.

The scannable documents were automatically fed into the image scanners where predefined processing criteria determined which fields were to be captured electronically. Open-ended (OE) response images were separated out for image-based scoring.

During scanning, a unique serial number was printed on each sheet of paper. This serial number was used to ensure document integrity and to maintain sequencing within a batch of books.

A monitor randomly displayed images, and the human operator adjusted or cleaned the scanner when the scanned image did not meet DRC's strict quality standards for image clarity.

All images passed through a process and a software clean-up program that despeckled, deskewed, and desmeared the images. A random sample of images was reviewed for image quality approval. If any document failed to meet image quality standards, the document was returned for rescanning.

Page-scan verification was performed to ensure that all predefined portions of the booklets were represented in their entirety in the image files. If a page was missing, the entire book was flagged for resolution.

After each batch was scanned, books were processed through a computer-based editing program to detect potential errors as a result of smudges, multiple marks, and omissions in predetermined fields. Marks that did not meet the predefined editing standards were routed to editors for resolution.

Experienced DRC Document Processing editing staff reviewed all potential errors detected during scanning and made necessary corrections to the data file. The imaging system displayed each suspected error. The editing staff then inspected the image and made any necessary corrections using the unique serial number printed on the document during scanning.

Upon completion of editing, quality control reports were run to ensure that all detected potential errors were reviewed again and a final disposition was determined.

Before batches of books were extracted for scoring, a final edit was performed to ensure that all requirements for final processing were met. If a batch contained errors, it was flagged for further review before being extracted for scoring and reporting.

During this processing step, the actual number of documents scanned was compared to the number of books assigned to the box during book receipt. Count discrepancies between book receipt and books scanned were resolved at this time.

Once all requirements for final processing were met, the batch was released for scoring and student level processing.

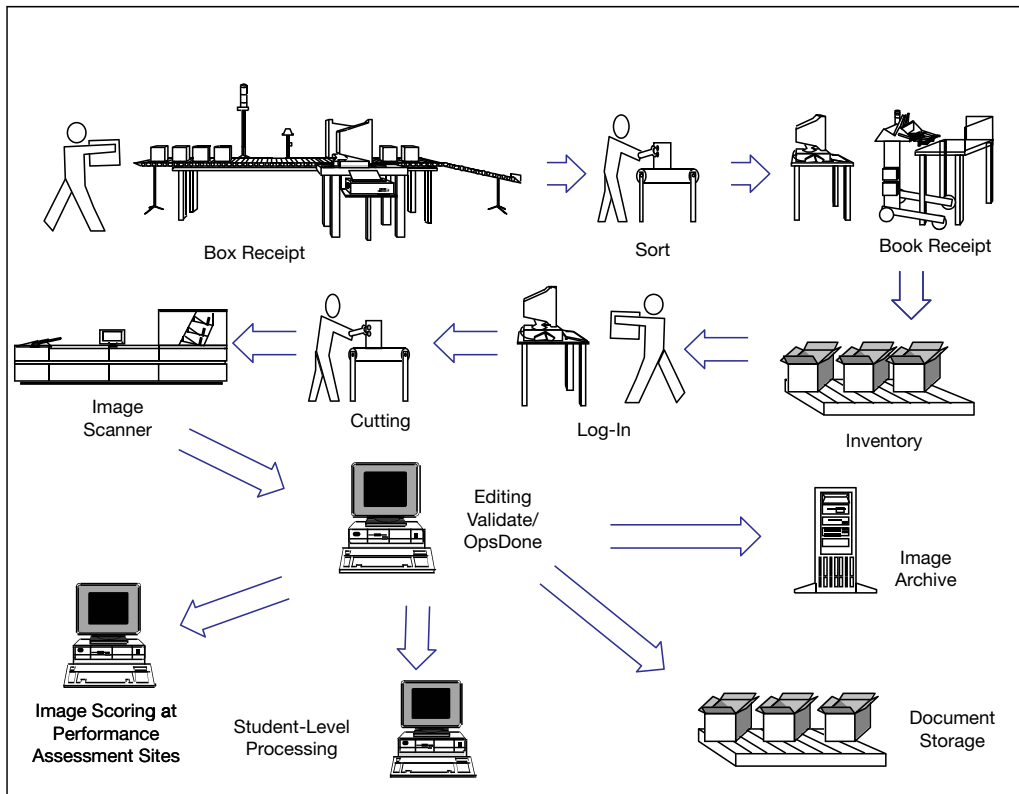
Table 8–1 shows the number of answer books received through book check-in, the number of books that contained student responses that were scanned and scored, the number of test books received, and the total number of books received for the Algebra I, Biology, and Literature Keystone Exams.

**Table 8–1. Counts of 2023 Keystone Exams Materials Received: Algebra I, Biology, and Literature**

<b>Exam</b>	<b>Answer Books Received</b>	<b>Used Answer Books Received</b>	<b>Test Books Received</b>	<b>Total Books Received</b>	<b>Total Books Shipped</b>
Algebra I (Winter)	87,198	42,568	87,198	174,396	174,444
Biology (Winter)	67,590	29,580	67,589	135,179	135,204
Literature (Winter)	57,847	26,933	57,848	115,695	115,718
Algebra I (Spring)	190,201	127,393	190,066	380,267	380,672
Biology (Spring)	146,752	100,091	146,737	293,489	293,710
Literature (Spring)	134,218	91,692	134,214	268,432	268,612
Algebra I (Summer)	987	84	987	1,974	1,974
Biology (Summer)	1,031	87	1,031	2,062	2,062
Literature (Summer)	910	47	910	1,820	1,820

Figure 8–1 illustrates the production workflow for DRC’s Ops MMS and Image Scanning and Scoring System from receipt of materials through all processing of materials and the presentation of scanned images for scoring.

**Figure 8–1. Workflow System**



## **MATERIALS STORAGE**

Upon completion of processing, student response documents were boxed for security purposes and final storage as follows:

- Project-specific box labels were created containing unique customer and project information, material type, batch number, pallet/box number, and the number of boxes for a given batch.
- Boxes were stacked on pallets that were labeled with the project information and a list of the pallet's contents before delivery to the Materials Distribution Center for final secure storage.

Materials will be destroyed one year after the contract year ends with PDE written approval.

## **ONLINE TESTING**

The DRC INSIGHT test engine runs on a custom web browser that is designed to ensure a fully secure environment during testing. The secure browser “locks down” the student’s testing device, preventing the student from accessing the desktop, the Internet, and other external programs. For non-secure testing such as practice and training sessions, students can use the Online Tools Training (OTT) environment, which runs on a standard web browser.

The custom browser software is downloaded from the DRC INSIGHTS Portal and installed onto student testing devices. The secure browser can be installed on computers individually, or it can be downloaded to a central location, copied, and distributed to multiple computers simultaneously using common network distribution tools. Everything needed for testing is found within the secure browser, eliminating the need for districts to coordinate updates to third-party software.

Prior to operational use, DRC’s quality assurance staff will perform full system-level tests in an independent test environment that simulates the production configuration. Tests are run on all supported computer platforms and browsers and include comprehensive review of system functionality, usability, reliability, security, and overall performance. Test content is also validated during this process.

Multiple methods are used to ensure secure data transfer, including encryption technologies and Secure Sockets Layer (SSL) protocol through Hypertext Transfer Protocol Secure (HTTPS). Test content is encrypted at the host server, and remains encrypted throughout all network transmissions; content is decrypted only once the student login is validated. Decrypted test content on the student workstation is stored only in memory during each test session. Once the session is ended (the test is completed or the student logs out), computer memory is purged to ensure security of test content is maintained.

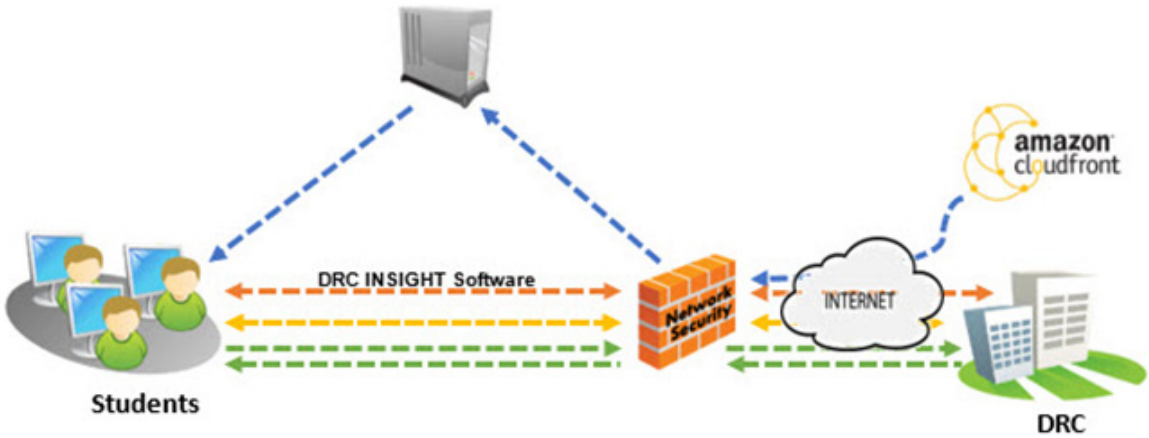
Responses are saved automatically every 45 seconds during testing, or when the student navigates away from an item or answers a selected-response item (whichever comes first). If a particular question takes the student longer than 45 seconds to answer, then the partial, incomplete responses are submitted at 45-second intervals until the student completes the item. This auto-save helps safeguard against students losing their work on longer items, such as constructed-response items. When the student returns to the test after a break or interruption, the student is returned to the point that they left off without having to navigate through all previously answered questions.

**Table 8–2. Counts of 2023 Keystone Exams Online Assessments**

Grade/Subject	Total Online Assessments Completed
Algebra I (Winter)	18,754
Biology (Winter)	15,853
Literature (Winter)	13,906
Algebra I (Spring)	48,017
Biology (Spring)	49,452
Literature (Spring)	45,597
Algebra I (Summer)	197
Biology (Summer)	139
Literature (Summer)	73

Figure 8–2 illustrates the secure transfer of online test responses between the student and DRC.

**Figure 8–2. Architecture of the Student Testing Experience**



## SCORING MULTIPLE-CHOICE ITEMS

For both online and paper-and-pencil modes, the scoring process included the scoring of multiple-choice (MC) items against the answer key and the aggregation of raw scores from the OE responses. A student's raw score is the actual number of points achieved for tested elements of an assessment. From the raw scores, the scale scores were calculated.

The student file was scored against the final and approved MC answer key. Items were scored as right, wrong, omitted, or double-gridded (more than one answer was bubbled for an item). Sections of the exam were evaluated as a whole, and an attempt status was determined for each student for each subject. The score program defined all data elements for reporting at the student level.

## RANGEFINDING

After student answer documents were received and processed, DRC's Performance Assessment Services (PAS) staff assembled groups of responses that exemplified the different score points for each subject. The score point ranges were represented by the following scoring guidelines:

- 0–3 item-specific scoring guidelines for Literature
- 0–4 item-specific scoring guidelines for Algebra I (some items were divided into separate parts that were scored on a 0–1, 0–2, or 0–3 point scale, but the sum of the parts always resulted in an overall score of 0–4 for each item)
- 0–3 item-specific scoring guidelines for Biology

Responses were pulled from the embedded field-test portion of the Keystone Exams for each subject. Once examples of all score points were selected for each item, sets were assembled for rangefinding and copies were made for each rangefinding participant. Rangefinding committees consisted of Pennsylvania educators, PDE staff members, DRC Test Development staff, and DRC Performance Assessment Services staff. The rangefinding meetings were held from July 10<sup>th</sup>-14<sup>th</sup> for Literature and Algebra 1, and from July 10<sup>th</sup> – July 13<sup>th</sup> for Biology, at the Lancaster Marriott at Penn Square in Lancaster, PA.

Each rangefinding meeting began in a joint session with a review of the history of the assessment as well as a discussion of the purpose of the rangefinding meeting and the role rangefinding plays within the item development process. The session then broke into subject/grade-specific committees. Sets of student responses were presented to the committees, one item at a time. Each committee initially reviewed and scored student responses as a group to ensure consistency in the interpretation of the scoring guidelines. Committee members then went on to score responses independently. For each student response, committee members' scores were discussed until a consensus was reached. Only those responses for which there was strong agreement among committee members were chosen for inclusion in training materials for DRC raters.

Discussions of student responses included the mandatory use of scoring guideline language. This ensured that committee members remained focused on the specific requirements of each score level. DRC PAS staff took notes addressing how and why the committees arrived at score point decisions, and this information was used by the scoring directors in rater training.

DRC and PDE discussed scoring guideline edits suggested by the rangefinding committees. Changes approved by PDE were incorporated into the scoring guidelines by DRC Test Development staff. The edited scoring guidelines were used in the preparation of materials and the training of raters.

## RATER RECRUITMENT/QUALIFICATIONS

DRC retains a number of raters from year to year; the overall return rate in 2023 was 48%. This pool of experienced raters was drawn from to staff the scoring of the 2023 Keystone Exams. To complete the rater staffing for this project, recruiting events were held and applications for rater positions were screened by DRC's recruiting staff. Candidates were personally interviewed by DRC staff. In addition, each candidate was required to provide an on-demand writing sample, an on-demand math sample, references, and proof of a four-year college degree. In this screening process, preference was given to candidates with previous experience scoring large-scale assessments and degrees emphasizing expertise in the subjects being scored. Staffing partners were used to

augment hiring using the same practices as those employed by DRC. The rater pool consisted of educators and other professionals with content-specific backgrounds. These individuals were valued for their content-specific knowledge, but they were required to set aside their own biases about student performance and accept the scoring standards of the Keystone Exams.

## **LEADERSHIP RECRUITMENT/QUALIFICATIONS**

Scoring directors and team leaders were selected from a pool of employees who displayed expertise as raters and leaders on previous DRC projects. These individuals had strong backgrounds in mathematics, English language arts, or science, and demonstrated organizational, leadership, communication, and management skills. All scoring directors had previous leadership experience working on large scale assessments. All scoring directors, team leaders, and raters were required to sign confidentiality agreements before handling secure materials.

Each group of raters was assigned a scoring director. All handscoring activities were led by a scoring director for the duration of the project. Scoring directors assisted in rangefinding, worked with supervisors to create training materials, conducted team leader training, and were responsible for training the raters. The scoring director made sure that handscoring reports were available and interpreted those reports for the raters. The scoring director also supervised the team leaders. Scoring directors were monitored by the project managers throughout the project.

Team leaders assisted the scoring director with rater training by answering individual questions that raters may not have felt comfortable asking in a large group. Once raters were qualified, team leaders were responsible for monitoring and maintaining the accuracy and workload of each team member. Ongoing monitoring identified those individuals having difficulty maintaining accuracy. These raters received one-on-one retraining from the team leader or scoring director. Any rater who could not be successfully retrained had his/her scores purged and was released from the project.

## **TRAINING**

As part of preparation for the 2023 Keystone Exams, DRC's PAS staff assembled the PDE-approved scoring guidelines and scored student responses approved by rangefinding committees into sets used for training raters. These item-specific scoring guidelines served as the raters' constant reference. Responses that were relevant in terms of the scoring concepts they illustrated were annotated and included in an anchor set. The full range of each score point was clearly represented and annotated in the anchor set, which was used for reference by raters throughout the project.

Training sets and qualifying sets contained student responses consensus-scored by rangefinding committee members. Raters were instructed on how to apply the scoring guidelines and were required to demonstrate a clear comprehension of each anchor set by performing well on the associated training materials. Responses were selected for training to show raters the range of each score point (e.g., high, mid, and low 2s). Examples of 0s were also included for all items. This process helped raters recognize the various ways that a student could respond in order to earn each score point outlined and defined in the item-specific scoring guidelines.

The scoring director conducted a team leader training session before training the raters. This session followed the same procedures as rater training but was more rigorous and in-depth due to the extra responsibilities required of team leaders. During team leader training, all pertinent training materials were reviewed and discussed. Team leaders were given access to fully annotated training materials with committee justifications from the rangefinding meetings. To facilitate scoring consistency, it was imperative that all team leaders imparted the same rationale for each response. Once the team leaders were qualified, leadership responsibilities were reviewed and team assignments were given. A ratio of one team leader per 7–10 raters ensured sufficient monitoring rates for team members.

Rater training began with the scoring director providing an intensive review of the scoring guidelines and anchor responses. Next, raters practiced by independently scoring the responses in the training sets. After each training set, the scoring director led a thorough discussion of the responses.

Once the scoring guidelines, anchor sets, and training sets were thoroughly discussed, each rater was required to demonstrate understanding of the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the true scores) on at least one of the qualifying sets. Raters who failed to achieve at least 70 percent exact agreement on the first qualifying set were given additional training, either individually or in a small group setting. Raters who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score any student responses. These individuals were removed from the pool of potential raters in DRC’s imaging system and released from the project.

DRC’s remote scoring is designed to very closely emulate the work that was done in our physical scoring locations. The platform, content, and expectations for quality remain the same, and interactive technology and content training and discussions are conducted live (virtually). The differences come with the method through which training is delivered (online), and in the modes of communication that are used (web screen sharing, webcast, video chat, and chat). Our scoring leaders are equipped with a variety of tools to ensure every scorer is successful in understanding and applying scoring criteria to student responses.

The 2023 assessment included the opportunity for students to respond in Spanish to Algebra I and Biology items. Rater training for the Spanish language response scoring was conducted by Tri-Lin Integrated Services in San Antonio, Texas, and was overseen by a DRC project manager, who is a Spanish language speaker with a strong handscoring background. All Spanish raters were bilingual and hired specifically to score the Spanish portion of the assessment and were required to meet the same standards set for raters of the English language version of the assessment.

**Table 8–3. Qualification Rates for 2023 Keystone Open-Ended Response Items – Winter**

Subject	% Qualifying	% That Did Not Qualify
Algebra I	100	0
Biology	99	1
Literature	100	0

**Table 8–4. Qualification Rates for 2023 Keystone Open-Ended Response Items – Spring**

Subject	% Qualifying	% That Did Not Qualify
Algebra I	99	1
Biology	98	2
Literature	100	0



**Table 8–5. Qualification Rates for 2023 Keystone Open-Ended Response Items – Summer**

<b>Subject</b>	<b>% Qualifying</b>	<b>% That Did Not Qualify</b>
Algebra I	100	0
Biology	100	0
Literature	100	0

## **HANDSCORING PROCESS**

Student responses were scored independently. All responses were scored once, and ten percent of the responses were scored a second time. The data collected from the ten-percent double-read portion was used to calculate the exact and adjacent agreement rates in the Scoring Summary Reports. The responses that were used for the ten percent read behind were randomly chosen by the imaging system at the item level. Additional read behinds by the team leaders and scoring directors were done to further ensure reliability.

Raters scored the imaged student responses remotely on either laptop or desktop computers.

Raters worked in a secure location in their homes. Image distribution was controlled, ensuring that student images were sent only to designated groups of raters qualified to score those items. Imaged student responses were electronically separated for routing to individual raters by item. Raters were only provided with student responses for items that they were qualified to score. Scores were keyed into DRC’s imaging system.

To handle possible alerts (i.e., student responses indicating potential issues related to students’ safety and well-being that sometimes require attention at the state or local level), DRC’s imaging system allows raters to forward responses needing attention to the scoring director. These alerts are reviewed by project management, who then notifies the students’ schools and PDE of the occurrences. PDE does not receive any identifying information about the students. At no point in the alerts process do raters, or other DRC handscoring staff, acquire any knowledge concerning a student’s personal identity.

## **HANDSCORING VALIDITY PROCESS**

One of the training tools PAS utilized to ensure rater accuracy was the validity process. The goal of the validity process is to ensure that scoring standards are maintained. Specifically, the objective is to make sure that raters score student responses in a manner consistent with statewide standards both within a single administration of the keystones and across consecutive administrations. In scoring the 2023 Keystone Exams, scoring consistency was maintained, in part, through the validity process.

The validity process began with the selection of scored responses. Forty validity responses were selected for each core open-ended (OE) item. These 40 responses were drawn from a pool of exemplars (responses that are representative of a particular score point and have been verified by the scoring director). The scores on validity responses are considered true scores.

The validity responses were then implemented to test rater accuracy. The responses were selected within the imaging system and dispersed intermittently to the raters. By the end of the project, raters had scored all 40 validity responses for any items they were qualified to score. Raters were unaware of when they were being dealt pre-scored validity responses and assumed that they were scoring live student responses. This helped bolster the internal validity of the process. All raters who received validity responses had already successfully completed the training/qualifying process.

The scores that the raters assigned to the validity responses were compared to the true scores in order to determine the validity of the raters' scores. For each item, the percentage of exact agreement as well as the percentage of high and low scores was computed. This data was accessed through the Validity Item Detail Report. The same sort of data was also computed for each specific rater. This data was accessed through the Validity Reader Detail Report. Both of these may be run as daily or cumulative reports.

The Validity Reader Detail Report was used to identify particular raters for retraining. If a rater on a certain day generated a lower rate of agreement on a group of validity responses, it was immediately apparent in the Validity Reader Detail Report. A lower rate of agreement was defined as anything below 70 percent exact agreement with the true scores. Any time a rater's validity agreement rate fell below 70 percent, the scoring director was cued to examine that rater's scoring. First, the scoring director attempted to ascertain what kind of validity responses the rater was scoring incorrectly. This was done to determine whether there was any sort of a trend (e.g., trending low on the 1–2 line). Once the source of the low agreement rate was determined, the rater was retrained. If it was determined that the rater had been scoring live responses inaccurately, then his/her scores were purged for that day, and the responses were re-circulated and scored by other raters.

The cumulative Validity Item Detail Report was utilized to identify potential group-wide trends in need of correction. For instance, if a particular validity response with a true score of 3 was given a score of 2 by a significant number of raters within the group, that trend would be revealed in the Validity Item Detail Report. To correct a trend of this sort, the scoring director would look for student responses similar to the validity response being scored incorrectly. Once located, these responses would be used in group-wide re-training, usually in the form of an annotated handout or a short set of responses without printed scores given to raters as a recalibration test.

Validity was employed on all core Algebra I, Biology, and Literature OE items. Each 40-response validity set was formulated to mirror the score point distribution that the item generated during its previous administration. Each validity set included at least five examples of each score point. Examples of different types of responses were included to ensure that raters were tested on the full spectrum of response types.

The exact rater agreement rate generated during the validity process was often higher than the inter-rater agreement rate for the same item. The reason for this discrepancy has to do with how validity sets are formulated. The 40 validity responses for each item are intended to cover the full breadth of each score point. For example, each validity set contains examples of high, mid, and low 2s. This scope ensures that the validity process is truly valid in terms of addressing the complete spectrum of response types. However, certain types of responses are generally not included in validity sets. These include line responses (i.e., examples of score points that are so close to the adjacent score point that raters are instructed to consult with a supervisor before assigning a score) and responses that, because of poor word choice/writing, are difficult to understand. The reason for these exclusions is that confusing/line/illegible responses often do not impart a teachable lesson. Since these types of responses are usually unique, any potential lesson the response might teach would apply only to that particular response. Conversely, the responses in validity sets are chosen because they represent common response types and teach lessons that can be applied to other similar responses. Due to this distinction, validity sets often generate a slightly higher agreement rate than is typically generated during operational scoring.

## **QUALITY CONTROL**

Rater accuracy was monitored throughout the scoring session by means of daily and on-demand reports. These reports ensured that an acceptable level of scoring accuracy was maintained throughout the project. Inter-rater reliability was tracked and monitored with multiple quality control reports that were reviewed by quality assurance analysts. These reports and other quality control documents were generated at the scoring centers, where they were reviewed by the scoring directors, team leaders, and project managers. The following reports and documents were used during the scoring of the open-ended items:

The Scoring Summary Report (includes two related reports).

- The Reader Monitor Report monitored how often raters were in exact agreement with one another and ensured that an acceptable agreement rate was maintained throughout the project. This report provided daily and cumulative exact and adjacent inter-rater agreement on the ten percent that was double read.

- The Score Point Distribution Report monitored the percentage of responses given each of the score points. For example, the Algebra I daily and cumulative reports showed what percentage of 0s, 1s, 2s, 3s, and 4s a rater—or group of raters—had given to all the responses scored at the time the report was produced. It also indicated the number of responses read by each rater so that production rates could be monitored.
- The Item Status Report monitored the progress of handscoring. This report tracked each response and indicated the status (e.g., not read, complete, awaiting supervisor review, etc.). This report ensured that all responses were scored by the end of the project.

The Reader Score Report identified all responses scored by an individual rater. This report was useful if any responses needed rescoring due to possible rater drift.

The Validity Reports (addressed in detail on previous page) tracked how raters performed by comparing pre-scored responses to raters' scores for the same responses. If a rater's scoring fell below the 70 percent determined agreement rate, remediation occurred. Raters who did not retrain to the required level of agreement were released from the project.

Read Behinds were used by the team leader/scoring director to monitor individual rater reliability. Team leaders read randomly-selected, scored responses from each team member on a daily basis. If the team leader disagreed with a rater's score, remediation occurred. This was a particularly effective form of feedback because it was performed in real time with live student responses scored by each rater.

Recalibration Sets were used throughout the scoring sessions to ensure accuracy by comparing each rater's scores with the true scores on a pre-selected set of responses. Recalibration sets helped to refocus raters on Pennsylvania scoring standards. These checks made sure there was no change in the scoring pattern as the project progressed. Raters failing to achieve 70 percent agreement with the recalibration true scores were given additional training to achieve the highest degree of accuracy possible. Raters who were unable to recalibrate were released from the project. The process used for creating and administering recalibration sets was similar to the process employed for creating and administering training sets.

**Table 8–6. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Winter 2023**

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Validity Agreement	% 0s	% 1s	% 2s	% 3s	% 4s	% B/NS
Alg. 1	1	897706	A	0-1	99	1	100	53	33	NA	NA	NA	14
Alg. 1	1	897706	B	0-1	100	0	100	81	5	NA	NA	NA	14
Alg. 1	1	897706	C	0-1	100	0	100	72	14	NA	NA	NA	14
Alg. 1	1	897706	D	0-1	100	0	100	80	5	NA	NA	NA	14
Alg. 1	1	969537	A	0-1	99	1	100	53	32	NA	NA	NA	16
Alg. 1	1	969537	B	0-1	100	0	100	83	1	NA	NA	NA	16
Alg. 1	1	969537	C	0-1	100	0	100	79	6	NA	NA	NA	16
Alg. 1	1	969537	D	0-1	100	0	98	78	7	NA	NA	NA	16
Alg. 1	1	795492		0-4	94	6	98	23	25	18	13	4	17
Alg. 1	2	739461	A	0-1	100	0	100	28	61	NA	NA	NA	11
Alg. 1	2	739461	B	0-1	100	0	100	77	11	NA	NA	NA	11
Alg. 1	2	739461	C	0-1	100	0	99	87	2	NA	NA	NA	11
Alg. 1	2	739461	D	0-1	100	0	99	79	10	NA	NA	NA	11
Alg. 1	2	714761	A	0-1	100	0	100	84	4	NA	NA	NA	12
Alg. 1	2	714761	B	0-1	100	0	99	77	11	NA	NA	NA	12
Alg. 1	2	714761	C	0-1	98	2	99	58	30	NA	NA	NA	12
Alg. 1	2	714761	D	0-1	100	0	100	85	3	NA	NA	NA	12
Alg. 1	2	820175		0-4	95	5	95	38	24	12	6	2	18
Bio.	1	966518		0-3	90	10	89	44	26	8	4	NA	17
Bio.	1	812926		0-3	93	7	87	30	26	18	8	NA	19
Bio.	1	869089		0-3	90	10	90	49	22	7	3	NA	19
Bio.	2	869033		0-3	91	9	83	40	28	11	4	NA	16
Bio.	2	737711		0-3	87	12	90	14	14	48	6	NA	18
Bio.	2	703529		0-3	97	3	94	58	13	5	2	NA	20
Lit.	1	928786		0-3	86	14	92	9	35	34	6	NA	17
Lit.	1	916572		0-3	85	15	91	19	35	20	6	NA	20
Lit.	1	916574		0-3	87	13	91	26	29	18	4	NA	23
Lit.	2	912094		0-3	85	15	88	9	37	29	7	NA	18
Lit.	2	912113		0-3	87	13	81	13	36	25	5	NA	20
Lit.	2	912223		0-3	86	14	91	12	39	23	4	NA	21

Notes. B = blank; N = non-scorable. NA= non-applicable. Algebra I responses received a possible total of 0–4 points. For some Algebra I items, raters applied a single score of 0, 1, 2, 3, or 4; however, many Algebra I items were divided into separate parts that were scored on 0–1, 0–2, or 0–3-point scales, the sum of which always resulted in an overall score of 0–4 points. For example, an Algebra I item might have a part A, a part B, a part C, and a part D, each of which was scored on a 0–1-point scale, resulting in a summed 0–4-point total score scale. Additionally, some Algebra I items with multiple parts could receive up to one point for “minimal understanding” (MU) even if the student did not receive a point, or points, for any of the item’s individual parts.

**Table 8–7. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Spring 2023**

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Validity Agreement	% 0s	% 1s	% 2s	% 3s	% 4s	% B/NS
Alg. 1	1	821550	A	0-1	100	0	98	16	76	NA	NA	NA	8
Alg. 1	1	821550	B	0-1	97	3	97	46	46	NA	NA	NA	8
Alg. 1	1	821550	C	0-1	99	1	97	84	8	NA	NA	NA	8
Alg. 1	1	821550	D	0-1	100	0	100	71	21	NA	NA	NA	8
Alg. 1	1	980115		0-4	87	13	84	17	49	13	6	1	15
Alg. 1	1	724685	A	0-1	99	1	96	53	35	NA	NA	NA	12
Alg. 1	1	724685	B	0-1	99	1	99	73	15	NA	NA	NA	12
Alg. 1	1	724685	C	0-1	100	0	98	85	3	NA	NA	NA	12
Alg. 1	1	724685	D	0-1	99	1	96	80	8	NA	NA	NA	12
Alg. 1	2	969452		0-4	85	14	86	35	24	15	10	6	11
Alg. 1	2	821012		0-4	85	14	84	44	13	10	13	7	13
Alg. 1	2	818209	A	0-1	99	1	99	48	40	NA	NA	NA	11
Alg. 1	2	818209	B	0-1	100	0	100	67	22	NA	NA	NA	11
Alg. 1	2	818209	C	0-1	100	0	100	78	10	NA	NA	NA	11
Alg. 1	2	818209	D	0-1	99	1	99	82	6	NA	NA	NA	11
Bio.	1	1035250		0-3	79	19	76	34	33	16	9	NA	9
Bio.	1	1058184		0-3	80	20	82	28	28	20	12	NA	11
Bio.	1	742581		0-3	90	10	93	22	39	19	6	NA	14
Bio.	2	983824		0-3	88	12	92	32	26	19	13	NA	9
Bio.	2	978657		0-3	93	7	90	41	21	13	12	NA	13
Bio.	2	1056094		0-3	85	15	82	28	27	21	10	NA	14
Lit.	1	994432		0-3	80	19	84	6	26	38	19	NA	11
Lit.	1	1067378		0-3	80	20	82	9	29	35	15	NA	12
Lit.	1	1067376		0-3	80	19	81	13	27	34	12	NA	14
Lit.	2	984231		0-3	82	18	82	7	36	34	10	NA	12
Lit.	2	1066357		0-3	79	21	83	12	30	31	12	NA	14
Lit.	2	1066782		0-3	79	21	79	10	33	30	11	NA	16

Notes. B = blank; NS = non-scorable. NA= non-applicable. Algebra I responses received a possible total of 0–4 points. For some Algebra I items, raters applied a single score of 0, 1, 2, 3, or 4; however, many Algebra I items were divided into separate parts that were scored on 0–1, 0–2, or 0–3-point scales, the sum of which always resulted in an overall score of 0–4 points. For example, an Algebra I item might have a part A, a part B, a part C, and a part D, each of which was scored on a 0–1-point scale, resulting in a summed 0–4-point total score scale. Additionally, some Algebra I items with multiple parts could receive up to one point for “minimal understanding” (MU) even if the student did not receive a point, or points, for any of the item’s individual parts.

**Table 8–8. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items Summer 2023**

Exam	Module	Item ID	Item Part	Score Point Range	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Validity Agreement	% 0s	% 1s	% 2s	% 3s	% 4s	% B/NS
Alg. 1	1	1064862	A	0-1	100	0	100	15	83	NA	NA	NA	3
Alg. 1	1	1064862	B	0-1	97	3	95	42	56	NA	NA	NA	3
Alg. 1	1	1064862	C	0-1	100	0	100	85	12	NA	NA	NA	3
Alg. 1	1	1064862	D	0-1	100	0	100	84	14	NA	NA	NA	3
Alg. 1	1	905408		0-4	89	11	91	28	44	11	9	4	5
Alg. 1	1	1064322	A	0-1	100	0	100	56	41	NA	NA	NA	2
Alg. 1	1	1064322	B	0-1	100	0	100	59	38	NA	NA	NA	2
Alg. 1	1	1064322	C	0-1	100	0	99	89	8	NA	NA	NA	2
Alg. 1	1	1064322	D	0-1	100	0	100	77	20	NA	NA	NA	2
Alg. 1	2	1023111	A	0-1	100	0	99	26	73	NA	NA	NA	1
Alg. 1	2	1023111	B	0-1	100	0	100	77	22	NA	NA	NA	1
Alg. 1	2	1023111	C	0-1	100	0	100	89	10	NA	NA	NA	1
Alg. 1	2	1023111	D	0-1	100	0	100	96	3	NA	NA	NA	1
Alg. 1	2	1063764		0-4	91	9	92	27	49	12	6	2	4
Alg. 1	2	820072	A	0-1	100	0	100	65	32	NA	NA	NA	3
Alg. 1	2	820072	B	0-1	100	0	100	89	8	NA	NA	NA	3
Alg. 1	2	820072	C	0-1	100	0	100	88	8	NA	NA	NA	3
Alg. 1	2	820072	D	0-1	100	0	100	89	8	NA	NA	NA	3
Bio.	1	1021841		0-3	93	7	96	11	24	41	21	NA	3
Bio.	1	979580		0-3	100	0	99	53	20	11	10	NA	6
Bio.	1	966416		0-3	100	0	98	29	27	26	12	NA	5
Bio.	2	1056105		0-3	85	15	93	39	29	19	10	NA	4
Bio.	2	978209		0-3	93	7	96	37	31	18	9	NA	5
Bio.	2	1059554		0-3	96	4	76	13	22	35	26	NA	3
Lit.	1	742085		0-3	84	16	97	7	42	43	1	NA	8
Lit.	1	1021903		0-3	84	16	98	10	46	34	1	NA	8
Lit.	1	1021911		0-3	81	19	98	10	51	25	1	NA	12
Lit.	2	704766		0-3	89	11	92	13	45	26	4	NA	11
Lit.	2	1065282		0-3	84	16	89	15	43	25	3	NA	14
Lit.	2	1065280		0-3	91	9	96	18	38	25	2	NA	16

Notes. B = blank; NS = non-scorable. NA= non-applicable. Algebra I responses received a possible total of 0–4 points. For some Algebra I items, raters applied a single score of 0, 1, 2, 3, or 4; however, many Algebra I items were divided into separate parts that were scored on 0–1, 0–2, or 0–3-point scales, the sum of which always resulted in an overall score of 0–4 points. For example, an Algebra I item might have a part A, a part B, a part C, and a part D, each of which was scored on a 0–1-point scale, resulting in a summed 0–4-point total score scale. Additionally, some Algebra I items with multiple parts could receive up to one point for “minimal understanding” (MU) even if the student did not receive a point, or points, for any of the item’s individual parts.

## CHAPTER NINE: DESCRIPTION OF DATA SOURCES

---

This section describes the filtering process and data sources used for the various analysis procedures discussed in the remaining sections of this report. Psychometric analyses were conducted at several points for the winter, spring, and summer. Pennsylvania Keystone Exams in Algebra I, Biology, and Literature: 1) key verification analyses for quality-control purposes; 2) pre-equating verification; 3) item analysis and calibration of field-test items embedded in the spring forms; and 4) analyses for this technical report.

### STUDENT FILTERING CRITERIA

Students' records included used for psychometric analyses needed to meet at least the following criteria:

- Module 1 Attempted Status = 1 (1 represents if the student attempted a minimum of five items in Module 1)
- Module 2 Attempted Status = 1 (1 represents if the student attempted a minimum of five items in Module 2)
- Module 1 Invalidated = N (N represents if the student's score was not invalidated)
- Module 2 Invalidated = N (N represents if the student's score was not invalidated)
- Student Duplication Status = N (N represents if no record duplication within subject)
- Module 1 Form Number = Module 2 Form Number (represents if the student took the same form for Module 1 and Module 2)
- Module 1 Form Name  $\neq$  01V (exclude students administered the VSL form in Module 1)
- Module 2 Form Name  $\neq$  01V (exclude students administered the VSL form in Module 2)

For each specific analysis conducted at different times, additional criteria might be needed to filter students. For example, the following criteria were used in addition to the ones listed above for the pre-equating verification, since the analyses were conducted during the scoring window:

- Module 1 Complete Status = 01
- Module 2 Complete Status = 01

The complete status of 01 represents the response string which includes scores on both the multiple-choice (MC) and constructed-response (CR) operational items. When the analyses were conducted using the final data files, these criteria were no longer necessary since all operational CR items had been scored.

Item analysis and calibration of embedded field-test items were conducted using the first-time testers only (i.e., retester = N). The classical item statistics for the field-test items analyzed by using the first-time testers were more comparable to the results of the spring 2011 Keystone Exams, which were given to the first-time test takers.

Because a large number of students took the spring Keystone Exams, only a representative sample of students' responses on field-test CR items was scored within each content area. For the item analysis of field-test CR items, the following additional criteria were used to select only those who were sampled for hand-scoring:

- Module 1 Complete Status = 02
- Module 2 Complete Status = 02

## KEY VERIFICATION DATA

The key verification data are mentioned only for completeness, as no formal results are provided in this technical document. Key verification is often conducted early in the scoring process to ensure the keys for the MC items are applied correctly. The data files used for the key verification analysis are usually (but not always) based on the student data from early-return schools. The sample representativeness is not required for this internal quality check. Available student data typically suffices as long as there is reasonable variability in the total-test scores of students. The details about the sample sizes for the winter, spring, and summer administrations can be found in Table 9–1.

## PRE-EQUATING VERIFICATION

The pre-equating verification data included all students who met the inclusion criteria and were scored by February 22, June 26, and August 29 for the winter, spring, and summer administrations, respectively. Pre-equating verification data included students who had testing accommodations, except those who utilized alternate video sign language forms.

## FINAL DATA

The final data files were used to conduct item analyses for the operational items and analyses conducted for Chapters Sixteen through Nineteen in this technical report. The final data contained students' responses to both the MC and CR items. All students' responses included in the analyses met the filtering criteria. The final sample sizes (or *n*-counts) can be found in the column labeled "Final" in Table 9–1.

**Table 9–1. Data Source *N*-Counts**

Administration	Content Area	Key Verification	Pre-Equating Verification	Final
Winter	Algebra I	16174	54701	54824
Winter	Biology	8740	40068	40077
Winter	Literature	12313	34743	35222
Spring	Algebra I	163044	157276	162971
Spring	Biology	137768	137770	137797
Spring	Literature	125523	125512	125553
Summer	Algebra I	274	274	274
Summer	Biology	210	210	210
Summer	Literature	111	111	111

## SPIRALING OF FORMS

During the administration of Keystone Exams, test forms were spiraled at the student level. The goal of spiraling is to achieve equivalent samples of students across forms so the classical statistics (e.g., *p*-value and point-biserial correlation) for all the field-test items can be compared. Given that the field-test items were embedded in the spring administration only, the equivalence of samples was checked for the spring administration instead of all administrations. When spiraling achieves randomly equivalent samples, the forms will have equal means (within sampling error) over the operational items.



Appendix H provides summary statistics for all the spring forms for each content area exam. The tables provide the form number (Form), number of students (N), test length in items (L), total points (Pts.), minimum (Min) score, maximum (Max) score, mean (Mean) score, median (Med) score, and standard deviation (SD). The extent to which the mean raw scores across forms are similar indicates the extent to which the student populations taking each form are of approximately equal ability. This equivalence of ability distributions across forms is the desired outcome of spiraling and allows for optimum analysis of the embedded field-test items.

Figure 9–1 presents the mean raw scores for all forms by content area and is based on the final data; in addition, accommodated forms (i.e., Spanish, Braille, large print, VSL) were removed to best represent comparisons among similar groups of test-takers. The spring form mean raw scores are plotted (circle-shaped marker) with standard error of mean lines. For each form, the standard error of mean was computed by taking the standard deviation of all student scores (assumed as the population standard deviation divided by the square root of the form *n*-count). The mean score across all forms is indicated by the red horizontal broken line. If the three-standard error band captures the horizontal line, then that suggests only random differences exist between the form mean and the population mean. This is true for all forms in all content areas.

**Figure 9–1. Spring Form Mean Scores with +/- Standard Error (SE) Bands**

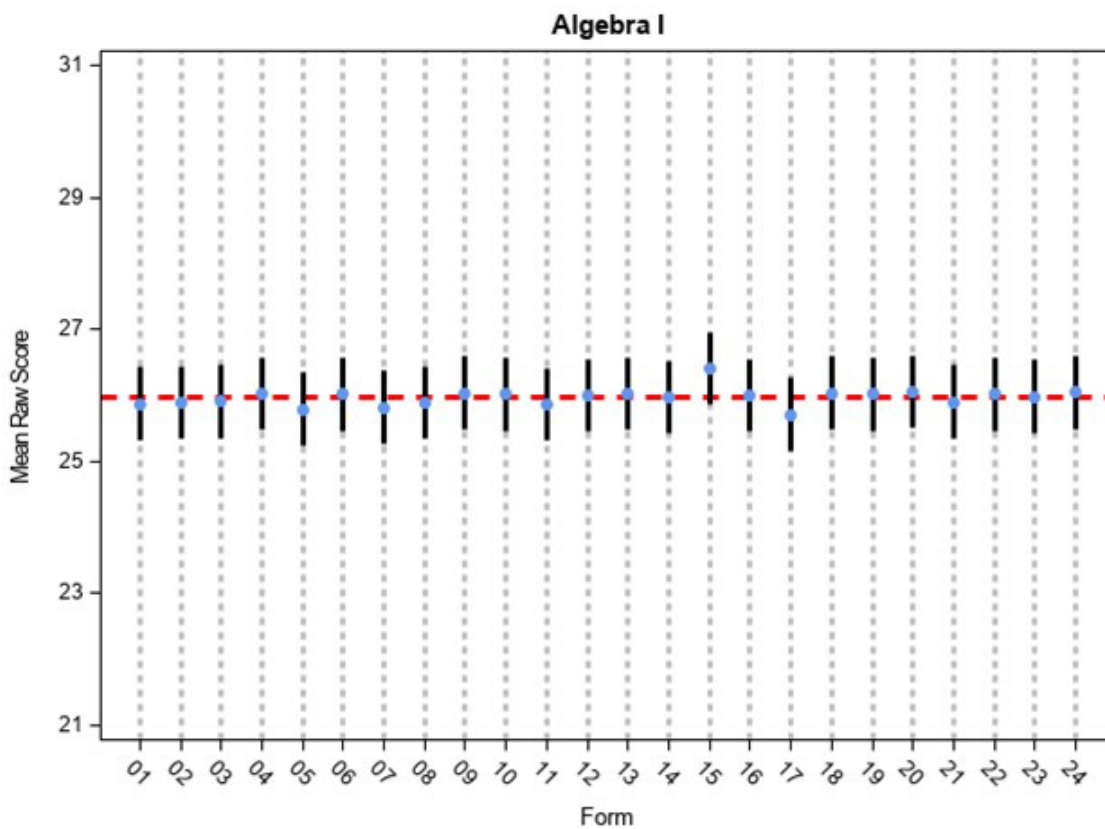
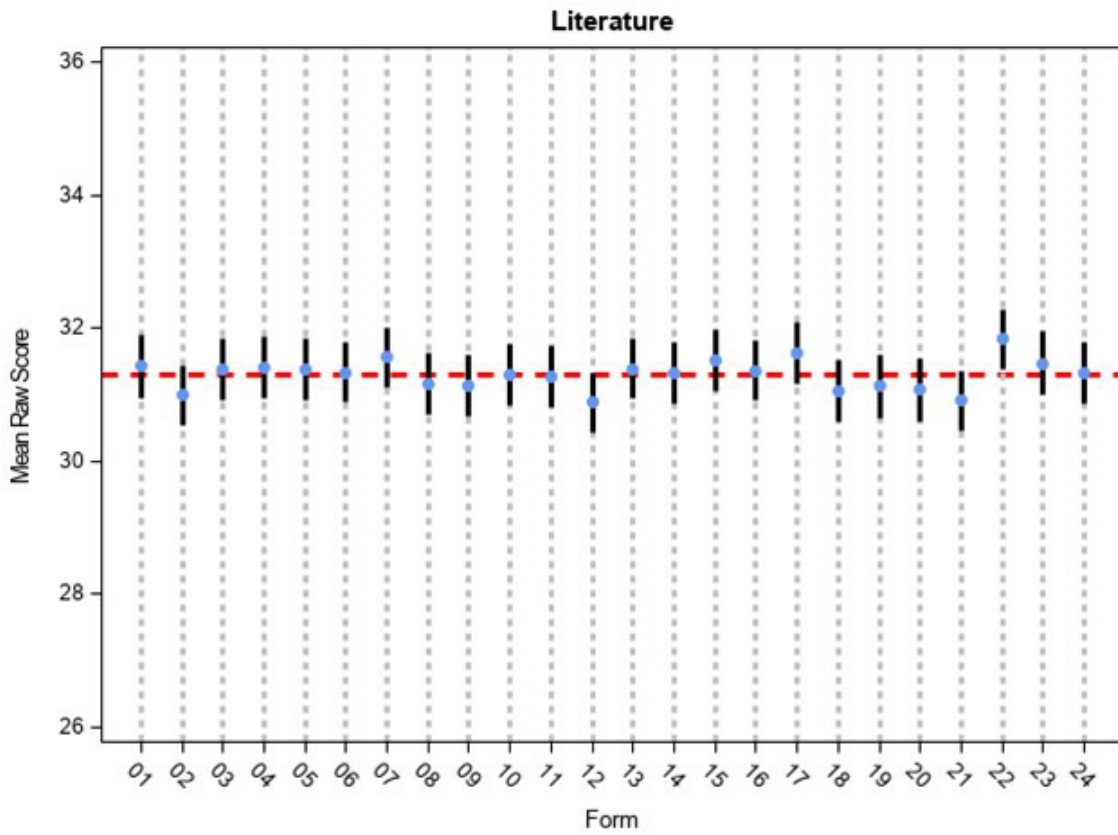
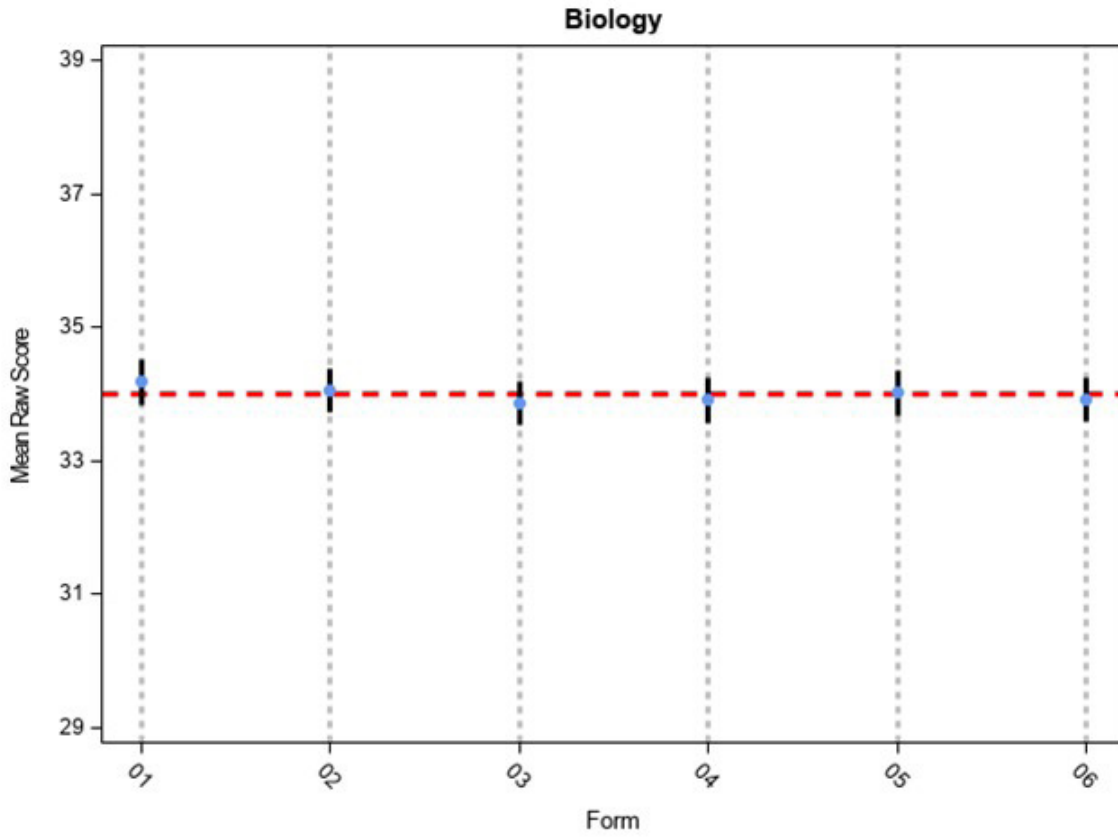


Figure 9–1 (continued). Spring Form Mean Scores with +/- Standard Error (SE) Bands



# CHAPTER TEN: SUMMARY DEMOGRAPHIC AND ACCOMMODATION DATA FOR SPRING 2023 KEYSTONE EXAMS

## ASSESSED STUDENTS

Students assessed on the Keystone Exams include students from public schools who are required to participate by virtue of being in the graduating class of 2023, students in a school district planning to use the Keystone Exams to meet graduation requirements, and students enrolled in an Algebra I, Biology, or Literature course during the 2022–2023 school year. The operational Keystone Exams were administered in both paper-and-pencil test (PPT) and computer-based test (CBT) formats.

Statistical tables and associated commentary embedded in this chapter are based on data from the Spring test administration of the Keystone Exams. There were two other administrations during the school year, each of which involved fewer students than the spring. One occurred during winter and the other in the summer. Tables summarizing results from these two administrations can be found in Appendix I.

Results for this chapter are presented in sets of tables for the three Keystone Exams administered in Spring (Algebra I, Biology, and Literature). Accompanying each numbered table is a letter (A, B, or L) to designate the content area. Tables 10–1A through 10–1L provides a summary of tests processed and scored, which are displayed separately by student grade level. The first two rows present the number processed for each administration mode (PPT and CBT). The total number of tests processed is presented on the third row. The fourth and fifth rows show the number and percentage of students with a Keystone Exam score, while the sixth and seventh row presents the number and percentage not receiving a score. Please note that the percent of students assessed (received a total score) is above 90% overall and is typically above 80% across grade levels.

**Table 10–1A. Students Assessed on the Spring 2023 Keystone Exam: Algebra I**

Description	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	23	204	4,293	21,193	50,156	30,482	16,296	245	122,892
Total number of CBT processed (Number)	4	99	2,238	8,993	19,649	11,382	5,562	90	48,017
Total number of tests processed (Number)	27	303	6,531	30,186	69,805	41,864	21,858	335	170,909
Total number of tests processed with a score (Number)	25	302	6,503	29,934	66,080	38,380	18,954	269	160,447
Total number of tests processed with a score (Percent)	92.6	99.7	99.6	99.2	94.7	91.7	86.7	80.3	93.9
Total number of tests processed without a score (Number)	2	1	28	252	3,725	3,484	2,904	66	10,462
Total number of tests processed without a score (Percent)	7.4	.3	.4	.8	5.3	8.3	13.3	19.7	6.1

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table 10–1B. Students Assessed on the Spring 2023 Keystone Exam: Biology**

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	12	216	34,603	44,975	15,556	294	95,656
Total number of CBT processed (Number)	2	60	17,717	25,214	6,332	127	49,452
Total number of tests processed (Number)	14	276	52,320	70,189	21,888	421	145,108
Total number of tests processed with a score (Number)	11	272	50,109	66,036	18,976	352	135,756
Total number of tests processed with a score (Percent)	78.6	98.6	95.8	94.1	86.7	83.6	93.6
Total number of tests processed without a score (Number)	3	4	2,211	4,153	2,912	69	9,352
Total number of tests processed without a score (Percent)	21.4	1.4	4.2	5.9	13.3	16.4	6.4

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table 10–1L. Students Assessed on the Spring 2023 Keystone Exam: Literature**

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	13	4	4,983	67,113	14,042	180	86,335
Total number of CBT processed (Number)	0	1	2,821	36,476	6,194	105	45,597
Total number of tests processed (Number)	13	5	7,804	103,589	20,236	285	131,932
Total number of tests processed with a score (Number)	7	5	6,909	98,872	17,387	225	123,405
Total number of tests processed with a score (Percent)	53.8	100	88.5	95.4	85.9	78.9	93.5
Total number of tests processed without a score (Number)	6	0	895	4,717	2,849	60	8,527
Total number of tests processed without a score (Percent)	46.2	0	11.5	4.6	14.1	21.1	6.5

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

## REASONS FOR STUDENT NON-ASSESSMENT

A small percent of students were not assessed. Although there are a variety of reasons for this, the major ones pertain to:

- Extended absence from school that continued beyond the assessment window.
- Failure to meet the attempt criteria on one or more test modules and no exclusion code marked by school personnel. The attempt criteria required a minimum of five items to be completed in each module.
- Medical emergency.
- Parental request in which the student’s parent/guardian reviewed the assessment, found it in conflict with his/her religious belief, and requested in writing that the student be excluded from participation.

- Parental request in which the student’s parent/guardian chose to have his/her child excluded from participation based on reasons other than conflict with religious belief, even though there is no provision for this exclusion in Pennsylvania regulation.
- Other reasons.

The number of students without a total test score for each of these reasons is provided in Tables 10–2A through 10–2L. Associated with this number is the percent of the total of non-assessed students in each column (grade level) attributed to a particular reason.

**Table 10–2A. Counts/Percentages of Students without Scores on the Spring 2023 Keystone Exam: Algebra I**

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	1	1	6	65	1,237	1,165	945	29	3,449
Extended absence from school (Percent)	50	100	21.4	25.8	33.2	33.4	32.5	43.9	33
Non-attempt (Number)	1	0	9	37	1,361	1,156	768	17	3,349
Non-attempt (Percent)	50	0	32.1	14.7	36.5	33.2	26.4	25.8	32
Medical emergency (Number)	0	0	4	39	104	87	61	0	295
Medical emergency (Percent)	0	0	14.3	15.5	2.8	2.5	2.1	0	2.8
Parental request - Chapter 4 (Number)	0	0	3	24	177	230	185	7	626
Parental request - Chapter 4 (Percent)	0	0	10.7	9.5	4.8	6.6	6.4	10.6	6
Parental request - Other reasons (Number)	0	0	5	49	176	205	263	2	700
Parental request - Other reasons (Percent)	0	0	17.9	19.4	4.7	5.9	9.1	3	6.7
Other reasons (Number)	0	0	1	38	670	641	682	11	2,043
Other reasons (Percent)	0	0	3.6	15.1	18	18.4	23.5	16.7	19.5
Total not assessed (Number)	2	1	28	252	3,725	3,484	2,904	66	10,462

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table 10–2B. Counts/Percentages of Students without Scores on the Spring 2023 Keystone Exam: Biology**

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	1	0	771	1,436	937	43	3,188
Extended absence from school (Percent)	33.3	0	34.9	34.6	32.2	62.3	34.1
Non-attempt (Number)	2	0	524	1,346	720	9	2,601
Non-attempt (Percent)	66.7	0	23.7	32.4	24.7	13	27.8
Medical emergency (Number)	0	1	70	125	70	0	266
Medical emergency (Percent)	0	25	3.2	3	2.4	0	2.8
Parental request - Chapter 4 (Number)	0	0	108	210	194	4	516
Parental request - Chapter 4 (Percent)	0	0	4.9	5.1	6.7	5.8	5.5
Parental request - Other reasons (Number)	0	0	89	269	286	4	648
Parental request - Other reasons (Percent)	0	0	4	6.5	9.8	5.8	6.9
Other reasons (Number)	0	3	649	767	705	9	2,133
Other reasons (Percent)	0	75	29.4	18.5	24.2	13	22.8
Total not assessed (Number)	3	4	2,211	4,153	2,912	69	9,352

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table 10–2L. Counts/Percentages of Students without Scores on the Spring 2023 Keystone Exam: Literature**

<b>Reason for Non-Assessment</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Extended absence from school (Number)	1	0	411	1,597	1,033	32	3,074
Extended absence from school (Percent)	16.7	0	45.9	33.9	36.3	53.3	36.1
Non-attempt (Number)	4	0	148	1,409	703	6	2,270
Non-attempt (Percent)	66.7	0	16.5	29.9	24.7	10	26.6
EL in first year in U.S. schools (Number)	0	0	0	1	1	0	2
EL in first year in U.S. schools (Percent)	0	0	0	0	0	0	0
Medical emergency (Number)	0	0	12	177	55	0	244
Medical emergency (Percent)	0	0	1.3	3.8	1.9	0	2.9
Parental request - Chapter 4 (Number)	0	0	22	411	173	6	612
Parental request - Chapter 4 (Percent)	0	0	2.5	8.7	6.1	10	7.2
Parental request - Other reasons (Number)	0	0	48	285	210	3	546
Parental request - Other reasons (Percent)	0	0	5.4	6	7.4	5	6.4
Other reasons (Number)	1	0	254	836	674	13	1,778
Other reasons (Percent)	16.7	0	28.4	17.7	23.7	21.7	20.9
Total not assessed (Number)	6	0	895	4,717	2,849	60	8,527

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

## **DEMOGRAPHIC CHARACTERISTICS OF STUDENTS RECEIVING TEST SCORES**

### **COMPOSITION OF SAMPLE USED IN SUBSEQUENT TABLES**

The following state summary statistic data analyses were completed using the final individual student data file containing records from the Spring administration, which are typically provided to the Pennsylvania Department of Education in July. State summary statistics were based on students who received a total test score on the Spring administration with the exception of students who attended non-public schools or those who were home schooled. Also excluded were students who were non-Keystone proficient.

Demographic data for students taking the Keystone Exams is presented separately for each subject (Tables 10–3A, 10–3B, 10–3L). Results for accommodations received were collected separately by subject and are presented in separate tables as well. For example, tables involving accommodations for Biology are found in Tables 10–4B, 10–5B, 10–6B, and 10–7B.

### **COLLECTION OF STUDENT DEMOGRAPHIC INFORMATION**

Data for analyses involving demographic characteristics were obtained primarily from information supplied by school district personnel through the Pennsylvania Information Management System (PIMS) and subsequently transmitted to DRC. Some data such as accommodation information are recorded by school personnel directly on the student answer document (PPT) or in the DRC INSIGHT Portal Test Setup (CBT) at the time a Keystone Exam is administered.

### **DEMOGRAPHIC CHARACTERISTICS**

Frequency data for each demographic category is presented in Tables 10–3A through 10–3L. Data is presented by grade level with PPT and CBT formats combined into a single composite. Shown at the bottom of the appropriate table is the number of assessed students contributing to summary statistics on which the column percentages are based.

**Table 10–3A. Demographic Characteristics of Students taking the Spring 2023 Keystone Exam: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	3	92	2,874	15,393	32,866	18,420	8,978	105	78,731
Female (Percent)	12	30.5	44.2	51.4	49.7	48	47.4	39	49.1
Male (Number)	6	210	3,626	14,532	33,205	19,957	9,973	164	81,673
Male (Percent)	24	69.5	55.8	48.5	50.2	52	52.6	61	50.9
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	10	39	116	68	46	0	279
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.2	.1	.2	.2	.2	0	.2
Asian (not Hispanic) (Number)	4	123	915	2,050	2,467	1,001	550	7	7,117
Asian (not Hispanic) (Percent)	16	40.7	14.1	6.8	3.7	2.6	2.9	2.6	4.4
Black or African American (not Hispanic) (Number)	0	0	129	1,580	11,065	7,898	4,916	49	25,637
Black or African American (not Hispanic) (Percent)	0	0	2	5.3	16.7	20.6	25.9	18.2	16
Hispanic (any race) (Number)	0	5	271	2,074	10,864	7,047	3,952	73	24,286
Hispanic (any race) (Percent)	0	1.7	4.2	6.9	16.4	18.4	20.9	27.1	15.1
Multi-Racial (not Hispanic) (Number)	0	12	268	1,158	2,995	1,835	902	16	7,186
Multi-Racial (not Hispanic) (Percent)	0	4	4.1	3.9	4.5	4.8	4.8	5.9	4.5
White (not Hispanic) (Number)	4	162	4,898	22,997	38,499	20,483	8,567	124	95,734
White (not Hispanic) (Percent)	16	53.6	75.3	76.8	58.3	53.4	45.2	46.1	59.7
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	9	27	64	44	18	0	162
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.1	.1	.1	.1	.1	0	.1
IEP (not gifted) (Number)	1	8	164	1,243	10,211	9,880	5,512	116	27,135
IEP (not gifted) (Percent)	4	2.6	2.5	4.2	15.5	25.7	29.1	43.1	16.9
Student exited IEP in last 2 years (Number)	0	3	75	361	663	299	142	0	1,543
Student exited IEP in last 2 years (Percent)	0	1	1.2	1.2	1	.8	.7	0	1
Title I (Number)	3	30	755	6,507	20,391	14,479	8,304	107	50,576
Title I (Percent)	12	9.9	11.6	21.7	30.9	37.7	43.8	39.8	31.5
Title III served (Number)	0	0	14	261	3,653	2,722	1,885	58	8,593
Title III served (Percent)	0	0	.2	.9	5.5	7.1	9.9	21.6	5.4
Title III not served (Number)	2	22	770	3,525	7,836	5,221	2,526	39	19,941
Title III not served (Percent)	8	7.3	11.8	11.8	11.9	13.6	13.3	14.5	12.4
Migrant student (Number)	0	0	1	9	117	68	58	4	257
Migrant student (Percent)	0	0	0	0	.2	.2	.3	1.5	.2
EL enrolled first year (Number)	0	0	0	7	60	69	86	0	222
EL enrolled first year (Percent)	0	0	0	0	.1	.2	.5	0	.1
EL enrolled not first year (Number)	0	0	17	303	3,790	2,808	1,903	57	8,878
EL enrolled not first year (Percent)	0	0	.3	1	5.7	7.3	10	21.2	5.5

**Table 10–3A (continued). Demographic Characteristics of Students taking the Spring 2023 Keystone Exam: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	9	66	256	119	43	1	494
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	.1	.2	.4	.3	.2	.4	.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	8	34	102	53	25	0	222
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	.1	.1	.2	.1	.1	0	.1
Former EL no longer monitored (Number)	0	2	73	342	879	582	276	0	2,154
Former EL no longer monitored (Percent)	0	.7	1.1	1.1	1.3	1.5	1.5	0	1.3
LIFE first year (Number)	0	0	0	0	4	4	5	0	13
LIFE first year (Percent)	0	0	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	0	1	30	41	44	4	120
LIFE not first year (Percent)	0	0	0	0	0	.1	.2	1.5	.1
Former EL exited and in 3rd year of monitoring (Number)	0	1	40	149	141	83	52	0	466
Former EL exited and in 3rd year of monitoring (Percent)	0	.3	.6	.5	.2	.2	.3	0	.3
Former EL exited and in 4th year of monitoring (Number)	0	1	34	229	432	105	53	0	854
Former EL exited and in 4th year of monitoring (Percent)	0	.3	.5	.8	.7	.3	.3	0	.5
Foreign exchange student (Number)	0	0	0	0	0	3	10	0	13
Foreign exchange student (Percent)	0	0	0	0	0	0	.1	0	0
Economically disadvantaged (Number)	0	21	844	7,933	34,007	22,667	11,655	169	77,296
Economically disadvantaged (Percent)	0	7	13	26.5	51.5	59.1	61.5	62.8	48.2
Historically Underperforming Subgroup (Number)	1	28	994	8,878	39,395	27,112	14,138	234	90,780
Historically Underperforming Subgroup (Percent)	4	9.3	15.3	29.7	59.6	70.6	74.6	87	56.6
Enrollment in school of residence after Oct 1 (Number)	0	4	33	313	2,643	1,986	1,342	45	6,366
Enrollment in school of residence after Oct 1 (Percent)	0	1.3	.5	1	4	5.2	7.1	16.7	4
Enrollment in district of residence after Oct 1 (Number)	0	3	27	287	2,270	1,739	1,216	42	5,584
Enrollment in district of residence after Oct 1 (Percent)	0	1	.4	1	3.4	4.5	6.4	15.6	3.5
Enrollment as PA resident after Oct 1 (Number)	0	0	20	147	1,312	909	670	20	3,078
Enrollment as PA resident after Oct 1 (Percent)	0	0	.3	.5	2	2.4	3.5	7.4	1.9
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	3	89	1,750	2,884	37,231	9,115	4,613	100	55,785



**Table 10–3A (continued). Demographic Characteristics of Students taking the Spring 2023 Keystone Exam: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	12	29.5	26.9	9.6	56.3	23.7	24.3	37.2	34.8
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	2	18	255	1,090	8,692	4,344	2,760	88	17,249
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	8	6	3.9	3.6	13.2	11.3	14.6	32.7	10.8
Military family (Number)	0	2	31	179	393	203	82	0	890
Military family (Percent)	0	.7	.5	.6	.6	.5	.4	0	.6
Homeless (Number)	0	0	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0	0	0
Foster (Number)	0	0	2	19	248	234	142	5	650
Foster (Percent)	0	0	0	.1	.4	.6	.7	1.9	.4
Home schooled (Number)	0	0	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	19	27	13	13	72
Court/agency placed (Percent)	0	0	0	0	0	.1	.1	4.8	0
Number of assessed students (Number)	25	302	6,503	29,934	66,080	38,380	18,954	269	160,447

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table 10–3B. Demographic Characteristics of Students Taking the Spring 2023 Keystone Exam: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	3	132	25,562	31,715	9,150	161	66,723
Female (Percent)	27.3	48.5	51	48	48.2	45.7	49.1
Male (Number)	2	140	24,538	34,318	9,824	191	69,013
Male (Percent)	18.2	51.5	49	52	51.8	54.3	50.8
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	67	128	43	1	239
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.1	.2	.2	.3	.2
Asian (not Hispanic) (Number)	0	20	3,061	2,406	610	9	6,106
Asian (not Hispanic) (Percent)	0	7.4	6.1	3.6	3.2	2.6	4.5
Black or African American (not Hispanic) (Number)	0	4	5,183	10,260	4,649	84	20,180
Black or African American (not Hispanic) (Percent)	0	1.5	10.3	15.5	24.5	23.9	14.9
Hispanic (any race) (Number)	1	6	4,739	10,228	3,888	90	18,952
Hispanic (any race) (Percent)	9.1	2.2	9.5	15.5	20.5	25.6	14
Multi-Racial (not Hispanic) (Number)	0	9	2,166	2,810	868	19	5,872
Multi-Racial (not Hispanic) (Percent)	0	3.3	4.3	4.3	4.6	5.4	4.3
White (not Hispanic) (Number)	4	233	34,837	40,136	8,901	149	84,260
White (not Hispanic) (Percent)	36.4	85.7	69.5	60.8	46.9	42.3	62.1
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	47	64	15	0	126
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.1	.1	.1	0	.1
IEP (not gifted) (Number)	2	6	5,970	13,181	4,775	126	24,060
IEP (not gifted) (Percent)	18.2	2.2	11.9	20	25.2	35.8	17.7
Student exited IEP in last 2 years (Number)	0	2	545	530	155	0	1,232
Student exited IEP in last 2 years (Percent)	0	.7	1.1	.8	.8	0	.9
Title I (Number)	0	55	11,616	18,904	7,803	170	38,548
Title I (Percent)	0	20.2	23.2	28.6	41.1	48.3	28.4
Title III served (Number)	0	0	1,082	3,336	1,896	55	6,369
Title III served (Percent)	0	0	2.2	5.1	10	15.6	4.7
Title III not served (Number)	1	57	5,832	9,300	2,497	45	17,732
Title III not served (Percent)	9.1	21	11.6	14.1	13.2	12.8	13.1
Migrant student (Number)	0	0	35	83	44	5	167
Migrant student (Percent)	0	0	.1	.1	.2	1.4	.1
EL enrolled first year (Number)	0	0	12	44	89	2	147
EL enrolled first year (Percent)	0	0	0	.1	.5	.6	.1
EL enrolled not first year (Number)	0	0	1,158	3,412	1,891	51	6,512
EL enrolled not first year (Percent)	0	0	2.3	5.2	10	14.5	4.8

**Table 10–3B (continued). Demographic Characteristics of Students Taking the Spring 2023 Keystone Exam: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	143	189	47	0	379
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	.3	.3	.2	0	.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	65	111	26	0	202
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	.1	.2	.1	0	.1
Former EL no longer monitored (Number)	0	2	780	1,276	284	2	2,344
Former EL no longer monitored (Percent)	0	.7	1.6	1.9	1.5	.6	1.7
LIFE first year (Number)	0	0	0	1	5	0	6
LIFE first year (Percent)	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	16	73	46	3	138
LIFE not first year (Percent)	0	0	0	.1	.2	.9	.1
Former EL exited and in 3rd year of monitoring (Number)	0	2	93	120	51	0	266
Former EL exited and in 3rd year of monitoring (Percent)	0	.7	.2	.2	.3	0	.2
Former EL exited and in 4th year of monitoring (Number)	0	0	257	151	44	1	453
Former EL exited and in 4th year of monitoring (Percent)	0	0	.5	.2	.2	.3	.3
Foreign exchange student (Number)	0	0	1	34	9	2	46
Foreign exchange student (Percent)	0	0	0	.1	0	.6	0
Economically disadvantaged (Number)	1	43	18,692	33,095	11,262	235	63,328
Economically disadvantaged (Percent)	9.1	15.8	37.3	50.1	59.3	66.8	46.6
Historically Underperforming Subgroup (Number)	3	48	21,779	39,091	13,501	292	74,714
Historically Underperforming Subgroup (Percent)	27.3	17.6	43.5	59.2	71.1	83	55
Enrollment in school of residence after Oct 1 (Number)	0	2	1,363	2,442	1,330	57	5,194
Enrollment in school of residence after Oct 1 (Percent)	0	.7	2.7	3.7	7	16.2	3.8
Enrollment in district of residence after Oct 1 (Number)	0	2	1,211	2,095	1,211	52	4,571
Enrollment in district of residence after Oct 1 (Percent)	0	.7	2.4	3.2	6.4	14.8	3.4
Enrollment as PA resident after Oct 1 (Number)	0	1	604	1,020	674	21	2,320
Enrollment as PA resident after Oct 1 (Percent)	0	.4	1.2	1.5	3.6	6	1.7
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	29	23,434	12,793	4,723	135	41,114

**Table 10–3B (continued). Demographic Characteristics of Students Taking the Spring 2023 Keystone Exam: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	10.7	46.8	19.4	24.9	38.4	30.3
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	13	5,194	5,650	2,837	91	13,785
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	4.8	10.4	8.6	15	25.9	10.2
Military family (Number)	0	4	360	337	99	1	801
Military family (Percent)	0	1.5	.7	.5	.5	.3	.6
Homeless (Number)	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0
Foster (Number)	0	0	131	286	126	6	549
Foster (Percent)	0	0	.3	.4	.7	1.7	.4
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	10	30	17	11	68
Court/agency placed (Percent)	0	0	0	0	.1	3.1	.1
Number of assessed students (Number)	11	272	50,109	66,036	18,976	352	135,756

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade.

**Table 10–3L. Demographic Characteristics of Students taking the Spring 2023 Keystone Exam: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	1	4	3,442	48,528	7,758	89	59,822
Female (Percent)	14.3	80	49.8	49.1	44.6	39.6	48.5
Male (Number)	3	1	3,467	50,333	9,625	136	63,565
Male (Percent)	42.9	20	50.2	50.9	55.4	60.4	51.5
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	9	168	42	0	219
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.1	.2	.2	0	.2
Asian (not Hispanic) (Number)	0	1	243	4,807	488	1	5,540
Asian (not Hispanic) (Percent)	0	20	3.5	4.9	2.8	.4	4.5
Black or African American (not Hispanic) (Number)	2	0	910	12,630	4,363	44	17,949
Black or African American (not Hispanic) (Percent)	28.6	0	13.2	12.8	25.1	19.6	14.5
Hispanic (any race) (Number)	0	0	936	12,202	3,670	57	16,865
Hispanic (any race) (Percent)	0	0	13.5	12.3	21.1	25.3	13.7
Multi-Racial (not Hispanic) (Number)	0	0	292	4,078	804	15	5,189
Multi-Racial (not Hispanic) (Percent)	0	0	4.2	4.1	4.6	6.7	4.2
White (not Hispanic) (Number)	2	4	4,512	64,873	8,001	108	77,500
White (not Hispanic) (Percent)	28.6	80	65.3	65.6	46	48	62.8
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	7	101	15	0	123
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.1	.1	.1	0	.1
IEP (not gifted) (Number)	2	0	1,165	16,050	5,254	87	22,558
IEP (not gifted) (Percent)	28.6	0	16.9	16.2	30.2	38.7	18.3
Student exited IEP in last 2 years (Number)	0	0	52	898	146	0	1,096
Student exited IEP in last 2 years (Percent)	0	0	.8	.9	.8	0	.9
Title I (Number)	0	0	1,950	23,804	7,160	112	33,026
Title I (Percent)	0	0	28.2	24.1	41.2	49.8	26.8
Title III served (Number)	0	0	171	3,000	1,761	22	4,954
Title III served (Percent)	0	0	2.5	3	10.1	9.8	4
Title III not served (Number)	0	0	1,109	13,247	2,336	39	16,731
Title III not served (Percent)	0	0	16.1	13.4	13.4	17.3	13.6
Migrant student (Number)	0	0	7	76	31	0	114
Migrant student (Percent)	0	0	.1	.1	.2	0	.1
EL enrolled first year (Number)	0	0	2	19	40	0	61
EL enrolled first year (Percent)	0	0	0	0	.2	0	0
EL enrolled not first year (Number)	0	0	191	3,158	1,792	24	5,165
EL enrolled not first year (Percent)	0	0	2.8	3.2	10.3	10.7	4.2

**Table 10–3L (continued). Demographic Characteristics of Students taking the Spring 2023 Keystone Exam: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	18	262	50	0	330
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	.3	.3	.3	0	.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	9	163	28	0	200
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	.1	.2	.2	0	.2
Former EL no longer monitored (Number)	0	0	89	2,053	258	1	2,401
Former EL no longer monitored (Percent)	0	0	1.3	2.1	1.5	.4	1.9
LIFE first year (Number)	0	0	0	1	2	0	3
LIFE first year (Percent)	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	2	48	39	0	89
LIFE not first year (Percent)	0	0	0	0	.2	0	.1
Former EL exited and in 3rd year of monitoring (Number)	0	0	10	171	53	0	234
Former EL exited and in 3rd year of monitoring (Percent)	0	0	.1	.2	.3	0	.2
Former EL exited and in 4th year of monitoring (Number)	0	0	15	232	42	2	291
Former EL exited and in 4th year of monitoring (Percent)	0	0	.2	.2	.2	.9	.2
Foreign exchange student (Number)	0	0	0	37	7	0	44
Foreign exchange student (Percent)	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	0	3,177	42,605	10,672	150	56,604
Economically disadvantaged (Percent)	0	0	46	43.1	61.4	66.7	45.9
Historically Underperforming Subgroup (Number)	2	0	3,655	50,002	13,004	182	66,845
Historically Underperforming Subgroup (Percent)	28.6	0	52.9	50.6	74.8	80.9	54.2
Enrollment in school of residence after Oct 1 (Number)	0	0	326	2,945	1,201	38	4,510
Enrollment in school of residence after Oct 1 (Percent)	0	0	4.7	3	6.9	16.9	3.7
Enrollment in district of residence after Oct 1 (Number)	0	0	266	2,510	1,075	35	3,886
Enrollment in district of residence after Oct 1 (Percent)	0	0	3.9	2.5	6.2	15.6	3.1
Enrollment as PA resident after Oct 1 (Number)	0	0	116	1,205	537	8	1,866
Enrollment as PA resident after Oct 1 (Percent)	0	0	1.7	1.2	3.1	3.6	1.5
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	2,755	16,936	4,131	64	23,886

**Table 10–3L (continued). Demographic Characteristics of Students taking the Spring 2023 Keystone Exam: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	39.9	17.1	23.8	28.4	19.4
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	898	6,634	2,521	56	10,109
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	13	6.7	14.5	24.9	8.2
Military family (Number)	0	0	36	529	81	0	646
Military family (Percent)	0	0	.5	.5	.5	0	.5
Homeless (Number)	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0
Foster (Number)	0	0	32	355	130	6	523
Foster (Percent)	0	0	.5	.4	.7	2.7	.4
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	36	40	19	16	111
Court/agency placed (Percent)	0	0	.5	0	.1	7.1	.1
Number of assessed students (Number)	7	5	6,909	98,872	17,387	225	123,405

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

### **PARTICIPATION BY ADMINISTRATION MODE**

Although there are two administration modes, paper/pencil test (PPT) and computer-based test (CBT), most students are administered PPT.

### **TEST ACCOMMODATIONS PROVIDED**

School personnel supplied information regarding accommodations that a student may have received while taking the Keystone Exams. Accommodations are classified in terms of presentation, response, setting, and timing to enable students to better manage disabilities that hinder their ability to learn and respond to assessments. An accommodations manual entitled, *Accommodations Guidelines* was updated for use with the PSSA and Keystone Exams. This manual may be found on the PDE website at [www.education.pa.gov](http://www.education.pa.gov). A glossary of accommodation terms as applied to the Keystone Exams is provided in Table 10–10 at the end of this chapter.

The frequency with which accommodations were utilized for PPT and CBT formats is summarized separately for each subject exam in Tables 10–4A through 10–7L. Tabled values are based on all students whose scores contributed to state summary statistics in a given Keystone Exam. Because of the very small incidence of usage of many accommodations, combined with the fact that a number of accommodations are primarily accessed by only one of the two administration modes, meaningful comparisons between modes are rather limited. In the following tables, a value of “N/A” denotes those instances in which a particular accommodation does not apply to one of the testing modes.

## **PRESENTATION ACCOMMODATIONS RECEIVED**

Presentation accommodations are those that provide alternate ways for students to access and process printed instructional material and assessments. These include auditory, tactile, visual, and combined auditory/visual modes of presentation. The number of presentation accommodations provided in the Keystone Exams varied by content area and test administration mode.

As depicted in Tables 10–4A through 10–4L, the actual frequencies were quite low, generally representing less than one-tenth of one percent of assessed students in each subject. The most notable exceptions, applicable to Algebra I and Biology only, were the read-aloud accommodations for all or some test items. Among accommodations specific to PPT, the Spanish accommodated form for Algebra I and Biology was the most frequently used. Among accommodations for specific to CBT, the use of audio was the most frequent for Algebra I and Biology. These accommodations are not offered for Literature. The percentages for Literature accommodations were low across all accommodations (less than one-tenth of one percent of assessed students).



**Table 10–4A. Incidence of Presentation Accommodations Received on the Spring 2023 Keystone Exam: Algebra I**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	6	N/A	6
Braille format (Percent)	0	N/A	0
Large print format (Number)	56	N/A	56
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	5	N/A	5
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	774	974	1,748
Some test items/questions read aloud (Percent)	.7	2.1	1.1
All test items/questions read aloud (Number)	452	824	1,276
All test items/questions read aloud (Percent)	.4	1.8	.8
Test items/questions signed (Number)	5	6	11
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for EL student (Number)	41	20	61
Test items/questions interpreted for EL student (Percent)	0	0	0
Amplification device (Number)	11	17	28
Amplification device (Percent)	0	0	0
Magnification device (Number)	7	9	16
Magnification device (Percent)	0	0	0
Color overlay (Number)	4	N/A	4
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	56	51	107
Other (per Accommodations Guidelines) (Percent)	0	.1	.1
Spanish version (Number)	1,786	N/A	1,786
Spanish version (Percent)	1.6	N/A	1.1
Audio (Number)	N/A	2,252	2,252
Audio (Percent)	N/A	4.8	1.4
Color Chooser (Number)	N/A	60	60
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	60	60
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast (Number)	N/A	28	28
Reverse Contrast (Percent)	N/A	.1	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	1	1
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	113,933	46,514	160,447

**Table 10–4B. Incidence of Presentation Accommodations Received on the Spring 2023 Keystone Exam: Biology**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	7	N/A	7
Braille format (Percent)	0	N/A	0
Large print format (Number)	44	N/A	44
Large print format (Percent)	.1	N/A	0
Computer Assistive Technology (Number)	4	N/A	4
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	516	990	1,506
Some test items/questions read aloud (Percent)	.6	2.1	1.1
All test items/questions read aloud (Number)	511	916	1,427
All test items/questions read aloud (Percent)	.6	1.9	1.1
Test items/questions signed (Number)	5	2	7
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for EL student (Number)	27	15	42
Test items/questions interpreted for EL student (Percent)	0	0	0
Amplification device (Number)	9	20	29
Amplification device (Percent)	0	0	0
Magnification device (Number)	3	13	16
Magnification device (Percent)	0	0	0
Color overlay (Number)	2	N/A	2
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	77	61	138
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Spanish version (Number)	1,299	N/A	1,299
Spanish version (Percent)	1.5	N/A	1
Audio (Number)	N/A	2,630	2,630
Audio (Percent)	N/A	5.5	1.9
Color Chooser (Number)	N/A	52	52
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	55	55
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast (Number)	N/A	29	29
Reverse Contrast (Percent)	N/A	.1	0
Refreshable Braille (Number)	N/A	1	1
Refreshable Braille (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	2	2
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	87,881	47,875	135,756

**Table 10–4L. Incidence of Presentation Accommodations Received on the Spring 2023 Keystone Exam: Literature**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	7	N/A	7
Braille format (Percent)	0	N/A	0
Large print format (Number)	44	N/A	44
Large print format (Percent)	.1	N/A	0
Computer Assistive Technology (Number)	6	N/A	6
Computer Assistive Technology (Percent)	0	N/A	0
Amplification device (Number)	11	11	22
Amplification device (Percent)	0	0	0
Magnification device (Number)	5	8	13
Magnification device (Percent)	0	0	0
Color overlay (Number)	3	N/A	3
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	62	30	92
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Color Chooser (Number)	N/A	37	37
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	40	40
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast (Number)	N/A	15	15
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	2	2
Refreshable Braille (Percent)	N/A	0	0
Number of assessed students (Number)	79,254	44,151	123,405

**RESPONSE ACCOMMODATIONS RECEIVED**

Response accommodations permit students to complete assignments, tests, and activities in different ways and to solve or organize problems using some type of assistive device or organizer. The number of response accommodations provided on the Spring Keystone Exams varied by subject.

The frequency with which these accommodations were utilized is summarized in Tables 10–5A through 10–5L. The actual frequencies are quite low, representing less than one-tenth of one percent of assessed students in nearly all instances, regardless of administration mode. The most common response accommodation for Algebra I and Biology was the use of a translation dictionary, however the percent of test-takers utilizing this accommodation was still very low (less than 0.3%).

**Table 10–5A. Incidence of Response Accommodations Received on the Spring 2023 Keystone Exam: Algebra I**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student’s direction (Number)	36	9	45
Test administrator marked multiple-choice responses at student’s direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student’s direction (Number)	38	16	54
Test administrator scribed open-ended responses at student’s direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	71	4	75
Test administrator transcribed student responses (Percent)	.1	0	0
Qualified interpreter translated, transcribed, and/or scribed student’s signed responses (Number)	13	1	14
Qualified interpreter translated, transcribed, and/or scribed student’s signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Number)	25	10	35
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	14	N/A	14
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	5	N/A	5
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	1	0	1
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for EL student (Number)	392	148	540
Translation dictionary for EL student (Percent)	.3	.3	.3
Other (per Accommodations Guidelines) (Number)	36	9	45
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	113,933	46,514	160,447

**Table 10–5B. Incidence of Response Accommodations Received on the Spring 2023 Keystone Exam: Biology**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student’s direction (Number)	20	9	29
Test administrator marked multiple-choice responses at student’s direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student’s direction (Number)	41	19	60
Test administrator scribed open-ended responses at student’s direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	59	10	69
Test administrator transcribed student responses (Percent)	.1	0	.1
Qualified interpreter translated, transcribed, and/or scribed student’s signed responses (Number)	10	1	11
Qualified interpreter translated, transcribed, and/or scribed student’s signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Number)	17	7	24
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	24	N/A	24
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	2	N/A	2
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	2	0	2
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for EL student (Number)	188	128	316
Translation dictionary for EL student (Percent)	.2	.3	.2
Other (per Accommodations Guidelines) (Number)	32	8	40
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	87,881	47,875	135,756

**Table 10–5L. Incidence of Response Accommodations Received on the Spring 2023 Keystone Exam: Literature**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student’s direction (Number)	42	6	48
Test administrator marked multiple-choice responses at student’s direction (Percent)	.1	0	0
Test administrator scribed open-ended responses at student’s direction (Number)	53	10	63
Test administrator scribed open-ended responses at student’s direction (Percent)	.1	0	.1
Test administrator transcribed student responses (Number)	94	4	98
Test administrator transcribed student responses (Percent)	.1	0	.1
Keyboard, word processor, or computer (Number)	46	N/A	46
Keyboard, word processor, or computer (Percent)	.1	N/A	0
Braille/Notetaker (Number)	3	N/A	3
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	2	0	2
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	19	10	29
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	79,254	44,151	123,405

**SETTING ACCOMMODATIONS RECEIVED**

Setting accommodations permit a change in the location in which a student receives instruction or participates in an assessment. In the Keystone Exam administration, there were four categories of setting accommodations, which applied to both administration modes and to each subject exam. As depicted in Tables 10–6A through 10–6L, the most common accommodation was small group setting for both PPT and CBT modes of administration, in which approximately 10% of assessed students were administered the test in a small group setting across subjects.

**Table 10–6A. Incidence of Setting Accommodations Received on the Spring 2023 Keystone Exam: Algebra I**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	14	3	17
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	206	43	249
One-on-one setting (Percent)	.2	.1	.2
Small group setting (Number)	10,102	5,227	15,329
Small group setting (Percent)	8.9	11.2	9.6
Other (per Accommodations Guidelines) (Number)	75	37	112
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Number of assessed students (Number)	113,933	46,514	160,447

**Table 10–6B. Incidence of Setting Accommodations Received on the Spring 2023 Keystone Exam: Biology**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	11	2	13
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	199	53	252
One-on-one setting (Percent)	.2	.1	.2
Small group setting (Number)	8,284	5,684	13,968
Small group setting (Percent)	9.4	11.9	10.3
Other (per Accommodations Guidelines) (Number)	59	36	95
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Number of assessed students (Number)	87,881	47,875	135,756

**Table 10–6L. Incidence of Setting Accommodations Received on the Spring 2023 Keystone Exam: Literature**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	16	1	17
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	175	55	230
One-on-one setting (Percent)	.2	.1	.2
Small group setting (Number)	7,526	5,011	12,537
Small group setting (Percent)	9.5	11.3	10.2
Other (per Accommodations Guidelines) (Number)	56	36	92
Other (per Accommodations Guidelines) (Percent)	.1	.1	.1
Number of assessed students (Number)	79,254	44,151	123,405

**TIMING ACCOMMODATIONS RECEIVED**

Timing accommodations involve a change in the allowable length of time to complete assignments or assessments, including the way in which time is organized. There were four categories of timing accommodations, which applied to both administration modes and to each subject exam. As depicted in Tables 10–7A through 10–7L, the most common accommodation was extended time for both PPT and CBT administration modes. For Algebra I and Literature, the percent of assessed testers receiving an accommodation of extended time was higher than that of Biology (7.8% for Algebra I, 6.8% for Literature, and 3.7% for Biology).

**Table 10–7A. Incidence of Timing Accommodations Received on the Spring 2023 Keystone Exam: Algebra I**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	8,984	3,566	12,550
Extended time (Percent)	7.9	7.7	7.8
Frequent breaks (Number)	748	1,400	2,148
Frequent breaks (Percent)	.7	3	1.3
Changed test schedule (Number)	318	37	355
Changed test schedule (Percent)	.3	.1	.2
Other (per Accommodations Guidelines) (Number)	26	8	34
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	113,933	46,514	160,447

**Table 10–7B. Incidence of Timing Accommodations Received on the Spring 2023 Keystone Exam: Biology**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	2,464	2,561	5,025
Extended time (Percent)	2.8	5.3	3.7
Frequent breaks (Number)	525	1,431	1,956
Frequent breaks (Percent)	.6	3	1.4
Changed test schedule (Number)	244	32	276
Changed test schedule (Percent)	.3	.1	.2
Other (per Accommodations Guidelines) (Number)	17	7	24
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	87,881	47,875	135,756

**Table 10–7L. Incidence of Timing Accommodations Received on the Spring 2023 Keystone Exam: Literature**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	5,457	2,885	8,342
Extended time (Percent)	6.9	6.5	6.8
Frequent breaks (Number)	521	1,224	1,745
Frequent breaks (Percent)	.7	2.8	1.4
Changed test schedule (Number)	206	24	230
Changed test schedule (Percent)	.3	.1	.2
Other (per Accommodations Guidelines) (Number)	27	6	33
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	79,254	44,151	123,405

### ACCOMMODATION RATE FOR NON-IEP AND IEP STUDENTS

A comparison between students without an IEP (non-IEP Students) and those with an IEP (IEP Students) with regard to having received an accommodation is provided in Table 10–8. In this data, accommodated means that a student received one or more of the total number of accommodations available for a given subject; however, this varies somewhat with administration mode. The category of non-accommodated indicates that a student did not receive any accommodations during testing.

As one may expect, the general pattern of findings reveals a consistent and substantially higher percentage of students with IEPs receiving an accommodation, in contrast to non-IEP Students. This same pattern holds true across administration mode. Moreover, the comparison between administration modes revealed that the accommodation rates for IEP students taking a PPT are somewhat lower than accommodation rates for IEP students taking a CBT for Algebra I and Biology, but not Literature.



**Table 10–8A. Accommodation Rate for Non-IEP and IEP Students on the Spring 2023 Keystone Exams: Algebra I**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	95,267	38,045	133,312
Non-Accommodated (Number)	86,290	35,968	122,258
Non-Accommodated (Percent)	90.6	94.5	91.7
Accommodated (Number)	8,977	2,077	11,054
Accommodated (Percent)	9.4	5.5	8.3
IEP Students (Number)	18,666	8,469	27,135
Non-Accommodated (Number)	9,534	3,089	12,623
Non-Accommodated (Percent)	51.1	36.5	46.5
Accommodated (Number)	9,132	5,380	14,512
Accommodated (Percent)	48.9	63.5	53.5

**Table 10–8B. Accommodation Rate for Non-IEP and IEP Students on the Spring 2023 Keystone Exams: Biology**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	72,847	38,849	111,696
Non-Accommodated (Number)	70,034	37,765	107,799
Non-Accommodated (Percent)	96.1	97.2	96.5
Accommodated (Number)	2,813	1,084	3,897
Accommodated (Percent)	3.9	2.8	3.5
IEP Students (Number)	15,034	9,026	24,060
Non-Accommodated (Number)	7,583	3,142	10,725
Non-Accommodated (Percent)	50.4	34.8	44.6
Accommodated (Number)	7,451	5,884	13,335
Accommodated (Percent)	49.6	65.2	55.4

**Table 10–8L. Accommodation Rate for Non-IEP and IEP Students on the Spring 2023 Keystone Exams: Literature**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	64,983	35,864	100,847
Non-Accommodated (Number)	59,776	34,606	94,382
Non-Accommodated (Percent)	92	96.5	93.6
Accommodated (Number)	5,207	1,258	6,465
Accommodated (Percent)	8	3.5	6.4
IEP Students (Number)	14,271	8,287	22,558
Non-Accommodated (Number)	7,229	3,464	10,693
Non-Accommodated (Percent)	50.7	41.8	47.4
Accommodated (Number)	7,042	4,823	11,865
Accommodated (Percent)	49.3	58.2	52.6

## THE INCIDENCE OF ACCOMMODATIONS AND IEP AND EL STATUS

Students with an IEP typically receive an accommodation of some type far more often than non-IEP students. Certain accommodations with very low frequencies are specific to particular disabilities while others are far more common and may also apply to students classified as English Learners (EL). Because the accommodations with the largest frequencies can potentially supply the most stable data when separated out for subgroup analysis, those in most common use are typically selected for display in Tables 10–9A through 10–9L. The most frequently occurring accommodations for assessed students were:

- Some test items/questions read aloud (Algebra I and Biology only)
- All test items/questions read aloud (Algebra I and Biology only)
- Small group setting
- Extended time
- Frequent breaks

Coding for IEP is dichotomous, as students are classified IEP and non-IEP. For purposes of this analysis, an English Learner (EL) is an assessed student classified EL and enrolled in a U.S. school fewer than 12 cumulative months. All other assessed students, including those who have exited an ESL/bilingual program and are in the first or second year of monitoring, are regarded as non-EL.

Customarily, a considerably larger percentage of IEP students receive a given accommodation than non-IEP students. Although less frequent, certain accommodations also have a high frequency rate for EL students. To separate out the effect of being classified IEP or EL, four possible combinations are presented in Tables 10–9A through 10–9L. These include general education students (who are neither IEP nor EL), students who are IEP but non-EL, students who are EL but non-IEP, and students who are both IEP and EL. The bottom row for each administration mode provides the total number of assessed students in each of the four classifications.

Please refer to Appendix I for winter and summer Keystone Exam results.

Across subjects and mode of administration, each subgroup analyzed had sufficient n-counts ( $N > 200$ ). Across subjects, IEP/EL and IEP/non-EL testers had the highest incidence of accommodations. The most common accommodation for these groups were small group setting, which was greater than 40% for CBT and greater than 25% for PPT. Meaning, the small setting accommodation is used in CBT more than PPT.

For the EL/non-IEP group, the most commonly used accommodation for was also small group setting, followed closely by extended time. While small group setting was used more in CBT than PPT, extended time was used more in PPT than CBT. Extended time percentages were somewhat higher for Algebra I than Biology and Literature.

For the general education (non-IEP/non-EL group), the most frequent accommodation used was extended time, and tended to be utilized more on PPT than CBT. Similar to the EL/non-IEP group. Based on the data in the series of 10-9 tables, there is certainly higher rates of use by the IEP and EL groups than the general population.

**Table 10–9A. Incidence of IEP and EL Students Receiving Selected Accommodations on the Spring 2023 Keystone Exam: Algebra I**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Some test items/questions read aloud (Number)	18	72	41	643
PPT - Some test items/questions read aloud (Percent)	3	1	0	3.6
PPT - All test items/questions read aloud (Number)	10	6	10	426
PPT - All test items/questions read aloud (Percent)	1.7	.1	0	2.4
PPT - Small group setting (Number)	169	631	1,067	8,235
PPT - Small group setting (Percent)	28.5	9.1	1.2	45.6
PPT - Extended time (Number)	62	554	6,662	1,706
PPT - Extended time (Percent)	10.4	8	7.5	9.4
PPT - Frequent breaks (Number)	18	48	85	597
PPT - Frequent breaks (Percent)	3	.7	.1	3.3
PPT - Number assessed (Number)	594	6,899	88,368	18,072
CBT - Some test items/questions read aloud (Number)	28	9	51	886
CBT - Some test items/questions read aloud (Percent)	11.9	.7	.1	10.8
CBT - All test items/questions read aloud (Number)	8	1	30	785
CBT - All test items/questions read aloud (Percent)	3.4	.1	.1	9.5
CBT - Small group setting (Number)	107	85	394	4,641
CBT - Small group setting (Percent)	45.3	6.2	1.1	56.4
CBT - Extended time (Number)	45	76	1,450	1,995
CBT - Extended time (Percent)	19.1	5.5	4	24.2
CBT - Frequent breaks (Number)	33	16	86	1,265
CBT - Frequent breaks (Percent)	14	1.2	.2	15.4
CBT - Number assessed (Number)	236	1,371	36,674	8,233
Total - Some test items/questions read aloud (Number)	46	81	92	1,529
Total - Some test items/questions read aloud (Percent)	5.5	1	.1	5.8
Total - All test items/questions read aloud (Number)	18	7	40	1,211
Total - All test items/questions read aloud (Percent)	2.2	.1	0	4.6
Total - Small group setting (Number)	276	716	1,461	12,876
Total - Small group setting (Percent)	33.3	8.7	1.2	48.9
Total - Extended time (Number)	107	630	8,112	3,701
Total - Extended time (Percent)	12.9	7.6	6.5	14.1
Total - Frequent breaks (Number)	51	64	171	1,862
Total - Frequent breaks (Percent)	6.1	.8	.1	7.1
Total - Number assessed (Number)	830	8,270	125,042	26,305

**Table 10–9B. Incidence of IEP and EL Students Receiving Selected Accommodations on the Spring 2023 Keystone Exam: Biology**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Some test items/questions read aloud (Number)	7	50	20	439
PPT - Some test items/questions read aloud (Percent)	1.6	1	0	3
PPT - All test items/questions read aloud (Number)	14	2	20	475
PPT - All test items/questions read aloud (Percent)	3.2	0	0	3.3
PPT - Small group setting (Number)	134	375	839	6,936
PPT - Small group setting (Percent)	30.2	7.8	1.2	47.5
PPT - Extended time (Number)	29	255	1,248	932
PPT - Extended time (Percent)	6.5	5.3	1.8	6.4
PPT - Frequent breaks (Number)	14	25	67	419
PPT - Frequent breaks (Percent)	3.2	.5	.1	2.9
PPT - Number assessed (Number)	443	4,838	68,009	14,591
CBT - Some test items/questions read aloud (Number)	25	11	22	932
CBT - Some test items/questions read aloud (Percent)	11.1	1	.1	10.6
CBT - All test items/questions read aloud (Number)	15	3	13	885
CBT - All test items/questions read aloud (Percent)	6.7	.3	0	10.1
CBT - Small group setting (Number)	118	81	409	5,076
CBT - Small group setting (Percent)	52.4	7	1.1	57.7
CBT - Extended time (Number)	44	41	461	2,015
CBT - Extended time (Percent)	19.6	3.6	1.2	22.9
CBT - Frequent breaks (Number)	30	7	97	1,297
CBT - Frequent breaks (Percent)	13.3	.6	.3	14.7
CBT - Number assessed (Number)	225	1,153	37,696	8,801
Total - Some test items/questions read aloud (Number)	32	61	42	1,371
Total - Some test items/questions read aloud (Percent)	4.8	1	0	5.9
Total - All test items/questions read aloud (Number)	29	5	33	1,360
Total - All test items/questions read aloud (Percent)	4.3	.1	0	5.8
Total - Small group setting (Number)	252	456	1,248	12,012
Total - Small group setting (Percent)	37.7	7.6	1.2	51.4
Total - Extended time (Number)	73	296	1,709	2,947
Total - Extended time (Percent)	10.9	4.9	1.6	12.6
Total - Frequent breaks (Number)	44	32	164	1,716
Total - Frequent breaks (Percent)	6.6	.5	.2	7.3
Total - Number assessed (Number)	668	5,991	105,705	23,392

**Table 10–9L. Incidence of IEP and EL Students Receiving Selected Accommodations on the Spring 2023 Keystone Exam: Literature**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Small group setting (Number)	125	214	825	6,362
PPT - Small group setting (Percent)	30	6.3	1.3	45.9
PPT - Extended time (Number)	45	189	3,971	1,252
PPT - Extended time (Percent)	10.8	5.6	6.4	9
PPT - Frequent breaks (Number)	12	17	62	430
PPT - Frequent breaks (Percent)	2.9	.5	.1	3.1
PPT - Number assessed (Number)	416	3,387	61,596	13,855
CBT - Small group setting (Number)	96	96	351	4,468
CBT - Small group setting (Percent)	48.5	7.8	1	55.2
CBT - Extended time (Number)	52	43	837	1,953
CBT - Extended time (Percent)	26.3	3.5	2.4	24.1
CBT - Frequent breaks (Number)	28	16	92	1,088
CBT - Frequent breaks (Percent)	14.1	1.3	.3	13.5
CBT - Number assessed (Number)	198	1,225	34,639	8,089
Total - Small group setting (Number)	221	310	1,176	10,830
Total - Small group setting (Percent)	36	6.7	1.2	49.4
Total - Extended time (Number)	97	232	4,808	3,205
Total - Extended time (Percent)	15.8	5	5	14.6
Total - Frequent breaks (Number)	40	33	154	1,518
Total - Frequent breaks (Percent)	6.5	.7	.2	6.9
Total - Number assessed (Number)	614	4,612	96,235	21,944

## GLOSSARY OF ACCOMMODATION TERMS

Table 10–10 provides a brief description of accommodation terms as used in the PSSA and Keystone Exams. Accommodation data was supplied by school personnel as noted in the left column of the table. The right column contains an explanation derived from the PDE publication, *Accommodations Guidelines*. This manual may be found on the PDE website at [www.education.pa.gov](http://www.education.pa.gov).

**Table 10–10. Glossary of Accommodation Terms as Applied in the PSSA and 2023 Keystone Exams**

Type of Testing Accommodation	Explanation
<b>Student used the following Online Presentation Accommodations</b>	
Braille format	Students may use a Braille format of the test. Answers must then be transcribed into the answer booklet without alteration.
Large print format	Students with visual impairments may use a large print format. Answers must then be transcribed into the answer booklet without alteration.
Magnification device	Devices to magnify print may be used for students with visual impairments and/or print disabilities.
Color overlay	Students with visual impairments may place a color overlay on a printed page of the test document to make text more readable.
Computer assistive technology (e.g., electronic screen reader) (PDE approval required)	Students with severe visual disabilities that prevent them from accessing instructional material or performing the skill may use computer assistive technology; however, PDE must approve the program and functions prior to the test window.
Test items/questions/text-dependent analysis signed	Deaf/hearing impaired students may receive test directions from a qualified interpreter. Signing is also permitted for PSSA ELA writing section multiple-choice items, and text-dependent analysis questions and all items in PSSA mathematics and science and for Keystone Algebra and Biology.
Test items/questions/text-dependent analysis interpreted for EL	A qualified interpreter may translate directions or clarify instructions for the assessments. The interpreter may translate but not define specific words or test questions on the PSSA mathematics, science, ELA writing section multiple-choice items, and text-dependent analysis questions and Keystone Algebra and Biology exams.
Some or all test items/questions/text-dependent analysis read aloud	Students unable to decode text visually may have items/questions read aloud for PSSA ELA writing section multiple-choice items, and text-dependent analysis questions and all items in PSSA mathematics and science and for Keystone Algebra and Biology; however, words may not be defined.
Amplification device	In addition to using hearing aids, an amplification device to enhance clarity may be required.
Other (PDE approval required)	Other presentation accommodations indicated in the <i>Accommodations Guidelines</i> may be provided; however, PDE approval is required prior to the test window.
Spanish version for PSSA (Math and Science) and Keystone (Algebra and Biology)	Students whose first language is Spanish and who have been enrolled in U.S. schools for fewer than three years may take this version.
<b>Student used the following Online Presentation Accommodations</b>	
Audio	The online test form reads permissible test directions and items for a student unable to decode text. The accommodation must be marked within the test engine system. The accommodation is available on PSSA mathematics, science, ELA writing section multiple-choice items, and text-dependent analysis questions and Keystone Algebra and Biology exams.
Video sign language (per <i>Accommodations Guidelines</i> )	Eligible students who use a sign language accommodation during instructional periods may use a VSL on the PSSA mathematics and science assessments, or Keystone Algebra I and Biology.
Color chooser or contrasting text chooser	The use of this accommodation enables a visually impaired student to change the background color or text color to make text more readable.
Refreshable Braille	This accommodation allows students to use a screen reader to produce a Braille translation output.

**Table 10–10 (continued). Glossary of Accommodation Terms as Applied in the PSSA and 2023 Keystone Exams**

Type of Testing Accommodation	Explanation
<b>Student used the following Response Accommodations</b>	
Braille/Note taker (per <i>Accommodations Guidelines</i> )	Students using this device as part of their regular instructional program may use it on the assessments; however, without thesaurus, spelling, or grammar checker.
Test administrator scribed <b>open-ended</b> responses at student's direction	A test administrator may record word-for-word exactly what a student dictated directly into the test booklet. This includes MC and OE responses Keystone Algebra, Biology, and Literature tests and PSSA mathematics and science.
Test administrator marked <b>multiple-choice</b> responses at student's direction	A test administrator may mark an answer booklet at the direction of a student (e.g., a student may point to an MC answer with the test administrator marking the response in the answer booklet).
Test administrator transcribed student responses (per <i>Accommodations Guidelines</i> )	A test administrator may transcribe (copy) a student's written, typed, or keyed response into a standard answer booklet.
Qualified Interpreter translated, transcribed, and/or scribed student's <b>signed</b> responses	A qualified interpreter may interpret a student's <b>signed</b> responses into written English for Keystone Algebra and Biology exams, and PSSA mathematics and science assessments. Interpreters are not permitted to make corrections or change the meaning of the response.
Qualified Interpreter translated, transcribed, and/or scribed EL student responses	A qualified interpreter may interpret a student's non-English oral responses into written English for Keystone Algebra and Biology exams, and PSSA mathematics and science assessments. Interpreters are not permitted to make corrections or change the meaning of the response.
Augmentative communication device	Students with severe communication difficulties may use a special device to convey responses, which must be transcribed into the answer booklet by the test administrator.
Keyboard, word processor, or computer (per <i>Accommodations Guidelines</i> )	This is an allowable accommodation as a typing function only for students with the identified need. Online test should be considered for students who prefer/need to type open-ended responses. Supports such as dictionaries, thesauri, spell checkers, and grammar checkers must be turned off. Answers must then be transcribed into the answer booklet without alteration.
Translation dictionary for EL student	A word-to-word dictionary that translates native language to English (or vice versa) without word definitions or pictures is allowed on any portion of the Keystone Algebra and Biology exams, and PSSA mathematics and science tests.
Computer assistive technology (e.g., electronic screen reader) (PDE approval required)	Students with blindness or extremely low vision may use dictate text into a computer. Responses must be transcribed verbatim into student's regular answer booklet.
Other (per <i>Accommodations Guidelines</i> or PDE approval)	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Documentation must be provided to PDE.
<b>Student used the following Setting Accommodations</b>	
Hospital/home testing	A student who is confined to a hospital or to home during the testing window may be tested in that environment.
One-on-one setting	One-on-one settings are necessitated in certain instances, such as to reduce distraction or in the use of certain devices. A separate room may be used to reduce distraction.
Small group setting	Some students may require a test setting with fewer students or a setting apart from all other students to minimize distraction.
Other (per <i>Accommodations Guidelines</i> or PDE approval)	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Documentation must be provided to PDE.

**Table 10–10 (continued). Glossary of Accommodation Terms as Applied in the PSSA and 2023 Keystone Exams**

Type of Testing Accommodation	Explanation
<b>Student used the following Timing Accommodations</b>	
Extended time	Extended time may be allotted for each section of the test as a planned accommodation to enable students to finish.
Frequent breaks	Frequent breaks (breaks within a test section) may be scheduled for the completion of each test section; however, a test section must be completed within one school day.
Changed test schedule	Students whose disabilities prevent them from following a regular, planned test schedule may follow an individual schedule that enables test completion.
Other (per <i>Accommodations Guidelines</i> or PDE approval)	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Documentation must be provided to PDE.



## CHAPTER ELEVEN: CLASSICAL ITEM STATISTICS

---

This chapter provides an overview of the two most familiar item-level statistics obtained from classical (traditional) item analysis: item difficulty and item discrimination. The following results pertain not only to the operational Keystone Exams items but also to the embedded field-test items. Other statistics such as Rasch item statistics and test-level statistics are discussed in Chapter Twelve and Chapter Seventeen, respectively.

### ITEM-LEVEL STATISTICS

Appendix J provides classical item statistics for all items (i.e., operational and embedded field-test items) in the Algebra I, Biology, and Literature Keystone Exams. Results are organized by administration and content area. These statistics represent the item characteristics most often used to determine whether an item functioned properly and/or how a group of students performed on a particular item. The item statistics in Appendix J include  $N$ , the number of students taking the test form for which there are valid test scores;  $p$ -values (denoted as  $P_{Val}$ ) for multiple-choice (MC) items and item means (denoted as Mean) for constructed-response (CR) items (indicators of item difficulty); proportions of students who chose each response option for MC items (denoted as  $P(A)$ ,  $P(B)$ ,  $P(C)$ , and  $P(D)$ ) or gained each score point for CR items (denoted as  $P(0)$ ,  $P(1)$ ,  $P(2)$ ,  $P(3)$ , and  $P(4)$ ); proportions of students who omitted an item (denoted as  $P(-)$  for MC items and  $P(B)$  for CR items); item-total correlations (denoted as Total, indicators of item discrimination); item-total correlations for each response option for MC items (denoted as  $Corr(A)$ ,  $Corr(B)$ ,  $Corr(C)$ ,  $Corr(D)$ ) and gained score point for CR items (denoted as  $Corr(1)$ ,  $Corr(2)$ ,  $Corr(3)$ ,  $Corr(4)$ ).

Appendix J also provides the Rasch measurement-based statistics in columns Rasch, Infit, and Outfit. Detailed explanations of these statistics can be found in Chapter Twelve. The differential item functioning (DIF) analysis on the embedded field-test items is provided as well. The detailed explanation of DIF codes can be found in Chapter Five.

### ITEM DIFFICULTY

At the most general level, an item's difficulty is indicated by its mean score in some specified group (e.g., grade level).

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

In the mean score formula above, the individual item scores ( $x_i$ ) are summed and then divided by the total number of students ( $n$ ). For MC items, student scores are represented by 0's and 1's (0 = wrong, 1 = right). With 0/1 scoring, the equation above also represents the number of students correctly answering the item divided by the total number of students. So, this is also the *proportion correct* for the item, or as it is better known, the  $p$ -value. In theory,  $p$ -values can range from 0.00<sup>1</sup> to 1.00 on the proportion-correct scale. For example, if an item has a  $p$ -value of 0.89, it means 89 percent of the students answered the item correctly. Additionally, this value might also suggest that the item is relatively easy and/or the students who attempted the item are relatively high achievers. In other words, item difficulty and student ability are somewhat confounded.

For CR items, mean scores can range from the minimum possible score (usually zero) to the maximum possible score (e.g., four points in the case of Algebra I CR items and three points in the case of Biology and Literature CR items). Sometimes a *pseudo p*-value is provided for a CR item by dividing the mean item score by the maximum possible item score.

---

<sup>1</sup> For MC items with four response options, pure random guessing would lead to an expected  $p$ -value of 0.25.

The minimum and maximum extremes of the difficulty scale are virtually never seen in applied practice. However, understanding what those values are helps illustrate that relatively lower values correspond to more difficult items and that relatively higher values correspond to easier items. (Because of this, some assert that this index would be better referred to as the item's *easiness*.)

Item difficulty is an important consideration for the Keystone Exams because of the various student achievement levels in Pennsylvania (Below Basic, Basic, Proficient, and Advanced). Items that are either very hard or very easy provide little information about student differences in achievement. However, an item answered correctly by a high percentage of students would suggest that the knowledge or skill the item taps has been mastered by most students. Conversely, an item answered correctly by a low percentage of students would suggest that few students have mastered the knowledge or skill the item taps. So, on a criteria-referenced test like the Keystone Exams, a test development goal is to include a wide range of item difficulties.

Utilizing the proportion of students who chose each option can be helpful for verifying keys. For example, if a large proportion of students chose a distractor instead of the correct answer of an MC item, it may indicate that the key is not correct. Proportion of students omitting or not reaching an item is useful for identifying issues related to testing time and item/test layout. Keystone Exams are not speed tests. Therefore, students should have enough time to take the exams. An omit rate greater than 5% for a single item could be an indication that students were not given enough time to take the test or an indication of an item/test layout problem. For example, some students might accidentally skip an item that follows a lengthy stem.

## ITEM DISCRIMINATION

At the most general level, item discrimination<sup>2</sup> indicates an item's ability to differentiate between high and low achievers. It is expected that students with high ability (i.e., those who perform well on the Keystone Exams overall) would be more likely to answer any given item correctly, while students with low ability (i.e., those who perform poorly on the Keystone Exams overall) would be more likely to answer the same item incorrectly. For the Keystone Exams, Pearson's product-moment correlation coefficient between item scores and test scores is used to indicate discrimination. As commonly practiced, Data Recognition Corporation (DRC) removes the item score from the total score so that the resulting correlations will not be spuriously high. The correlation coefficient can range from negative 1.0 to positive 1.0. If the aforementioned expectation is met (high-scoring students tend to get the item right while low-scoring students do not), the correlation between the item score and the total test score will be both positive and noticeably large in its magnitude (i.e., well above zero) indicating that the item is a good discriminator between high- and low-ability students.

Item-total correlation for each option is another indicator of an item's ability to differentiate between high and low achievers. It is expected that students with high ability would be less likely to choose any distractors, while students with low ability would be more likely to choose a distractor. In other words, the item-total correlations for the distractors are expected to be negative.

In summary, the correlation will be positive in value when the mean test score of the students answering the item correctly is higher than the mean test score of the students answering the item incorrectly.<sup>3</sup> In other words, students who did well on the total test tended to do well on the item as well. However, an interaction can exist between item discrimination and item difficulty. Items answered correctly (or incorrectly) by a large proportion of examinees (i.e., items with extreme *p*-values) can have reduced power to discriminate and thus can have lower correlations.

Discrimination is an important consideration for the operational Keystone Exams because the use of more discriminating items on a test is associated with more reliable test scores. This in turn means that score estimates will be more precise (i.e., there will be smaller confidence intervals around the scores) and, perhaps more importantly, that more accurate performance level placements will be made. The issues of reliability, confidence intervals, and performance level classifications are further discussed in Chapter Eighteen.

---

<sup>2</sup> As noted earlier, the discrimination index for dichotomous MC items is typically referred to as the *point-biserial correlation coefficient*. For CR items, the term *item-test correlation* is sometimes used.

<sup>3</sup> It is legitimate to view the point-biserial correlation as a standardized mean. A positive value indicates that students who chose that response had a higher mean score than the average score; a negative value indicates that students who chose that response had a lower mean score than the average score.

## SCATTER PLOTS OF ITEM DISCRIMINATION AND DIFFICULTY

Figure 11–1 contains a series of scatter plots showing item discrimination (i.e., item-total correlation on  $y$ -axis) on the item difficulty (i.e.,  $p$ -value on  $x$ -axis) for the operational items on each exam by test administration. These plots provide information about item discrimination and difficulty in a single visual image for each Keystone Exam. The  $x$ - and  $y$ -axes visually represent many important distributional indices, including the distribution of  $p$ -values on the  $x$ -axis, the distribution of item discrimination on the  $y$ -axis, and the relationship between  $p$ -values and discrimination as the scatter plot.

It should be noted that pseudo  $p$ -values are used for CR items in these plots. Of course, the bivariate relationship between discrimination and difficulty is also presented. One does not usually expect any type of trend here. However, as noted earlier, it is often the case that items with extreme difficulties can have lower discrimination values, so this can be revealed in such a plot.

**Figure 11–1. Scatter Plots of Item Discrimination and Difficulty**

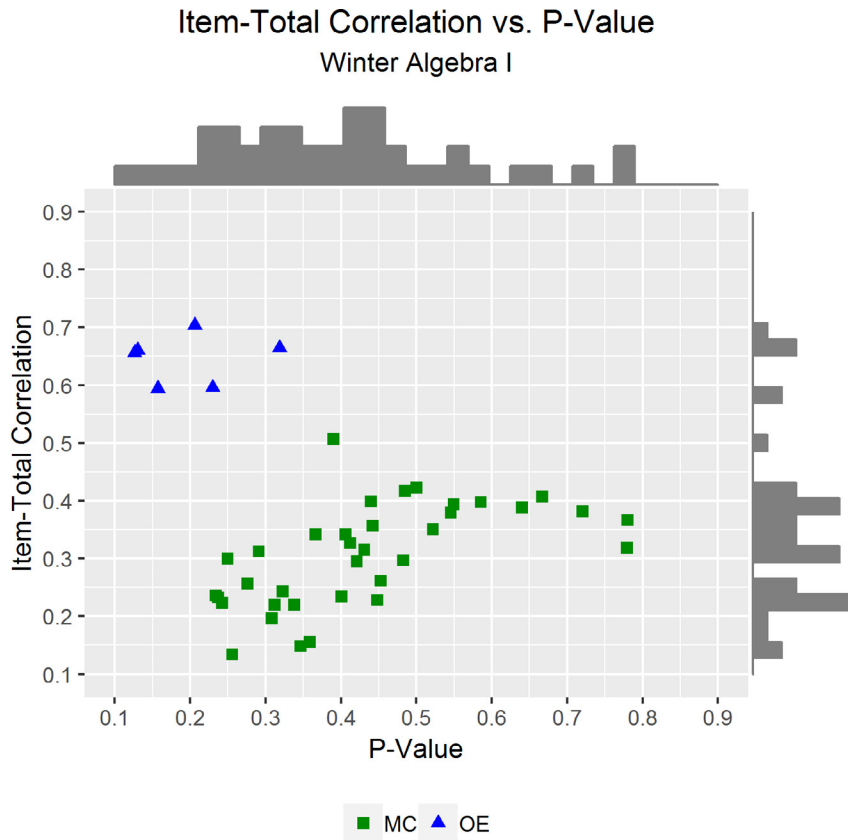
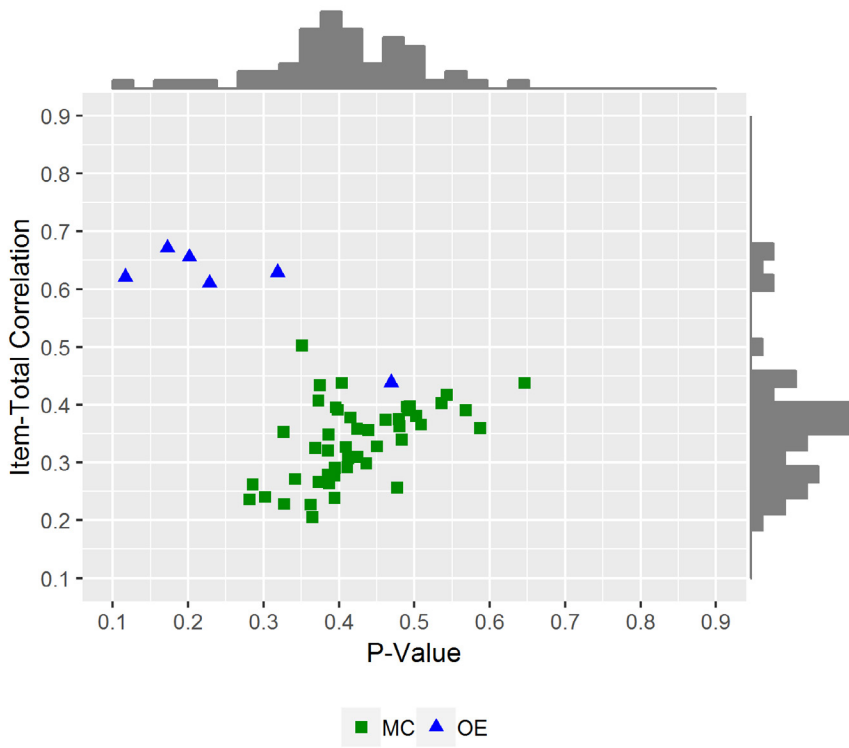


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty

Item-Total Correlation vs. P-Value

Winter Biology



Item-Total Correlation vs. P-Value

Winter Literature

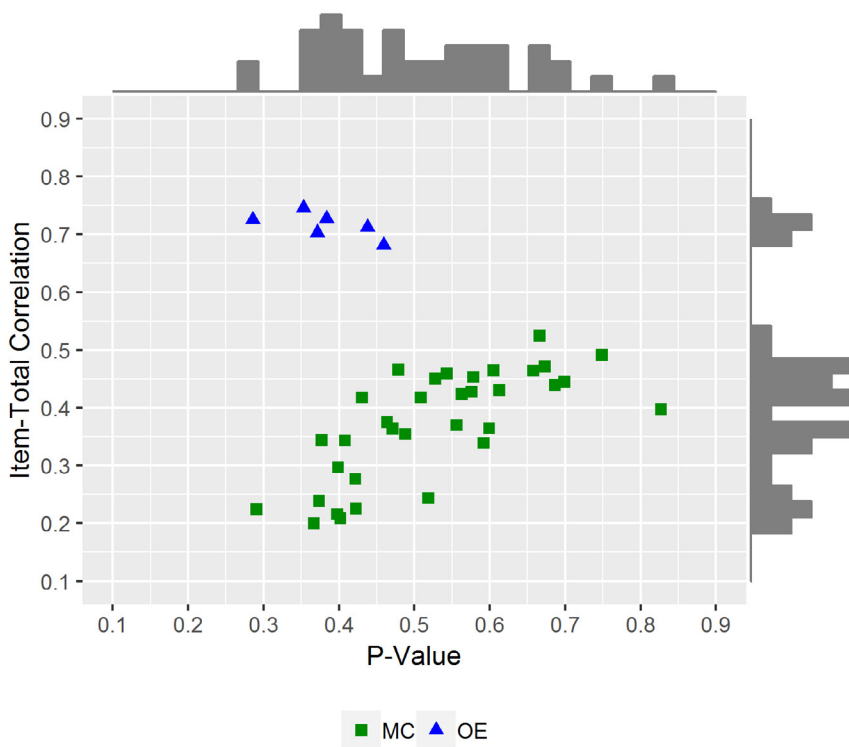
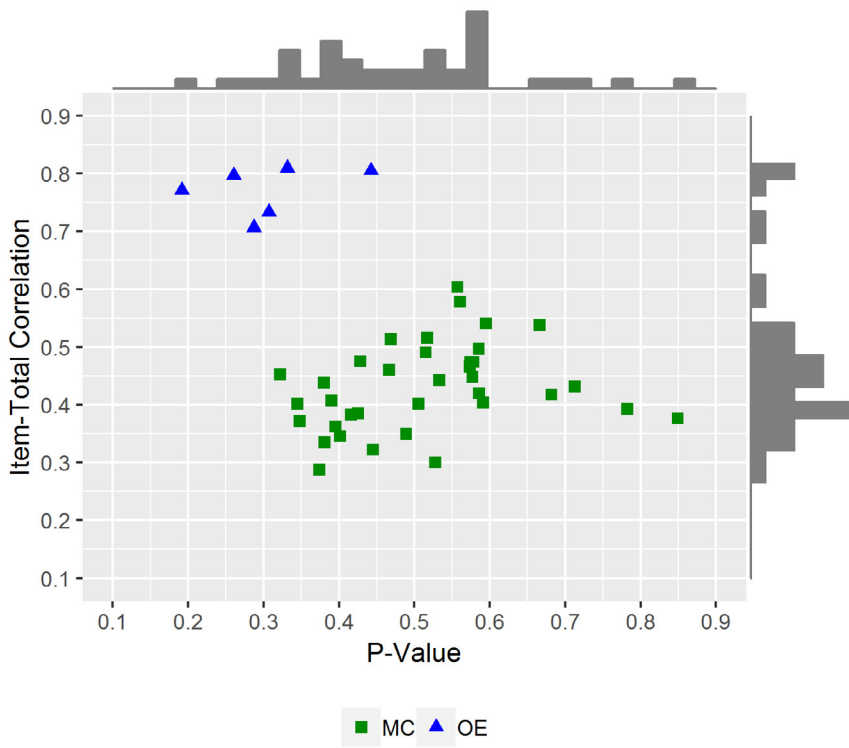


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty

Item-Total Correlation vs. P-Value

Spring Algebra I



Item-Total Correlation vs. P-Value

Spring Biology

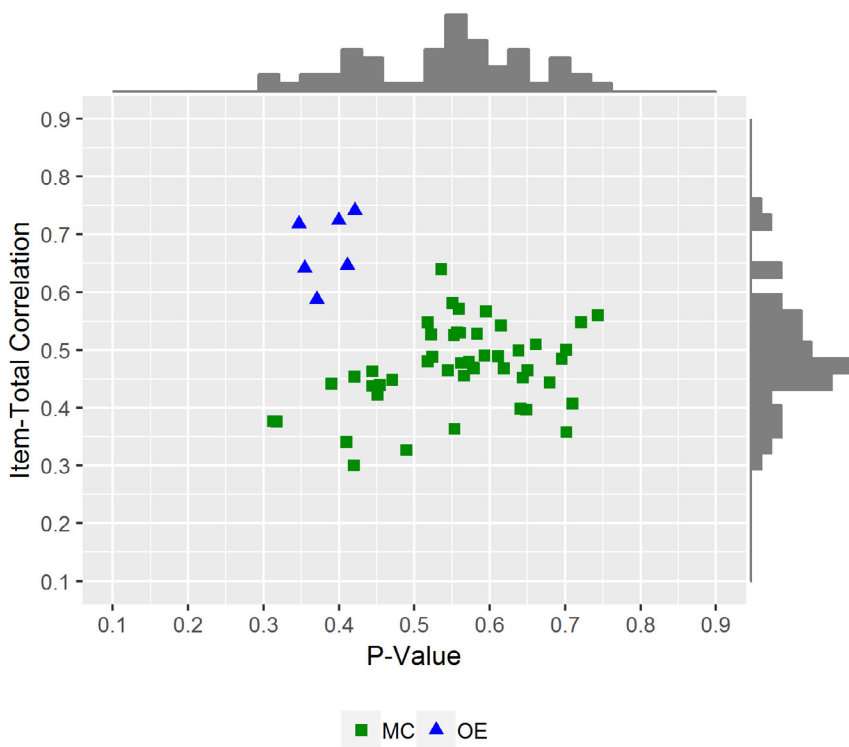
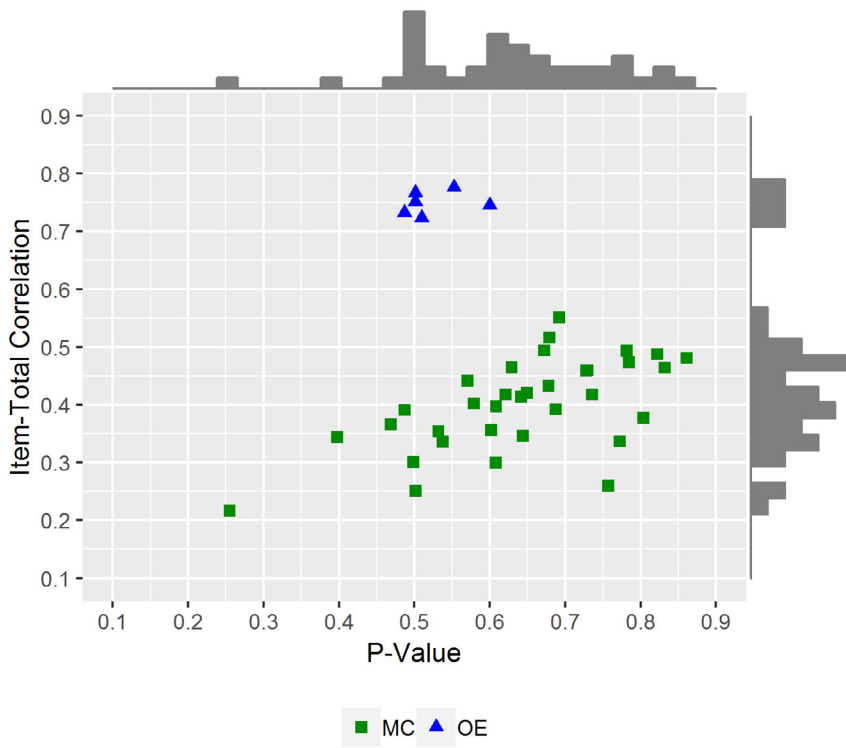


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty

Item-Total Correlation vs. P-Value

Spring Literature



Item-Total Correlation vs. P-Value

Summer Algebra I

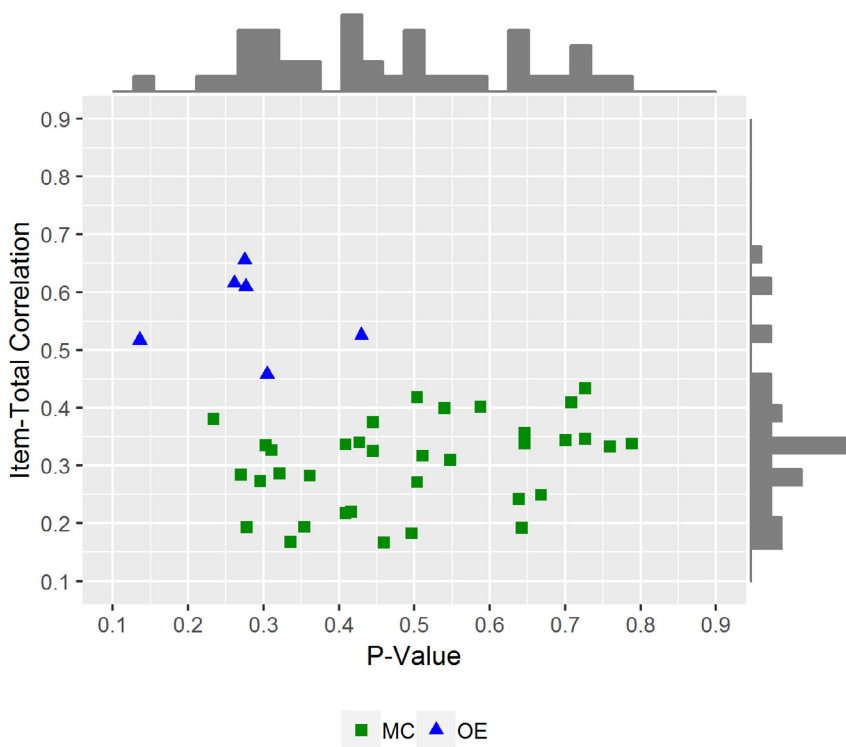
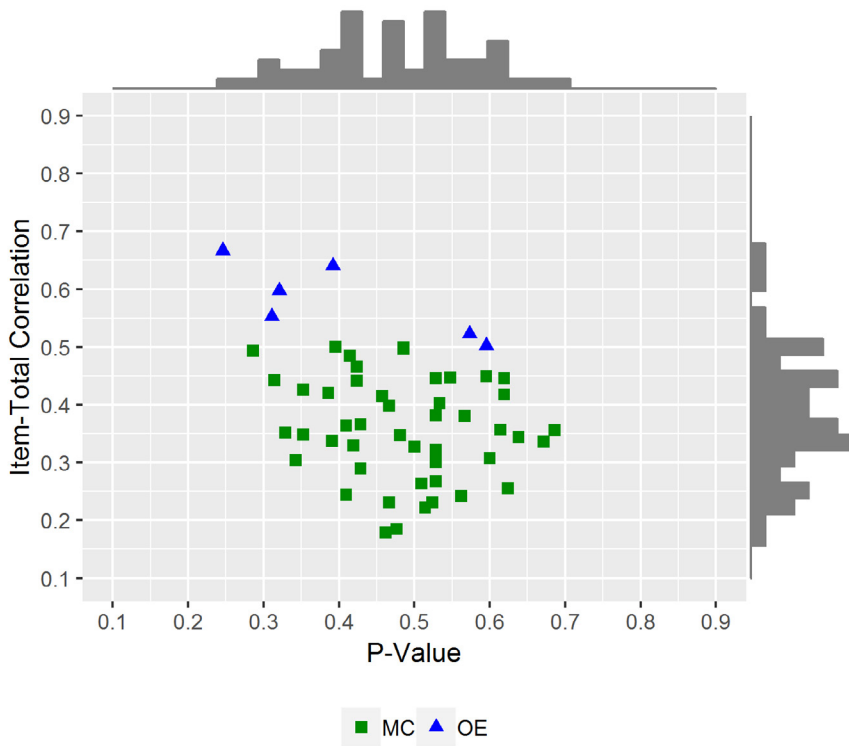


Figure 11–1 (continued). Scatter Plots of Item Discrimination and Difficulty

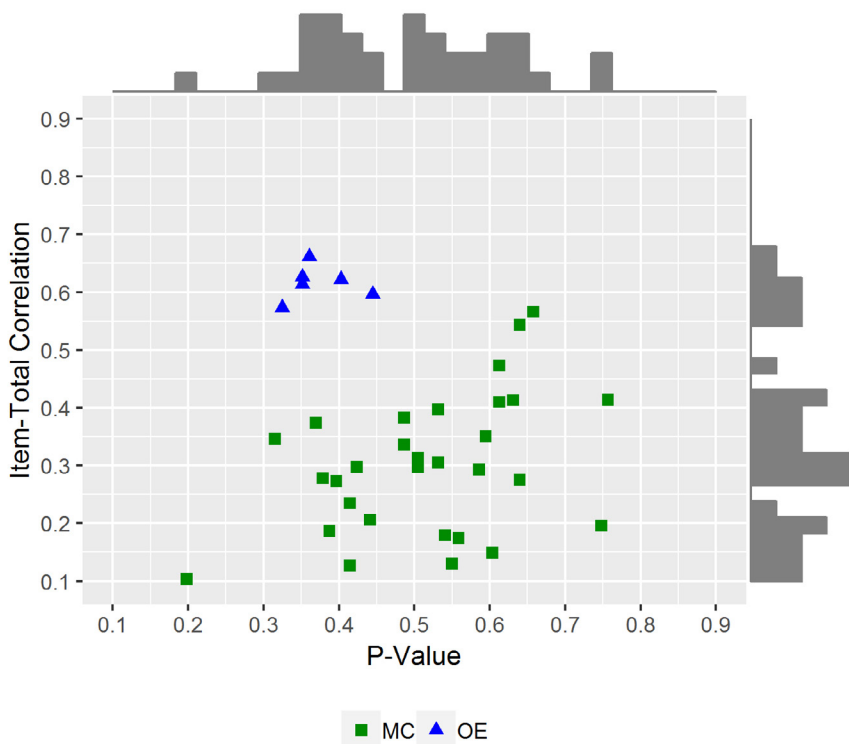
Item-Total Correlation vs. P-Value

Summer Biology



Item-Total Correlation vs. P-Value

Summer Literature



## OBSERVATIONS AND INTERPRETATIONS

Table 11–1 provides the mean and median  $p$ -values and median<sup>4</sup> item-total correlations for the operational MC and CR items in each content area. The mean  $p$ -value for the operational MC items ranged from 0.42 to 0.64 with standard deviation (SD) ranging from 0.08 to 0.16, whereas the mean  $p$ -values for the CR items ranged from about 0.19 to 0.53 with standard deviation ranging from 0.03 to 0.15. The median item-test correlations ranged from 0.28 to 0.47 and 0.57 to 0.78 for the MC and CR items, respectively. The CR correlations tended to be higher than the MC correlations, which is not surprising because the CR items include more score points and tend to be better at discriminating between low and high achieving students than MC items.

It is impossible to make global conclusions about the overall test quality from these item statistics alone. With that caveat in mind, the results presented in this chapter indicate that the item difficulties and discriminations were within expected and acceptable ranges.

**Table 11–1. Mean and Median Statistics for Operational MC and CR Items**

Administration	Content Area	MC Items Mean $p$ -Value	MC Items SD $p$ -Value	MC Items Median $p$ -Value	MC Items Median I-T Corr.	CR Items Mean $p$ -Value	CR Items SD $p$ -Value	CR Items Median $p$ -Value	CR Items Median I-T Corr.
Winter	Algebra I	0.43	0.15	0.42	0.30	0.19	0.07	0.18	0.66
Winter	Biology	0.42	0.08	0.41	0.34	0.25	0.13	0.21	0.62
Winter	Literature	0.53	0.12	0.52	0.39	0.38	0.06	0.38	0.72
Spring	Algebra I	0.52	0.12	0.52	0.43	0.30	0.08	0.30	0.78
Spring	Biology	0.56	0.10	0.56	0.47	0.38	0.03	0.38	0.68
Spring	Literature	0.64	0.13	0.65	0.41	0.53	0.04	0.51	0.75
Summer	Algebra I	0.49	0.16	0.48	0.31	0.28	0.09	0.28	0.57
Summer	Biology	0.48	0.10	0.48	0.35	0.41	0.15	0.36	0.58
Summer	Literature	0.49	0.14	0.50	0.28	0.37	0.04	0.36	0.62

*Note.* I-T Corr. is the item-total test score correlation; SD represents the standard deviation.

<sup>4</sup> Given that the value of the item-total correlation coefficient is not a linear function of the magnitude of the relation between the item and total test scores, the median instead of the mean of the item-total correlation was calculated for this statistic.



## CHAPTER TWELVE: RASCH ITEM CALIBRATION

---

The particular item response theory (IRT) model used for the Keystone Exams is based on the work of Georg Rasch. Rasch models have had a long-standing presence in applied testing programs and have been the methodology continually used to calibrate the Pennsylvania System of School Assessment (PSSA) items in recent history. Consequently, this model was also chosen for the Keystone Exams. IRT has several advantages over classical test theory, so it has become the standard procedure for analyzing item response data in large-scale assessments. However, IRT models make several strong assumptions related to dimensionality, local item independence, and model-data fit. Resulting inferences derived from any application of IRT rest strongly on the degree to which the underlying assumptions are met.

This chapter outlines the procedures used for calibrating the operational Keystone Exams items. Generally, item calibration is the process of assigning a difficulty-parameter estimate to each item on an assessment so that they are placed on a common scale. This chapter briefly introduces the Rasch model, reports the results from evaluations of the adequacy of the Rasch assumptions, and summarizes the Rasch item statistics for the Keystone Exams in Algebra I, Biology, and Literature.

### DESCRIPTION OF THE RASCH MODEL

The Rasch partial credit model (RPCM; Wright & Masters, 1982) was used to calibrate Keystone Exams item response data because both multiple-choice (MC) and constructed-response (CR) items were part of the assessment. The RPCM extends the Rasch model (Rasch, 1960) for dichotomous (0, 1) items to accommodate polytomous CR items. Under the RPCM, for a given item  $i$  with  $m_i$  score categories, the probability of person  $n$  scoring  $x$  ( $x = 0, 1, 2, \dots, m_i$ ) is given by:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^x (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\beta_n - \delta_{ij})}, \quad x = 0, 1, \dots, m_i$$

where  $\beta_n$  represents a student's proficiency (ability) level, and  $\delta_{ij}$  is the step difficulty of the  $j$ th step on item  $i$ . For dichotomous MC items, the RPCM reduces to the standard Rasch model and the single step difficulty is referred to as the item's difficulty. The Rasch model predicts the probability of person  $n$  getting item  $i$  correct as follows:

$$\Phi_{mi}(X = 1 | \beta_n) = \frac{\exp(\beta_n - \delta_{ij})}{1 + \exp(\beta_n - \delta_{ij})}.$$

The Rasch model places both student ability and item difficulty (estimated in terms of log-odds or logits) on the same continuum. When the model assumptions are met, it also provides person ability estimates that are independent of the items employed in the assessment, and, conversely, estimates item difficulty independently of the sample of examinees. (As noted in Chapter Eleven, interpretation of item  $p$ -values confounds item difficulty and student ability.)

## **SOFTWARE AND ESTIMATION ALGORITHM**

Item calibration was implemented via WINSTEPS computer program (Linacre, 2019), which employs unconditional (UCON), joint-maximum-likelihood estimation (JMLE).

## **SAMPLE CHARACTERISTICS**

The characteristics of calibration samples are reported in Chapter Nine. These samples only include the students who attempted the tests. All omitted and multiple responses (more than one response selected) for MC items were scored as incorrect answers (coded as 0s) for calibration purposes.

## **CHECKING RASCH ASSUMPTIONS**

Because the Rasch model was the basis of all calibration and equating analyses associated with the Keystone Exams, the validity of the inferences from these results depends on the degree to which the assumptions of the model are met and how well the test data fits the model. Therefore, it is important to check these assumptions. This section evaluates the dimensionality of the data, local item independence, and model-data fit at the item level. Though a variety of methods are available for assessing these issues, the Rasch analyses and criteria available from WINSTEPS were used here. It should be noted that only operational items were analyzed since they are the basis of student scores.

Given Keystone Exams use a pre-equating design (see details in Chapter Fifteen), a pre-equating verification process was used to evaluate the stability of item parameters (refer to Chapter Fifteen on Equating for additional information). After reviewing the analyses results for the winter, spring, and summer administrations, a decision was made to use the item difficulty estimated from the field-test data to generate the raw-to-scaled score conversion tables. In this chapter, the adequacy of the Rasch calibration assumptions was checked with all the item difficulties anchored to the pre-equated values.

## **UNIDIMENSIONALITY**

Rasch models assume that one dominant dimension determines the difference in students' performances. WINSTEPS provides results from a principal components analysis (PCA) that can be used to assess the unidimensionality assumption. Different from standard applications of PCA, WINSTEPS conducts its PCA on the response residuals, not the original observations. That is, the primary dimension from the Rasch model is removed first and then the residual variance is analyzed. The purpose of the analysis is to verify whether any other dominant components exist among the residuals (i.e., they account for a practically significant amount of residual variance). If any other dimensions are found, the unidimensionality assumption would be violated.

For Keystone Exams, the standardized residuals were used to conduct the PCA because simulation studies indicate that it gives the most accurate reflection of secondary dimensions in the items (Linacre, 2019). Table 12–1 presents the PCA results by administration for each content area. The results include the eigenvalues and variance explained by each component. As can be seen from the table, the eigenvalues for the first component are much larger than those for the rest of the components. The first component explained about 25.8 to 45.9 percent of the total variance. The rest of the components explained only a small percentage of variance (less than 3% in most cases). These results suggest that each of the Keystone Exams essentially measure a single dominant dimension.

**Table 12–1. Results from PCA of Residuals in WINSTEPS**

<b>Administration/ Content Area</b>	<b>Component</b>	<b>Eigenvalue</b>	<b>Variance Explained</b>
Winter Algebra I	1	30.4	42.0%
Winter Algebra I	2	2.0	2.7%
Winter Algebra I	3	1.6	2.2%
Winter Algebra I	4	1.3	1.8%
Winter Algebra I	5	1.3	1.7%
Winter Biology	1	21.1	28.1%
Winter Biology	2	1.6	2.1%
Winter Biology	3	1.5	2.0%
Winter Biology	4	1.4	1.9%
Winter Biology	5	1.3	1.7%
Winter Literature	1	21.7	35.1%
Winter Literature	2	2.9	4.6%
Winter Literature	3	1.4	2.3%
Winter Literature	4	1.4	2.2%
Spring Algebra I	1	35.6	45.9%
Spring Algebra I	2	1.8	2.3%
Spring Algebra I	3	1.5	1.9%
Spring Algebra I	4	1.3	1.7%
Spring Algebra I	5	1.3	1.6%
Spring Biology	1	29.7	35.5%
Spring Biology	2	1.7	2.0%
Spring Biology	3	1.5	1.7%
Spring Biology	4	1.3	1.5%
Spring Biology	5	1.2	1.4%
Spring Literature	1	23.3	36.8%
Spring Literature	2	3.0	4.8%
Spring Literature	3	1.5	2.3%
Spring Literature	4	1.3	2.1%
Spring Literature	5	1.2	1.9%
Summer Algebra I	1	25.8	38.1%
Summer Algebra I	2	2.1	3.0%
Summer Algebra I	3	1.9	2.7%
Summer Algebra I	4	1.8	2.6%
Summer Algebra I	5	1.6	2.4%
Summer Biology	1	22.7	29.6%
Summer Biology	2	2.3	3.0%
Summer Biology	3	2.2	2.9%
Summer Biology	4	2.0	2.6%
Summer Biology	5	1.9	2.4%
Summer Literature	1	13.9	25.8%
Summer Literature	2	3.8	7.1%
Summer Literature	3	2.3	4.2%
Summer Literature	4	2.1	3.9%
Summer Literature	5	2.0	3.7%

## LOCAL INDEPENDENCE

Local independence (LI) is a fundamental assumption of IRT. No relationship should exist between examinees' responses to different items after accounting for the abilities measured by a test. In formal statistical terms, a test  $X$  that is composed of items  $X_1, X_2, \dots, X_I$  is locally independent with respect to the latent variable  $\delta^n$  if, for all  $\mathbf{x} = (x_1, x_2, \dots, x_I)$  and  $\delta^n$ ,

$$P_n(\mathbf{X} = \mathbf{x} | \delta^n) = \prod_{i=1}^I P(X_i = x_i | \delta^n).$$

This formula essentially states that the probability of any pattern of responses across all items ( $\mathbf{x}$ ), after conditioning on the abilities ( $\delta^n$ ) measured by the test, should be equal to the product of the conditional probabilities across each item (cf. the multiplication rule for independent events where the joint probabilities are equal to the product of the associated marginal probabilities).

The equation above shows the condition after satisfying the strong form of local independence. A *weak form* of local independence (WLI) was proposed by McDonald (1979). The distinction is important as many indicators of local dependency are actually framed by WLI. The requirement here would be for the conditional covariances of all pairs of item responses, conditioned on the abilities, to be equal to zero. When this assumption is met, the joint probability of responses to an item pair, conditioned on abilities, is the product of the probabilities of responses to these two items, as shown below. (This is a *weaker* form because higher-order dependencies among items are allowed.) Based on the WLI, the following expression can be derived:

$$P(X_i = x_i, X_j = x_j | \delta^n) = P(X_i = x_i | \delta^n)P(X_j = x_j | \delta^n).$$

Marais and Andrich (2008) pointed out that local item dependence in the Rasch model can occur in two ways that some may not distinguish. The first way occurs when the assumption of unidimensionality is violated. Here, other nuisance dimensions besides a dominant dimension also determine students' performance (this can be called *trait dependence*). The second violation occurs when responses to an item depend on responses to another. This is a violation of statistical independence and can be called *response dependence*. Many people treat the assumptions of *unidimensionality* and *local independence* as one phenomenon and believe that once unidimensionality holds, that local independence also holds. By distinguishing the two sources of local dependence, one can see that while local independence can be related to unidimensionality, the two are different assumptions, and therefore, require different tests.

Residual item correlations estimated in WINSTEPS for each item pair were used to assess the local dependence among the Keystone Exams items. In general, these residuals are computed as follows. First, expected item performance based on the Rasch model is determined using ability and item parameter estimates. Next, deviations (residuals) between the examinees' expected and observed performance is determined for each item. Finally, for each item pair, a correlation between the respective deviations is computed.

Two types of residual correlations are available in WINSTEPS: raw and standardized residuals. It should be noted that the raw score residual correlation essentially corresponds to Yen's  $Q_3$  index (Yen, 1993), a popular LI statistic. The expected value for the  $Q_3$  statistic is approximately  $-1/(k-1)$  when no local dependence exists, where  $k$  is test length. Thus, the expected  $Q_3$  values should be approximately -0.026 or larger for the Keystone Exams (since Literature is the shortest test with 40 items). Index values that are greater than 0.20 indicate a degree of local dependence that probably should be examined by test developers (Chen & Thissen, 1997). Since the two residual correlations are very similar, the default *standardized residual correlation* in WINSTEPS was used for these analyses. Table 12–2 shows the summary statistics—mean, SD, minimum (Min), maximum (Max), and several percentiles ( $P_{10}$ ,  $P_{25}$ ,  $P_{50}$ ,  $P_{75}$ ,  $P_{90}$ )—for all the residual correlations for each test. The total number of item pairs ( $N$ ) and the number of pairs with absolute residual correlations greater than 0.20 are also reported in this table. The mean residual correlations were very close to the criteria of -0.02. The correlations were tended to be smaller for Algebra and Biology than Literature, which was likely influenced by the passage-based structure of the examination. There was a higher number of items flagged in the summer administration, which could be related to the sample of test-takers.

**Table 12–2. Summary of Item Residual Correlations**

Administration	Content Area	N	Mean	SD	Min	P10	P23	P50	P75	P90	Max	Ct > 0.20
Winter	Algebra I	861	-0.02	0.03	-0.11	-0.06	-0.05	-0.03	0.00	0.02	0.22	1
Winter	Biology	1431	-0.02	0.02	-0.12	-0.04	-0.03	-0.02	-0.01	0.01	0.11	0
Winter	Literature	780	-0.02	0.05	-0.15	-0.09	-0.05	-0.02	0.00	0.02	0.33	10
Spring	Algebra I	861	-0.02	0.03	-0.11	-0.06	-0.04	-0.02	0.00	0.01	0.08	0
Spring	Biology	1431	-0.02	0.02	-0.11	-0.04	-0.03	-0.02	-0.01	0.01	0.07	0
Spring	Literature	780	-0.02	0.06	-0.15	-0.10	-0.05	-0.02	0.00	0.02	0.31	11
Summer	Algebra I	861	-0.02	0.06	-0.24	-0.10	-0.07	-0.03	0.02	0.06	0.24	4
Summer	Biology	1431	-0.02	0.07	-0.26	-0.11	-0.06	-0.02	0.03	0.07	0.25	13
Summer	Literature	780	-0.02	0.11	-0.37	-0.17	-0.11	-0.03	0.05	0.12	0.51	58

### ITEM FIT

WINSTEPS estimates two item-fit statistics (infit and outfit) for evaluating the degree to which the Rasch model predicts the observed item responses. Each fit statistic can be expressed as a mean square (MnSq) statistic or on a standardized metric (Z-standardized with mean = 0 and variance = 1). Mean-square values are more oriented toward practical significance, while Z-standardized values are more oriented toward statistical significance. Though both are informative, the Z-standardized values are very likely too sensitive to the large sample sizes observed on the Keystone Exams. In this situation it is recommended that the Z-standardized values be ignored if the mean-square values are acceptable (Linacre, 2018).

Both infit and outfit mean-square represent the average of standardized residual variance (the difference between the observed score and the Rasch estimated score divided by the square root of the Rasch model variance). The difference is that the outfit statistic gives all examinees equal weight in computing the fit and tends to be affected more by unexpected responses far from the person, item, or rating scale category measure (i.e., it is more sensitive to outlying, off-target, low-information responses). The infit statistic is weighted by the examinee locations relative to item difficulty and tends to be affected more by unexpected responses close to the person, item, or rating scale category measure (i.e., informative, on-target responses). Some feel that extreme infit values are a greater threat to the measurement process than extreme outfit values since most tests intend to measure the on-target population rather than extreme outliers.

The expected mean-square value is 1.0, and it can range from 0 to infinity. Deviation in excess of the expected value can be interpreted as noise or lack of fit between the items and the model. Values lower than the expected value can be interpreted as item redundancy or overfitting items (too predictable, too much redundancy), and values greater than the expected value indicate underfitting items (too unpredictable, too much noise). Rules of thumb regarding practically significant mean-square values vary. More conservative users might prefer items with mean-square values that range from 0.8 to 1.2. Others believe reasonable test results can be achieved with values from 0.5 to 1.5. In the results below, values outside of 0.7 to 1.3 are given practical importance.

Table 12–3 presents the summary statistics of infit and outfit mean-square statistics for the Keystone Exams in Algebra I, Biology, and Literature, including the mean, SD, and minimum and maximum values. The number of items within the range of [0.7, 1.3] is also reported in Table 12–3. As can be seen, the mean values for both fit statistics were close to 1.00 for all the exams. Most of the items had mean-squared fit statistics ranging from 0.7 to 1.3.

**Table 12–3. Summary of Infit and Outfit Mean-Square (MnSq) Statistics**

Admin	Content Area	N	Infit MnSq Mean	Infit MnSq SD	Infit MnSq Min	Infit MnSq Max	Infit MnSq [0.7, 1.3]	Outfit MnSq Mean	Outfit MnSq SD	Outfit MnSq Min	Outfit MnSq Max	Outfit MnSq [0.7, 1.3]
Winter	Algebra I	42	1.03	0.16	0.68	1.44	37	1.07	0.25	0.60	1.69	34
Winter	Biology	54	1.03	0.15	0.69	1.51	49	1.04	0.19	0.63	1.55	47
Winter	Literature	40	1.03	0.18	0.69	1.43	35	1.08	0.28	0.68	1.75	28
Spring	Algebra I	42	1.02	0.14	0.70	1.31	41	1.04	0.19	0.66	1.39	39
Spring	Biology	54	0.97	0.09	0.78	1.24	54	0.96	0.12	0.67	1.32	52
Spring	Literature	40	1.03	0.13	0.73	1.33	39	1.10	0.22	0.73	1.61	34
Summer	Algebra I	42	0.97	0.13	0.74	1.24	42	0.97	0.16	0.70	1.32	40
Summer	Biology	54	0.99	0.10	0.78	1.24	54	0.99	0.15	0.78	1.47	52
Summer	Literature	40	0.99	0.18	0.56	1.26	36	1.01	0.23	0.56	1.59	32

## RASCH ITEM STATISTICS

As noted earlier, the Rasch model expresses item difficulty (and student ability) in units referred to as logits, rather than on the percent-correct metric. In the simplest case, a logit is a transformed  $p$ -value with the average  $p$ -value becoming a logit of zero. In this form, logits resemble  $z$ -scores or standard normal deviates; a very difficult item might have a logit of +4.0 and a very easy item might have a logit of –4.0. However, they have no formal relationship to the normal distribution.

The logit metric has several mathematical advantages over  $p$ -values. Logits have an interval scale, meaning that two items with logits of 0.0 and +1.0, respectively, are the same distance apart as two items with logits of +3.0 and +4.0. Logits are not dependent on the ability level of the students. For example, a test form can have a mean logit of zero, regardless of if the average item  $p$ -value for the student sample is 0.8 or 0.3.

The standard Rasch calibration procedure arbitrarily sets the mean difficulty of the items on any form at zero. Under normal circumstances where all students are administered the same set of items, any item with a  $p$ -value lower than the average item on the form receives a positive logit and any item with a  $p$ -value higher than the average receives a negative logit. Consequently, the logits for any calibration relate to an arbitrary origin defined by the center of items on that form. Logits for both item difficulties and student abilities are placed on the same scale and relate to the same mean item difficulty.

There are a number of other choices that could be made for centering the item difficulties. Rather than using all the items, the origin could be defined by content. For the Keystone Exams, all test forms in a particular content area share the same operational item set. All items on each form can then be easily adjusted to a single origin by defining the origin as the mean of the operational items. With this done, the origins for all the forms will be statistically equal. For example, items on any two forms that are equally difficult will now have statistically equal logit difficulties.

Appendix J reports the item statistics including classical and Rasch logit difficulties for all the operational items and the field-test items embedded in the spring forms. Table 12–4 summarizes the Rasch logit difficulties of the operational items on each test for each administration. The mean of MC item difficulty was no longer equal to zero as it was for the 2011 administration. This is because all the item parameter estimates were anchored to the pre-equated values in order to be estimated on the baseline scale. The mean item difficulties for MC items were less than that of CR items. Table 12–4 also shows the mean standard errors (SE) of the item difficulties, which were relatively small, suggesting that items were calibrated with very small errors. The minimum (Min) and maximum (Max) values and standard deviations (SD) suggest the Keystone Exams items covered a relatively wide range of difficulties.

**Table 12–4. Summary of Rasch Item Difficulties**

Administration/Content Area	Item Types	N	Mean Item Difficulty	Mean SE	SD Item Difficulty	Min Item Difficulty	Max Item Difficulty
Winter Algebra I	All	42	0.39	0.01	0.93	-1.82	2.17
Winter Algebra I	MC	36	0.19	0.01	0.83	-1.82	1.38
Winter Algebra I	CR	6	1.58	0.01	0.51	0.82	2.17
Winter Biology	All	54	0.22	0.01	0.52	-0.96	1.22
Winter Biology	MC	48	0.13	0.01	0.48	-0.96	1.21
Winter Biology	CR	6	0.93	0.01	0.29	0.55	1.22
Winter Literature	All	40	0.35	0.01	0.73	-1.41	1.47
Winter Literature	MC	34	0.25	0.01	0.74	-1.41	1.47
Winter Literature	CR	6	0.93	0.01	0.09	0.82	1.08
Spring Algebra I	All	42	0.48	0.01	0.75	-1.62	1.95
Spring Algebra I	MC	36	0.31	0.01	0.67	-1.62	1.34
Spring Algebra I	CR	6	1.46	0.00	0.37	0.87	1.95
Spring Biology	All	54	0.19	0.01	0.49	-0.85	1.10
Spring Biology	MC	48	0.12	0.01	0.48	-0.85	1.10
Spring Biology	CR	6	0.73	0.00	0.09	0.59	0.85
Spring Literature	All	40	0.28	0.01	0.77	-1.27	1.71
Spring Literature	MC	34	0.17	0.01	0.78	-1.27	1.71
Spring Literature	CR	6	0.86	0.00	0.29	0.44	1.21
Summer Algebra I	All	42	0.40	0.13	0.76	-1.10	2.26
Summer Algebra I	MC	36	0.24	0.13	0.68	-1.10	1.54
Summer Algebra I	CR	6	1.35	0.07	0.48	0.90	2.26
Summer Biology	All	54	0.32	0.14	0.45	-0.88	1.25
Summer Biology	MC	48	0.27	0.15	0.44	-0.88	1.25
Summer Biology	CR	6	0.68	0.09	0.46	0.03	1.10
Summer Literature	All	40	0.30	0.20	0.68	-1.00	1.84
Summer Literature	MC	34	0.20	0.21	0.69	-1.00	1.84
Summer Literature	CR	6	0.86	0.13	0.16	0.71	1.16

## ITEM DIFFICULTY-STUDENT ABILITY MAP

The distributions of the Rasch item logits (item difficulty estimates) are shown on the item difficulty-student ability maps presented in Figure 12–1. In each item-student map, the top bar graph displays the student distribution on the logit scale, and the bottom displays markers of item difficulty parameter estimates. MC items are represented by a circle and CR items are represented by a square. As noted earlier, the Rasch model enables placement of both items and students on the same scale. Consequently, one can easily visualize information regarding the relationship between the distributions of item difficulty and student ability. The vertical red lines (from left to right) represent the performance cut-points: below basic/basic, basic/proficient, and proficient/advanced. On the top plot, the logit represents lower abilities (negative values) to higher abilities (positive values), whereas on the bottom plot the logit represents easier items (negative values) to harder items (positive values). To achieve precise measures of student ability, the student distribution should mirror the item distribution.

The common pattern seen across all maps was that the item difficulties were comparable to the student ability levels. It is also important to understand where the items are providing more-accurate measurement. This issue is addressed more fully in Chapter Eighteen (see Figure 18–2).

**Figure 12–1. Item Difficulty-Student Ability Maps**

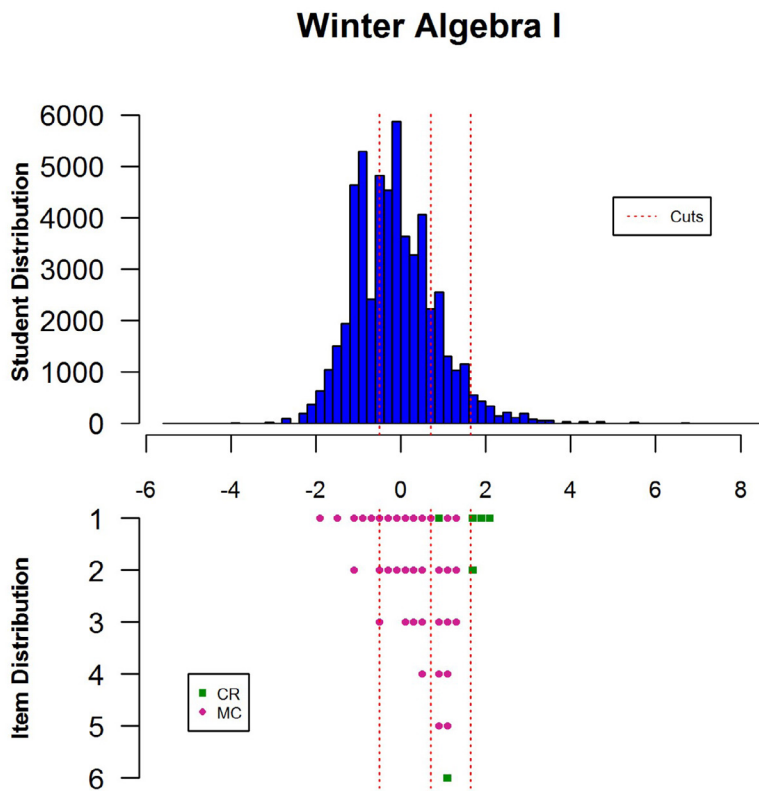
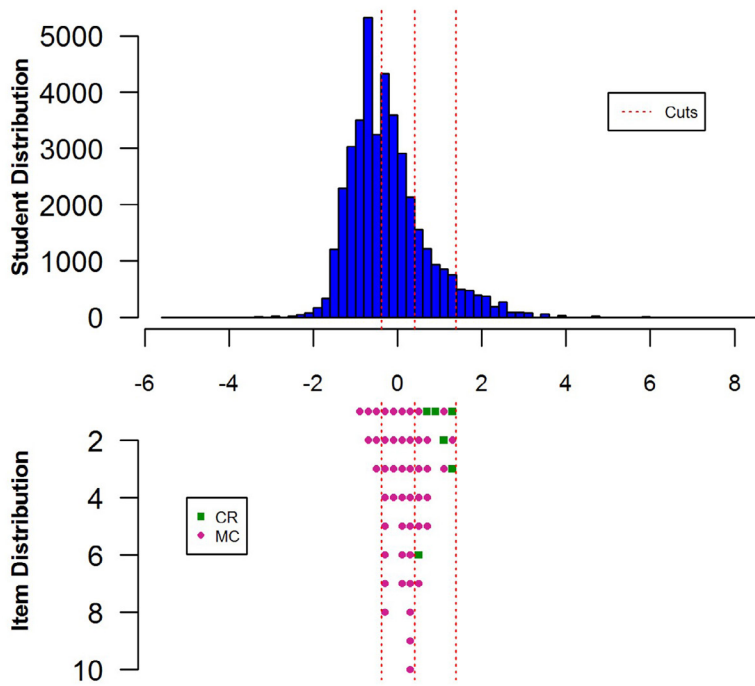




Figure 12–1 (continued). Item Difficulty-Student Ability Maps

### Winter Biology



### Winter Literature

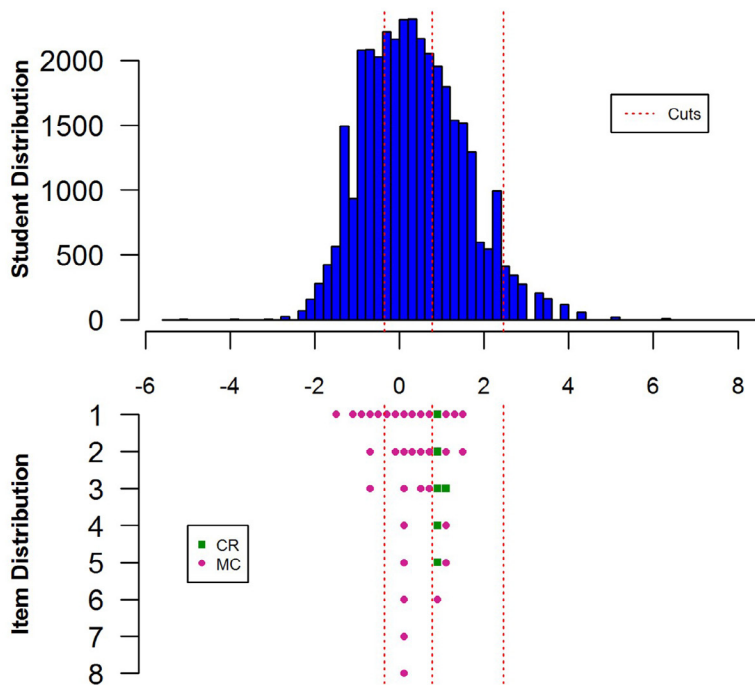
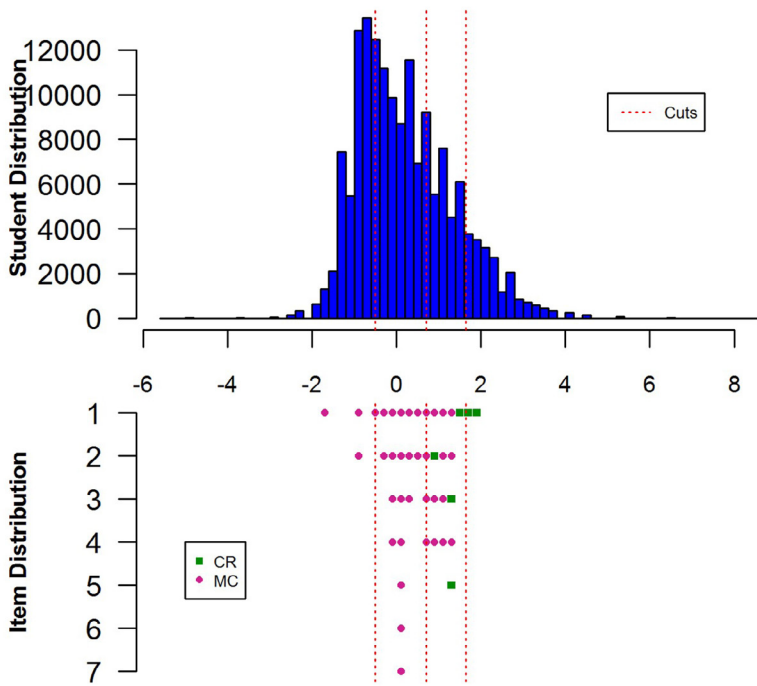


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

### Spring Algebra I



### Spring Biology

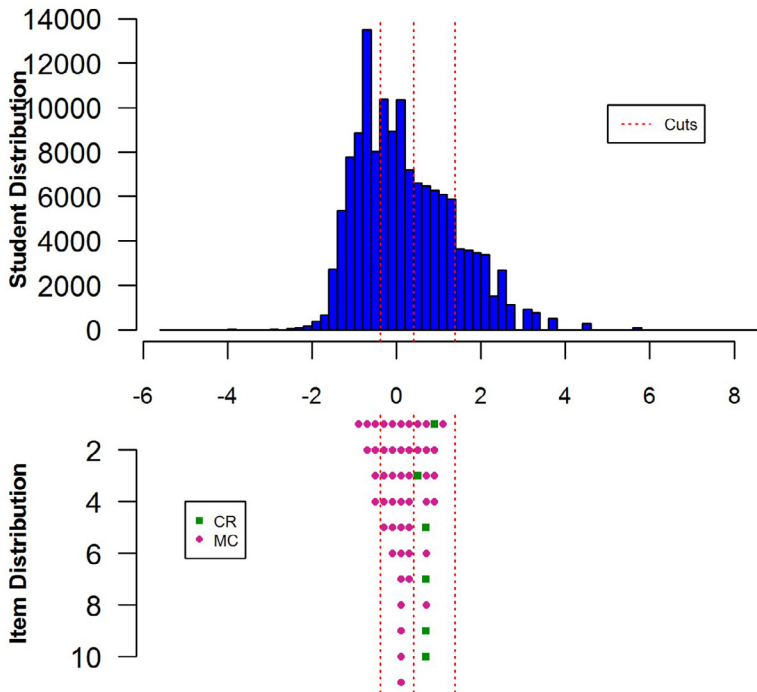
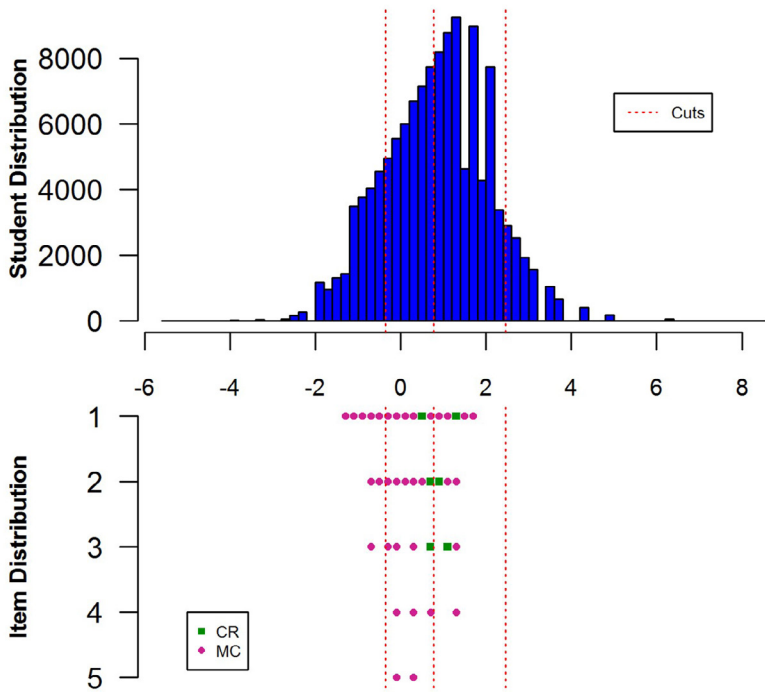


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

### Spring Literature



### Summer Algebra I

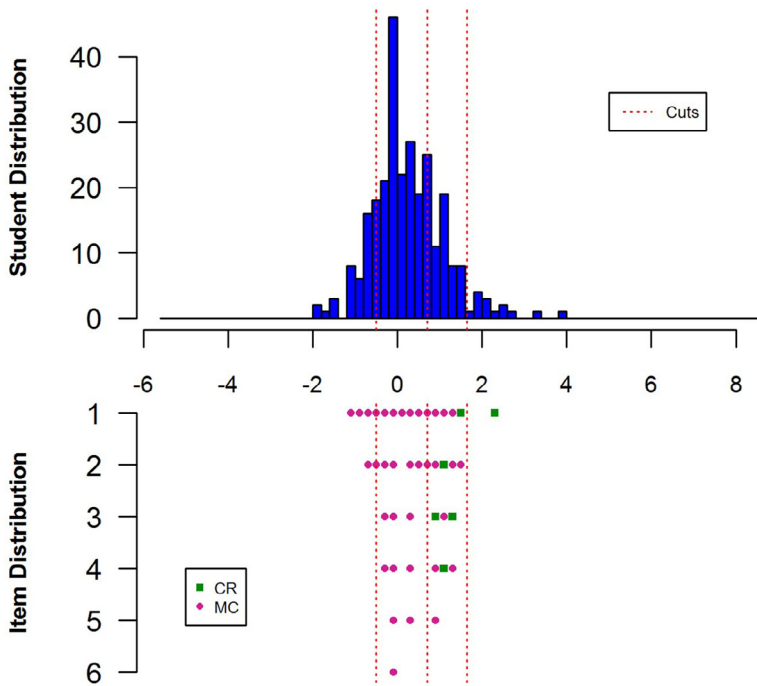
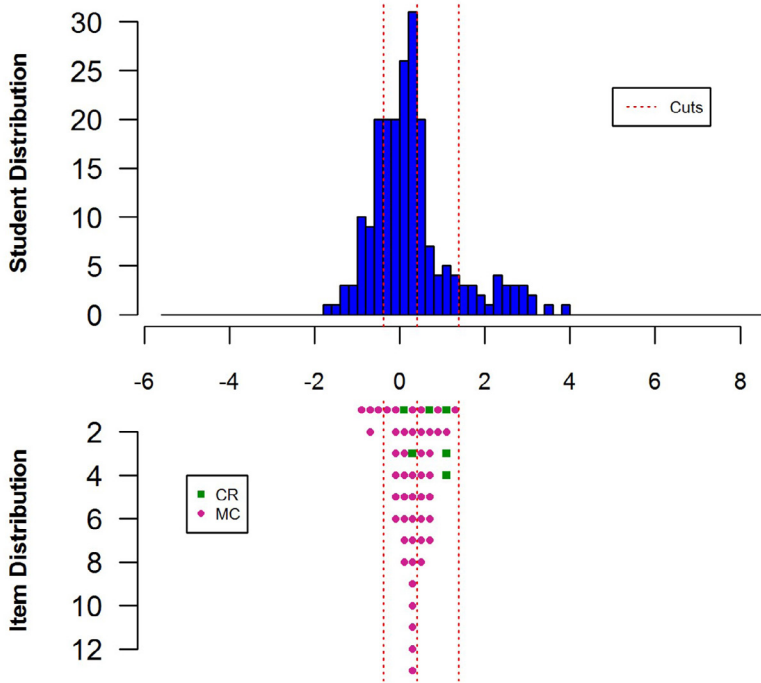
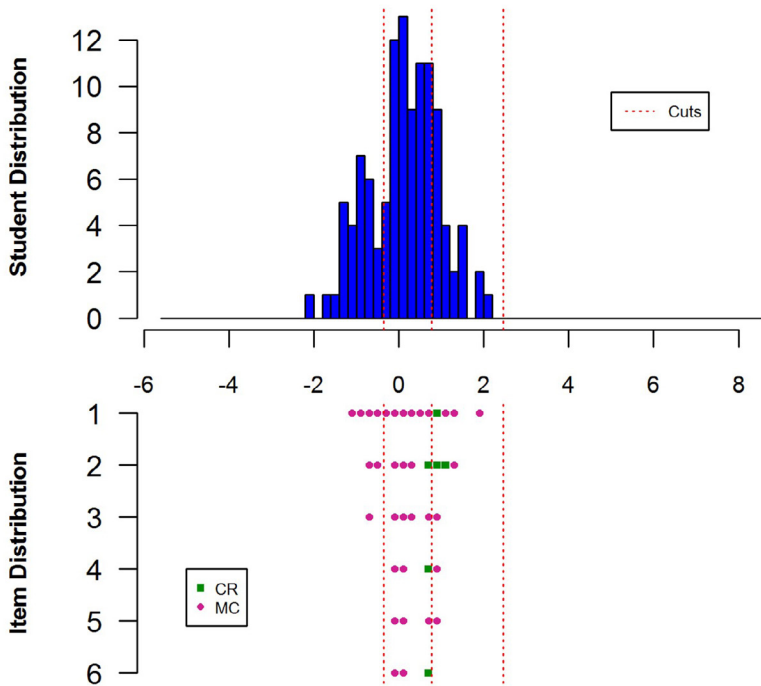


Figure 12–1 (continued). Item Difficulty-Student Ability Maps

### Summer Biology



### Summer Literature



## CHAPTER THIRTEEN: STANDARD SETTING

---

### STANDARD SETTING AND PERFORMANCE LEVEL DESCRIPTORS

The Keystone Performance Level Descriptors (PLDs) are paragraphs that describe the knowledge and skills expected at different performance levels with respect to the content standards (Pennsylvania Keystone Exams Assessment Anchor Content Standards and Eligible Content) for each of the Keystone Exams. Descriptors must be clearly written to ensure that all stakeholders have a common understanding of what describes expected performance at the various levels (i.e., Below Basic, Basic, Proficient, and Advanced). PLDs were developed, reviewed, and finalized by the PDE/QRT<sup>1</sup> and committees of Pennsylvania educators as required by the Chapter 4 Regulations. After the development and final review by PDE/QRT and Pennsylvania educators, the descriptors were prepared for use during the standard setting workshop. During this meeting, the descriptors were used to guide the standard setting process for each of the Keystone Exams. They were instrumental to the validity and defensibility of the standard setting process.

The standard setting for the Algebra I, Biology, and Literature Keystone Exams was conducted by Data Recognition Corporation (DRC) using a Bookmark procedure (Lewis, Mitzel, & Green, 1996) during a workshop held in Harrisburg, Pennsylvania, June 23–24, 2011. After the standard setting event, the descriptors were finalized. Along with the recommended cut scores, final PLDs for each of the Keystone Exams were submitted to the Pennsylvania Board of Education for final approval.

Below is a summary of the process that was used to guide the development of the Keystone Exams PLDs and a summary of the methodology and results of the standard setting workshop. Additional details about the standard setting event can be found in the *Keystone Standard Setting Technical Report* (Pennsylvania Department of Education, 2011).

### DEVELOPMENT OVERVIEW FOR THE PERFORMANCE LEVEL DESCRIPTORS

The Keystone Exams PLDs were developed by Pennsylvania educators during two meetings. The goal of the first meeting was to have Pennsylvania educators review the general Pennsylvania Policy Definitions that describe, at a high level, performance expected for each level and complete an in-depth analysis of the Keystone Exams Assessment Anchors and Eligible Content in order to create a bulleted list describing, in detail, what students are expected to know and be able to do at each performance level. The goal of the second meeting was to have committees of Pennsylvania educators review the Pennsylvania Policy Definitions again and draft general descriptors (paragraphs) that build upon and/or summarize the information from the bulleted lists of what students are expected to know and be able to do at each performance level.

Guiding documents were prepared for each meeting. The guiding documents included the following:

- PowerPoint training presentations
- Meeting agendas
- Assessment Anchors and Eligible Content documents
- Policy definitions
- Other relevant materials as needed to help guide the work of the committees

All meeting materials were submitted to PDE/QRT for review and approval before each Keystone Exams meeting following an agreed-upon development schedule. The following section provides specific information concerning each meeting.

---

<sup>1</sup> The PDE/QRT includes the representatives from the Pennsylvania Department of Education, members of the Quality Review Team, and/or others appointed by the Quality Review Team.

## **ROLE OF FACILITATORS AND OBSERVERS FOR THE MEETINGS**

The role of the facilitators was to ensure that a fair and orderly consensus process was followed for each meeting, that the committee members' work was adequately documented, and that the process stayed on schedule. The facilitators developed the agenda, prepared all meeting materials such as the PowerPoint training presentations and the task-guiding documents, and provided the initial training on the development of the specific descriptors (meeting 1) and the general descriptors (meeting 2). PDE/QRT members supported the facilitation process and/or served as observers of the process.

The facilitators also served as a resource, answering questions pertaining to the content of the standards (Assessment Anchor Content Standards and Eligible Content) and the documents developed to guide the process. Facilitators also summarized the results of each meeting, finalized the results, and prepared the specific descriptors/bulleted lists (meeting 1) and the general descriptors (meeting 2) for PDE/QRT review and approval.

## **PERFORMANCE LEVEL DESCRIPTORS MEETING 1**

### **CREATING SPECIFIC LISTS DESCRIBING WHAT STUDENTS SHOULD KNOW AND BE ABLE TO DO AT EACH PERFORMANCE LEVEL**

The first PLD meeting for Algebra I, Biology, and Literature Exams was held May 18–19, 2010, in Harrisburg, Pennsylvania. The purpose of the first meeting was to guide Pennsylvania educators in understanding the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature for what the Commonwealth of Pennsylvania determined students should know and be able to do for a given Keystone Exam subject. Committee members applied this understanding to the development of a bulleted list of specific determinations as to the level of knowledge and skills deemed necessary for each performance level. The section below describes the process used in the first meeting.

### **TRAINING**

Pennsylvania educators received general training on how to develop specific PLDs, including training on how to describe student performance in relation to the Keystone Exams Assessment Anchors and Eligible Content. The training also provided educators with a general overview of the Standards Aligned System (SAS) and the high-level plan for the Keystone Exams. Definitions of key terms (e.g., Assessment Anchor Content Standards, Eligible Content, Performance Level Descriptors) were provided along with information on the background and purpose of the Keystone Exams. Keystone Exams content-specific materials (e.g., Assessment Anchor Content Standards, Eligible Content, other guiding documents) were distributed. The PDE/QRT also provided information on the policy definitions for existing Pennsylvania assessments.

### **ANALYZING THE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT AND THE GENERAL POLICY DEFINITIONS FOR PENNSYLVANIA ASSESSMENTS**

Following the introductory training, educators were divided into groups according to each Keystone Exam. Each group focused specifically on the task at hand—developing the specific PLDs for a given Keystone Exam. Committee members were informed of the format of the specific descriptors (bulleted list) and the number of proposed performance levels for each Keystone Exam (Below Basic, Basic, Proficient, and Advanced). Committee members were then given time to familiarize themselves with the policy definitions and the Assessment Anchors and the Eligible Content for a given Keystone Exam. They were provided with PDE/QRT-approved guiding documents to facilitate the process. Beginning with Proficient, committee members were asked to draft, in bulleted-list format, each performance level for Basic, Proficient, and Advanced, making sure to consider the knowledge and skills required or deemed necessary for each performance level. Note: Educators were not asked to create a specific descriptor for Below Basic.

## DRAFTING SPECIFIC DESCRIPTORS

Outlined below is the sequence of steps taken to develop specific descriptors. The sequence was not always followed exactly. For example, some steps occurred simultaneously; other steps were repeated as needed or reordered as necessary.

1. The committee began with the development of the bulleted list for Proficient to serve as a model for the work during the remainder of the development process. As a formative first task using the Assessment Anchors and Eligible Content and the Pennsylvania Policy Definition for Proficient, the committee was asked to discuss, deliberate, and reach consensus on its initial bulleted list of the knowledge and skills needed to be considered Proficient. During this process, members were encouraged to consult all available resources and guiding documents. Particular emphasis was placed on the alignment of the knowledge and skills necessary for Proficient performance with what students are expected to know and be able to do as defined by the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature.
2. Once the committee drafted a bulleted list of the knowledge and skills needed to describe Proficient performance based upon the Assessment Anchors and Eligible Content, a group discussion took place. In reviewing the bulleted list for Proficient, the educators were specifically asked to determine whether all members agreed that the list included the appropriate knowledge and skills from the Assessment Anchors and Eligible Content to describe the Proficient performance level and that all Assessment Anchors and Eligible Content were sufficiently addressed.
3. The results of the discussion were summarized, and suggested revisions were made. The summary feedback was presented to the committee for additional consideration. An open discussion followed. Committee consensus was reached.
4. Following development of the bulleted list of the knowledge and skills needed for the Proficient performance level as determined by the committee, the committee began the development of the bulleted lists describing the specific knowledge and skills needed for Basic and Advanced. To complete the task, the committee members followed the procedures analogous to those used to develop the specific bulleted list for the Proficient performance level. These procedures included, as a formative first task, the committee's use of the Assessment Anchors and Eligible Content and the Pennsylvania Policy Definitions (e.g., Basic, Advanced) to discuss, deliberate, and reach consensus on its initial bulleted list of the knowledge and skills needed for Basic and then Advanced. This order of development—Proficient first, followed by Basic and then Advanced—was followed throughout the remainder of the process.

Once the initial drafts of the bulleted lists for Basic, Proficient, and Advanced were developed, a group discussion took place. To guide the discussion, the following questions were used to evaluate each specific descriptor (bulleted list) for a given performance level (Basic, Proficient, Advanced):

- Is the description of the performance level appropriate? If not, what revisions need to be made?
  - Is the description of the specific Keystone Exam inappropriate because the list of knowledge and skills included in the description of the performance level is too demanding? If so, what revisions need to be made?
  - Is the description inappropriate because the knowledge and skills included in the description of the performance level is inconsistent with the expectation of the high standards as reflected in the Policy Definition? If so, what revisions need to be made?
  - Is the description inappropriate because the knowledge and skills included in the description of the performance level might be too easy? If so, what revisions need to be made?
5. The results of the discussion were summarized, and suggested revisions were listed. The summary feedback was presented to the committee for additional consideration. An open discussion followed. Depending upon the degree of concurrence, the facilitators proposed revisions based on the committee members' feedback to the specific descriptors (bulleted lists) for each descriptor. Committee consensus was reached.

6. Once consensus was reached, the bulleted lists or specific descriptions for each performance level were reviewed once again to confirm that all Assessment Anchors and Eligible Content were sufficiently addressed for each performance level and that the lists showed a clear progression from one performance level to the next. The results of the discussion were summarized, and suggested revisions were listed. The summary feedback was presented to the committee for additional consideration. An open discussion followed. Depending upon the degree of concurrence, the facilitators proposed revisions to the lists for each descriptor based on the committee members' feedback. Committee consensus was reached.
7. Following completion of the committee's work, the specific PLDs or bulleted lists of the knowledge and skills needed for each descriptor were collected. The bulleted lists were prepared for final review by the PDE/QRT. Upon approval by the PDE/QRT, the bulleted lists of the knowledge and skills describing each performance level were posted on the PDE website for additional review and feedback.

## PERFORMANCE LEVEL DESCRIPTORS MEETING 2

### CREATING GENERAL DESCRIPTIVE PARAGRAPHS DESCRIBING WHAT STUDENTS SHOULD KNOW AND BE ABLE TO DO AT EACH PERFORMANCE LEVEL

The second meeting for Algebra I, Biology, and Literature Exams took place in Harrisburg, Pennsylvania, on April 27–28, 2011. The second meeting built upon the work completed at the first meeting. The purpose of the second meeting was to guide the committee of Pennsylvania educators in developing general PLDs (paragraphs) for each of the performance levels (Basic, Proficient, and Advanced). These paragraphs were clearly written to ensure all stakeholders have a common understanding of what describes expected performance at the various levels. The paragraphs were not to be as specific as the bulleted lists but were to be aligned to the bulleted lists. In order to complete the task, the educators reviewed the Pennsylvania Policy Definitions for the performance levels.

**Table 13–1. Pennsylvania Policy Definitions**

Level	Description
Advanced	The Advanced Level reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
Proficient	The Proficient Level reflects satisfactory academic performance. Proficient work indicates a solid understanding and adequate display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
Basic	The Basic Level reflects marginal academic performance. Basic work indicates a partial understanding and limited display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. This work is approaching satisfactory performance, but has not been reached. There is a need for additional instructional opportunities to achieve the Proficient Level.
Below Basic	The Below Basic Level reflects inadequate academic performance. Below Basic work indicates little understanding and minimal display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. There is a major need for additional instructional opportunities to achieve the Proficient Level.

The committee members then reviewed the specific bulleted list describing the knowledge and skills for Proficient based upon the Assessment Anchors and Eligible Content to determine whether the list of knowledge and skills provided in the bulleted list was still in alignment with the Policy Definition for Proficient. This review by the committee also included an in-depth analysis of the Assessment Anchors and Eligible Content. The section below describes, in detail, the process used in the second meeting.

## TRAINING

Pennsylvania educators received general training on how to develop general descriptors (paragraphs) that describe performance at the various levels, including training on how to describe student performance in relation to the Keystone Exams Assessment Anchors and Eligible Content. The training also included providing Pennsylvania educators with a general overview of the SAS and the high-level plan for the Keystone Exams. Definitions of key terms (e.g., Assessment Anchor Content Standard, Eligible Content, specific and general Performance Level Descriptors) were provided along with information on the background and purpose of the Keystone Exams. A review of the Pennsylvania Policy Definitions was also included in the training, including a discussion of how the policy definition for Proficient relates to what it means to be Proficient on a given Keystone Exam. Content-specific materials (e.g., Policy Definitions, Assessment Anchor Content Standards and Eligible Content, specific descriptors or bulleted lists from the first meeting, other guiding documents) were also distributed.



## **ANALYZING THE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT AND THE POLICY DEFINITION FOR PROFICIENT**

Following the introductory training, Pennsylvania educators were divided into groups according to Keystone Exam. Each group focused specifically on the task at hand—developing the general PLD paragraphs (Basic, Proficient, and Advanced) for a given Keystone Exam. To begin the process, educators reviewed the Pennsylvania Policy Definition for Proficient.

### **DRAFTING GENERAL DESCRIPTOR PARAGRAPHS**

Once the committee reviewed the bulleted list for alignment to the Policy Definition for Proficient, committee members were asked to describe, in general terms, the knowledge and skills deemed necessary for each performance level (Basic, Proficient, and Advanced), beginning with Proficient. As a formative first task, committee members were instructed to refer to the bulleted list of the knowledge and skills required or deemed necessary for each performance level. Outlined below is the sequence of steps for the process used to develop the general PLD paragraphs. The sequence was not always followed exactly. For example, some steps occurred simultaneously; other steps were repeated as needed or reordered as necessary.

The committee began with the development of the general descriptor paragraph for the Proficient performance level. This general descriptor served as a model for the committee’s work during the remainder of the development process. Using the Assessment Anchors and Eligible Content, the specific descriptors (bulleted list) for Proficient, and the Pennsylvania Policy Definition for Proficient, the committee was asked to discuss, deliberate, and reach consensus on a written description of the knowledge and skills needed for Proficient. During the process, members were encouraged to consult all available resources and guiding documents. Particular emphasis was placed on the alignment of the knowledge and skills necessary for the Proficient performance descriptor to the Assessment Anchors and Eligible Content for the given Keystone Exam.

Note: In order to help guide educators in the development of the general descriptor paragraph for Proficient, samples of descriptor paragraphs for Algebra I, Biology, and Literature (e.g., Georgia, North Carolina) were provided. The committee members were encouraged to approach the task by noting how the sample general descriptors must provide the right words to define performance—having a balance between keeping the description of Proficient general enough yet not as specific as the bulleted list. Committee members were also encouraged not to focus too heavily upon style, grammar, and mechanics at this stage. In other words, committee members were not to serve as “wordsmiths.”

1. Once an initial draft paragraph summarizing the knowledge and skills needed to describe Proficient performance was developed, a group discussion took place. Committee members were asked to review the draft paragraph and determine whether the paragraph provided a clear description of what it means to be Proficient on a given Keystone Exam and the Policy Definition for Proficient. The goal of the discussion was to reach consensus.
2. Following development of the general paragraph describing Proficient, the committee began the development of the general paragraph describing Basic performance and the general paragraph describing Advanced performance. To complete the task, the committee members followed the procedures analogous to those used to develop the general paragraph describing Proficient performance on a given Keystone Exam. This process included, as a formative first task, using the Assessment Anchors and Eligible Content, the specific descriptors (bulleted lists), and the Pennsylvania Policy Definitions for a given level (e.g., Basic, Advanced) and discussing, deliberating, and reaching consensus on the knowledge and skills needed for Basic and the knowledge and skills needed for Advanced. This order of development—Proficient first, followed by Basic and then Advanced—was followed throughout the remainder of the process.
3. Once the initial draft paragraphs were developed for the other performance levels, a group discussion took place. In reviewing the state of development of the general PLD paragraphs at this stage, the committee members were asked to consider the following questions:
4. Does each paragraph clearly summarize the knowledge and skills required for a given performance level (Basic, Proficient, and Advanced)? If not, what revisions need to be made?

5. Does each paragraph provide for an appropriate description of the performance level? In other words, does each paragraph provide an overview or summary of the knowledge and skills appropriate for a given performance level? If not, what revisions need to be made?
6. Does any paragraph provide information that should not be included in the description of the performance level? If so, what revisions need to be made?
7. Is there information in any PLD paragraph that does not align well with the Pennsylvania Policy Definitions for a given performance level? If so, what revisions need to be made?
8. Do any paragraphs include information that might be inconsistent with the knowledge and skills defined by the Assessment Anchors and Eligible Content? If so, what revisions need to be made?
9. Does any paragraph include information describing performance that might be too demanding or too easy? If so, what revisions need to be made?
10. The results of the discussion were summarized, and revisions to each general PLD paragraph were made. Committee consensus was reached.
11. Once consensus was reached, the paragraphs describing performance at each level were reviewed again by the committee to confirm the following:
12. The PLD paragraphs show a clear progression from one performance level to the next level.
13. The PLD paragraphs are consistent with the Pennsylvania Policy Definitions.
14. The PLD paragraphs are aligned to the Assessment Anchors and Eligible Content.
15. The results of the discussion were summarized, suggested revisions were made, and committee members' feedback was incorporated into the paragraphs. Committee consensus was reached.
16. Following completion of the committee's work, the general PLD paragraphs were provided to PDE/QRT for final review and feedback. Upon approval by PDE/QRT, the general PLD paragraphs were used to guide the standard setting process.

## **STANDARD SETTING**

A major purpose in the design of the standard setting workshop for the Keystone Exams is to establish procedures to set the performance cuts for the newly developed exams and, at the same time, adhere to the framework required by federal guidelines (USED, 2004) for setting performance levels. Federal guidelines (USED, 2004: Sect 2) specify that the setting of performance standards must involve the following elements:

- Formal adoption of performance categories that comprise at least three levels
- Pluralistic representation by education stakeholders, to include, for example, members of the public, school teachers and administrators, special education teachers, etc.
- Performance standards based primarily on expert judgment regarding content-based expectations of student achievement, but including the consideration of student impact data
- Descriptions of the competencies associated with each performance level

Accordingly, the standard setting workshop is designed to satisfy the following goals:

- A defensible and federally acceptable standard setting methodology that emphasizes a content-based approach for recommending the new performance standards
- The incorporation of PLDs developed by Pennsylvania educators into the standard setting process. (The larger goal around the incorporation of PLDs into the process is to help ensure the alignment of Pennsylvania's content standards to performance expectations as established by the recommended cut scores.)

The panelists were informed that the results from this meeting would be presented to the Board for review and possible adoption.

## **PANELIST RECRUITMENT**

PDE selected committee members for the Algebra I, Biology, and Literature standard setting workshop mostly from members who participated in the May 2010 and April 2011 Performance Level Advisory committees. These committee members were selected as the starting pool because they represented the diversity of the Commonwealth of Pennsylvania, had a mix of teaching and committee experience, and, most importantly, were familiar with the PLDs of the Keystone Exams. From this list, PDE selected a subset of 25 members for Algebra I, 25 members for Biology, and 23 members for Literature to serve as eligible candidates. DRC, in collaboration with PDE and its Technical Advisory Committee (TAC), established a target of 15 to 20 participants for each of the Keystone Exams in Algebra I, Biology, and Literature.

Between March and June 2011, a great effort was made to recruit enough panelists to meet the target number of participants. In accordance with federal guidelines for representative committees and TAC's recommendation of recruiting a few committee members with higher education experience, the following background factors were applied in the recruitment decision:

- Gender
- Ethnicity
- Grade level and higher education experience
- Content expertise
- Geographic location
- Specializations
- Experience in developing state academic standards, state assessments, and other related activities

However, due to the unavailability of and the cancellation by some committee members, a total of 15, 13, and 11 panelists attended the standard setting workshop for Algebra I, Biology, and Literature, respectively. Table 13–2 contains the summary information about the characteristics of the selected panelists for each content area based on their self-reported responses to the Participant Survey. As can be seen from this table, there were committee members who considered themselves minority in the Algebra I and Literature groups. There were also committee members with administration and/or teaching experience in higher education, special education, and/or individualized education plan (IEP); those with experience working in different regions; and those with different lengths of teaching experience.

**Table 13–2. Self-Reported Demographic Composition of Panelists by Content Area**

<b>Demographic Information</b>	<b>Algebra I Number</b>	<b>Algebra I Percent</b>	<b>Biology Number</b>	<b>Biology Percent</b>	<b>Literature Number</b>	<b>Literature Percent</b>
Gender: Male	9	60.0%	5	38.5%	5	45.5%
Gender: Female	6	40.0%	8	61.5%	6	54.5%
Ethnicity: Asian	1	6.7%	0	0.0%	0	0.0%
Ethnicity: American Indian	0	0.0%	0	0.0%	0	0.0%
Ethnicity: Black	1	6.7%	0	0.0%	1	9.1%
Ethnicity: Latino	0	0.0%	0	0.0%	0	0.0%
Ethnicity: Multi-Race	0	0.0%	0	0.0%	0	0.0%
Ethnicity: White	13	86.7%	13	100.0%	10	90.9%
Role: Classroom Teacher	8	53.3%	9	69.2%	4	36.4%
Role: Educator	3	20.0%	0	0.0%	0	0.0%
Role: Higher Education Educator	3	20.0%	1	7.7%	4	36.4%
Role: Other	1	6.7%	3	23.1%	3	27.3%
Special Education: Yes	7	46.7%	7	53.8%	4	36.4%
Special Education: No	2	13.3%	4	30.8%	3	27.3%
Special Education: N/A	6	40.0%	2	15.4%	4	36.4%
LEP: Yes	4	26.7%	5	38.5%	2	18.2%
LEP: No	4	26.7%	6	46.2%	5	45.5%
LEP: N/A	6	40.0%	2	15.4%	4	36.4%
LEP: Missing	1	6.7%	0	0.0%	0	0.0%
Region: Urban	3	20.0%	2	15.4%	2	18.2%
Region: Suburban	7	46.7%	5	38.5%	5	45.5%
Region: Rural	4	26.7%	6	46.2%	3	27.3%
Region: Other	1	6.7%	0	0.0%	1	9.1%
Experience: Less than 10 years	0	0.0%	5	38.5%	0	0.0%
Experience: 10–20 years	4	26.7%	2	15.4%	4	36.4%
Experience: 20–30 years	8	53.3%	4	30.8%	3	27.3%
Experience: More than 30 years	3	20.0%	2	15.4%	4	36.4%

## MATERIALS PREPARATION

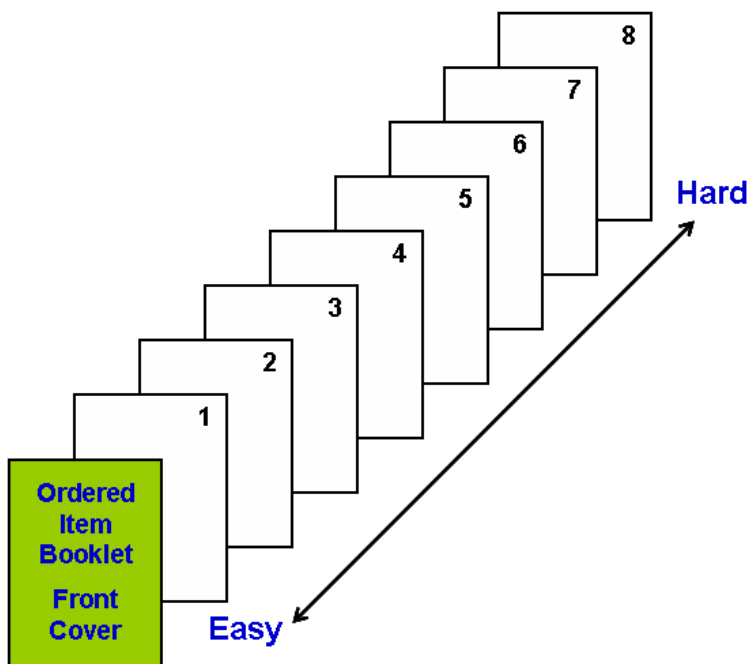
Workshop materials were developed and printed by DRC. The following is a list of materials that were available to panelists during the workshop:

- Item Map
- Item Separation Map
- Ordered Item Booklet (OIB)
- Passages
- Scoring Rubrics
- 2011 Operational Test Form
- PLDs
- Content Standards
- Participant Rating Form
- Participant Survey
- Readiness Form
- Evaluation Form
- Adhesive bookmarks, pens, highlighters, etc.

**Item Map.** The item map is a summary document displaying relevant information regarding each item. It contains the OIB page number, the original test sequence, item type, key, and content standard. The item map is ordered by difficulty in the same manner as the ordered item booklet. The item separation map is a graphical display of the relative difficulty of each item.

**Ordered Item Booklet.** The ordered item booklet is composed of all the operational items included in the test given to students in Spring 2011. Items are ordered from the easiest to the hardest. Each page contains an item and a page number. For constructed-response (CR) items, each score point with a sample response has a unique location in the OIB. A visual illustration of the OIB is provided in Figure 13–1.

**Figure 13–1. Illustration of Ordered Item Booklet**



To ensure there was no item difficulty gap for the items in an OIB, a few field test items were added to the OIBs. Table 13–3 shows the number of items supplemented into the OIBs by content area.

**Table 13–3. Number of Score Points in OIB and Number of Items Supplemented**

Exam	Number of Score Points in OIB	Number of Items Supplemented
Algebra I	63	3
Biology	69	3
Literature	54	2

Details of all other materials can be found in the *Keystone Standard Setting Technical Report* (Pennsylvania Department of Education, 2011).

### DATA PREPARATION

In Bookmark standard setting (Lewis et al., 1996), the locations of items are typically rescaled to produce better alignment with the task of asking panelists what a student should know and be able to do. A probability of 0.67 is often used to find the corresponding item location during rescaling because this probability aligns better with the likelihood panelists use to make their judgment on whether a borderline student should answer the item correctly or receive a score point or higher. For Keystone Exams, the multiple-choice (MC) items were calibrated using the familiar form of the dichotomous Rasch model. The CR items were calibrated using another model in the Rasch family, Master’s partial-credit model (Wright & Masters, 1982). The latter model parameterizes each threshold needed to obtain the maximum score on the task. Consequently, there is one item difficulty parameter for each of the  $n - 1$  score transitions (0/1, 1/2, etc.), or thresholds. Using the equated item parameters, the locations of items were rescaled to a response probability of 0.67 (i.e.,  $RP=0.67$ ). For MC items, the item locations were found by solving

$$\Phi_{ni} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

for the value of  $\beta_n$  that gives  $\Phi_{ni} = 0.67$ .  $\Phi_{ni}$  is the probability that person  $n$  scores 1 on item  $i$ ;  $\beta_n$  is the ability of person  $n$ ; and  $\delta_i$  is the difficulty of item  $i$ .

For CR items, the probability of person  $n$  scoring  $x$  on item  $i$  is

$$\pi_{nix} = \frac{\exp \sum_{j=0}^x (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\beta_n - \delta_{ij})}, \quad x = 0, 1, \dots, m_i$$

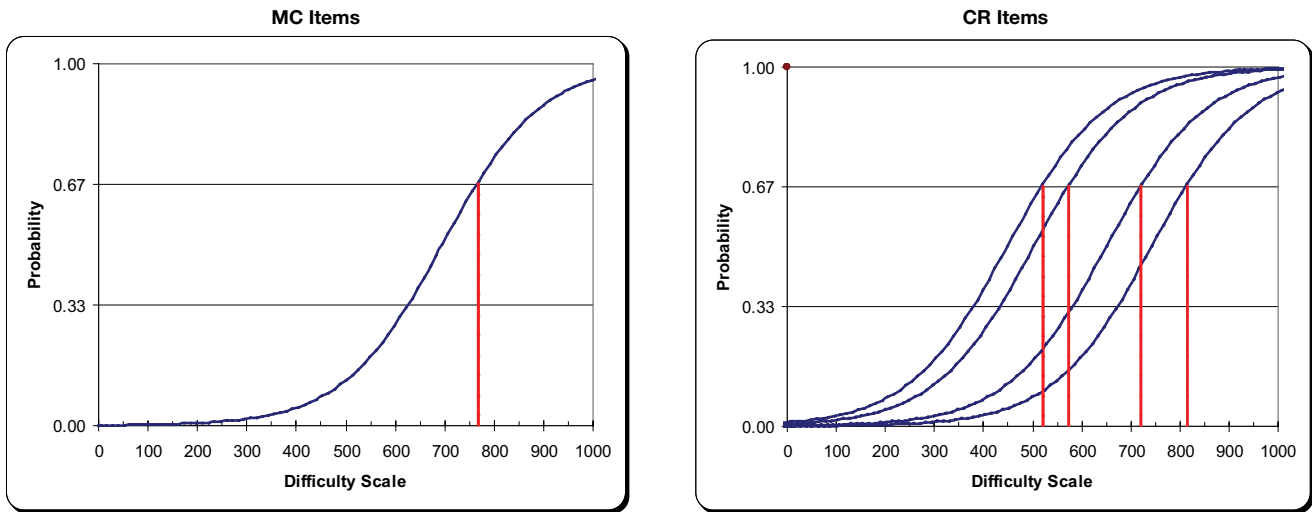
where  $m_i$  is the number of thresholds and, for notational convenience,

$$\exp \sum_{j=0}^0 (\beta_n - \delta_j) = 1.$$

This equation expresses the probability of person  $n$  scoring  $x$  on the  $m_i$  threshold of item  $i$  as a function of the person’s measure ( $\beta_n$ ) and the threshold difficulties ( $\delta_{ij}$ ) of the  $m_i$  thresholds for item  $i$ . The observation  $x$  is a count of the successfully completed item thresholds. The item location for a score point is determined by finding the  $\beta_n$  for the person who has a 0.67 probability of earning this score point or higher.

The figure below shows how the difficulty values of MC items and score values for the CR items were treated in determining their respective OIB placements. For an MC item (left plot), the difficulty is the point on the scale at which the examinees have a 0.67 probability of answering the item correctly. For the CR item (right plot), the four illustrated values (e.g., on a 0–1000 scale) indicate where the examinees have a 0.67 probability of earning a particular score point or higher. The item difficulty for the MC item is 768, and the four threshold values for the CR items are 521, 575, 723, and 815. The value of 521 is the location on the scale where examinees have a 0.67 probability of earning a score of 1 or higher (i.e., 2, 3, or 4). The value of 575 is where examinees have a 0.67 probability of earning a score of 2 or higher (i.e., 3 or 4). The value of 723 is where examinees have a 0.67 probability of earning a score of 3 or 4.

**Figure 13–2. Example of Obtaining Item Difficulties for MC and CR Items**



## TRAINING

The overall training was conducted the first morning of the workshop. Participants were informed that they were to

- be responsible for all secure materials,
- verify their individual placements for each round of judgments, and
- participate in a discussion as a large group.

Content-specific training was conducted after content area groups assembled in different rooms. These training materials included the following:

- Item Map
- Item Separation Map
- OIBs
- Training Rubrics and/or Passages
- PLDs
- Rating Form

Panelists were told that the process includes iterations (rounds) of individual judgments, group discussions, and opportunities to revise judgments. In addition, impacts were presented (percentage of students in each performance level) based on the large groups' results and external data.

## BOOKMARK PROCEDURE

DRC utilized a Bookmark method to set the performance standards. Bookmark is one in a broad category of methods commonly referred to as item mapping that focus on items rather than examinees. To begin the process, participants were asked to visualize the knowledge and skills of a student who is at the borderline between two performance levels based on the PLDs. Thereafter, participants were given an ordered item booklet (with items ordered from easiest to most difficult) and asked to assess whether this borderline student has a reasonably high probability of answering each item correctly. "Reasonably high" was defined as 0.67. In addition, an item map was presented that contained the response key, the content objective, and the item sequence in the test booklets. An item separation map was also presented that showed the relative difficulty of each item. Panelists were given a rating form to record their individual placements for all performance levels in each round. Before each round, panelists were asked to fill out a readiness form in order to proceed.

**Round 1.** The Bookmark procedure proceeded in three rounds. Round 1 began following the review and discussion of PLDs facilitated by a DRC test development specialist. Participants then reviewed the OIBs independently. During this review, they were asked to determine what academic knowledge, skills, and competencies were required for a barely Proficient, Basic, or Advanced student to respond correctly to each successively more difficult item.

Training by the overall psychometric lead during the bookmark placement session emphasized the following points:

- The bookmark represents a judgment of the divide between items that a student at the borderline of a performance level should master and those that are not necessary to master.
- Bookmark placement should not be thought of as separating two items but rather two groups of items. In other words, a placement should not hinge on distinctions drawn for adjacent items without some compelling reason, such as a large gap in content difficulty.
- Students with a scaled score at a given cut score should have approximately a 0.67 probability of correctly responding to a MC item or receiving a certain score point and higher for a CR item at the cut score. These same students should have a higher probability of success on easier items (before the bookmark placement) and a lower probability of success on harder items (after the bookmark placement).



- While placing their bookmarks, panelists should consider what students should know and be able to do in the context of the skills implied by the PLDs and the item content.
- Panelists could start with placing the Basic/Proficient cut point, next the Below Basic/Basic cut point, and finally the Proficient/Advanced cut point.

Panelists were asked to record their bookmark placements on the rating form after they filled out a readiness form, which indicated they had completed the training and understood the standard setting process and their roles. Panelists' judgments were entered into a spreadsheet program. The median ratings of all panelists were calculated. The median placements were treated as the recommended cut scores. In addition, the standard errors associated with the recommended bookmark placements were calculated and associated impact data were determined.

**Round 2.** Round 2 started with a discussion of Round 1 results. The individual panelists' Round 1 bookmark placements, the median bookmark placements, and the percentage of students in each performance level were presented. Panelists were instructed to verify the ratings entered into the program as correct. A large-group discussion followed. The panelists compared their results with others by considering questions such as why they made their Round 1 placements at the locations where they did and what skills and knowledge were required to answer the items. After that, the impact data, based on the median bookmark placement from Round 1 (using the Spring 2011 operational test score distributions), were provided to help panelists frame the effects of their judgments. During Round 2 discussion, there was no attempt by the facilitators to reach consensus.

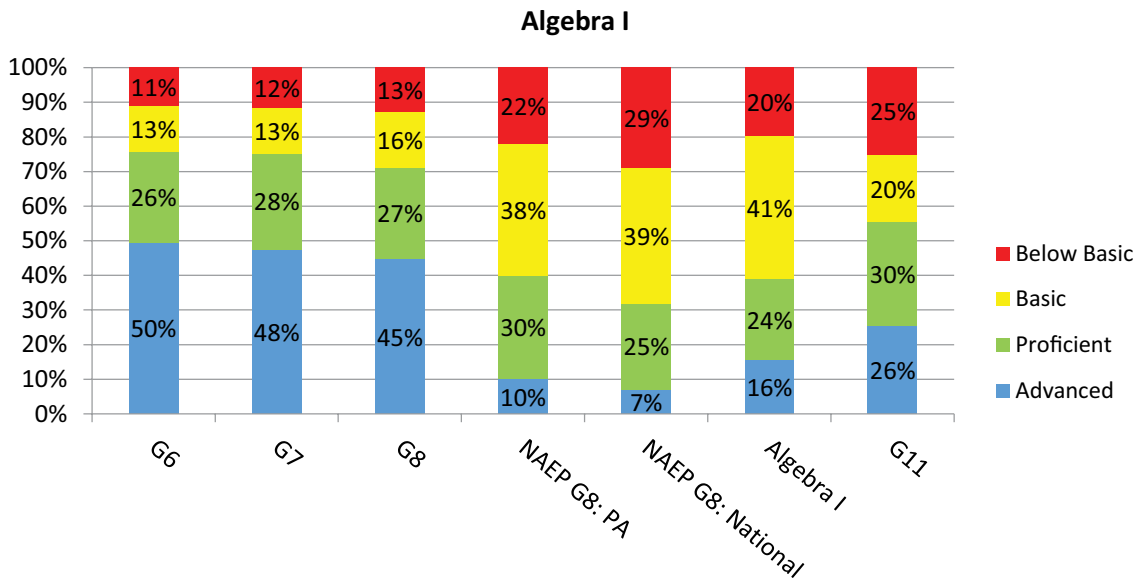
After Round 2 discussion, panelists were asked to make a second set of bookmark placements. Before they revised their Round 1 placements, they were asked to fill out the readiness form to make sure they understood how to adjust their placements (if they desired to do so) based on Round 1 information. The judgments were entered into the spreadsheet program to calculate the median cut scores for each table and the full panel. The associated impact data were also calculated.

**Round 3.** Round 3 began with a discussion of Round 2 results. The process followed in Round 2 was used. More specifically, the individual panelist's Round 2 bookmark placements, the median bookmark placements from Round 2, and the percentage of students in each performance level were presented. Panelists were instructed to verify the ratings entered into the program as correct. A table discussion followed. Panelists compared their results with others by considering questions such as why they made their Round 2 placements at the locations where they did and what skills and knowledge were required to answer the questions. The impact data, based on the median bookmark placement from Round 2, were provided to help panelists frame the effects of their judgments.

The Keystone Exams are one component of Pennsylvania's new system of high school graduation requirements. Because of the high-stakes consequences, the TAC strongly recommended bringing in external impact data to provide panelists with a reference outside of the Keystone Exams. The intent was to achieve reasonableness of results rather than to use the external data in a directive manner. DRC investigated Pennsylvania students' performance on the Pennsylvania System of School Assessment (PSSA), National Assessment of Educational Progress (NAEP), and Student Achievement Test (SAT) and presented external data as shown in Figures 13–3A to 13–3C before panelists made their Round 3 judgments. The panelists were informed of the following points:

- The PSSA and NAEP results were based on students' performance in 2009. The PSSA results were from grades 6–8 and 11. The NAEP results were from grade 8.
- All students in grades 6–8 and 11 in Pennsylvania took the PSSA. A sample that represents the Pennsylvania grade 8 students took the NAEP tests.
- The SAT results were based on the performance of students who took the SAT in 2010 or prior years.
- About 99% of students in the 2010 SAT data file indicated their expected graduation dates were in 2010; most of these students were in grade 11 in 2009. Therefore, the 2010 SAT data and the 2009 PSSA data were matched.
- Based on the matched sample, it was found that students with higher PSSA scores were more likely to take the SAT. To represent the full population in terms of demographics and PSSA scores, the matched sample was weighted by students' demographics and PSSA scores when calculating the impacts.

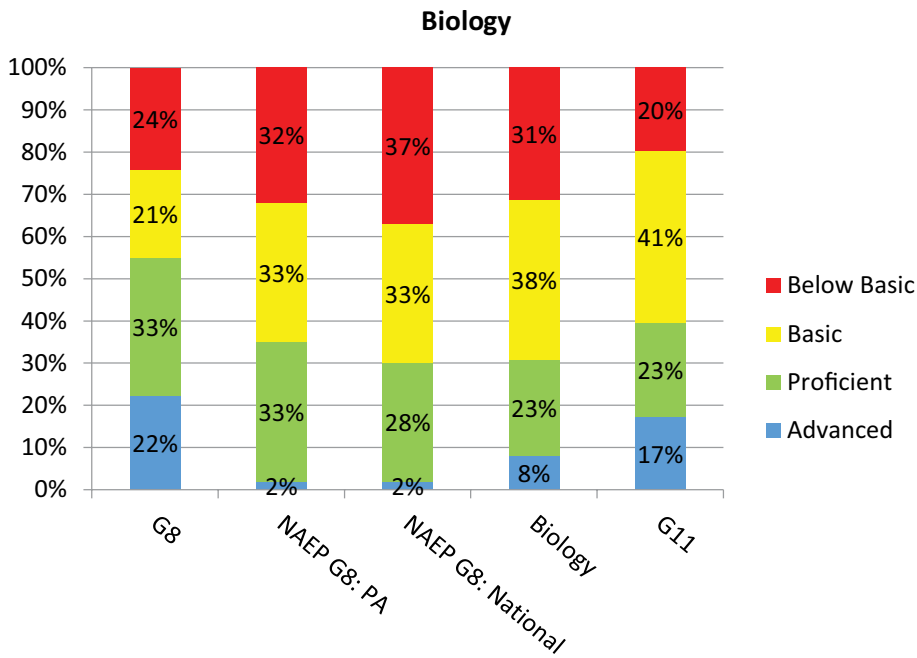
**Figure 13–3A. External Impact Data: Algebra I**



**Table 13–4A. Number of Score Points in OIB and Number of Items Supplemented: Algebra I**

Performance Level	PSSA G6	PSSA G7	PSSA G8	NAEP G8: PA	NAEP G8: National	KE Alg. I	PSSA G11	College Ready Yes Projected	SAT: College Ready Yes: PA	SAT: College Ready Yes: National
Below Basic	11.1%	11.6%	12.8%	22.0%	29.0%	19.8%	25.1%	0.9%	51.7%	54.0%
Basic	13.2%	13.1%	16.0%	38.0%	39.0%	41.2%	19.5%	7.8%	51.7%	54.0%
Proficient	26.2%	27.8%	26.6%	30.0%	25.0%	23.5%	29.8%	42.2%	51.7%	54.0%
Advanced	49.5%	47.5%	44.7%	10.0%	7.0%	15.5%	25.5%	91.2%	51.7%	54.0%
Below Basic + Basic	24.3%	24.7%	28.8%	60.0%	68.0%	61.0%	44.8%	4.0%	51.7%	54.0%
Proficient + Advanced	<b>75.7%</b>	<b>75.3%</b>	<b>71.3%</b>	<b>40.0%</b>	<b>32.0%</b>	<b>39.0%</b>	<b>55.3%</b>	<b>64.9%</b>	<b>51.7%</b>	<b>54.0%</b>
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	38.1%	51.7%	54.0%
Total N	128,421	132,803	135,909	3,600	161,700	93,703	135,676	61,118	65,426	N/A

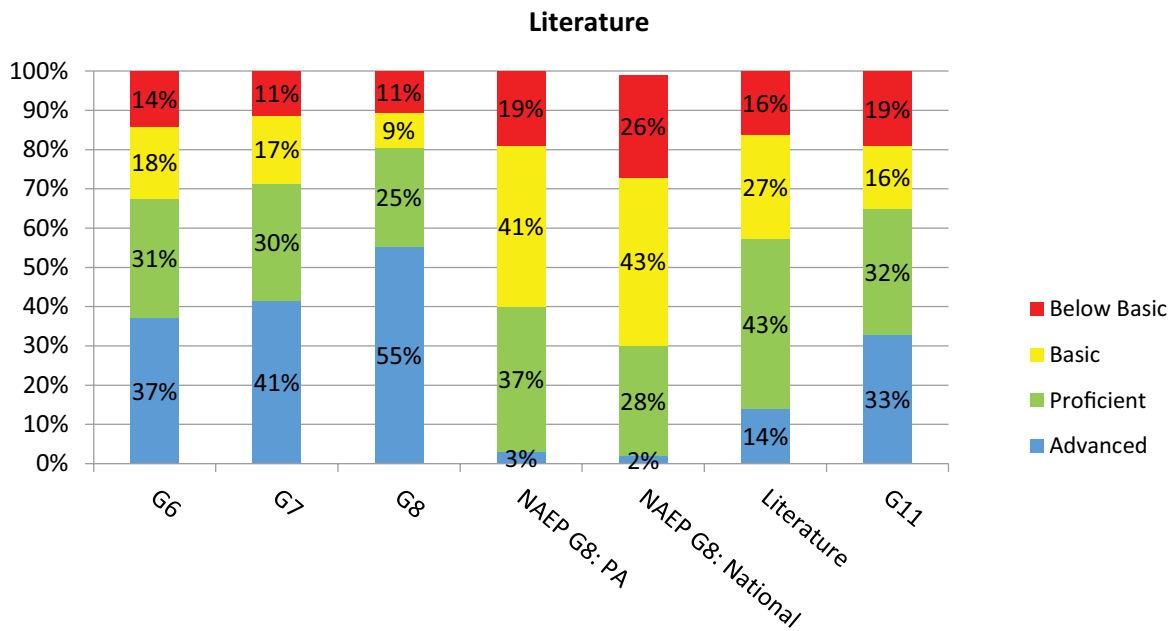
**Figure 13–3B. External Impact Data: Biology**



**Table 13–4B. Number of Score Points in OIB and Number of Items Supplemented: Biology**

Performance Level	PSSA G8	NAEP G8: PA	NAEP G8: National	KE Biology	PSSA G11	College Ready Yes Projected	SAT Math: College Ready Yes: PA	SAT Math: College Ready Yes: National
Below Basic	24.0%	32.0%	37.0%	31.4%	19.8%	1.4%	38.7%	43.0%
Basic	21.1%	33.0%	33.0%	37.9%	40.5%	9.1%	38.7%	43.0%
Proficient	32.5%	33.0%	28.0%	22.6%	22.5%	45.9%	38.7%	43.0%
Advanced	22.3%	2.0%	2.0%	8.1%	17.2%	86.6%	38.7%	43.0%
Below Basic + Basic	45.1%	65.0%	70.0%	69.3%	60.3%	6.6%	38.7%	43.0%
Proficient + Advanced	<b>54.8%</b>	<b>35.0%</b>	<b>30.0%</b>	<b>30.7%</b>	<b>39.7%</b>	<b>63.8%</b>	<b>38.7%</b>	<b>43.0%</b>
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	29.4%	38.7%	43.0%
Total N	134,969	3,600	151,100	46,394	131,534	60,311	65,426	N/A

**Figure 13–3L. External Impact Data: Literature**



**Table 13–4L. Number of Score Points in OIB and Number of Items Supplemented: Literature**

Performance Level	PSSA G6	PSSA G7	PSSA G8	NAEP G8: PA	NAEP G8: National	KE Literature	PSSA G11	College Ready Yes: Projected	SAT Math: College Ready Yes: PA	SAT Math: College Ready Yes: National
Below Basic	14.0%	11.2%	10.6%	19.0%	26.0%	16.2%	19.0%	1.2%	46.4%	50.0%
Basic	18.4%	17.4%	8.9%	41.0%	43.0%	26.5%	16.0%	4.7%	46.4%	50.0%
Proficient	30.5%	30.0%	25.2%	37.0%	28.0%	43.3%	32.1%	24.8%	46.4%	50.0%
Advanced	37.1%	41.4%	55.3%	3.0%	2.0%	14.0%	32.9%	76.7%	46.4%	50.0%
Below Basic + Basic	32.4%	28.6%	19.5%	60.0%	69.0%	42.7%	35.1%	2.7%	46.4%	50.0%
Proficient + Advanced	<b>67.6%</b>	<b>71.4%</b>	<b>80.5%</b>	<b>40.0%</b>	<b>30.0%</b>	<b>57.3%</b>	<b>64.9%</b>	<b>51.1%</b>	<b>46.4%</b>	<b>50.0%</b>
Total Percentage	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	34.5%	46.4%	50.0%
Total N	128,284	132,641	135,739	3,500	160,900	42,292	135,470	61,081	65,426	N/A

The Keystone Exams and PSSA results were presented to the panelists first. Panelists were encouraged to compare the impact data and discuss whether the results for the Keystone Exams were reasonable. The NAEP results were added next, and the SAT results were introduced last for comparison and discussion. While discussing the external data, panelists were reminded that all these tests were created for different purposes and might cover different content standards.

Before panelists provided their final judgments, they were instructed to fill out the readiness form to make sure they understood how to adjust their placements (if they desired to do so) based on the Round 2 information and external impact data. After their individual bookmark placements, panelists filled out the evaluation form. The judgments were entered into the spreadsheet program to calculate the median placements for the full panel. The associated impact data were also calculated. The Round 3 results were presented to the panelists for their information after the lunch break.

### PANELISTS' RECOMMENDATIONS

Table 13–5 provides a summary of each round’s median, minimum, and maximum ratings (i.e., bookmark page numbers) of the group.

**Table 13–5. Summary of Panelists’ Ratings for Each Round**

Exam	Round	Bookmark Page Number Median	Bookmark Page Number Min.	Bookmark Page Number Max.	Basic/ Proficient Median	Basic/ Proficient Min.	Basic/ Proficient Max.	Proficient/ Advanced Median	Proficient/ Advanced Min.	Proficient/ Advanced Max.
Alg. I	1	11	6	19	28	14	42	45	33	56
Alg. I	2	11	6	13	26	17	33	42	40	46
Alg. I	3	11	10	12	26	18	30	46	41	46
Bio.	1	9	4	15	26	20	30	56	43	62
Bio.	2	8	7	14	24	21	30	54	50	60
Bio.	3	8	7	12	22	20	30	54	50	60
Lit.	1	8	5	14	27	12	34	47	38	52
Lit.	2	9	8	15	23	15	34	46	38	48
Lit.	3	9	8	15	25	17	34	48	38	48

### CUT POINTS AND STANDARD ERRORS

Each bookmark page number is associated with a bookmark difficulty (i.e., logit value). The logit cut is the bookmark difficulty corresponding to the median OIB page number minus one. The logit cut and the standard error (SE) of median logit based on panelists’ Round 1 rating were used to establish the 1 and 2 SE confidence intervals. By bracketing the median cut score by 2 SEs, the 95% confidence interval was identified; the confidence interval can be used to estimate the effects of false positives (passing students who may not actually have sufficient knowledge and skills) or false negatives (failing students who do have sufficient knowledge and skills). PDE can use these standard errors to identify the appropriate cut score by taking into consideration the variance in the human judgments. Table 13–6 summarizes the logit cuts associated with Round 3 median ratings, median +/-1 SE, and median +/-2 SE. The corresponding impacts (percentages in performance level) are provided in this table as well. Note that BB represents Below Basic; B represents Basic; P represents Proficient; and A represents Advanced.

**Table 13–6. Summary of Logit Cuts and Impacts**

Exam	Stats	Logit Cut BB/B	Logit Cut B/P	Logit Cut P/A	Percentage in Performance Level (%) BB	Percentage in Performance Level (%) B	Percentage in Performance Level (%) P	Percentage in Performance Level (%) A	Percentage in Performance Level (%) P+A
Alg. 1	Median-2SE	-0.7273	0.4291	1.3694	14.7	37.3	30.2	17.8	48.0
Alg. 1	Median-1SE	-0.6181	0.5659	1.5041	17.2	37.8	31.6	13.4	45.0
Alg. 1	Median	-0.5090	0.7027	1.6388	19.8	41.2	27.5	11.5	39.0
Alg. 1	Median+1SE	-0.3999	0.8395	1.7735	22.4	44.4	23.5	9.7	33.2
Alg. 1	Median+2SE	-0.2907	0.9763	1.9082	25.2	47.0	19.7	8.1	27.8
Bio.	Median-2SE	-0.5977	0.3098	1.2500	22.5	38.9	27.9	10.7	38.6
Bio.	Median-1SE	-0.4933	0.3564	1.3205	28.3	33.1	29.3	9.3	38.6
Bio.	Median	-0.3888	0.4029	1.3910	31.4	32.7	27.8	8.1	35.9
Bio.	Median+1SE	-0.2843	0.4494	1.4615	34.5	32.2	25.2	8.1	33.3
Bio.	Median+2SE	-0.1799	0.4960	1.5320	40.7	26.0	26.4	6.9	33.3
Lit.	Median-2SE	-0.6561	0.2338	1.7014	12.4	20.9	46.3	20.4	66.7
Lit.	Median-1SE	-0.5545	0.4116	1.9551	14.2	25.3	46.5	14.0	60.5
Lit.	Median	-0.4530	0.5894	2.2088	16.2	26.5	48.4	8.9	57.3
Lit.	Median+1SE	-0.3515	0.7672	2.4625	18.3	31.4	43.6	6.7	50.3
Lit.	Median+2SE	-0.2499	0.9450	2.7162	20.5	36.5	38.2	4.8	43.0

**FINAL RESULTS**

After reviewing the results in Table 13–6 and considering panelists’ discussions at the standard setting workshop, PDE recommended using the logit cut scores associated with the median of panelists’ Round 3 ratings for Algebra I and Biology. For Literature, PDE recommended the logits cuts associated with the Round 3 median plus 1 SE.

To avoid negative values on the logit scale, the scaling constants were determined next to linearly convert the logit values to scaled scores. The scaled cut scores for each performance level were obtained by linearly transforming the logit cuts. Details of the scaling process can be found in Chapter Fourteen. A brief description is below.

For Keystone Exams, the linear transformation from logits or Rasch measures to scaled scores was established by anchoring the logit cut for Basic/Proficient to a scaled score 1500 and fixing the slope constant to 50. The intercept constant was calculated next based on the known values 1500, 50, and the logits cut for Basic/Proficient for each content area. In addition, the bottom of the scale was truncated at the lowest obtainable scaled score (LOSS), 1200. The top of the scaled scores was truncated at the highest obtainable scaled score (HOSS), 1800. The recommended scaled score cuts and the corresponding impacts were provided to the Board on July 20, 2011, for approval. Table 13–7 presents the final scaling constants and the Board-approved scaled-score ranges for each performance level.

**Table 13–7. Summary of Scaled-Score Ranges and Scaling Constants**

Exam	Performance Level Below Basic	Performance Level Basic	Performance Level Proficient	Performance Level Advanced	Scaling Constants Slope	Scaling Constants Intercept
Algebra I	1200–1438	1439–1499	1500–1545	1546–1800	50	1464.365
Biology	1200–1459	1460–1499	1500–1548	1549–1800	50	1479.355
Literature	1200–1443	1444–1499	1500–1583	1584–1800	50	1461.140

The Keystone Exams are reported by total and modules. Although the panelists made recommendations based on the total test only, the Basic/Proficient cut for the total test is applied directly in setting the passing cut score for each module. In this case, the passing scaled score cut at module level is 1500.

## CHAPTER FOURTEEN: SCALING

Scaling is used to transform test score values (i.e., raw scores) onto a scale that can be interpreted by users easily and correctly. Raw scores cannot be used to compare students' achievement across administrations because they depend on the difficulty of the tests. The same student can score higher on an easy test than on a difficult test. To overcome the limitation of raw scores, the scaled scores are introduced to report students' achievement in each Keystone Exam. This chapter describes the two major steps to convert a raw score to a scaled score (SS) and some key considerations for establishing the score scale for Keystone Exams.

### RAW SCORES TO RASCH ABILITY ESTIMATES

The pre-equated item parameter estimates for the operational items (further discussed in Chapters Twelve and Fifteen) were used to obtain Rasch person ability estimates and asymptotic standard errors of measurement for each possible raw score value for the overall test, as well as each module. The generation of this raw score-to-Rasch ability was accomplished through application of the fundamental formulas in the Rasch measurement model. The combination of both dichotomously scored multiple-choice (MC) items as well as polytomously scored constructed-response (CR) items requires the use of a partial-credit model (RPCM; Wright & Masters, 1982). The Newton-Raphson iterative procedure is used to obtain precise ability estimates:

$$b_r^{(t+1)} = b_r^t - \frac{r - \sum_i^L \sum_{k=1}^m k P_{rik}^{(t)}}{-\sum_i^L \left[ \sum_{k=1}^m k^2 P_{rik}^{(t)} - \left( \sum_{k=1}^m k P_{rik}^{(t)} \right)^2 \right]} \quad r = 1, \dots, M - 1,$$

where  $b_r^t$  is the estimated ability of the student with score  $r$  after  $t$  iterations,  $k$  is the number of thresholds,  $L$  is the number of items,  $M = \sum_i^L m_i$ , and  $P_{rik}^{(t)}$  is the probability,  $\pi_{nix}$ , defined earlier in Chapter Twelve:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^x (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^x (\beta_n - \delta_{ij})}, \quad x = 0, 1, \dots, m_i$$

The asymptotic standard error of measurement (SEM) was estimated from the denominator of the final iteration:

$$SE(b_r) = \left[ \sum_i^L \left[ \sum_{k=1}^m k^2 P_{rik}^{(t)} - \left( \sum_{k=1}^m k P_{rik}^{(t)} \right)^2 \right] \right]^{-1/2}.$$

The Rasch ability estimates and the corresponding SEMs are then transformed to scaled scores and SEMs of scaled scores as discussed in the following section.

### ZERO AND PERFECT SCORES

A direct ability estimate for zero (no points earned) or perfect (all points earned) raw scores can't be achieved. Thus, a default procedure for estimating such extreme scores was used for the Keystone Exams. Essentially, a fractional raw score (a value less than one, e.g., 0.3) was added to zero scores and subtracted from perfect scores to determine the corresponding logit values for these extreme scores.

## RASCH ABILITY ESTIMATES TO SCALED SCORES

Generally, scaled scores are preferred over Rasch ability estimates for reporting purposes. One issue is that Rasch ability estimates are on a scale that includes negative and decimal values. By transforming the Rasch ability estimates to scaled scores, all reported values can become positive integers, which allows for better interpretation by parents and students. Since Rasch ability estimates are comparable after equating to the base administration/year, the transformed scaled scores have a common scale across administrations, even though the corresponding raw scores may differ. Refer to Chapter Fifteen for additional details on equating.

Scaled scores are obtained through a linear transformation of Rasch ability estimates that are associated with each raw score point. Before the linear equation is established for each content area, a few points were considered for the Keystone Exams:

- Avoid scales that might be confused with scores for other types of assessment, for example:
  - Scaled scores ranging from 0 to 100 (because this might be confused with percentage correct scores or percentile ranks)
  - Scaled scores ranging from 200 to 800 (because this might be confused with SAT scores)
  - Scaled scores with similar ranges as the ones for the Pennsylvania System of School Assessment (PSSA) or Classroom Diagnostic Tools (CDT)
- Avoid scales similar to raw scores from a base form.
- Avoid scales that might suggest the scores are more precise than they actually are (i.e., suggesting more precision than can actually be supported by the test scores).
- Avoid scales with negative numbers and decimals.

In terms of industry standard practice, a common perspective is that scaled scores should facilitate score interpretation while at the same time minimize misinterpretation and unwarranted inferences. Often this is done by attaching meaning to the scores<sup>1</sup> (Peterson, Kolen, and Hoover, 1989). The incorporation of content meaning is one way to facilitate score interpretation. This might be done in several different ways. For example, the current PSSA scaled scores, like those of many other state assessments, try to input some content meaning by having the PSSA performance level cut scores have known values on the scaled-score metric. Such an approach appears to make good sense given the purposes of a criterion-referenced test like the PSSA.

As a result, a scaled score range of 1200 to 1800 and the Proficient scaled cut-score of 1500 was used to establish the scales for each of the Keystone Exams in Algebra I, Biology, and Literature. It is worth noting that, although careful considerations were given to the selection of these values, they are completely arbitrary based on previously discussed considerations. For example, the label of 1500 could have been called 100 or any other value or letter without affecting any of the relationships among schools, administrations, students, or items. In other words, changing the scale would simply be changing the labels on the axis of a graph without moving any of the points.

### LINEAR TRANSFORMATION FORMULAS

The scaled scores for the Keystone Exams are obtained through a linear transformation of the Rasch ability estimates ( $\hat{\beta}$ ). Specifically,

$$SS = m\hat{\beta} + b,$$

where  $m$  is the slope and  $b$  is the intercept. The linear transformation for the Keystone Exams was derived by anchoring the Proficient cut (i.e., Rasch ability estimate) recommended by the panelists at the standard-setting workshop to the scaled score 1499.5 (i.e., 1500 after rounding), and then set the slope of the line. There could be many lines with different slopes going through the anchor point. However, the slope of the line has influence over the variability of the scaled scores. For Keystone Exams, the slope of 50 was chosen because it results in desired scaled score standard deviation. Once the scaled score, slope, and Rasch ability estimate are determined, the

---

<sup>1</sup> Not everyone agrees with this sentiment. Some have argued the opposite point, that is, any attempt to add meaning to test scores actually predisposes the scores to be misinterpreted (Angoff, 1984).



intercept  $b$  can be derived by the equation above. The final slopes and intercepts for deriving scaled scores for the Keystone Exams are provided in Table 14–1.

**Table 14–1. Scaling Constants by Content Area**

Exam	Scaling Constants Slope	Scaling Constants Intercept
Algebra I	50	1,464.365
Biology	50	1,479.355
Literature	50	1,461.140

## ROUNDING

The linearly transformed scaled scores are always rounded to the nearest integer value for reporting purposes. Values greater than or equal to 0.50 are rounded up. Values less than 0.50 are rounded down.

## LOWEST OBTAINABLE SCALED SCORES

The Keystone Exams in Algebra I, Biology, and Literature have a lowest obtainable scaled score (LOSS) of 1200. Any derived scaled score less than 1200 is truncated to this minimum value. The selection of a LOSS is mainly based on two considerations: 1) extreme low scaled scores may have an impact on the average of the scaled scores at school/district level and 2) score truncation makes sense from a score precision perspective given measurement errors at the extremes are large. The LOSS value 1200 is established by giving consideration to *chance* performance over the MC items (e.g., if 40 four-option MCs were on a test, approximately 10 points might be earned on guessing alone) and considering the percentage of students who would be awarded the LOSS values.

## HIGHEST OBTAINABLE SCALED SCORES

A highest obtainable scale score (HOSS), 1800, is set for the Keystone Exams for the same reasons described for the LOSS value. However, unlike the LOSS value, which is set initially by giving consideration to guessing over MC items, it is somewhat more difficult to determine what rules should be applied to establish the HOSS. Based on the empirical results, the value 1800 corresponds to a logit value (or Rasch ability estimate) that ranged from 6 to 7, and 0 percent of students received this score.

## RAW-TO-SCALED-SCORE TABLES

The final raw-to-scaled score conversion tables can be found in Appendix K. Note that only the raw-to-scaled score tables for each single administration were reported. In other words, these tables cannot be used to look for a student's highest total scaled score to date if it is combined from two different administrations. The conditional standard error of measurement (CSEM, see Chapter Eighteen for detailed discussion) and corresponding confidence intervals based on 1 CSEM are also provided in these tables.

## CHAPTER FIFTEEN: EQUATING

---

Equating is a statistical process that is used to adjust scores on test forms so that scores from two different forms can be used interchangeably (Kolen & Brennan, 2004) even though the test forms consist of different items. In large-scale testing programs, it is a common practice to have different item sets appear in different test forms across administrations. Students' raw scores (or number-correct scores) cannot be compared across forms or administrations because the scores depend on the difficulty of the items on a form. The same student can score higher on an easy test than on a difficult test. Although there are various equating methods available for different psychometric paradigms (IRT and CTT), the Keystone utilizes an IRT approach aligned with the assumptions of the Rasch model, the IRT pre-equating method. The first step in any IRT equating method is to conduct scale linking, in which item difficulties from independent calibrations are transformed onto the same scale (Kolen & Brennan, 2014). Once scale linking is conducted, we can proceed with any IRT-based equating methods.

Since the Keystone's inception, a pre-equating design has been implemented due to the many advantages it offers. Specifically, employing a pre-equating method allows for a shortened turn-around time for score reporting due to the use of previously linked item parameters for test construction and development of raw-to-scaled-score tables. In this chapter, we provide a brief comparison of pre- and post-equating, an explanation of the procedure implemented for the Keystone examinations, and a description of the evaluation of pre-equated and post-equated solutions. Summary results are also presented. The equating design and analyses were conducted for all administrations (winter, spring, and summer) of the Keystone examinations.

### PRE- VS. POST-EQUATING

As with other Pennsylvania assessment programs, the Rasch model is used to guide the test design, form construction, calibration, scaling, and equating of the Keystone Exams. The key step of equating test forms using the Rasch model is to place the item parameters from different administrations on the same scale, also referred to as scale linking. Once the item parameters from different operational test forms are on the same scale, the Newton Raphson procedure can be used to convert number-correct scores to scaled scores as described in Chapter Fourteen. As a result, the scaled scores can be compared and used interchangeably within and across administrations.

As is the case with many K–12 large-scale assessment programs, all scored items are field tested prior to operational use. Once the field-test items' difficulties are placed on the base scale or common metric, in theory, one should not expect the Rasch item difficulties for these items to change (except within a reasonable range of measurement error) after they are administered in an operational test, provided the Rasch model fits the data. Based on this theoretical advantage of using the Rasch model, equating can be conducted using the item parameters that were previously calibrated. This statistical procedure is referred to as pre-equating.

In contrast, post-equating requires data from the current administration to be calibrated. Then, newly estimated item parameters are linked and placed on the same scale as banked item parameters, and scores are equated. With this in mind, pre-equating is advantageous because much of the work is completed before test administration, allowing more time for quality control; whereas post-equating relies on the same given timeframe for calibration, scale linking, equating scores, and implementing quality control procedures.

Although in theory the two equating procedures should provide identical results when the model fits the data, each has its own advantages and disadvantages. The use of pre-equating can facilitate the operational process in terms of rapid score reporting, more time for quality control, and more flexibility in the assessment. One successful application of pre-equating is for computer-adaptive tests (CATs) where test questions are tailored to the student's achievement as the test progresses, such as the Pennsylvania Classroom Diagnostic Tools (CDT). The CDT is designed to provide diagnostic information about student performance and is available throughout the school year at no cost. CATs require automated scoring for all item types (including constructed-response) and allow for immediate score reporting upon completion of the test. However, a variety of issues need to be considered when using pre-equating in practice. For example, students may not be motivated to take the field tests, especially standalone field tests, which may make the items appear harder in the field test than in the operational test (Eignor, 1985; Eignor & Stocking, 1986; Stocking & Eignor, 1986; Kolen & Harris, 1990). Other concerns for the field-test items include item context, item position, and sample size. In contrast, the use of post-equating, when applicable, does not have the same motivational concerns as with pre-equating. Also, post-equating uses post-administration

data and is thought to yield more accurate analysis results, given that the number of students who take the operational tests is usually large. On the other hand, when the reporting window is extremely tight, as is the case with some graduation or end-of-course exams in various states, post-equating must occur within a very short time, allowing less time for the equating analyses and quality control.

## **EQUATING DESIGN FOR KEYSTONE EXAMS**

The Keystone Exams, like many other graduation or end-of-course exams, require quick turnaround of test results. After the exams are administered, the bulk of the time is consumed by various data-processing steps. As a result, the equating analyses must be produced in a short period of time, which introduces risk of delays in score reporting. In addition, a student's best-score-to-date is calculated as the combination of performance across any administrations (see Chapter Sixteen for details), thus increasing the complexity of equating analyses and score reporting for future administrations. To control the quality and timeliness of post administration processing, pre-equating, one of the most promising applications of Rasch model or item response theory (see Lord, 1980, chap. 13), has been implemented for the Keystone Exams.

To implement the pre-equating model in the Keystone Exams, efforts have been made to enhance the accuracy of pre-equating results based on the findings from the literature. For example, to address the concerns regarding students' motivation to take field tests, stand-alone field tests were not used; rather field-test items were embedded throughout the test so that students would perceive no differences between field-test items and operational items. This approach allows Rasch item difficulty estimates to be used for future pre-equating purposes and is based on the assumption that students should be equally motivated to take the operational and embedded field-test items, especially when they are not aware of which item is a field-test item. To minimize item context and item position effects (i.e., lack of motivation and fatigue), field-test items were interspersed within the operational sections. With this design, students have a smaller chance of knowing the field-test item positions. Fatigue effects due to field-test items being placed in the last section of the operational test can be mitigated in this design as well. To improve the accuracy of the Rasch item difficulties and parameters estimated from the field-test data, DRC scored all MC items and a large sample of CR items given that larger sample sizes can increase the estimation accuracy. The test designs for the operational PSSA mathematics, ELA, and science assessments used multiple test forms that shared several common elements. The operational items were the same on all forms and for all students. Student total raw scores and scaled scores, as well as accountability reporting, were based exclusively on the operational items.

## **SCALE LINKING**

Keystone utilizes a concurrent calibration linking design to obtain item parameters for field-tested items (administered during the spring administration). Results from scale linking are item parameters (Rasch difficulties and thresholds) for field-tested (FT) items that are on the base scale. The chain originates from the scale of measurement defined for each test's base form, which is used as the reference for calibrating all items in the item pool. The base form is usually the form upon which the cut scores were established (see Chapter Thirteen). In the case of the Keystone, scales and cut scores were established for all subjects in 2011; therefore, the examinations are linked to the scales set in 2011.

The Rasch Partial Credit Model (RPCM) is used for the calibrating data, given its flexibility for dichotomously scored (i.e., MC) and polytomously scored (i.e., CR) item types (Masters, 1982). The RPCM is discussed in detail in Chapter Twelve. Without employing scale linking, Rasch difficulties for the field-tested items would not be directly comparable to other items on the base scale. A partially anchored concurrent calibration was employed to estimate all 2023 field-test item parameters for each test on its respective base scale. The following steps are conducted following each spring administration to estimate the FT item parameters.

1. Calibrate OP and MC FT items in a partially-anchored concurrent design to estimate FT MC item parameters.
  - a. Anchor OP item parameters to the banked values.
  - b. Include all OP and MC FT items.
  - c. Include only students who have completed the test and met other criteria described in Chapter Nine.

2. Calibrate OP and FT items in a partially-anchored concurrent design to estimate FT CR item parameters.
  - a. Anchor OP item parameters to the banked values.
  - b. Anchor MC FT item parameters to those estimated in Step 1.
  - c. Include all OP and FT items.
  - d. Include only students who have completed the test and met other criteria described in Chapter Nine.
  - e. Include students who were selected as a part of the FT sample for handscoring.
  - f. The resulting item parameters are banked for future use.

## PRE-EQUATING VERIFICATION

Although extra care has been taken to guarantee the success of pre-equating during the test design, form construction, and calibration of embedded field-test items, DRC also ensured that the pre-equated results had reasonable data-model fit during the pre-equating verification process. Once sufficient data was available, pre-equating verification was conducted to assess data-model fit and allow the parameters of any misfitting items to be freely calibrated. Any misfitting item was identified, and parameters were freely estimated in a subsequent calibration (using a partially anchored design) to improve data-model fit. All calibration was conducted using WINSTEPS (Linacre, 2019). Both sets of item parameters were then used to estimate student abilities, which were then transformed to scaled scores. (Transformation formulas are provided in Chapter Fourteen.) The data and results presented in this section refer to misfitting items as those in which infit mean-square values exceeded a criterion of 1.3. The number of items identified during pre-equating verification for each content area is shown in Table 15–1. In both Algebra and Literature, three misfitting items were identified in winter, one item was identified in the spring and no items were identified in the summer. For Biology, four items were identified across all administrations. Once items were identified, item parameters were freely estimated in a subsequent calibration and differences in scoring tables were analyzed between fully anchored pre-equated results (hereinafter “pre-equated”) and partially anchored pre-equated results (hereinafter “post-equated”). Pre-equating verification analyses were conducted at the item level, person level, and form level. Additional results from the pre-equating verification analyses can be found in Appendix L.

**Table 15–1. Number of Misfitting Items Identified during Pre-Equating Verification by Administration and Content Area**

Administration	Algebra I N	Biology N	Literature N
Winter	3	4	3
Spring	1	0	1
Summer	0	0	0

At the same time, DRC test development specialists reviewed all misfitting items (infit mean-square values greater than 1.3) and items with large displacement values (absolute value greater than 0.5) to ensure that items were presented in the same manner as their prior administration.

## ITEM-LEVEL ANALYSES

Item-level analyses indicate whether the data fit the Rasch model with respect to item-fit statistics and whether the item parameters showed displacement. For item misfit, this included the number of items that had reasonable fit statistics (e.g., greater than 0.7 and less than 1.3) supported by prior literature (Wright & Linacre, 1994). Table 15–2 shows the item fit statistics comparison for each content area and administration. For each administration and content area, both pre-equated and post-equated solutions showed similar fit to the model in terms of infit and outfit statistics. The results support that the data fit the pre-equated solution well.

**Table 15–2. Item Infit and Outfit Mean-Square Statistics by Administration and Content Area**

Admin.	Content Area	Method	N	Infit Mean	Infit SD	Infit Min	Infit Max	Infit Range [0.7,1.3]	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit Range [0.7,1.3]
Winter	Algebra I	Pre	54814	1.03	0.16	0.68	1.44	37/42	1.07	0.24	0.60	1.69	34/42
Winter	Algebra I	Post	54814	1.02	0.13	0.69	1.25	40/42	1.04	0.20	0.60	1.42	37/42
Winter	Biology	Pre	40073	1.03	0.15	0.69	1.51	49/54	1.04	0.19	0.63	1.55	47/54
Winter	Biology	Post	40073	1.02	0.12	0.71	1.25	54/54	1.02	0.16	0.63	1.34	51/54
Winter	Literature	Pre	35208	1.03	0.18	0.69	1.43	35/40	1.08	0.27	0.68	1.75	28/40
Winter	Literature	Post	35208	1.02	0.16	0.68	1.30	38/40	1.06	0.25	0.67	1.52	27/40
Spring	Algebra I	Pre	157276	1.02	0.14	0.70	1.31	41/42	1.04	0.19	0.66	1.39	39/42
Spring	Algebra I	Post	157276	1.02	0.13	0.71	1.22	42/42	1.03	0.19	0.66	1.38	39/42
Spring	Biology	Pre	137770	0.97	0.09	0.78	1.24	54/54	0.96	0.12	0.67	1.32	52/54
Spring	Biology	Post	137770	0.97	0.09	0.78	1.24	54/54	0.96	0.12	0.67	1.32	52/54
Spring	Literature	Pre	125512	1.03	0.13	0.73	1.33	39/40	1.10	0.22	0.73	1.61	34/40
Spring	Literature	Post	125512	1.03	0.13	0.74	1.27	40/40	1.08	0.21	0.73	1.54	35/40
Summer	Algebra I	Pre	274	0.97	0.13	0.74	1.24	42/42	0.97	0.16	0.70	1.32	40/42
Summer	Algebra I	Post	274	0.97	0.13	0.74	1.24	42/42	0.97	0.16	0.70	1.32	40/42
Summer	Biology	Pre	210	0.99	0.10	0.78	1.24	54/54	0.99	0.15	0.78	1.47	52/54
Summer	Biology	Post	210	0.99	0.10	0.78	1.24	54/54	0.99	0.15	0.78	1.47	52/54
Summer	Literature	Pre	111	0.99	0.18	0.56	1.26	36/40	1.01	0.23	0.56	1.59	32/40
Summer	Literature	Post	111	0.99	0.18	0.56	1.26	36/40	1.01	0.23	0.56	1.59	32/40

In addition, the results from the fully-anchored pre-equated calibration were assessed to identify any issues with item parameter stability. Displacement refers to the shift in item parameters between anchored values and values that would have been freely estimated in an unanchored calibration. The items with an absolute displacement value greater than or equal to 0.5 were identified and reviewed by DRC test development specialists. Table 15–3 summarizes the outliers flagged by displacement. These items were reviewed by DRC content specialists, but no obvious reasons were found to explain the shift in item parameters. It is somewhat expected to see fewer items identified in the spring than the winter and summer administrations because the spring administration has the highest volume of test-takers and it is the most representative of the sample used for field-testing and subsequent item parameter estimation.

**Table 15–3. Count of Items Flagged for Displacement by Administration and Content Area**

Administration	Algebra I N	Biology N	Literature N
Winter	4	4	4
Spring	0	0	2
Summer	4	1	4

## PERSON-LEVEL ANALYSES

The second set of analyses conducted consisted of analyzing person-level fit statistics, which can be another indicator of whether the data fit the model. Table 15–4 summarizes the overall person-level infit and outfit statistics by administration and content area for both the pre-equated and post-equated solutions. The table specifies the mean, standard deviation (SD), minimum (Min), maximum (Max), and proportion of persons who had reasonable fit statistics (e.g., greater than 0.5 and less than 1.5)<sup>1</sup> for both infit and outfit statistics. The results in the tables indicate that person-level fit does not vary by equating method.

Furthermore, Appendix L includes the results for the pre-equating verification, including the person infit boxplots for all administrations and content areas for both pre-equated and post-equated solutions. Appendix L also provides boxplots disaggregated by ethnicity, gender, English Learners (ELs), and students with individualized educational programs (IEPs). The person infit plots indicate that the data fits the pre- and post-equated solutions similarly.

**Table 15–4. Person Infit and Outfit Mean-Square Statistics by Administration and Content Area**

Admin.	Content Area	Method	N	Infit Mean	Infit SD	Infit Min	Infit Max	Infit Range [0.5,1.5]	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit Range [0.5,1.5] <sup>+</sup>
Winter	Algebra I	Pre	54814	1.00	0.20	0.37	2.75	98.5%	1.07	0.27	0.08	9.66	92.9%
Winter	Algebra I	Post	54814	0.99	0.20	0.35	2.76	98.5%	1.04	0.25	0.08	9.50	94.7%
Winter	Biology	Pre	40073	1.02	0.12	0.49	2.34	99.8%	1.04	0.13	0.20	2.29	99.5%
Winter	Biology	Post	40073	0.99	0.12	0.50	2.29	99.8%	1.02	0.12	0.19	2.30	99.8%
Winter	Literature	Pre	35208	0.99	0.21	0.32	2.27	98.2%	1.08	0.32	0.15	5.53	89.7%
Winter	Literature	Post	35208	0.98	0.21	0.30	2.27	98.1%	1.06	0.30	0.15	5.81	91.3%
Spring	Algebra I	Pre	157276	1.00	0.21	0.23	3.63	97.2%	1.04	0.23	0.10	9.90	96.0%
Spring	Algebra I	Post	157276	0.99	0.21	0.24	3.62	97.20%	1.03	0.22	0.09	9.90	96.20%
Spring	Biology	Pre	137770	0.97	0.15	0.51	2.89	99.30%	0.96	0.15	0.26	3.69	99.10%
Spring	Biology	Post	137770	0.97	0.15	0.51	2.89	99.30%	0.96	0.15	0.26	3.69	99.10%
Spring	Literature	Pre	125512	1.00	0.24	0.35	2.52	96.30%	1.10	0.38	0.12	9.90	87.30%
Spring	Literature	Post	125512	1.00	0.24	0.34	2.53	96.20%	1.08	0.37	0.11	9.90	88.00%
Summer	Algebra I	Pre	274	0.96	0.23	0.55	2.23	97.40%	0.97	0.17	0.47	1.68	98.90%
Summer	Algebra I	Post	274	0.96	0.23	0.55	2.23	97.40%	0.97	0.17	0.47	1.68	98.90%
Summer	Biology	Pre	210	0.97	0.13	0.66	1.42	100.00%	0.99	0.11	0.61	1.66	99.50%
Summer	Biology	Post	210	0.97	0.13	0.66	1.42	100.00%	0.99	0.11	0.61	1.66	99.50%
Summer	Literature	Pre	111	0.93	0.17	0.63	1.35	100.00%	1.01	0.23	0.55	1.91	95.50%
Summer	Literature	Post	111	0.93	0.17	0.63	1.35	100.00%	1.01	0.23	0.55	1.91	95.50%

<sup>1</sup> While items and persons are on the same scale, items tend to be more stable. As such, stricter rules are applied to item-fit statistics than person-fit statistics in determining reasonable fit.

## NORMALIZED SCALED SCORE DIFFERENCES

On the form level, we evaluated differences between pre-equated and post-equated results. Normalized differences were calculated as the difference between the scaled score divided by the average CSEM of pre- and post-equated results at each raw score point (see Equation below). Table 15–5 displays the descriptive statistics of the normalized scaled score differences as well as the normalized scaled score difference at the proficient cut-score. A difference at the proficient cut-score may indicate a difference in performance level classification. Results indicated that normalized differences were all within reasonable expectations (min = -0.06, max = 0.23), where the largest difference was observed for winter Biology. The plots for normalized scaled score differences are displayed in Appendix L.

$$\text{Normalized Scale Score Difference} = \frac{(SS_{Pre} - SS_{Post})}{((CSEM_{Pre} + CSEM_{Post})/2)}$$

**Table 15–5. Normalized Scaled Score Differences Summary by Administration and Content Area**

Administration	Content Area	SS Difference Mean	SS Difference Min	SS Difference Max	SS Difference at Prof. Cut
Winter	Algebra I	0.070	0.000	0.143	0.071
Winter	Biology	0.138	0.000	0.231	0.231
Winter	Literature	0.074	0.000	0.133	0.063
Spring	Algebra I	0.020	-0.063	0.071	0.071
Spring	Biology	0.000	0.000	0.000	0.000
Spring	Literature	0.030	0.000	0.067	0.000
Summer	Algebra I	0.000	0.000	0.000	0.000
Summer	Biology	0.000	0.000	0.000	0.000
Summer	Literature	0.000	0.000	0.000	0.000

## PERFORMANCE LEVEL CLASSIFICATION

Pre-equated solutions were considered reasonable if classification consistency did not change more than 5%. Table 15–6 shows the consistency of classifications with respect to performance levels. The three numeric values within each cell refer to the proportion of students that do not agree at each of the three cuts (Basic, Proficient, and Advanced). If a numeric entry is followed by a negative sign, then pre-equating resulted in a lower percentage of students in the adjacent performance level when compared to post-equating. On the other hand, if the numeric entry is followed by a positive sign, then pre-equating resulted in a higher percentage of students in the adjacent performance level when compared to post-equating.

Performance level classification was identical between the pre- and post-equated solutions for all administrations and content areas except winter Biology and Literature. There was the most movement in winter Biology performance level classifications between the pre-equated and post-equated solution. Yet these results indicate that the pre-established criterion, in which no more than 5% of performance level classifications changed between the pre- and post-equated solutions, was met. After comparing and evaluating the results, the percentage of students classified differently was less than 5% within each classification. The TAC agreed if classification consistency was less than 5%, then pre-equated solutions should be accepted.

**Table 15–6. Performance Level Impact Summary between Pre- and Post-Equated Solutions by Administration and Content Area**

Administration	Algebra I	Biology	Literature
Winter	Exact	(4+,1+,1+)	(3+,0,0)
Spring	Exact	Exact	Exact
Summer	Exact	Exact	Exact

## SCALE STABILITY AND MAINTENANCE

Scale stability is a critical component of any testing program. The 2014 Standards of Educational and Psychological Testing state that “Testing programs that attempt to maintain a common scale over time should conduct periodic checks of the stability of the scale on which the scores are reported” (p.103). Conducting item parameter checks, ensuring that item parameters do not drift over time, and potentially updating operational item parameters are a few ways in which testing programs can maintain scale stability. Although many of these aspects are checked during the pre-equating verification process, it is also important to analyze student performance and scale stability following each administration.

## TEST CHARACTERISTIC CURVES AND LOGIT PLOTS

Figures 15–1 and 15–2 help one visualize the across-administration differences in the difficulties of operational items. In Figure 15–1, the test characteristic curves (TCCs) for each administration are presented for each content area. These figures show the winter, spring, and summer TCCs and indicate alignment among the 2023 test forms in terms of difficulty in the logit metric. TCCs that are closely aligned translate into similar raw-score cut points and similar test difficulty across years. The three dotted vertical lines represent the Basic, Proficient, and Advanced cut-scores on the logit (theta) scale. All content areas showed very small differences across administrations.



Figure 15–1. Test Characteristic Curves

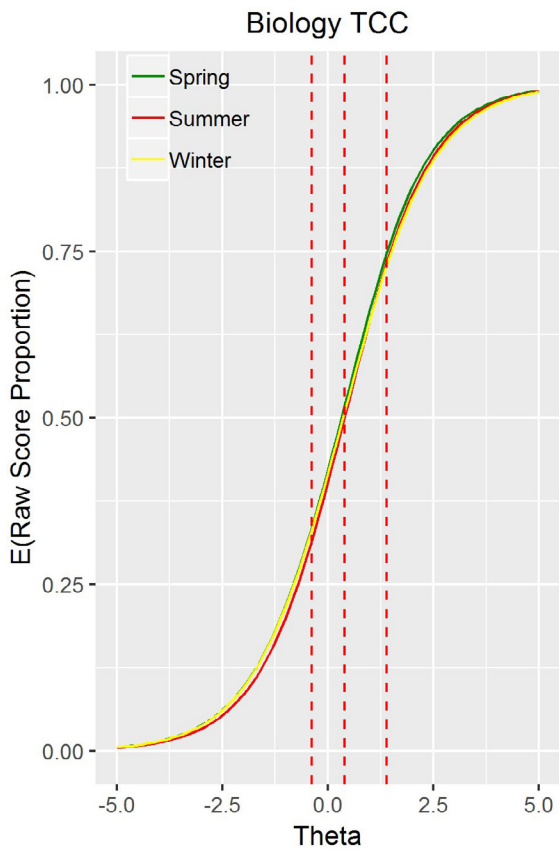
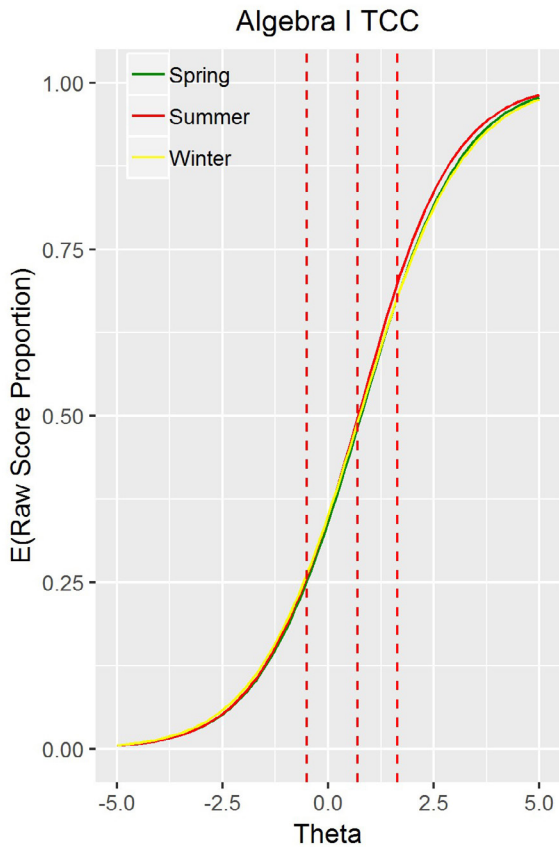


Figure 15–1 (continued). Test Characteristic Curves

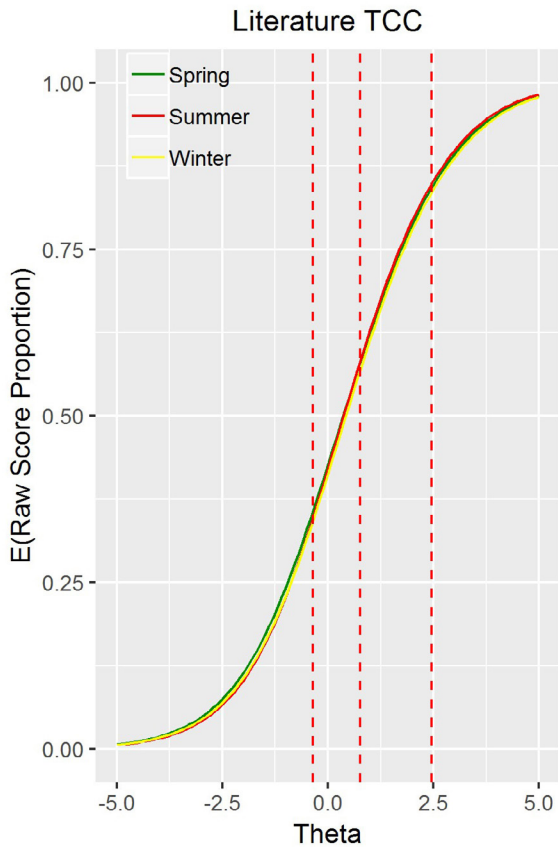


Figure 15–2 displays the relationship between the pre-equated item difficulties (x-axis) and the post-equated item difficulties (y-axis) on the logit (theta) scale. The black line represents the identity line; if points fall on the identity line, it indicates that there is no difference between the pre-equated and post-equated item difficulty. Points that do not fall on the identity line indicate items that were identified as misfitting during the pre-equating verification process and were freely estimated in a subsequent calibration. In all cases, the item difficulties that were freely estimated under the post-equated model were close to the identity line. The plots provide evidence of reasonable across-year stability of item difficulty, meaning the pre-equated item difficulties were similar to the post-equated item difficulties.

Figure 15-2. Logit Plots

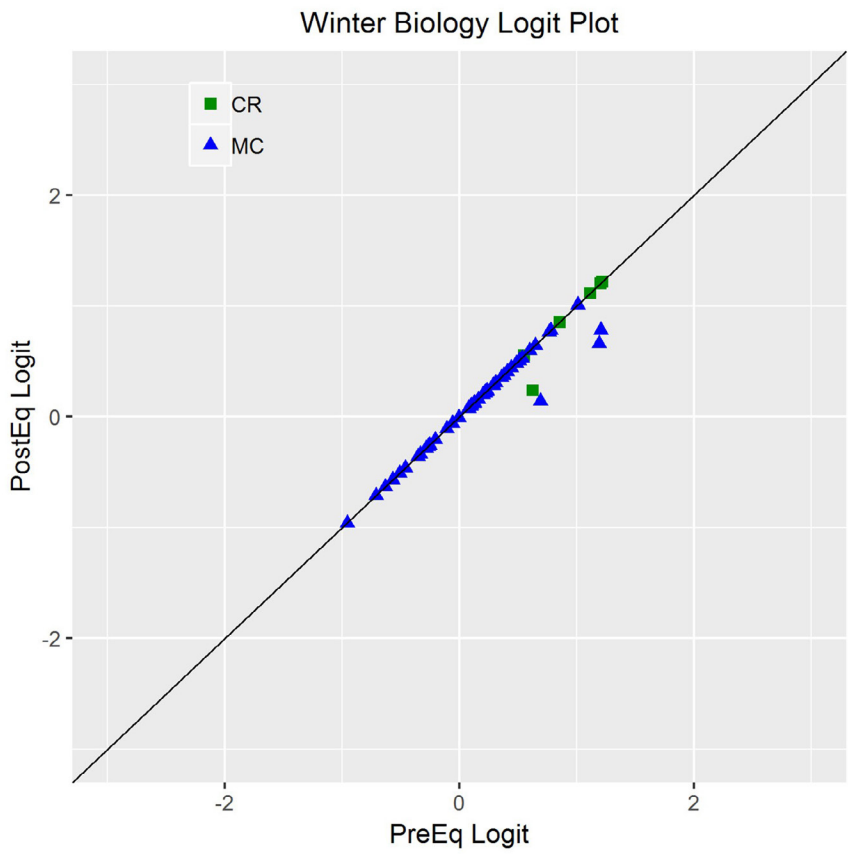
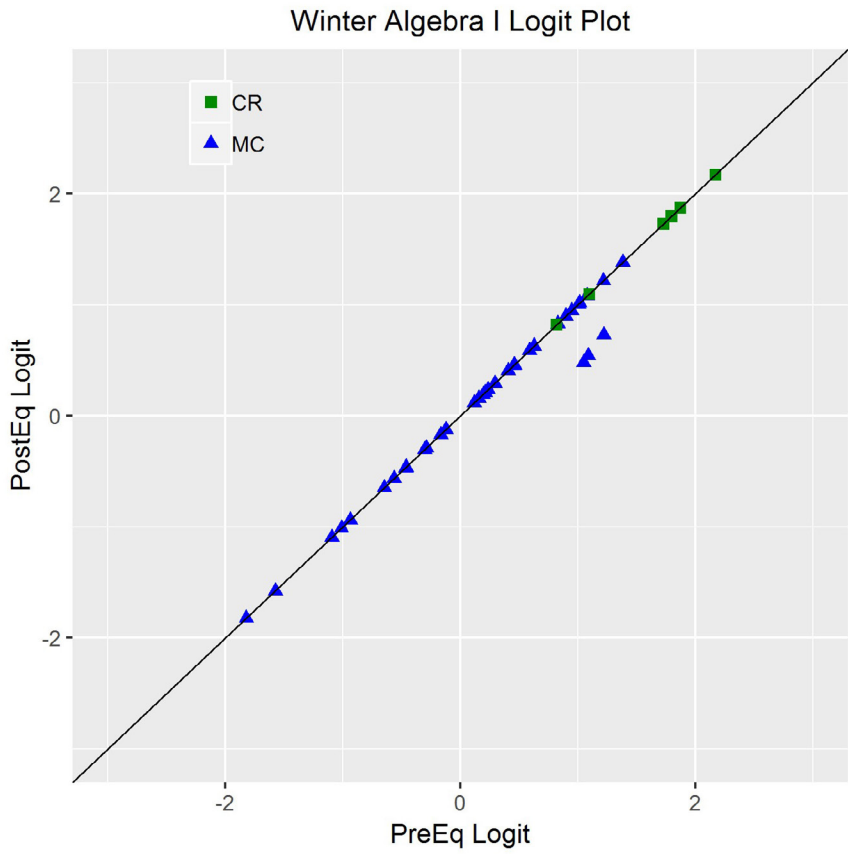


Figure 15-2 (continued). Logit Plots

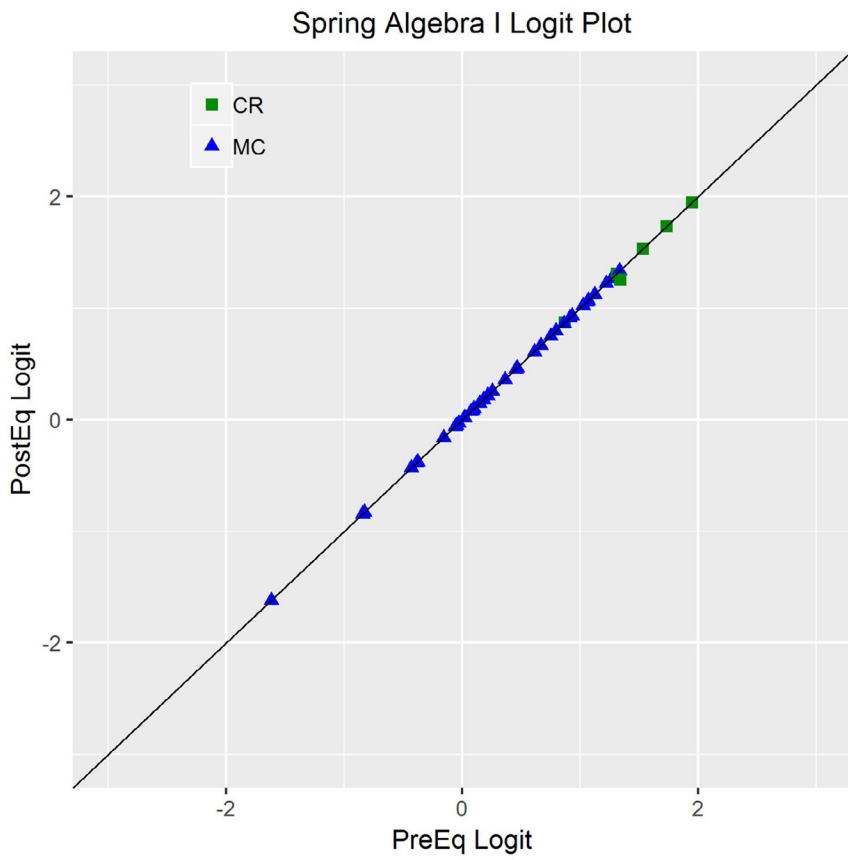
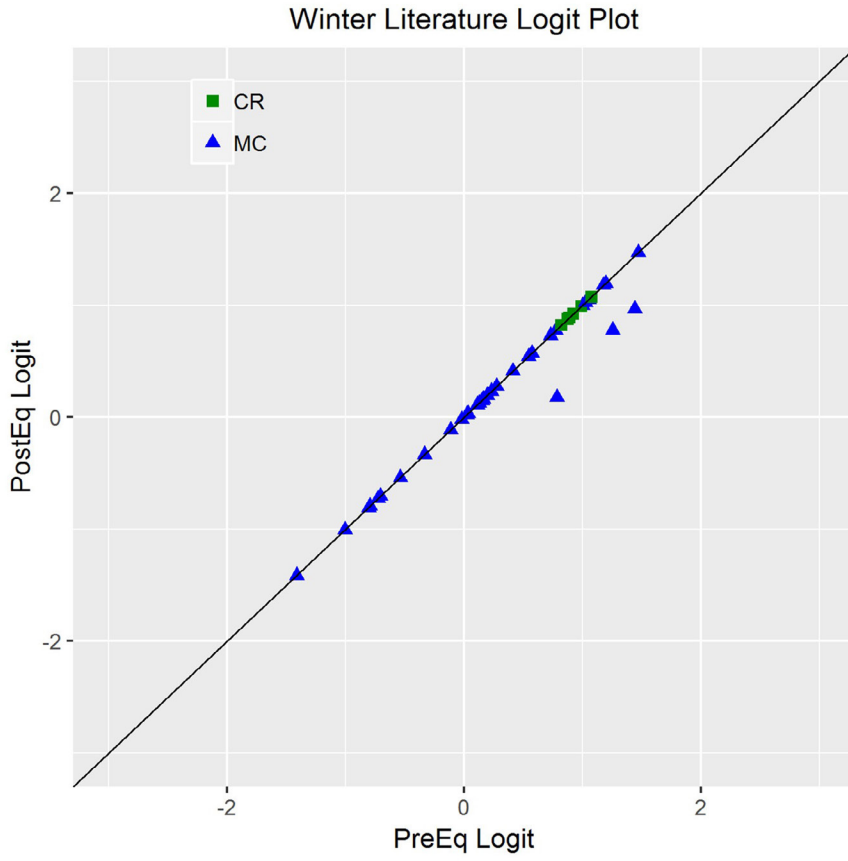


Figure 15-2 (continued). Logit Plots

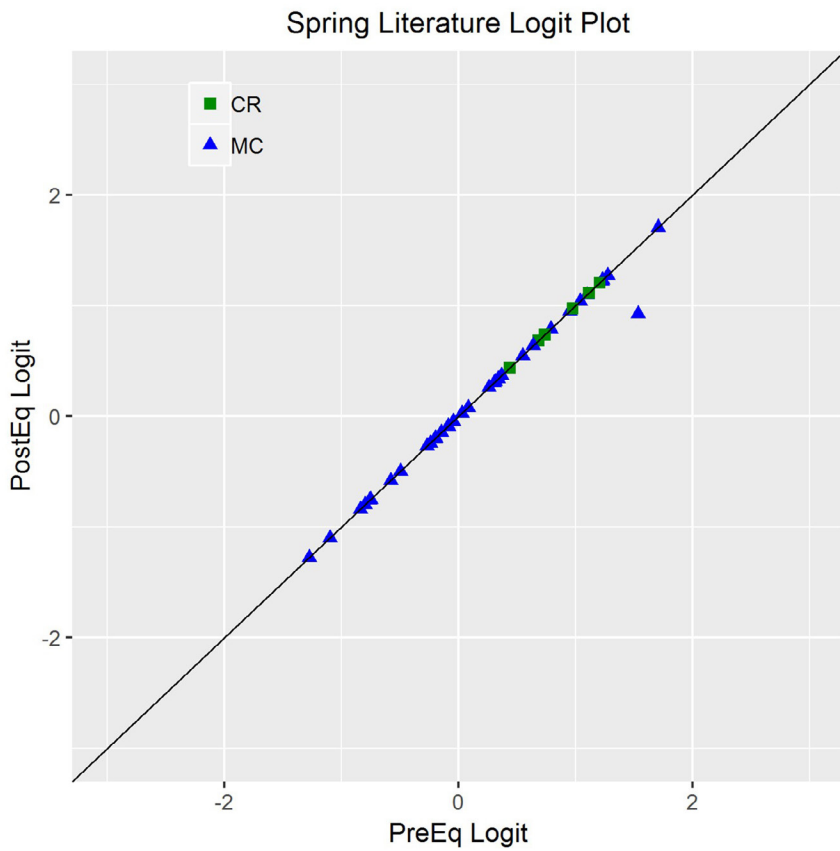
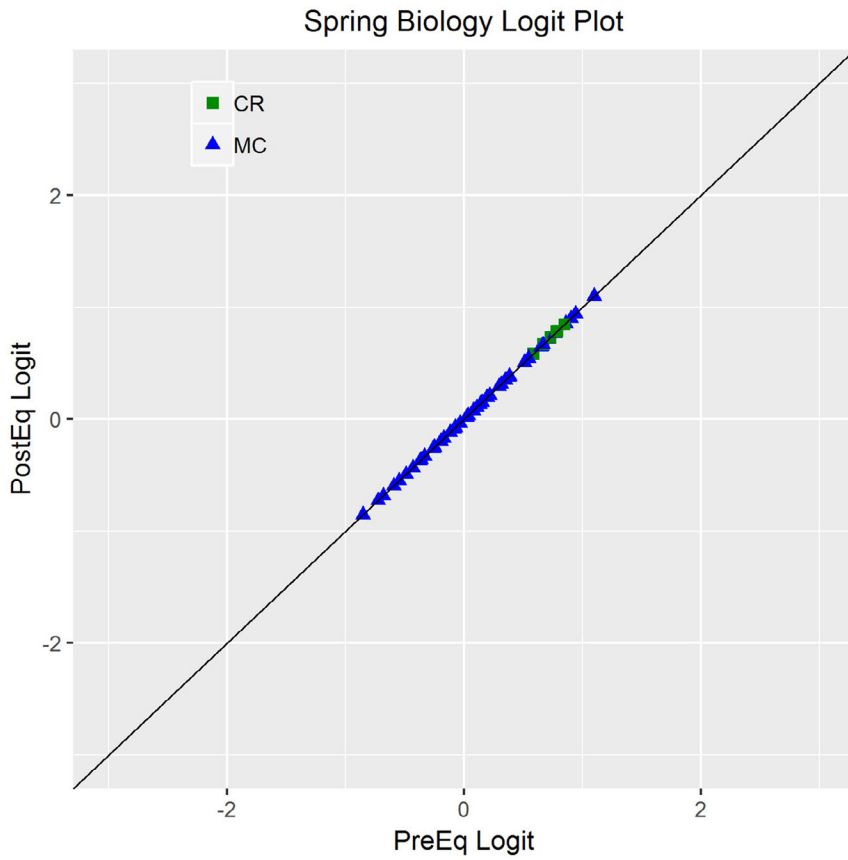


Figure 15-2 (continued). Logit Plots

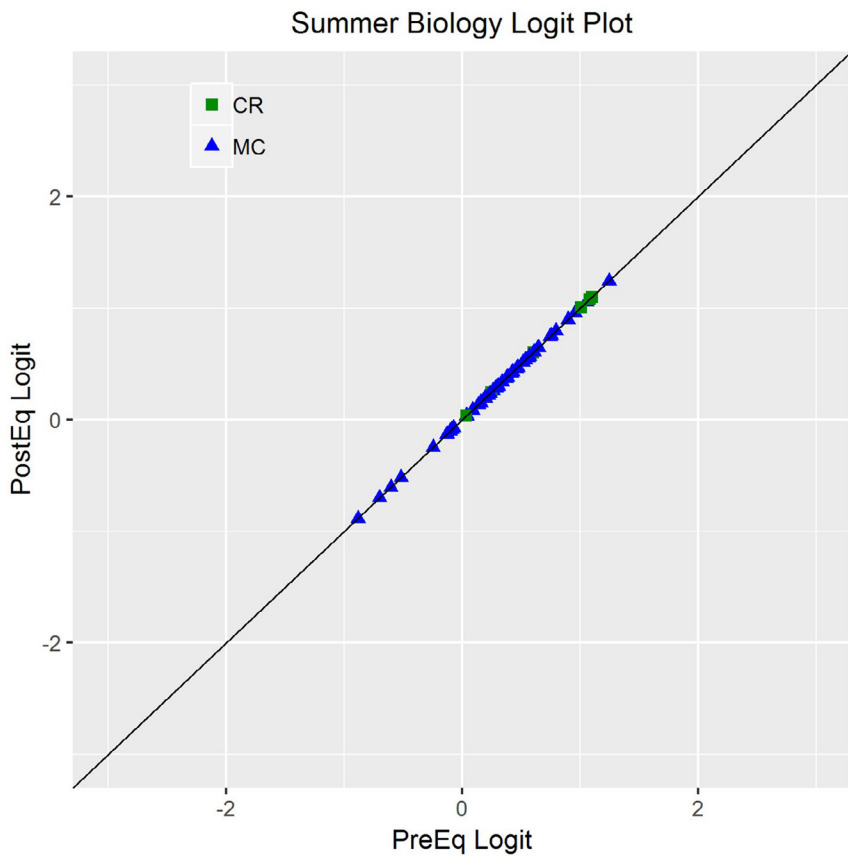
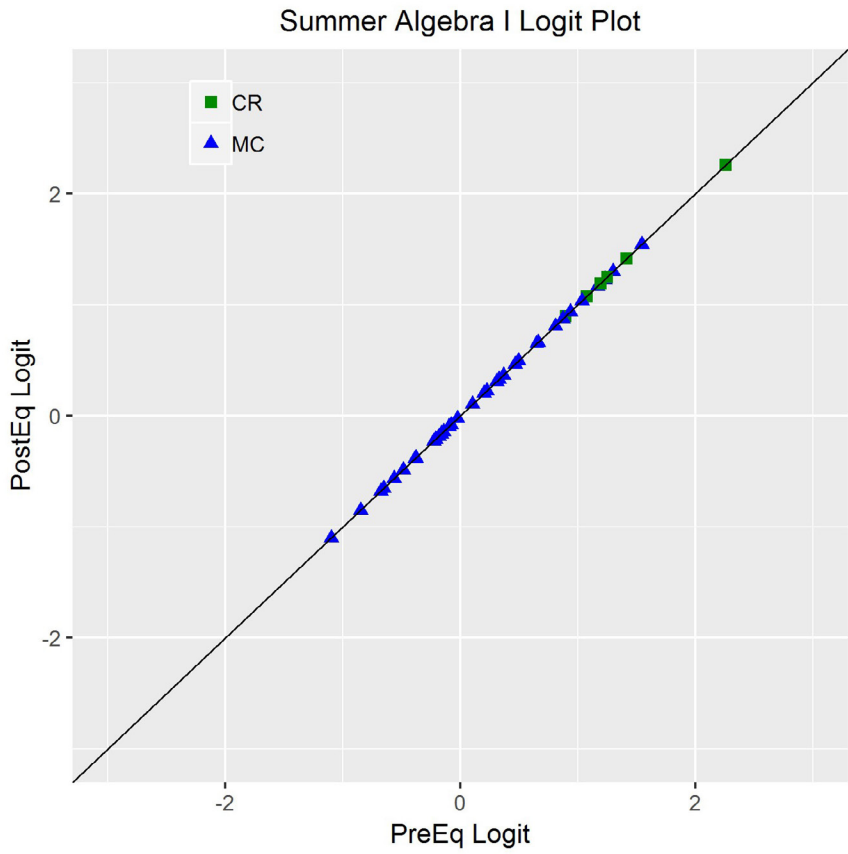
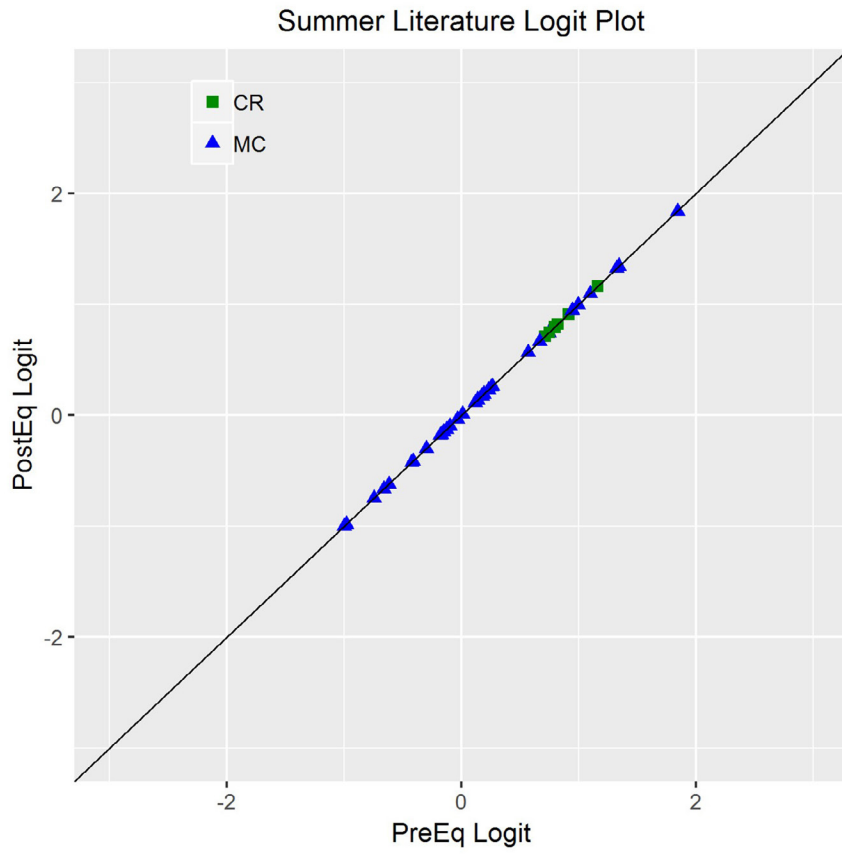


Figure 15-2 (continued). Logit Plots



## CHAPTER SIXTEEN: SCORES AND SCORE REPORTS

This chapter provides information about the scores provided for the Pennsylvania Keystone Exams (e.g., scaled scores, performance levels, and module scores), how the scores are presented on score reports, and appropriate and inappropriate uses of the scores.

### SCORING

Keystone Exams items include both multiple-choice (MC) and constructed-response (CR) items. Each correct response to an MC item receives a score of 1. Incorrect responses receive a score of 0. Scores on CR items range from 0 to 4, depending on the content area. Table 16–1 summarizes the types of items used in each content-area exam.

**Table 16–1. Item Types Used by Content Area**

Exam	Item Type MC (1 point)	Item Type CR (3 point)	Item Type CR (4 point)
Algebra I	■		■
Biology	■	■	
Literature	■	■	

### DESCRIPTION OF TOTAL-TEST SCORES

Different types of scores have been developed for Keystone Exams reporting. Since the underlying properties of these scores are not necessarily the same, the particular scores used depend on the purposes for which the test has been given. The following types of scores are provided for reporting overall performance on each Keystone Exam:

- Raw scores
- Scaled scores
- Highest total test scaled score to date
- Performance levels

### RAW SCORES

A raw score (or number-correct score) is the number of points a student earned over all the operational MC and CR items. By itself, the raw score has very limited utility. One limitation is that it can only be interpreted with reference to the total number of items on a specific exam (e.g., a raw score of 15 on a 20-item exam is different from a raw score of 15 on a 30-item exam). In addition, raw scores depend on the difficulty of test items across test forms (e.g., a raw score of 15 on a test with 20 easy items is different from a raw score of 15 on a test with 20 difficult items). Because the difficulty of the items on a test can change from administration to administration, raw scores should not be compared across administrations.

### SCALED SCORES

Scaled scores were introduced in Chapter Fourteen. In the simplest sense, a scaled score is a transformed number-correct score. The specifics of the transformation processes for the Keystone Exams were also discussed in Chapter Fourteen. When all students take the same test items, as with the operational items on the Keystone Exams, the more points the student earns, the higher the associated scaled score will be.

The value of switching to the more abstract scaled-score metric is that it produces more general, interpretable, and equitable results. As noted above, a raw score of 30 is meaningless unless the maximum raw score is known. The difficulty of the test items was also mentioned as an additional challenge with interpreting raw scores. Number-correct scores are transformed to scaled scores to remove the effects of test length and item difficulty. (Strictly speaking, transformation of number-correct scores to percent-correct scores would also remove the effect of test length, but it would do nothing to adjust for the difficulty of the items.)



Another advantage of scaled scores is that they lend themselves to interpretations at what is referred to as an interval level, while raw scores do not. Interval-level scales allow an interpretation of a scaled score difference of 5 points to be the same whether the scores are 1295 vs. 1300 or 1445 vs. 1450. Raw-score differences, in this context, cannot be interpreted in this manner and are thus neither generalizable nor equitable.

A scaled score of 1500—or any other value for a particular content-area exam, such as Algebra I—should have the same absolute meaning in the current administration as it had in previous administrations when test scores are properly equated across administrations. More importantly, a significant increase in the scaled score from the previous administration to the current administration means that student performance improved<sup>1</sup>; it does not say anything about whether this administration's exam is easier or harder than last administration's exam. To make these interpretations requires no information about the length or the difficulty of the exam in either administration, although these variables are essential for the process of deriving the scaled scores.

There is considerable auxiliary information presented in this report that might aid in further contextualizing Keystone Exams scaled scores:

- Chapter Fourteen provides information on the development of the Keystone Exams scaled-score system, including transformation formulas, rounding rules, and general scale characteristics (e.g., minimum values).
- Chapter Seventeen provides total-test score statistics. In particular, Table 17–2 lists the scaled-score means and standard deviations for the testing results.

## HIGHEST TOTAL TEST SCALED SCORE TO DATE

After an administration, every test-taker earns an administration-specific scaled score based on their overall performance, as well two scaled scores based on their performance on Module 1 and Module 2. As previously discussed, each examination consists of two modules that measure distinct topics within each content area. For example, for Literature Module 1 represents fiction-based content and Module 2 represents non-fiction-based content. If a student takes a Keystone Exam multiple times, they also receive a highest total test score to date that represents the best combination of performance on Module 1 and Module 2. The advantage of this calculation is to combine Module-level performance from different administrations that may represent a higher total score and performance level (i.e., proficient or advanced). Both the administration-specific scaled scores (Module 1, Module 2, and total) and the highest total test scaled score are included on the score reports (see Figures 16–1 and 16–2).

In this technical report, both administration-specific and highest test scaled score to date information is presented. In general, information about raw scores always represents the administration-specific performance. Meaning, results represent the raw scores earned on that specific administration. On the other hand, information about scaled score performance represents the highest scaled score to date calculation unless otherwise specified. Moreover, information about performance level classifications also represents the performance level that corresponds to the highest total test scaled score to date. Meaning when scaled scores and performance levels are presented, results represent the best combination of Module 1 and Module 2 that yield the highest overall score and performance level classification.

## PERFORMANCE LEVELS

Keystone Exams results are also reported using four performance levels: Below Basic, Basic, Proficient, and Advanced. The cut scores on the scaled-score metric (i.e., the lowest possible scaled score to enter the Basic, Proficient, and Advanced levels) were presented earlier in this report. However, the information is repeated below (Table 16–2) for convenience.

---

<sup>1</sup> This example is not an endorsement of conducting a trend analysis with just two years of results. Further, small differences may not be statistically or practically significant.

**Table 16–2. Scaled Score Cuts for Each Performance Level by Content Area**

Exam	Min	Scaled Score Cuts BB/B	Scaled Score Cuts B/P	Scaled Score Cuts P/A	Max
Algebra I	1,200	1,439	1,500	1,546	1,800
Biology	1,200	1,460	1,500	1,549	1,800
Literature	1,200	1,444	1,500	1,584	1,800

Note. BB = Below Basic; B = Basic; P = Proficient; and A = Advanced

Performance level descriptors (PLDs) are another way to attach meaning to the scaled-score metric. They associate precise quantitative ranges of scaled scores with verbal, qualitative descriptions of student status. While much less precise, the qualitative description of the levels is one way for parents and teachers to interpret the student scores. They are also useful in assessing the status of the school. The Pennsylvania General PLDs developed by Pennsylvania Department of Education (PDE) and teacher panels are given below. These are also included on student score reports.

- **Advanced:** Superior academic performance indicating an in-depth understanding and exemplary display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
- **Proficient:** Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content.
- **Basic:** Marginal academic performance indicating work approaching, but not yet reaching, satisfactory performance. Performance indicates a partial understanding and limited display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. The student may need additional opportunities and/or increased student academic commitment to achieve the Proficient level.
- **Below Basic:** Inadequate academic performance indicating little understanding and minimal display of the skills included in the Keystone Exams Assessment Anchors and Eligible Content. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.

## DESCRIPTION OF MODULE SCORES

Each of the Keystone Exams in Algebra I, Biology, and Literature contains two modules. A module score describes performance of a student, school, or district on a particular module (content standard defined in the exam). The following types of scores are provided for Keystone Exams at module level:

- Raw scores
- Scaled scores
- Performance levels

## MODULE RAW SCORES

Raw scores at module and assessment anchor levels were reported in different summary reports. As described earlier, a raw score is the number of points a student earned over all the operational MC and CR items; it depends on the difficulty and length of the test form; and it should not be compared across administrations. In the summary reports, the school, district, and/or state median points earned were reported at module and assessment anchor levels. These raw scores can provide some diagnostic information when they are compared with the minimum estimated points needed to pass. The latter is calculated by summing the probabilities of a barely proficient student answering the items included in a module or assessment anchor correctly. The sum is rounded up to the nearest integer. The probability is derived using the Rasch models discussed in Chapter Twelve.

## **MODULE SCALED SCORES**

The module scaled scores were provided in the individual student report. For the Keystone Exams, the module scaled score represents a student's achievement on each module. They can be compared across administrations because they are statistically equated. However, it is not advisable to compare scores across modules because each module contains varying item content and difficulty. This variation is also the reason the total scaled score is not the average of the two modules' scaled scores.

## **MODULE PERFORMANCE LEVELS**

Based on the testing results at the module level, students can be classified as Passed or Not Passed. The derived scaled score cut is 1500 for both modules. This cut score is determined by panelists' recommendations for the proficient cut of the corresponding total test. Note that a student who does not pass a module can still be Proficient or above on the total test if the student performs very well on the other module. If a student is not proficient on the total test but passes one module, although it is recommended that this student take both modules during retesting, the student can choose to take just the non-passed module because the final score is based on the highest combination of module scores.

## **APPROPRIATE SCORE USE**

### **INDIVIDUAL STUDENTS**

Scaled scores on the Keystone Exams indicate a student's achievement with respect to the Keystone Exams Assessment Anchors and Eligible Content. Scaled scores are primarily used to determine student performance level classifications (i.e., a criterion-referenced inference). Scaled scores that are based on Rasch models are typically assumed to be of the interval type, so comparisons may be made on differences in scaled scores. If this assumption holds, then it would be safe to infer for Algebra I that the ability difference between 1410 and 1420 represents the same ability difference that separates 1550 and 1560. Scaled scores can also be used to compare the performance of an individual student to the performance of a similar demographic or subgroup at a school or district. Test score standard errors (discussed in Chapter Eighteen) should be considered.

### **GROUPS OF STUDENTS**

Test results can be used to evaluate performance over time. Mean scaled scores can be compared across administrations within the same content area to indicate whether a student's performance is improving across years. Generally, such trend analyses benefit from using mean results from as many test administrations as possible. Different cohorts of students are used (i.e., the same student or students are not tracked across grade levels). All scores can be analyzed within the same content area for any single administration to determine which demographic or program group had, for example, the highest average performance or the highest percentage of students at or above Proficient.

Module scores can help evaluate academic areas for relative strengths or weaknesses. These module scores provide information to identify areas where further diagnosis is warranted. Generalizations from test results may be made to the specific content domain represented by the academic standards measured in the Keystone Exams. However, all instruction and program evaluations should include as much information from other sources as possible to provide a complete picture of performance.

## **CAUTIONS FOR SCORE USE**

### **EXTREME ERROR FOR EXTREME SCORES**

Student scores toward the minimum or maximum ends of the score range have very large standard errors of measurement (SEM) and, therefore, should be interpreted very cautiously. The maximum scaled score only provides a very rough estimate of a student's ability. For example, if a student achieved the maximum scaled score (i.e., 1800), it cannot be determined whether this student could have achieved an even higher scaled score. If the test were 10 items longer, a different estimate might have been obtained. Similarly, if the items on a new test are more difficult than the items on a previous administration, the maximum scaled score would likely be higher on the new test because it would take a greater level of achievement to answer the items correctly. In this manner, extreme scaled scores may vary from one administration to the next even if the number of test items does not change. The fluctuation of extreme scaled scores complicates the comparisons of students with scaled scores at the extreme ends of the score distribution. To minimize confusion and potential misinterpretation, the minimum and maximum

scaled scores possible on the Keystone Exams have been fixed (see Table 16–2) so they do not change between administrations.

## **UNIQUE SCALE FOR EACH CONTENT AREA**

Scaling was conducted for each content-area exam separately. Therefore, the scaled scores should be interpreted only within each content area. The scaled scores are not status indicators in the same sense as percentile ranks (or scales that are essentially transformations of percentile ranks) and therefore cannot be used to profile relative strengths and weaknesses across content areas. As an example, the scaled scores of 1450 in Algebra I and 1400 in Biology gained by a student do not necessarily imply that the student performed better in Algebra I than in Biology.

## **USING KEYSTONE EXAMS RESULTS FOR OTHER PURPOSES**

Other uses or inferences based on Keystone Exams results may or may not be valid as the validity evidence and arguments provided in Chapter Nineteen may not necessarily support other score uses and interpretations. According to the *Standards for Educational and Psychological Tests* (AERA, APA, & NCME, 2014), if a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary. Finally, a universal caveat for any test’s result is that it should not be used for placement and educational planning alone. Instead, other information about the student (e.g., other test performance data) should be included.

## **REPORT DEVELOPMENT**

Several months prior to the first release of reports for the Keystone Exams, PDE and DRC conducted focus groups with Pennsylvania educators and parents/guardians. In the focus groups, educators and parents/guardians provided feedback on sample report for the Keystone Exams. Feedback from the focus groups was used to inform the design and content of the Keystone individual and summary reports. The focus groups targeted educator and parent/guardian constituencies in three geographic regions of the state—the Pittsburgh area, the Harrisburg area, and the Philadelphia area.

Two preliminary educator groups were convened in Harrisburg on November 15 and 17, 2010. These groups, totaling 34 educators, reviewed the student report and provided feedback using both a survey and group discussion. Substantive changes to the individual student report were made as a result of these meetings, with two different versions of the report emerging from these reviews. These two groups did not review the summary reports.

A second set of focus groups were conducted in December 2010 to review the updated reports. For the December meetings there were 35 panelists (22 educators & 13 parents) for six focus groups in Pittsburgh (December 3), Harrisburg (December 6), and King of Prussia (December 7). The three educator groups reviewed the two versions of the student report and the one version of the school summary report. The three parent groups reviewed the two versions of the student report.

Feedback from these two focus groups was taken into consideration during final report development. For more information about the focus groups, please refer to the *Keystone Exams Score Report Focus Group Findings* (Pennsylvania Department of Education, 2011).

## **REPORTS**

The following score reports are provided to students, schools, and districts for the Keystone Exams in Algebra I, Biology, and Literature:

- Parent Letter
- Individual student report
- School summary report
- District summary report
- State summary report
- Report interpretation guide

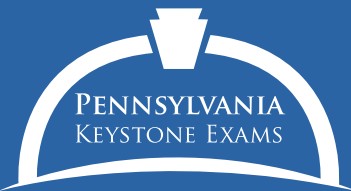
## **PARENT LETTER**

Parent letters were delivered to Pennsylvania districts when district files were posted after each Keystone administration. This score report provided parents and students with their first glimpse of performance on the Keystone Exams. This report provides results at the student level.

## **INDIVIDUAL STUDENT REPORT**

A student report is provided for all students who took the Keystone Exams. Two copies of the individual student report for all Keystone Exams were sent to each school district and charter school for distribution to parents, teachers, guidance counselors, and/or principals. School districts and charter schools may publish the results of the Keystone Exams school-level reports. This report is a two-page color document that provides the types of scores explained earlier in this chapter. Screenshots of the two pages from a sample individual student report are provided in Figures 16–1 and 16–2. Figure 16–2 displays the second page of the Individual Student Report, including scale score on each module, and the highest total test scale score to date.

Figure 16–1. Page 1 of the Individual Student Report



# Student Report

**Student Name:** SAMPLE STUDENT 1


**PA Student ID:**

**School:**

**District:**

**Test Date:**

**Grade:**

**Content Area:** 

Biology

Student's Keystone Exam Result			
		Goal Range	
Below Basic	Basic	Proficient	Advanced
		✓	

Dear Parent/Guardian:

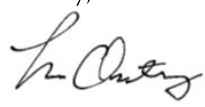
On behalf of the Pennsylvania Department of Education, I would like to thank you for taking the time to review this report. Academic success begins at home. Children with parents/guardians who are engaged in their child's academics are more likely to enjoy the learning process and succeed.

This report provides information about your child's recent performance on a Pennsylvania test known as the Keystone Exam. On this page, you can see your child's overall performance – below basic, basic, proficient or advanced.

Within this report, you will find specific information about your child's performance on the Biology Keystone Exam. It displays your child's Highest Total Test Scale Score to Date for Module 1 and Module 2. Module 1 assesses Cells and Cell Processes, and Module 2 assesses Continuity and Unity of Life.

For detailed information about the Keystone exams, please visit the Pennsylvania Department of Education's Standards Aligned System website at [www.pdesas.org](http://www.pdesas.org), or contact your child's school.

Sincerely,



Noe Ortega  
Acting Secretary of Education

**About the Keystone Exams**


The Keystone Exams are end-of-course assessments designed to evaluate student performance on academic content. The purpose of the Algebra I, Biology, and Literature Keystone Exams is to measure student, educator, and school accountability. Keystone Exams are designed to be administered to students at or near the end of a Keystone-related course. Students' results are banked until their junior year for accountability purposes. Keystone Exams are one component of Pennsylvania's system of high school graduation requirements affecting students in the class of 2023 and beyond.

These tests were developed by Pennsylvania educators and were aligned to the standards adopted by the Pennsylvania State Board of Education. The results help students, parents, and educators understand how well rigorous expectations for student achievement in core subject areas are being met.

A Report Interpretation Guide is available at [www.education.pa.gov](http://www.education.pa.gov). Type "Keystone Interpretation Guide" in the search field or consult the local school district or school.

[www.pdesas.org](http://www.pdesas.org)

Biology



pennsylvania

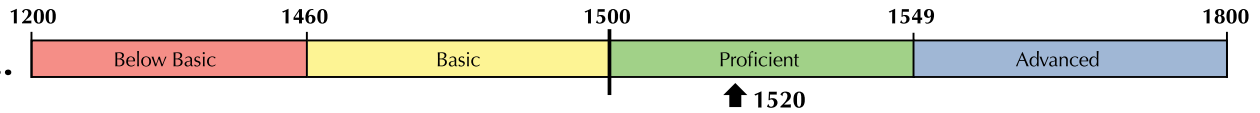
DEPARTMENT OF EDUCATION

155200-000001-13021
Page 1

Figure 16–2. Page 2 of the Individual Student Report

## Performance Level on Total Test

Student’s highest total test scale score to date is indicated by the ( ↑ ). If this student were to test again under similar circumstances, his/her score would likely remain in the following range: 1507-1533.



Inadequate academic performance that indicates little understanding and minimal display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.

Marginal academic performance, work approaching, but not yet reaching, satisfactory performance. Performance indicates a partial understanding and limited display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content. The student may need additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.

Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content.

Superior academic performance indicating an in-depth understanding and exemplary display of the skills included in the Keystone Exams Assessment Anchors & Eligible Content.

### Biology



### SAMPLE STUDENT 1

	Module 1 Cells and Cell Processes			Module 2 Continuity and Unity of Life			Total Test	
	Result	Scale Score	Test Date	Result	Scale Score	Test Date	Scale Score	Performance Level
<b>Highest Total Test Scale Score to Date</b> <sup>1</sup>	Passed	1542		Passed	1501		1520	Proficient
<b>Scores of Three Most Recent Test Events</b>	Passed	1542		Passed	1501			

<sup>1</sup>The highest total test scale score to date is the highest score computed from all possible combinations of module 1 and module 2. The total scale score is not the simple average of the module scale scores. Rather, it is weighted by the relative difficulty of questions from each module.

## **SUMMARY REPORTS**

Summary reports are provided at the school, district, and state levels. These reports contain summary information about the percentage of students in each of the four performance levels. Raw scores are also provided by assessment anchor to allow schools or districts to identify strengths and weaknesses at the content-strand level.

## **REPORT INTERPRETATION GUIDE**

A report interpretation guide is provided to help parents and other Keystone Exams stakeholders better understand test-result information presented in the individual student report. The report interpretation guide can be found on the PDE website at [www.education.pa.gov](http://www.education.pa.gov).



## CHAPTER SEVENTEEN: OPERATIONAL TEST STATISTICS

This chapter presents various summary statistics for the total-test scores based on the final data file described in Chapter Nine. Related information covered elsewhere in this report includes the item-level statistics that were presented in Chapters Eleven (classical item statistics) and Twelve (Rasch item statistics). Please refer to these chapters for additional consideration as item difficulty distributions can affect total score distributions.

### PERFORMANCE LEVEL STATISTICS

Table 17–1 presents performance level percentages by test administration, content area, and student type. Performance level classifications are based on the highest total scaled score to date. As can be seen from the table, the overall percentage in each performance level varied from administration to administration, depending on the ratio of the first-time testers and retesters. In general, retesters had a lower percentage of students classified in the Proficient and Advanced performance levels than first-time testers.

**Table 17–1A. Performance Level Percentages: All Testers**

Administration	Content Area	N	Below Basic (%)	Basic (%)	Proficient (%)	Advanced (%)
Winter	Algebra I	54,824	25.5	55.2	15.2	4.1
Winter	Biology	40,077	33.1	45.5	14.4	7.0
Winter	Literature	35,222	18.7	45.1	31.6	4.6
Spring	Algebra I	162,971	24.3	45.9	17.4	12.4
Spring	Biology	137,797	30.1	30.7	23.2	16.0
Spring	Literature	125,553	14.9	31.6	44.6	8.9
Summer	Algebra I	274	4.7	64.2	25.9	5.1
Summer	Biology	210	11.0	52.4	24.3	12.4
Summer	Literature	111	18.9	55.9	25.2	0.0

**Table 17–1B. Performance Level Percentages: First-time Testers**

Administration	Content Area	N	Below Basic (%)	Basic (%)	Proficient (%)	Advanced (%)
Winter	Algebra I	12,437	32.4	33.7	19.3	14.6
Winter	Biology	14,111	31.4	25.2	24.1	19.4
Winter	Literature	16,588	16.4	28.1	46.1	9.4
Spring	Algebra I	120,395	24.3	39.1	20.1	16.4
Spring	Biology	113,460	28.3	25.8	26.6	19.3
Spring	Literature	108,505	14.1	27.6	48.1	10.2
Summer	Algebra I	36	8.3	30.6	30.6	30.6
Summer	Biology	46	17.4	10.9	17.4	54.3
Summer	Literature	25	36.0	32.0	32.0	0.0

**Table 17–1C. Performance Level Percentages: Retesters**

Administration	Content Area	N	Below Basic (%)	Basic (%)	Proficient (%)	Advanced (%)
Winter	Algebra I	42,387	23.4	61.6	13.9	1.0
Winter	Biology	25,966	34.0	56.6	9.1	0.3
Winter	Literature	18,634	20.8	60.2	18.7	0.3
Spring	Algebra I	42,568	24.3	64.9	9.7	1.1
Spring	Biology	24,336	38.7	53.6	7.3	0.4
Spring	Literature	17,046	19.9	57.2	22.5	0.4
Summer	Algebra I	238	4.2	69.3	25.2	1.3
Summer	Biology	164	9.1	64.0	26.2	0.6
Summer	Literature	86	14.0	62.8	23.3	0.0

## SCALED SCORES

Table 17–2 provides the scaled score means and standard deviations by test administration, content area, and student type (all testers, first-time testers and retesters). The descriptive statistics are based on the highest total scaled score to date following each administration. As can be seen from the table, first-time testers earned higher scaled scores than retesters.

**Table 17–2. Means and Standard Deviations of Scaled Scores**

Administration	Content Area	All Testers Mean	All Testers SD	First-Time Testers Mean	First-Time Testers SD	Retesters Mean	Retesters SD
Winter	Algebra I	1467.4	43.1	1476.0	63.9	1464.8	34.2
Winter	Biology	1478.6	41.0	1494.2	56.8	1470.1	25.1
Winter	Literature	1487.7	52.2	1505.6	61.5	1471.8	35.5
Spring	Algebra I	1478.1	55.0	1483.8	59.8	1462.0	33.0
Spring	Biology	1493.3	54.7	1498.9	57.6	1467.3	25.6
Spring	Literature	1505.2	58.0	1510.0	59.2	1474.7	37.2
Summer	Algebra I	1489.1	35.8	1515.6	66.8	1485.1	26.4
Summer	Biology	1500.2	45.4	1545.2	71.0	1487.5	22.7
Summer	Literature	1478.5	39.7	1463.5	56.4	1482.9	32.5

## RAW SCORES

### SUMMARY STATISTICS

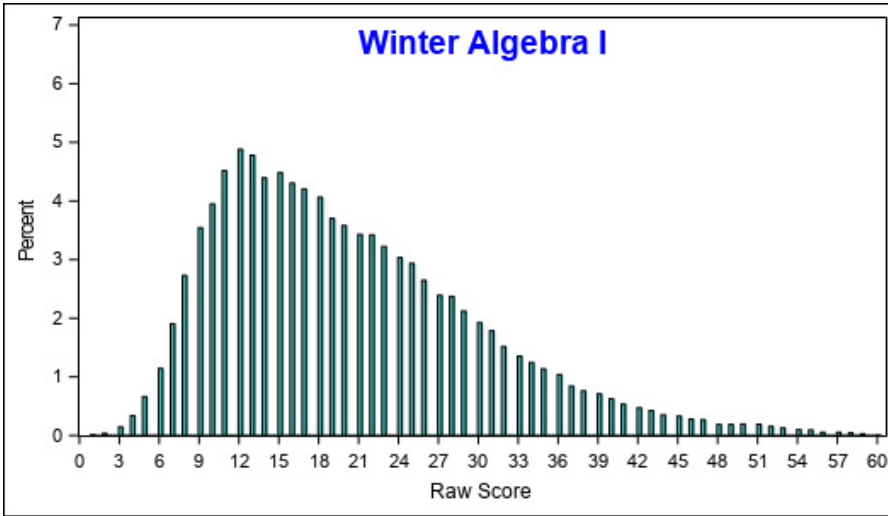
The reader is referred to Appendix M to review summary statistics for the operational raw scores. The statistics reported include number of points possible (Pts.), number of items (Len.), number of students tested (*N*), mean raw score (Mean), standard deviation of raw score (SD), reliability (*r*), and traditional standard error of measurement (SEM).

### SCORE DISTRIBUTIONS

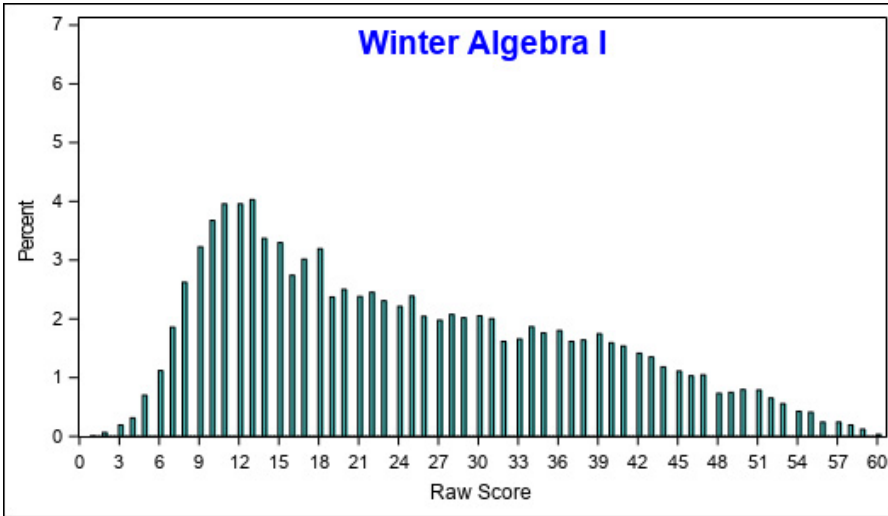
Raw score distributions for each administration and content area are provide in Figure 17–1. Graphs are shown for all test-takers, first-time test-takers, and retesters . As can be seen from the graphs, the retesters tended to score lower than the first-time testers. For all testers, the distribution of raw scores for Algebra I and Biology are positively skewed while the distributions of raw scores for Literature are negatively skewed.

Figure 17–1. Raw Score Distributions

All Testers



First-Time Testers



Retesters

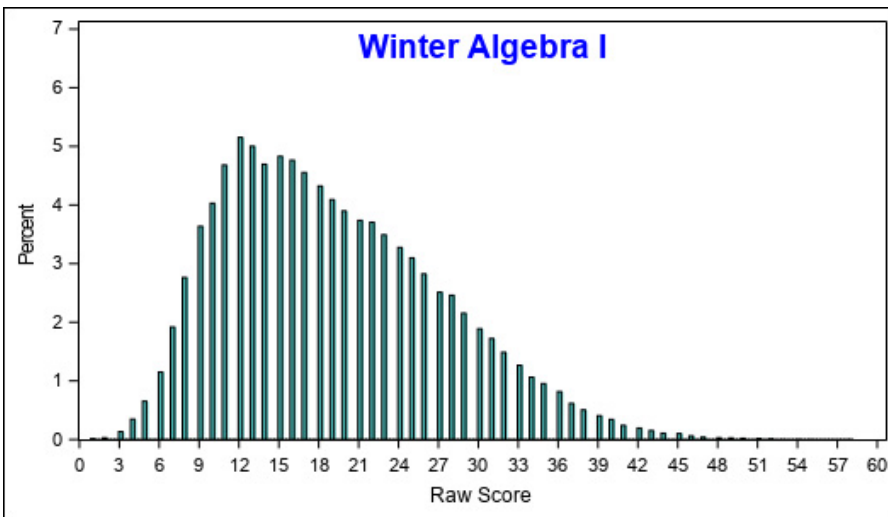
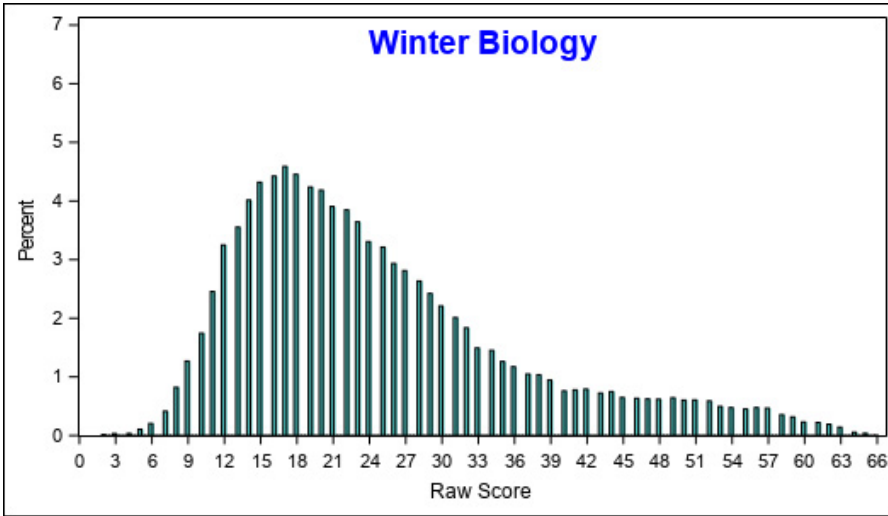
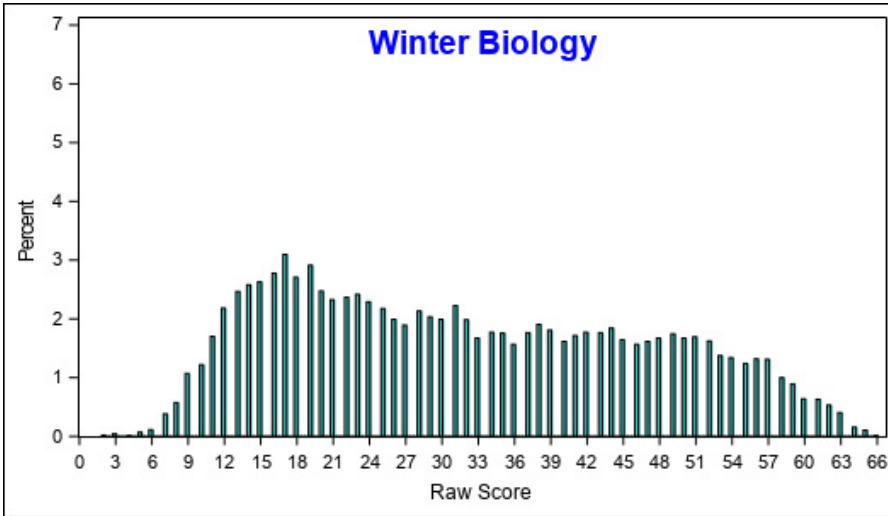


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

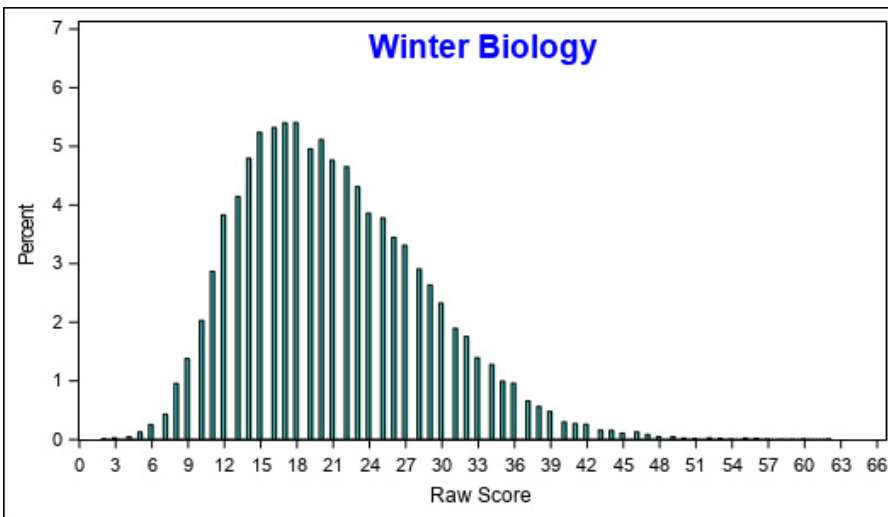
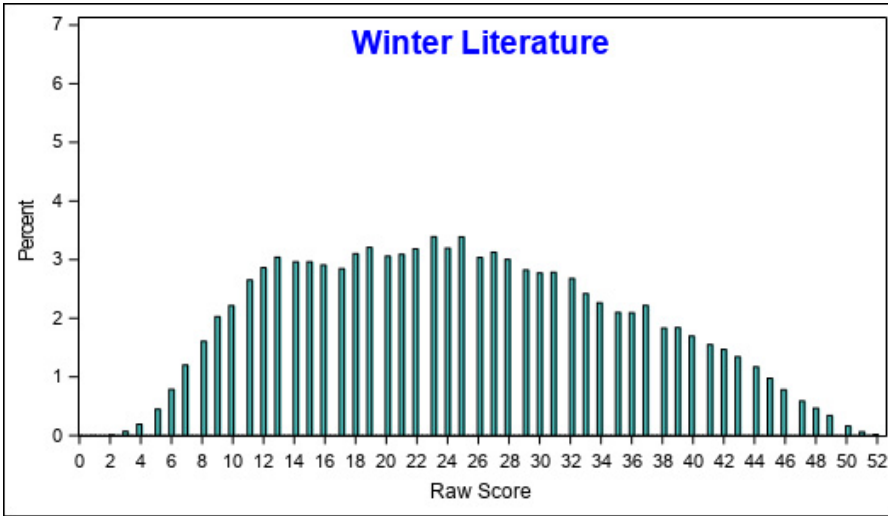
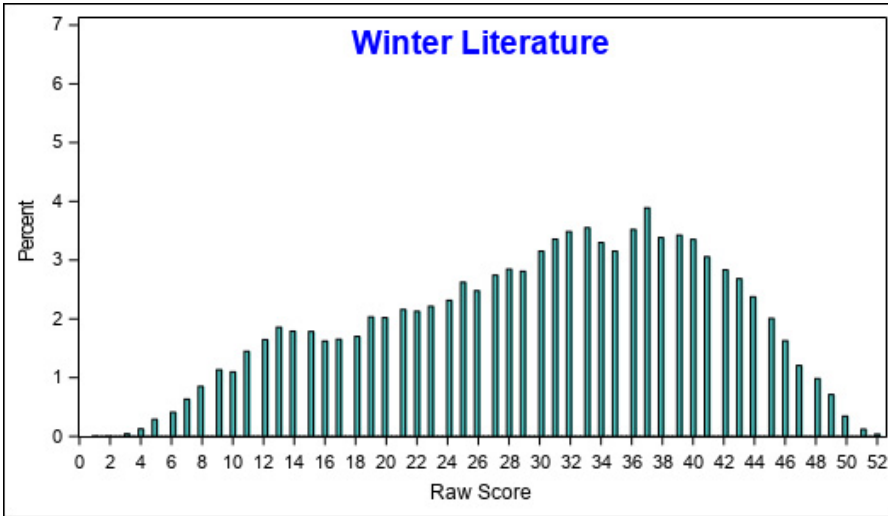


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

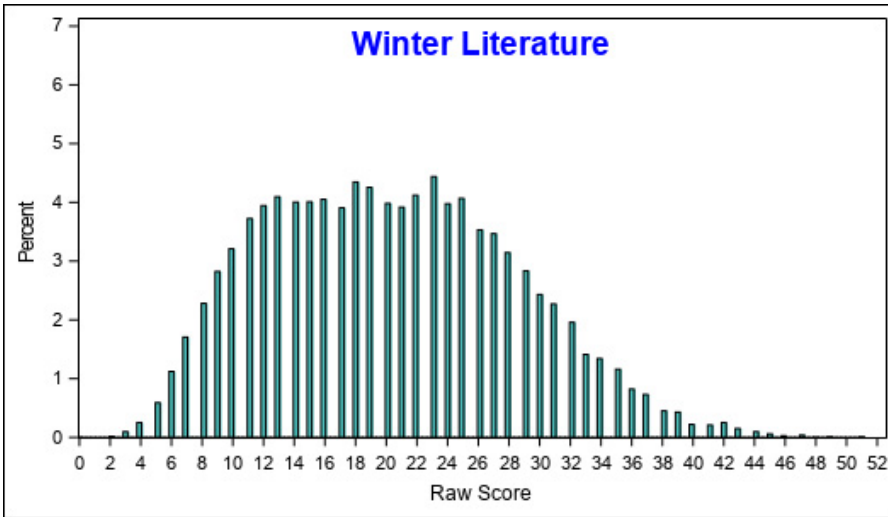
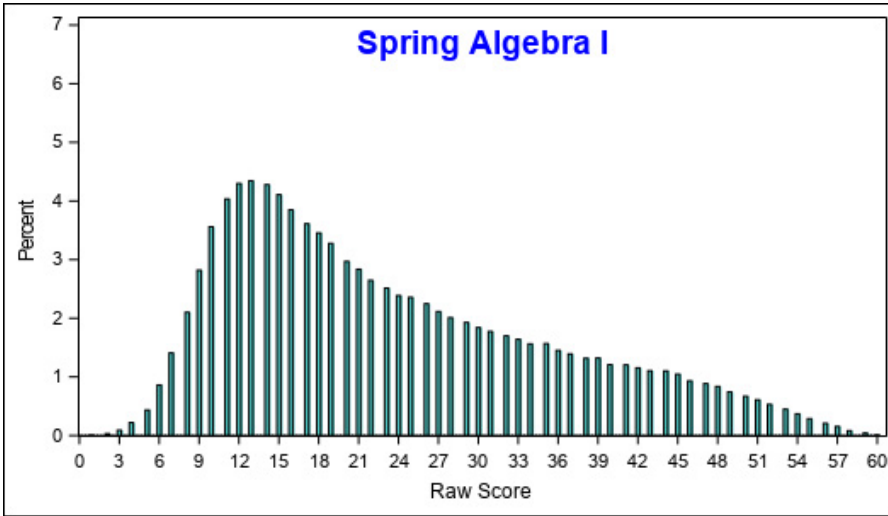
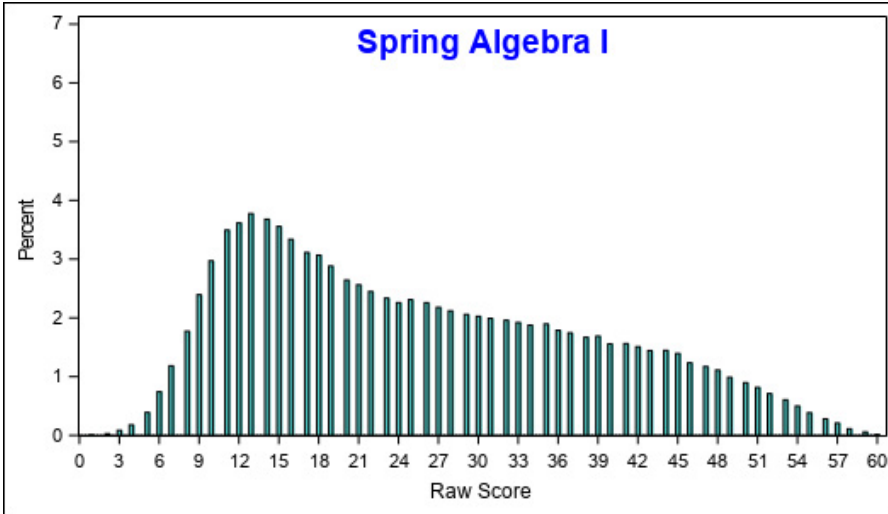


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

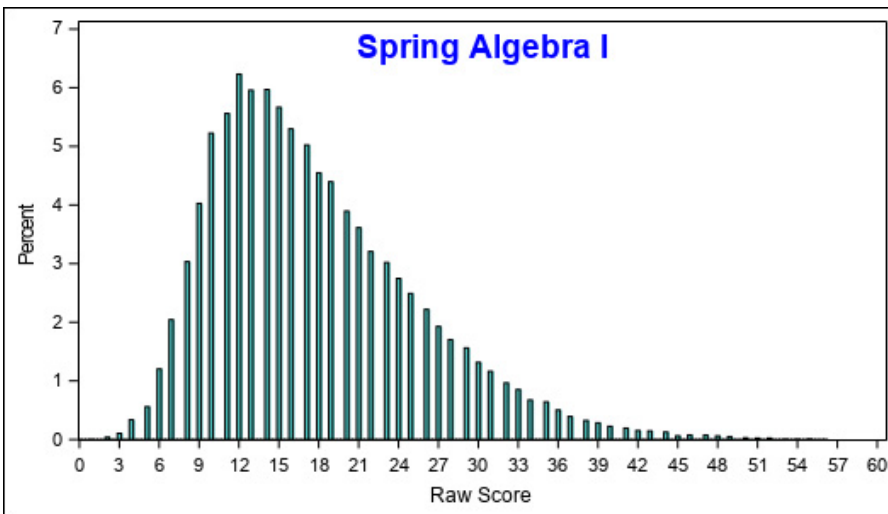
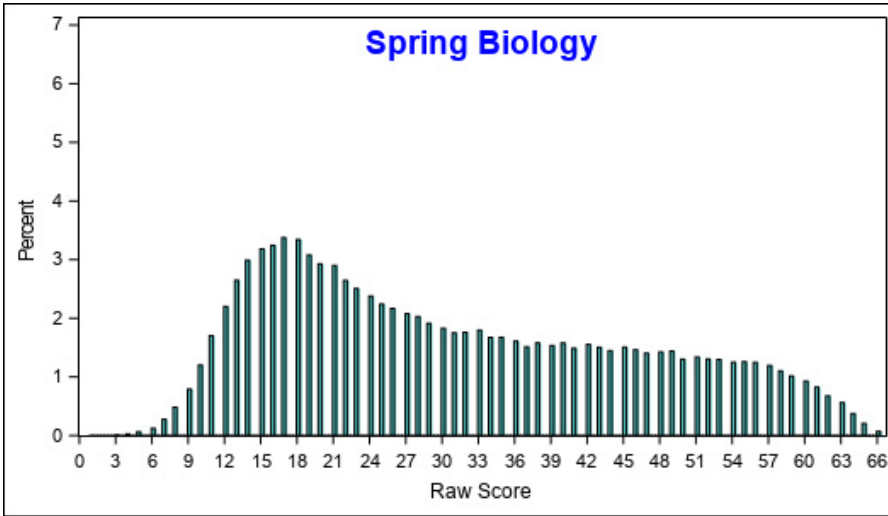
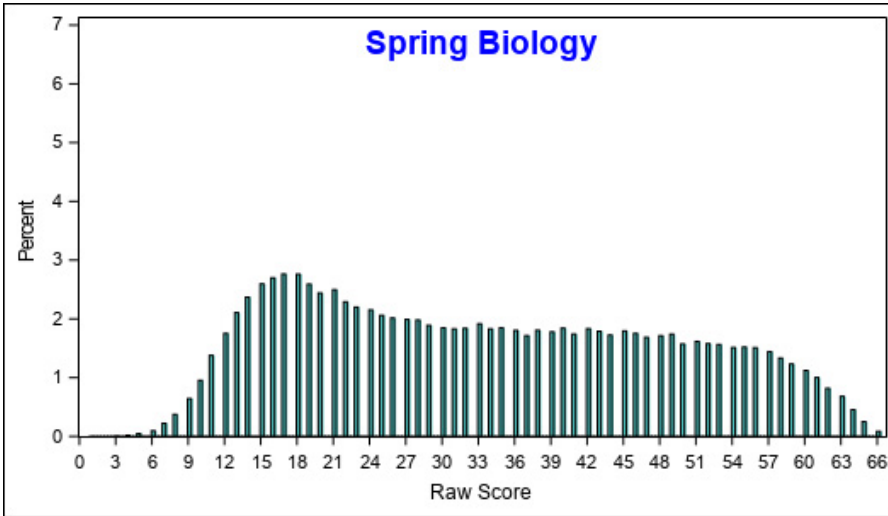


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

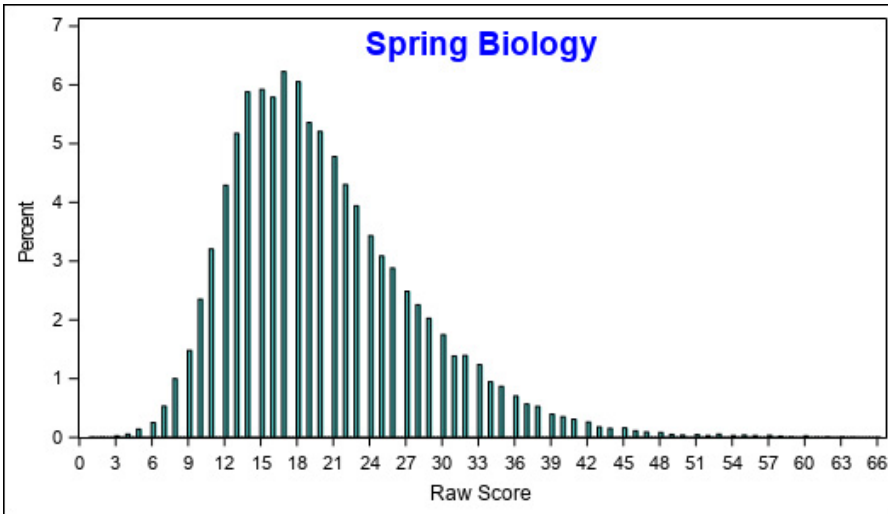
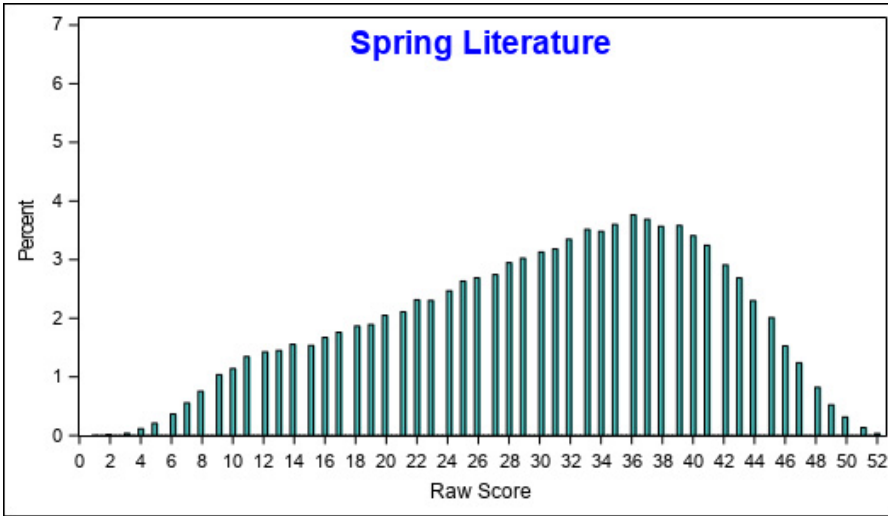
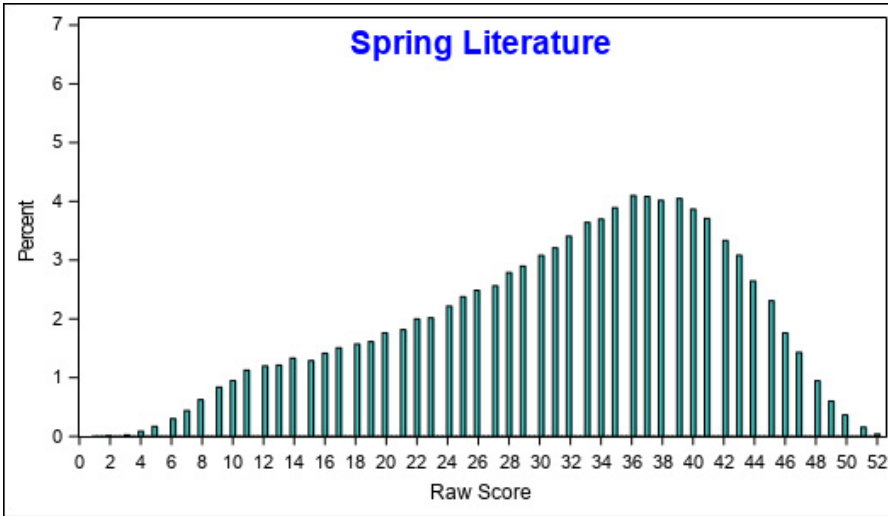


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

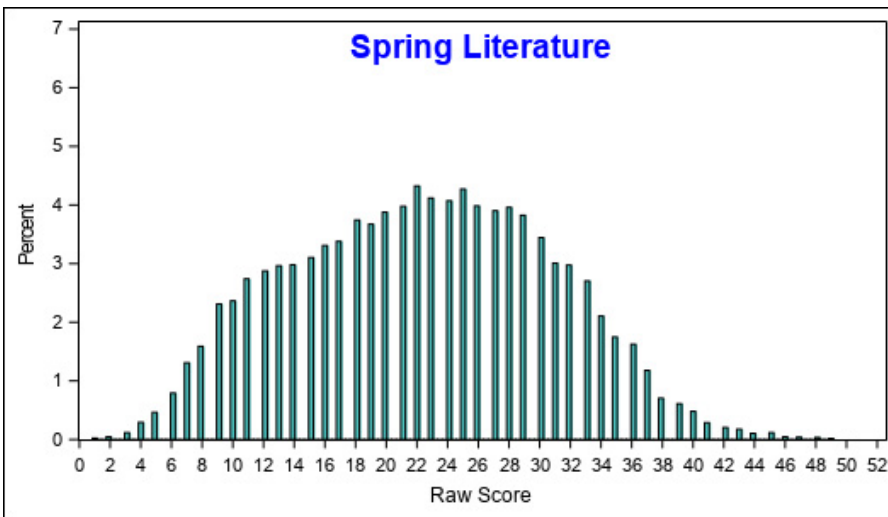
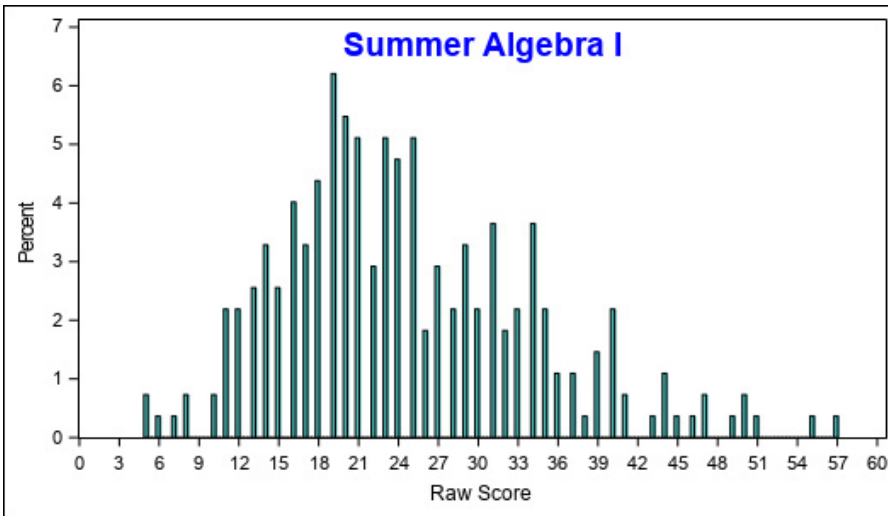


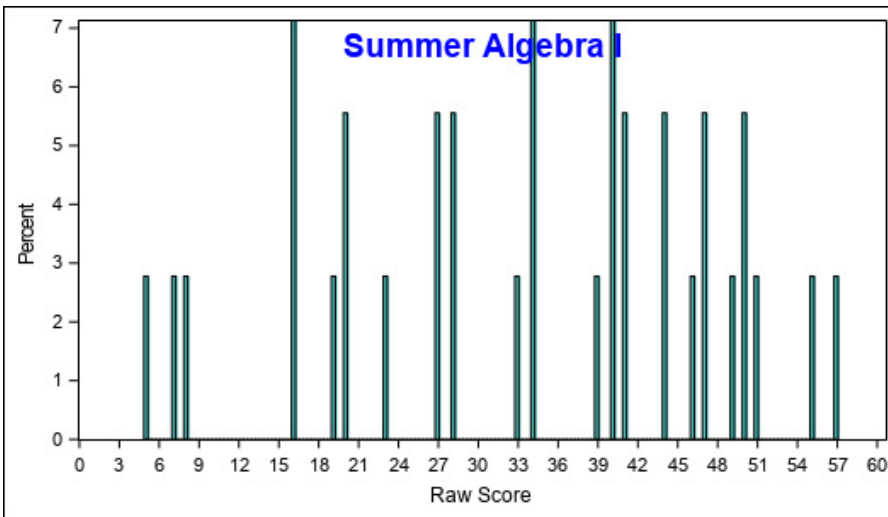


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

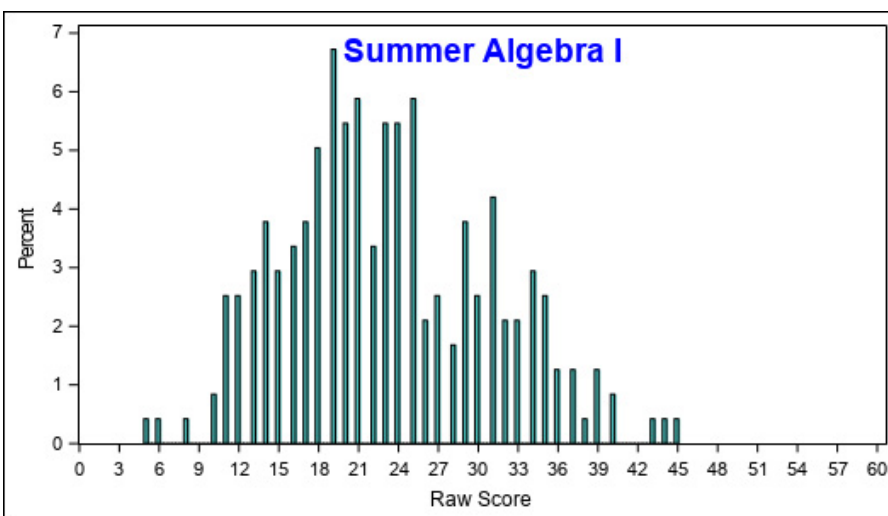
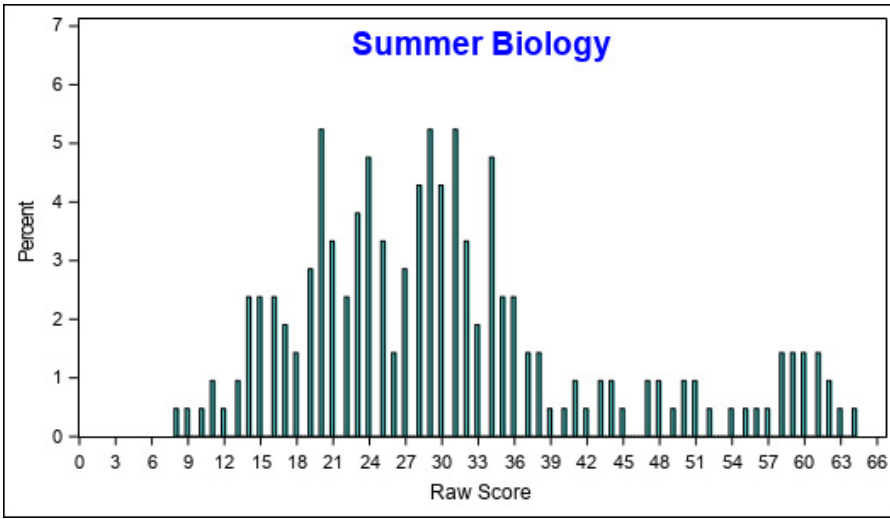
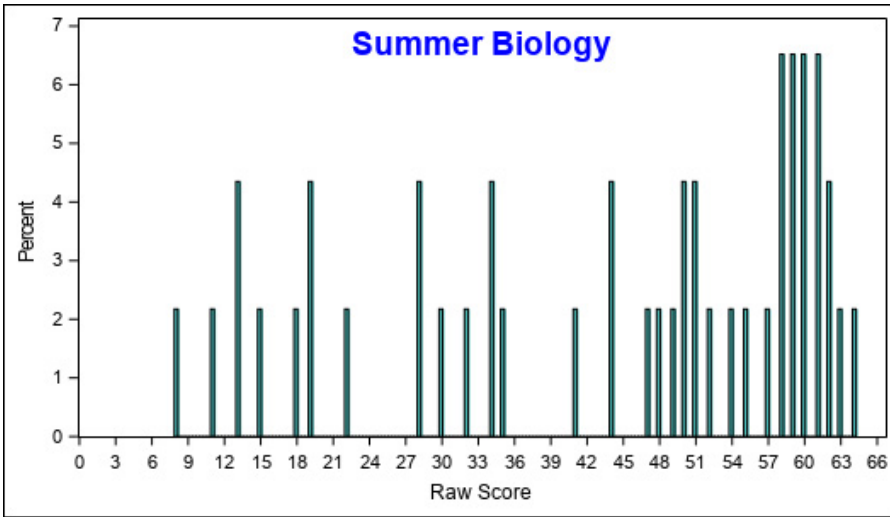


Figure 17–1 (continued). Raw Score Distributions

All Testers



First-Time Testers



Retesters

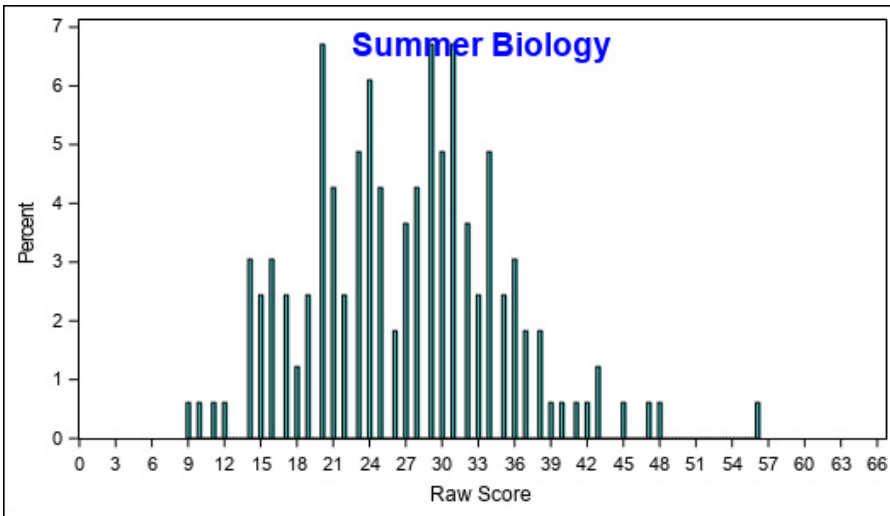
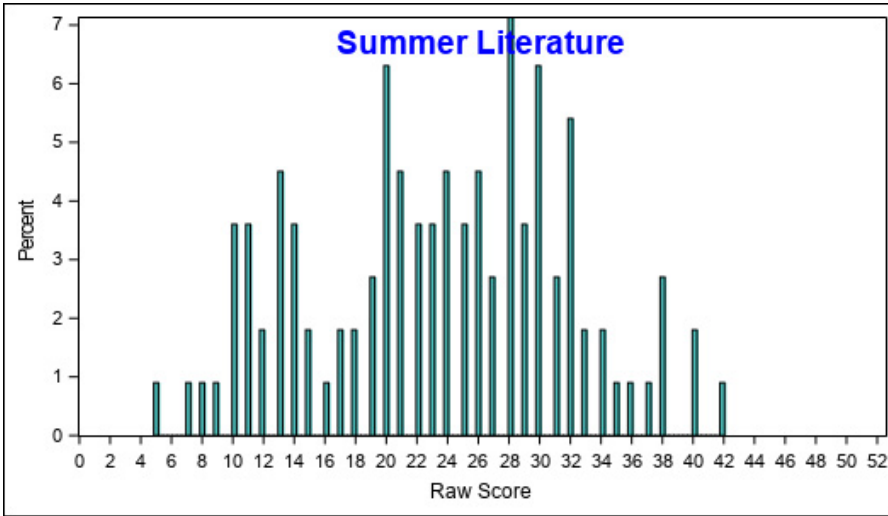
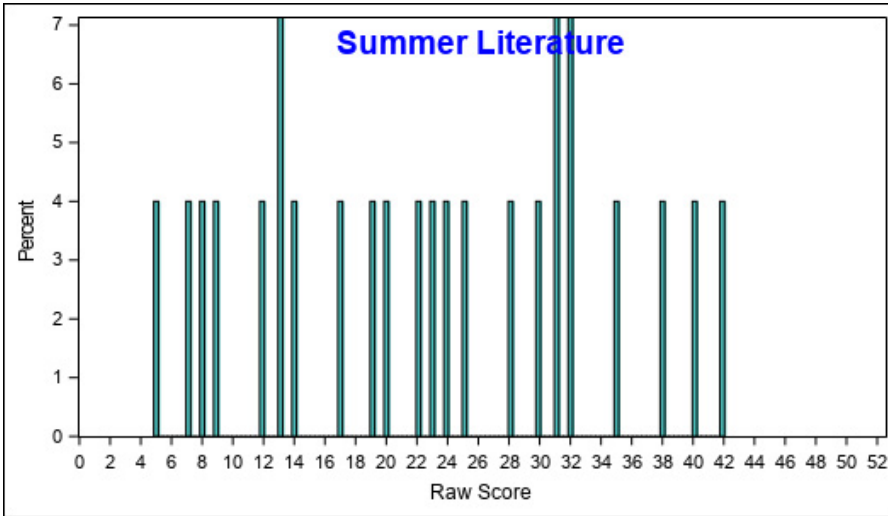


Figure 17–1 (continued). Raw Score Distributions

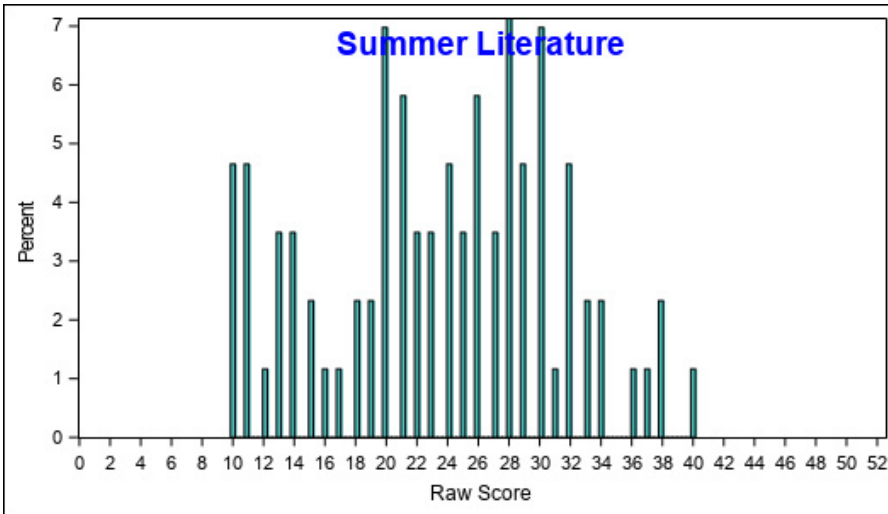
All Testers



First-Time Testers



Retesters



## CHAPTER EIGHTEEN: RELIABILITY

---

This chapter addresses the reliability of Pennsylvania Keystone Exams test scores. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), reliability refers to

the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual test taker; the degree to which scores are free of random errors of measurement for a given group (p. 222).

Frisbie (2005) highlighted several elements of this definition. First, reliability is a property of test scores, not of a test itself. Many may appreciate this distinction, but in casual usage, individuals frequently make reference to a reliable test. While reliability concerns test scores (and not the test specifically), it's important to appreciate the fact that test scores can be affected by characteristics of the instrument. For example, all other things being equal, tests with more items/points tend to be more reliable than tests with fewer items/points. Second, reliability coefficients are group specific. Reliabilities tend to be higher in populations that are more heterogeneous and lower in populations that are more homogeneous. Consequently, both test length and population heterogeneity should be considered when evaluating reliability.

There are other reliability considerations that may be less evident from the definition above yet are still important for test users to understand. While freedom from measurement error is highlighted in the definition, reliability is specifically concerned with random sources of error. Indeed, the degree of inconsistency due to random error sources is what determines reliability: less consistency is associated with lower reliability and more consistency is associated with higher reliability. Of course, systematic error sources also exist. These can artificially increase reliability and decrease validity. (Validity is further discussed in Chapter Nineteen.)

Another noteworthy issue is that multiple sources of error exist (e.g., the day of testing, the items used, the raters who score the items). However, most widely used reliability indices only reflect a single type of error. Consequently, it is important for test users to understand which specific type of error is being considered in a reliability study, and equally, if not more importantly, which types are not.

Understanding the distinction between relative error and absolute error is important because many reliability indices only reflect relative error. Relative error is of interest whenever the relative ordering of individuals with respect to their test performance is of interest. When specific score values are considered important (e.g., if cut scores are used), then absolute error is of interest, too. Generally, there is more error variance when considering the absolute scores of examinees, which, in turn, suggests lower reliability. Understanding examinee rank-order stability is also important; however, such stability might be well achieved even when the specific score values are considerably different.

As the above discussion suggests, reliability is a complex, nonunitary notion that cannot be adequately represented by a single number. There are several reliability indices available, and these may not provide the same results (Frisbie, 2005). The remainder of this chapter covers the following:

- Reliability coefficients and their interpretation
- Unconditional and conditional standard errors of measurement
- Decision consistency
- Rater agreement

## RELIABILITY INDICES

As shown below, the reliability coefficient expresses the consistency of test scores as the ratio of true score variance to total score variance. The total variance contains two components: variance in true scores and variance due to the imperfections in the measurement process. Put differently, total variance equals true score variance plus error variance.<sup>1</sup>

$$\rho_X^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

Reliability coefficients indicate the degree to which differences in test scores reflect true differences in the attribute being tested rather than random fluctuations. Total test score variance (i.e., individual differences) is partly due to real differences in the attribute (true variance) and partly due to random error in the measurement process (error variance).

Reliability coefficients range from 0.0 to 1.0. If all test score variance were true, the index would equal 1.0. The index would be 0.0 if none of the test score variance were true. Such scores would be pure random noise—that is, all measurement error. If the index had a value of 1.0, scores would be perfectly consistent—that is, contain no measurement error. Although values of 1.0 are never achieved in practice, it is clear that larger coefficients are more desirable as they indicate that test scores are less influenced by random error. (How big is big enough and how small is too small are issues considered in a later section.)

As noted in the introduction, there are several different indices that can be used to estimate this ratio. One approach is referred to as internal consistency, which is derived from analyzing the performance consistency of individuals over the items within a test. As discussed below, these internal consistency indices do not take into account other sources of error, such as day-to-day variations (e.g., student health, testing environment) or rater inconsistency.

## COEFFICIENT ALPHA

Although a number of reliability indices exist, perhaps the most frequently reported for achievement tests is coefficient alpha. Consequently, this index is the one reported for the Keystone Exams (see the column with title “r” in Appendix M). Alpha indicates the internal consistency over the responses to a set of items measuring an underlying trait, in this case, academic achievement, in content areas such as Algebra I, Biology, and Literature.

Alpha is an internal consistency index. It can be conceptualized as the extent to which an exchangeable set of items from the same domain would result in a similar rank ordering of students. Note that relative error is reflected in this index. Variation in student performance from one sample of items to the next should be of particular concern for any achievement test user. Consider two hypothetical vocabulary tests intended for the same group of students. Each test contains different sets of unique words that are believed to be randomly equivalent, perhaps like the ones shown below:

**Table 18–1. Two Hypothetical Vocabulary Tests**

Test One	Test Two
Abase	Abate
Boon	Bilk
Capricious	Circuitous
Deface	Debase
....	....
Zealous	Zenith

<sup>1</sup> A covariance term is not required as true scores and error are assumed to be uncorrelated in classical test theory.

If a representative group of students could take both tests, the correlation between the scores obtained would represent the parallel-forms reliability of the test scores. However, such data-collection designs are impractical in large-scale settings, and fatigue and practice effects are likely to affect the results. Internal-consistency reliability indices arose in part to provide reliability measures using the data from just a single test administration. So, if students only took Test One and the coefficient alpha index for those test scores were high, this would suggest that Test Two would provide a very similar rank ordering of the students if they had taken it instead. If coefficient alpha were low, dissimilar rank orderings would likely be observed—again, relative-error variance is reflected in alpha.

## FORMULA

Consider the following data matrix representing the scores of persons (rows) on items (columns):

**Table 18–2. Person × Item Score ( $X_{pi}$ ) Infinite (Population-Universe) Matrix**

Person	Item 1	Item 2	... l	Item ... k
1	$Y_{11}$	$Y_{12}$	... $Y_{1i}$	... $X_{1k}$
2	$Y_{21}$	$Y_{22}$	... $Y_{2i}$	... $X_{2k}$
.....				
P	$Y_{p1}$	$Y_{p2}$	... $Y_{pi}$	... $X_{pk}$
.....				
N	$Y_{N1}$	$Y_{N2}$	... $Y_{Ni}$	... $X_{Nk}$

Note. Adapted from Cronbach and Shavelson (2004).

Then, a general computational formula for alpha is as follows:

$$\alpha = \frac{N}{N - 1} \left( 1 - \frac{\sum_{i=1}^N \sigma_{Y_i}^2}{\sigma_X^2} \right)$$

where  $N$  is the number of parts (items or testlets),  $\sigma_X^2$  is the variance of the observed total test scores, and  $\sigma_{Y_i}^2$  is the variance of part  $i$ .

## FURTHER INTERPRETATIONS

### RULES OF THUMB

Which reliability values are considered high enough? Which values are considered too low? Although frequently asked for, any rules of thumb for interpreting the magnitude of reliability indices are mostly arbitrary. Another approach is to research the reliabilities from similar testing instruments to see what values are commonly observed. For the Keystone Exams, comparisons to tests of similar lengths that were administered to similar student populations from other large-scale assessment programs would be relevant. For many other state assessment programs, reliabilities in the low 0.90s are usually the highest ever observed, and reliabilities in the high 0.80s are very common.

The lower a given reliability coefficient, the greater the potential for over-interpretation of the associated results. As suggested earlier, there is no firm guideline regarding how low is too low. However, as an informative point of reference, a reliability coefficient of 0.50 would mean that there is as much error variance as true-score variance in the scores.

## IS ALPHA A LOWER LIMIT TO RELIABILITY?

According to Brennan (1998), the conventional wisdom that coefficient alpha is a lower limit to reliability is based largely on a misunderstanding. In reflecting on the 50th anniversary of his seminal 1951 article, Cronbach—in Cronbach and Shavelson (2004)—expressed similar misgivings about this conventional wisdom:

one could argue that alpha was almost an unbiased estimate of the desired reliability. . . the *almost* in the preceding sentence refers to a small mathematical detail that causes the alpha coefficient to run a trifle lower than the desired value. This detail is of no consequence and does not support the statement made frequently in textbooks or in articles that alpha is a lower value to the reliability coefficient. That statement is justified by reasoning that starts with the definition of the desired coefficient as the expected consistency among measurements that had a higher degree of parallelism than the random parallel concept implied.

The assumptions for three common parallelism models are presented in Table 18–3. Alpha’s assumptions come from the Essentially-Tau-Equivalent model, which does not require equal means or equal variances across test parts. Based on this, Brennan (1998) asserts that the lower-limit issue, as conceptualized by many, provides an answer to a question that is of minimal importance. Reframed differently, the goal of selecting a reliability coefficient is not to find the one that provides the highest coefficient, but the one that most accurately reflects the test data under study.

It is important to note that there are factors encountered in practice that may legitimately make coefficient alpha an underestimate of reliability. However, there are also factors that might make coefficient alpha an overestimate of reliability. Both possibilities are discussed further below and generally arise when the Essentially-Tau-Equivalent assumptions are strained.

**Table 18–3. Summary of Expectations/Observable Relationships for Different Parallelism Models**

Relationship	Degree of Measurement Parallelism* Classically Parallel	Degree of Measurement Parallelism* Essentially-Tau Equivalent	Degree of Measurement Parallelism* Congeneric
Content Similarity	Yes	Yes	Yes
Equal Means across Parts	Yes	No	No
Equal Variances across Parts	Yes	No	No
Equal Covariances across Parts	Yes	Yes	No
Equal Covariances with other Variables	Yes	Yes	No

\*Note. Other models exist but are not considered here due to their limited application in practice.

## BIASES THAT MIGHT MAKE ALPHA AN UNDERESTIMATE OF RELIABILITY

There are factors that might negatively bias coefficient alpha, making the apparent reliability lower than it may actually be. In practice, two situations frequently encountered that might cause this include tests that are composed of mixed item types (e.g., MC and CR items) and tests that include a planned stratification of the test items according to topics or subdomains.

Although both situations strictly violate the assumptions used in deriving the coefficient alpha (i.e., the tests are not based on equal part lengths in the former case and are not randomly parallel in the latter case), neither necessarily guarantees that the reliability will be markedly lower. In the latter case, reliability will be underestimated only when strand items are homogeneous enough for the average covariance within strata to exceed the average covariance between strata. Although both are potential influences for the Keystone Exams, the total test score reliabilities are reported in Appendix M for each administration, content area, and subgroup (ethnicity, gender, EL, IEP, and economically disadvantaged).

## BIASES THAT MIGHT MAKE ALPHA AN OVERESTIMATE OF RELIABILITY

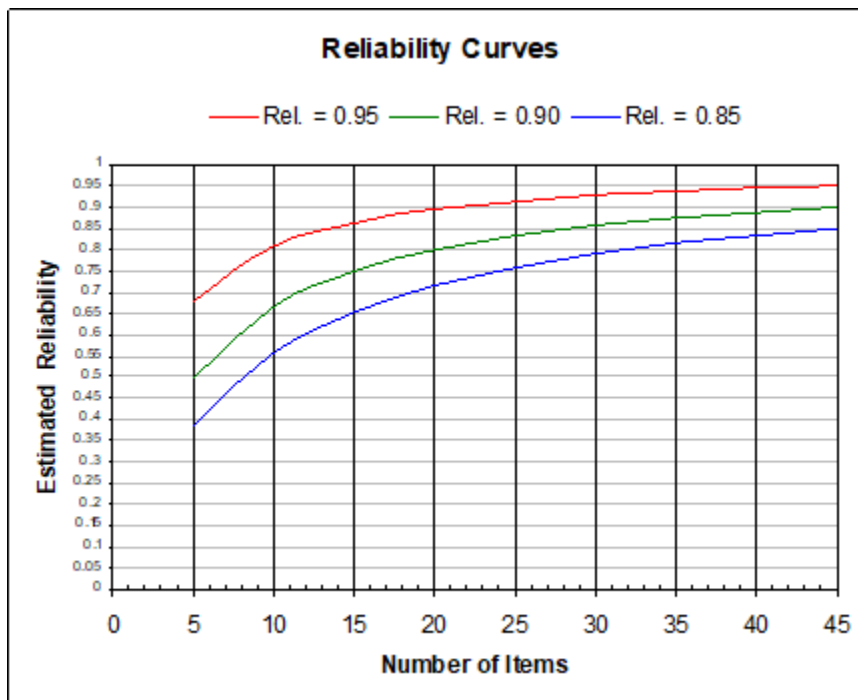
As emphasized in earlier sections, coefficient alpha only accounts for measurement error that arises from the selection of items used on a particular test form. There are other sources of random inaccuracy. One is due to the occasion of testing. Examples of other various random conditions that might affect students on any particular testing occasions include illness, fatigue, and anxiety. Also, when a test includes CR items, another source of random fluctuation can be the CR item scorers. In a sense, alpha may be positively biased because it does not account for other important sources of random error. Actually, any internal consistency reliability index might understate the overall problem of measurement error because they all ignore such sources of random error.

Another positive bias can occur when items are associated (clustered) with a common stimulus. Item bundles and testlets are other frequently used terms for this situation. One concrete example is when multiple reading comprehension items are associated with a common passage selection. Again, such a situation does not guarantee that the reliability estimate will be markedly affected, but the potential exists.

## MODULE SCORE RELIABILITY

As noted in the introduction, reliabilities tend to go up in value with an increase in test length and go down in value with a decrease in test length. Figure 18–1 illustrates this relationship for a hypothetical 45-point test with three total score reliabilities: 0.95, 0.90, and 0.85. As an example, the curve for reliability equal to 0.90 suggests that a 15-item module would be expected to have a score reliability of 0.75. The use of the Spearman-Brown prophecy formula assumes all items are exchangeable, which in practice they may not be. While such a chart may not perfectly model actual module correlations, the intent is only to illustrate the substantial impact that limited numbers of items impact that limited numbers of items in a module can have on module-score reliability. One should not be surprised that module scores with more points tend to show higher reliability coefficients and those with fewer points tend to show lower reliability coefficients.

**Figure 18–1. Example of the Relationship Between Test Length and Reliability**



As can be seen in Appendix M, the reliability coefficients at the module level ranged from 0.72 to 0.88 and were always less than the total score reliabilities, likely because the number of items at the module level is half of the number of items in the total test.



## STANDARD ERROR OF MEASUREMENT

The reliability coefficient is a unit-free indicator that reflects the degree to which scores are free of measurement error. It always ranges between 0.0 and 1.0 regardless of the test's scale. Reliability coefficients best reflect the extent to which measurement inconsistencies may be present or absent in a group. However, they are not that useful for helping users interpret test scores. The standard error of measurement (SEM) is another indicator of test score precision that is better suited for determining the effect of measurement inconsistencies on the scores obtained by individual examinees. This is particularly so for conditional SEMs (CSEM) discussed further below.

### TRADITIONAL STANDARD ERROR OF MEASUREMENT

A precise, theoretical interpretation of the SEM (see Appendix M) is somewhat unwieldy. A beginning point for understanding the concept is as follows. If everyone being tested had the same true score,<sup>2</sup> there would still be some variation in observed scores due to imperfections in the measurement process, such as random differences in attention during instruction or concentration during testing or the sampling of test items. The standard error is defined as the standard deviation<sup>3</sup> of the distribution of observed scores for students with identical true scores. Because the SEM is an index of the random variability in test scores in actual score units, it represents very important information for test score users.

The SEM formula is provided below.

$$SEM = SD\sqrt{1-reliability}$$

It indicates that the value of the SEM depends on both the reliability coefficient and the standard deviation of test scores. If the reliability were equal to 0.00 (the lowest possible value), the SEM would be equal to the standard deviation of the test scores. If test reliability were equal to 1.00 (the highest possible value), the SEM would be 0.0. In other words, a perfectly reliable test has no measurement error (Harvill, 1991). Additionally, the value of the SEM takes the group variation (i.e., score standard deviation) into account. Consider that an SEM of 3.0 on a 10-point test would be very different from an SEM of 3.0 on a 100-point test.

### TRADITIONAL SEM CONFIDENCE INTERVALS

The SEM is an index of the random variability in test scores in actual score units, which is why it has such great utility for test score users. SEMs allow statements regarding the precision of individual tests scores. SEMs help place reasonable limits (Gulliksen, 1950) around observed scores through construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores,  $X$ , and adding and subtracting a multiplicative factor of the SEM. As an example, students with a given true score will have observed scores that fall between  $\pm 1$  SEM about two-thirds of the time.<sup>4</sup> For  $\pm 2$  SEM confidence intervals, the percentage increases to about 95 percent.

### FURTHER INTERPRETATIONS

#### ONE SEM FOR ALL TEST SCORES

The SEM approach described above only provides a single numerical estimate for constructing the confidence intervals for examinees regardless of their score levels. However, such confidence intervals vary according to one's score. Consequently, care should be taken when using the SEM for students with extreme scores. An alternate approach that conditions the SEM on a student's score estimate is described in the next sections.

#### GROUP SPECIFIC

As noted in the introduction, reliabilities are group specific. The same is true for SEMs because both score reliabilities and score standard deviations vary across groups.

---

<sup>2</sup> True score is the score the person would receive if the measurement process were perfect.

<sup>3</sup> The standard deviation of a distribution is a measure of the dispersion of the observations. For the normal distribution, about 16 percent of the observations are more than one standard deviation above the mean.

<sup>4</sup> Some prefer the following interpretation: if a student were tested an infinite number of times, the  $\pm 1$  SEM confidence intervals constructed for each score would capture the student's true score 68 percent of the time.

## RAW SCORE METRIC

The SEM approach is calculated using raw scores, and as such, the resulting confidence interval bands are on the raw score metric. Error bands on the scaled score metric are considered in the next section.

## TYPE OF ERROR REFLECTED

The interpretation of the SEM should be driven by the type of score reliability that underpins it. So, the Keystone Exams SEMs involve the same source of error relevant to internal consistency indices. As noted earlier, a precise technical explanation of the SEM (and resulting confidence intervals) can be unwieldy. Because of this, score users are often provided less complex interpretations.

One simpler description sometimes used is that a confidence interval represents the possible score range that one would observe if a student could be tested twice with the same instrument. Taking the same test on a different day implies the only source of random error being considered is related to the occasion of testing—such as a student might be sleepier one day than another, might be sick, or might not have eaten a good breakfast. There is a reliability index that captures this source of random error, and it is referred to as the test-retest reliability coefficient. This is not the type of reliability computed for the Keystone Exams. When internal consistency reliability estimates are used, such an explanation blurs the fact that random error based on the occasion of testing is not considered.

When SEMs are derived from internal consistency reliability estimates, a better approach is to describe the confidence interval as providing reasonable bounds for the range of scores that a student might receive if he or she took an equivalent version of the test. (That is, the student took a test that covered exactly the same content but included a different set of items.) As an example, if the Algebra I score was 1450 and the SEM band was 1435 to 1465, then a student would be likely to receive a score somewhere between 1435 and 1465 if he or she took a different version of the test.

## **RESULTS AND OBSERVATIONS**

Coefficient alpha results and associated (traditional) SEMs for various Keystone Exam scores are documented in Appendix M. Values were derived using the final data file (see Chapter Nine). The results are organized by administration and then content area. Each table also breaks out the modules and groups of interest such as the total student population (overall), gender, ethnicity, English learners (EL), students with an individualized education plan (IEP), and economically disadvantaged (ED) students. The statistics reported include the number of points possible (Pts.), number of items (Len.), number of students tested (N), mean raw score (Mean), standard deviation of raw score (SD), reliability ( $r$ ), and traditional standard error of measurement (SEM).

Note that these tables report the standard deviations of observed scores. Assuming normally distributed scores, one would expect about two-thirds of the observations to be within one standard deviation of the mean. An estimate of the standard deviation of the true scores can be computed as

$$\hat{\sigma}_r = \sqrt{\hat{\sigma}_x^2 - \hat{\sigma}_x^2(1 - \hat{\rho}_x)}$$

For all subjects, the overall test score reliability values are higher than the reliabilities at the module level, likely due to the fact that each module contains half as many items as the entire test. The test score reliabilities for the spring administration were the highest, ranging from 0.91 to 0.94. The reliabilities ranged from 0.88 to 0.91 and 0.86 to 0.91 for the winter and summer administrations, respectively. The reliability coefficients indicate highly consistent test scores for these instruments.

It was also noted that reliabilities tend to go up in value with an increase in population heterogeneity and go down in value with a decrease in more homogeneous populations. Once again, there is no firm guideline regarding how low is too low. The lower a given reliability coefficient, the greater the potential for over-interpretation. As a point of reference, a reliability coefficient of 0.50 would suggest that there is as much error variance as true-score variance in the scores. It should be noted that the reliability of group mean scores (e.g., school or district means) tends to be higher than that of individual scores, suggesting interpretation of module scores at these aggregate levels is reasonable.

## RASCH CONDITIONAL STANDARD ERRORS OF MEASUREMENT

The CSEM also indicates the degree of measurement error but does so in scaled-score units and varies as a function of a student's actual scaled score. Therefore, the CSEM may be especially useful in characterizing measurement precision in the neighborhood of a score level used for decision-making—such as cut scores for identifying students who meet a performance standard.

Technically, when a Rasch model is applied, the CSEM at any given point on the ability continuum is defined as the reciprocal of the square root of the test information function derived from the Rasch scaling model:

$$CSEM(\hat{\beta}_n) = \frac{1}{\sqrt{I(\hat{\beta}_n)}}$$

where  $CSEM(\hat{\beta}_n)$  is conditional standard error of measurement and  $I(\hat{\beta}_n)$  is test information function. Test information depends on the sum of the corresponding information functions for the test items. Item information depends on each item's difficulty and conditional item score variance. The formula above utilizes the Rasch ability  $\beta_n$  metric. The conditional standard error on the scaled-score (SS) metric is determined simply by multiplying the  $CSEM(\hat{\beta}_n)$  by the slope (multiplicative constant,  $m$ ) of the linear transformation equation used to convert the Rasch ability estimates to scaled scores.

$$CSEM(SS) = CSEM(\hat{\beta}_n) * m$$

Chapter Fourteen provides the linear transformation formulas for each of the Keystone Exams.

## RASCH CSEM CONFIDENCE INTERVALS

CSEMs also allow statements regarding the precision of individual tests scores. And like SEMs, they help place reasonable limits around observed scaled scores through construction of an approximate score band. The confidence intervals are constructed by adding and subtracting a multiplicative factor of the CSEM and may be interpreted as described in the earlier section.

## FURTHER INTERPRETATIONS

### DIFFERENT CSEMS FOR DIFFERENT TEST SCORES

The CSEM approach provides different numerical estimates for constructing the confidence intervals for examinees depending on their specific score levels. The magnitude of the CSEM values is U-shaped, with larger CSEM values associated with lower and higher scores.

### GROUP SPECIFIC

Assuming reasonable model-data fit—as explored in Chapter Twelve—the Rasch-based CSEMs (conditioned on score level) should not vary across groups.

### SCALED-SCORE METRIC

The CSEM and associated confidence interval bands are on the scaled-score metric.

### TYPE OF ERROR REFLECTED

The CSEMs documented on the Keystone Exams score reports are the Rasch-based conditional standard errors of measurement described above. These are provided by the program WINSTEPS described in Chapter Twelve. As noted earlier, these CSEMs are based on the concept of statistical information. For the purpose of providing a simpler explanation of CSEMs to test score users, the earlier description of SEMs framed using the idea of internal consistency reliability was provided in the Keystone Exams score report interpretive guide.<sup>5</sup> Score report content is considered in greater detail in Chapter Sixteen.

---

<sup>5</sup> Because Rasch CSEMs are based on statistical information, it is questionable whether they account for error variance due to items. However, it seems difficult to construct a simple explanation of Rasch CSEMs for the general public.

## RESULTS AND OBSERVATIONS

Figure 18–2 shows the Rasch CSEMs associated with each scaled-score level. (This information is also provided in tabular form in Appendix K.) Values were derived using the calibration data file described in Chapter Nine. The scaled scores and associated CSEMs represent the administration-specific scores. The values are fairly consistent across a noticeably large range of the scaled scores, as demonstrated by the relatively flat bottoms of most plots. The values increase at both extremes (i.e., at the lowest and highest observed scaled scores), giving these figures their typical U-shaped pattern. The three red-dashed vertical lines represent the Basic, Proficient, and Advanced scaled cut-scores, respectively, moving from lower to higher scaled score values. CSEM values at the cut score lines are associated with smaller values, indicating more precise measurement occurs at these cuts.

**Figure 18–2. Conditional Standard Error Plots for Each Administration and Content Area**

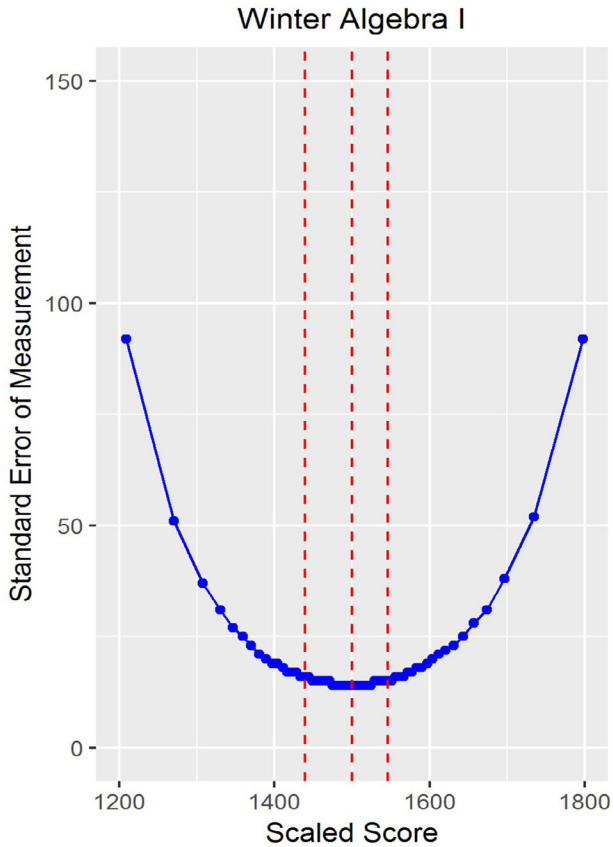


Figure 18–2 (continued). Conditional Standard Error Plots for Each Administration and Content Area

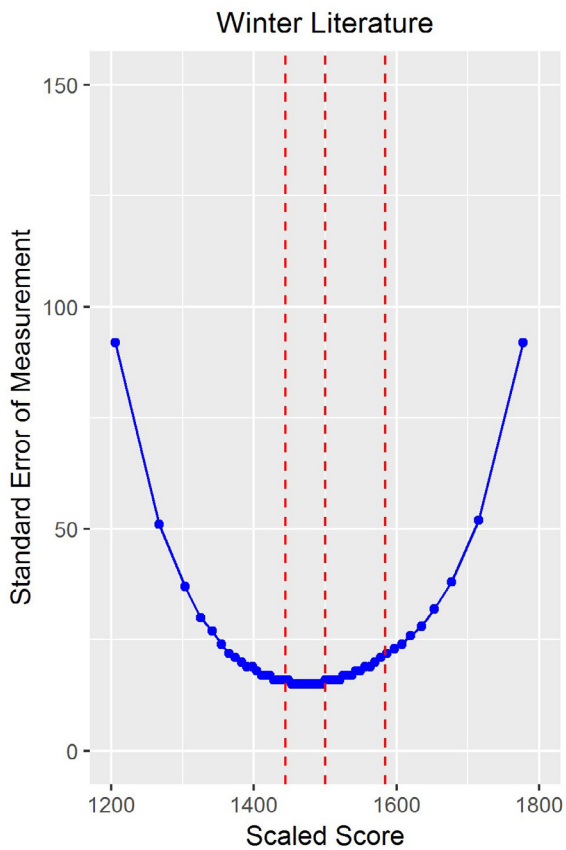
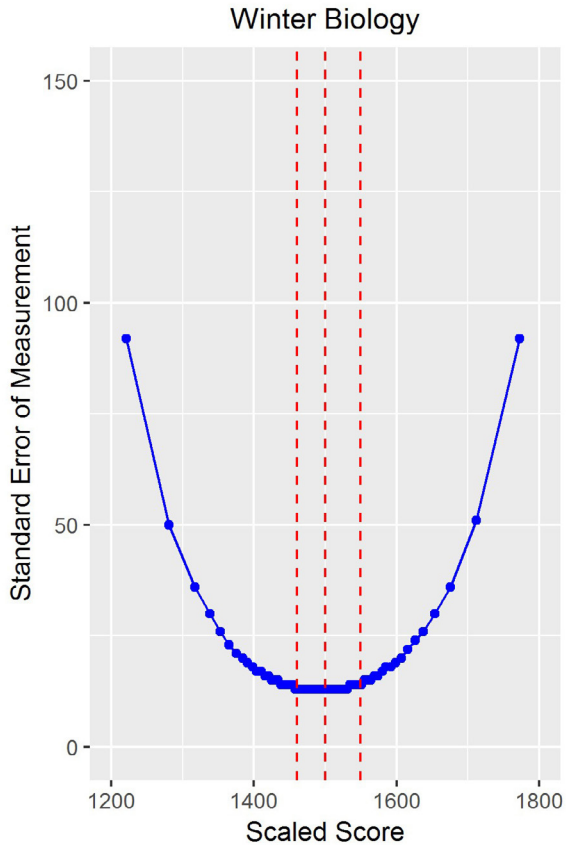


Figure 18–2 (continued). Conditional Standard Error Plots for Each Administration and Content Area

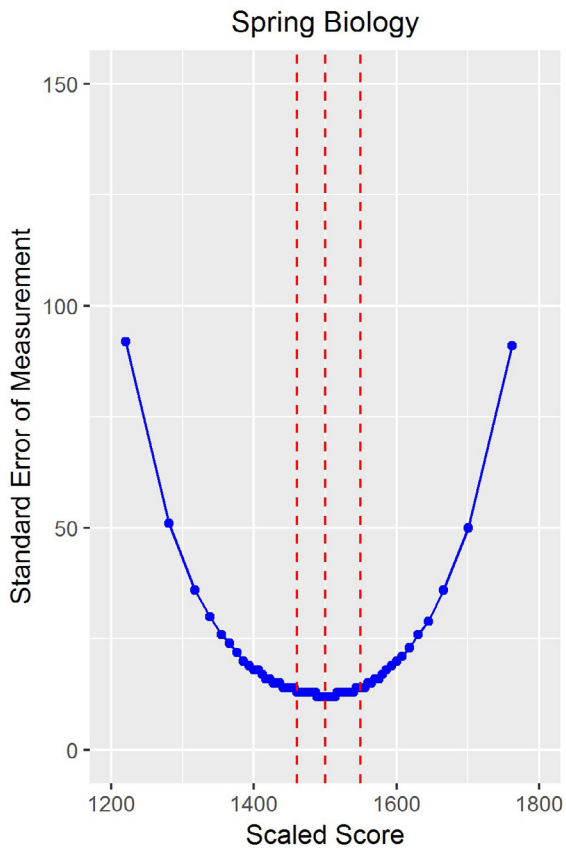
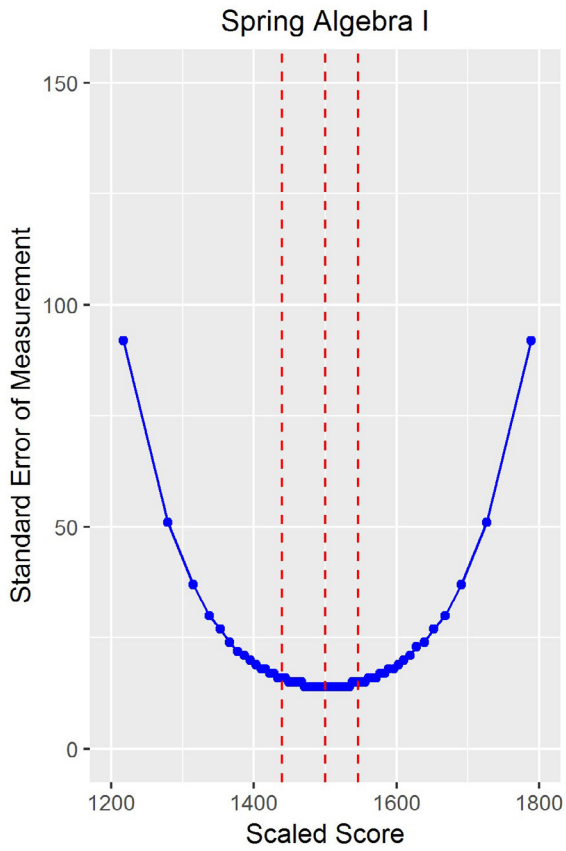


Figure 18–2 (continued). Conditional Standard Error Plots for Each Administration and Content Area

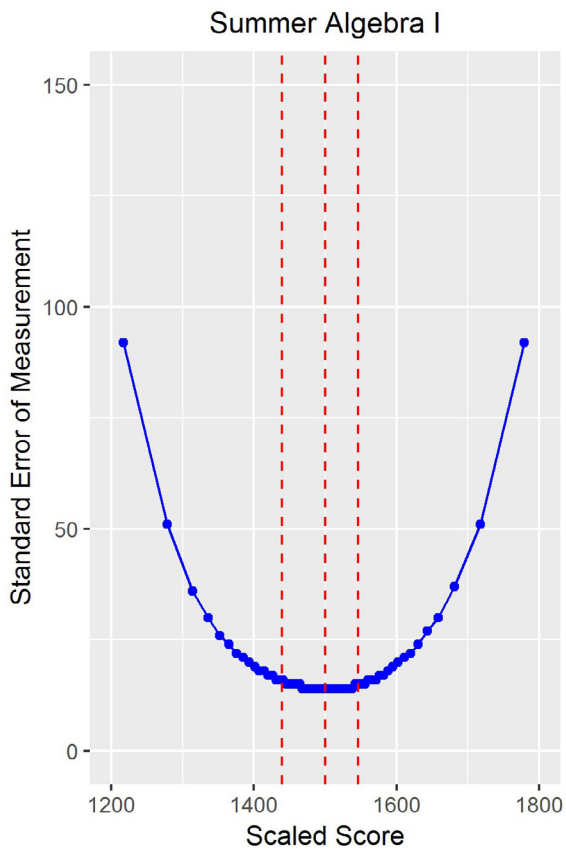
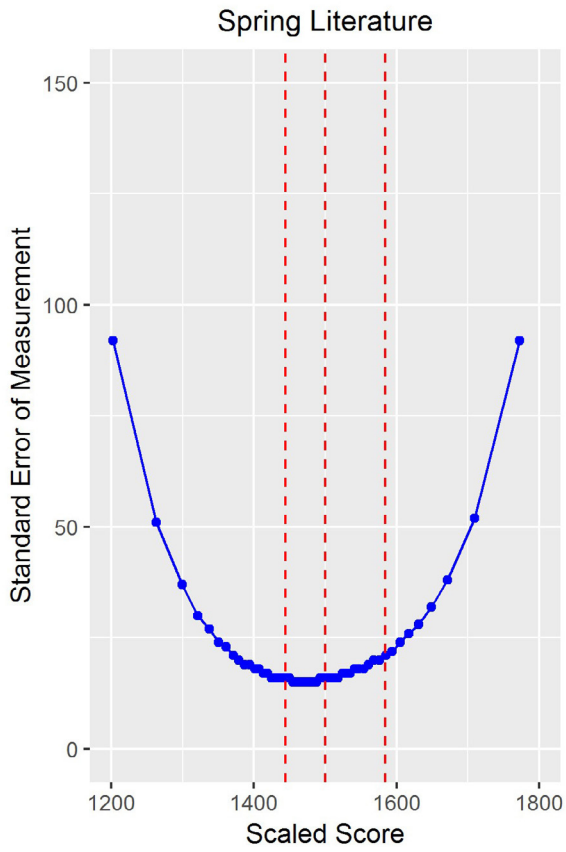
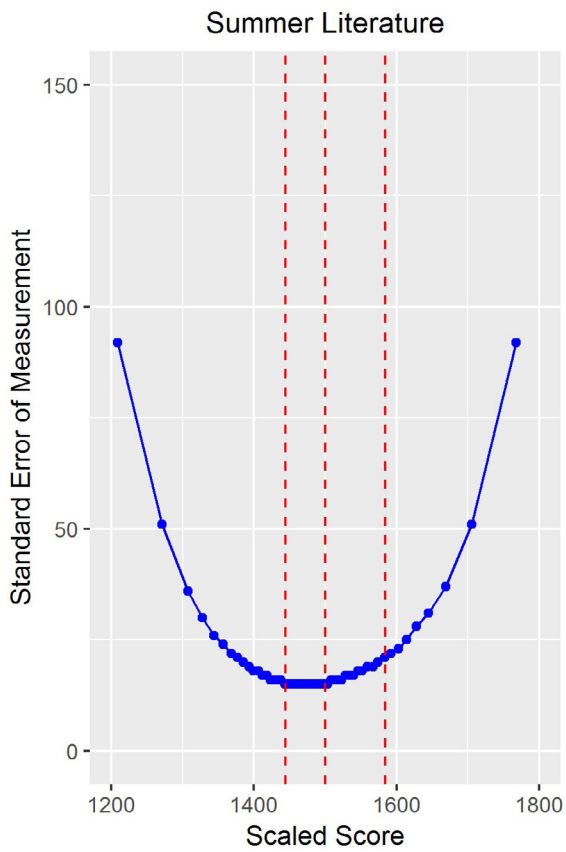
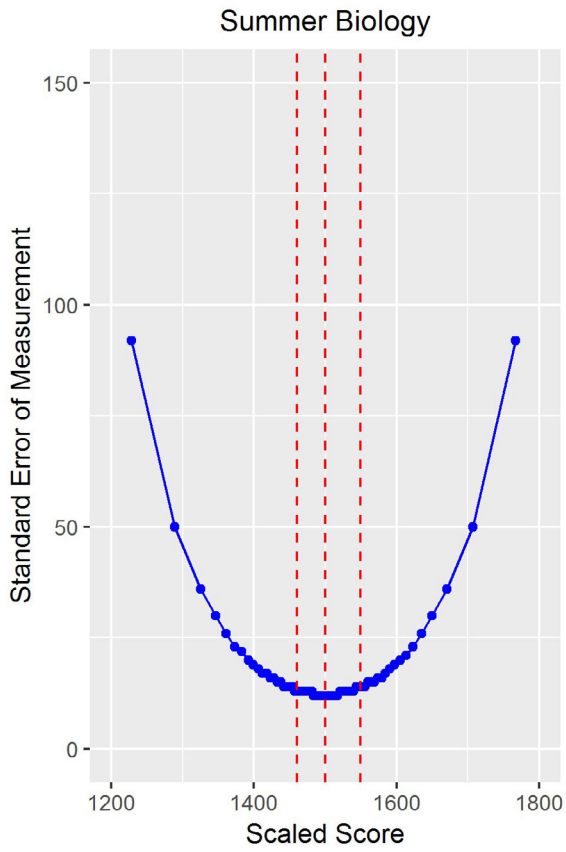


Figure 18–2 (continued). Conditional Standard Error Plots for Each Administration and Content Area





## RELIABILITY OF PERFORMANCE LEVEL CLASSIFICATION DECISIONS

Student performance on the Keystone Exams is classified into one of four achievement levels using the cut scores described in Chapter Thirteen. The reliability of the classification decisions can be assessed by two statistics: decision accuracy and decision consistency.

### DECISION ACCURACY

Decision accuracy describes the extent to which performance level classification decisions based on the administered test form would agree with the decisions that would be made on the basis of a perfectly reliable test (i.e., if it was possible to know each examinee's true score). Decision accuracy answers the question: How does the actual classification of test takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores were somehow known?

### DECISION CONSISTENCY

Decision consistency describes the extent to which classification decisions based on the administered test form would agree with the decisions made if a parallel alternate form had been administered. Decision consistency answers the question: What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test?

Since the true scores are unknown and it is not feasible to repeat the Keystone Exams to estimate the proportion of students who would be reclassified in the same performance levels, a statistical model needs to be imposed on the data in order to project the consistency of classifications solely using data from the available administration (Hambleton and Novick, 1973). Although a number of procedures are available, two well-known methods were developed by Hanson and Brennan (1990) and Livingston and Lewis (1995), utilizing specific true score models. These approaches are fairly complex, and the cited sources contain details regarding the statistical models used to calculate the decision accuracy and consistency from a single administration.

For Keystone Exams, given the two approaches provide similar results, true scores and single-form scores on forms parallel to the one actually given are estimated following the Livingston and Lewis (1995) method. The decision accuracy is estimated using an estimated joint distribution of reported performance-level classifications on the current form of the exam and the performance-level classifications based on the true score. Decision consistency is estimated using an estimated joint distribution of reported performance-level classifications on the current form of the exam and performance-level classifications on the parallel alternate form. In each case, the proportion of performance-level classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the joint distribution. Reliability of classification at each performance-level cut score is estimated by collapsing the joint distribution at the passing score boundary into a 2-by-2 table and summing the two entries.

Several factors might affect the classification decision accuracy and consistency. One important factor is the reliability of the scores. All other things being equal, more reliable test scores tend to result in more similar reclassifications. Another factor is the location of the cut score in the score distribution. More consistent classifications are observed when the cut scores are located away from the mass of the score distribution. For example, when scores are close to being normally distributed, the mass is concentrated in the middle of the distribution, and thus, classifications tend to become more consistent when cut scores go up from 70 percent to 80 percent, or, alternatively, go down from 30 percent to 20 percent. The number of performance levels is also a consideration. Consistency indices for four performance levels should be lower than for those based on two categories. This is not surprising since classification using four levels would allow more opportunity to change achievement levels. Hence, there would be more classification errors with four achievement levels, resulting in lower consistency indices.

The results—derived using the program *BB-Class* (Brennan, 2004)—for the overall accuracy and consistency across all four performance levels as well as for the dichotomies created by the three cut scores are presented in Table 18–4.

Across all administrations and content areas, the overall decision accuracy ranged from 0.73 to 0.83 and the decision consistency ranged from 0.65 to 0.76. Dichotomous decisions have the higher accuracy and consistency values than the overall. The decision accuracy of the Basic/Proficient cut scores ranged from 0.90 to 0.95 and the decision consistency ranged from 0.86 to 0.93. Specifically for the spring administration, the decision accuracy results indicate that 91% to 94% of students meeting or exceeding the Proficient cut score would receive the same classification if their true scores were known. If a parallel test were administered, 88% to 92% of students meeting or exceeding the Proficient cut score would be classified in the same way.

**Table 18–4. Reliability of Performance-Level Classification Decisions**

Administration	Content Area	Statistic	Overall	Below Basic/ Basic	Basic/Proficient	Proficient/ Advanced
Winter	Algebra I	accuracy	0.80	0.88	0.94	0.98
Winter	Algebra I	consistency	0.72	0.83	0.92	0.97
Winter	Biology	accuracy	0.82	0.89	0.95	0.98
Winter	Biology	consistency	0.75	0.85	0.93	0.97
Winter	Literature	accuracy	0.82	0.92	0.92	0.98
Winter	Literature	consistency	0.74	0.88	0.89	0.97
Spring	Algebra I	accuracy	0.79	0.89	0.94	0.96
Spring	Algebra I	consistency	0.71	0.84	0.91	0.95
Spring	Biology	accuracy	0.83	0.92	0.94	0.96
Spring	Biology	consistency	0.76	0.89	0.92	0.95
Spring	Literature	accuracy	0.80	0.95	0.91	0.94
Spring	Literature	consistency	0.72	0.92	0.88	0.91
Summer	Algebra I	accuracy	0.77	0.88	0.91	0.98
Summer	Algebra I	consistency	0.68	0.84	0.88	0.96
Summer	Biology	accuracy	0.73	0.82	0.94	0.97
Summer	Biology	consistency	0.65	0.77	0.91	0.96
Summer	Literature	accuracy	0.80	0.90	0.90	1.00
Summer	Literature	consistency	0.72	0.86	0.86	1.00

## RATER AGREEMENT

Because CR items are included on the Keystone Exams, another source of random error is related to the scorers of those items. Frisbie (2005) noted that “test score reliability differs from scorer reliability” and that “the need for one kind of estimate cannot be satisfied by the other.” Additionally, the data most easily obtainable that captures this information comes from the “10 percent read behinds” collected during the scoring process. Partly because of the way these data are obtained and reported (i.e., it’s **not** a ratio of true score variance over observed score variance), the term *rater agreement* is used here, not *rater reliability* or *inter-rater reliability* as these terms are somewhat misleading.

The rater agreements for the Keystone Exams are presented in Tables 18–5 to 18–7. In addition, the percentages awarded to each score point are also presented in these tables. As the table shows, the exact inter-rater agreement percentages ranged from 87 to 100 percent for winter and 79 to 100 percent for spring, and 81 to 100 for summer. Overall, Algebra I had the highest exact agreements while Literature tended to have lower exact agreement. The percentages of exact and adjacent agreement for all content areas were nearly 100.

**Table 18–5. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items: Winter**

Content Area	Item	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Exact + Adjacent Agreement	% Awarded for Score Point 0	% Awarded for Score Point 1	% Awarded for Score Point 2	% Awarded for Score Point 3	% Awarded for Score Point 4	% Awarded for Score Point B/NS
Algebra I	1A	99	1	100	53	33	NA	NA	NA	14
Algebra I	1B	100	0	100	81	5	NA	NA	NA	14
Algebra I	1C	100	0	100	72	14	NA	NA	NA	14
Algebra I	1D	100	0	100	80	5	NA	NA	NA	14
Algebra I	2A	99	1	100	53	32	NA	NA	NA	16
Algebra I	2B	100	0	100	83	1	NA	NA	NA	16
Algebra I	2C	100	0	100	79	6	NA	NA	NA	16
Algebra I	2D	100	0	100	78	7	NA	NA	NA	16
Algebra I	3	94	6	100	23	25	18	13	4	17
Algebra I	4A	100	0	100	28	61	NA	NA	NA	11
Algebra I	4B	100	0	100	77	11	NA	NA	NA	11
Algebra I	4C	100	0	100	87	2	NA	NA	NA	11
Algebra I	4D	100	0	100	79	10	NA	NA	NA	11
Algebra I	5A	100	0	100	84	4	NA	NA	NA	12
Algebra I	5B	100	0	100	77	11	NA	NA	NA	12
Algebra I	5C	98	2	100	58	30	NA	NA	NA	12
Algebra I	5D	100	0	100	85	3	NA	NA	NA	12
Algebra I	6	95	5	100	38	24	12	6	2	18
Biology	1	90	10	100	44	26	8	4	NA	17
Biology	2	93	7	100	30	26	18	8	NA	19
Biology	3	90	10	100	49	22	7	3	NA	19
Biology	4	91	9	100	40	28	11	4	NA	16
Biology	5	87	12	99	14	14	48	6	NA	18
Biology	6	97	3	100	58	13	5	2	NA	20
Literature	1	86	14	100	9	35	34	6	NA	17
Literature	2	85	15	100	19	35	20	6	NA	20
Literature	3	87	13	100	26	29	18	4	NA	23
Literature	4	85	15	100	9	37	29	7	NA	18
Literature	5	87	13	100	13	36	25	5	NA	20
Literature	6	86	14	100	12	39	23	4	NA	21

*Note.* Some of the Algebra I CR items were scored by part. For example, 2A in the second column means part A of item 2. B/NS in the last column represents blank/non-scorable. NA means not applicable.

**Table 18–6. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items: Spring**

Content Area	Item	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Exact + Adjacent Agreement	% Awarded for Score Point 0	% Awarded for Score Point 1	% Awarded for Score Point 2	% Awarded for Score Point 3	% Awarded for Score Point 4	% Awarded for Score Point B/NS
Algebra I	1A	100	0	100	16	76	NA	NA	NA	8
Algebra I	1B	97	3	100	46	46	NA	NA	NA	8
Algebra I	1C	99	1	100	84	8	NA	NA	NA	8
Algebra I	1D	100	0	100	71	21	NA	NA	NA	8
Algebra I	2	87	13	100	17	49	13	6	1	15
Algebra I	3A	99	1	100	53	35	NA	NA	NA	12
Algebra I	3B	99	1	100	73	15	NA	NA	NA	12
Algebra I	3C	100	0	100	85	3	NA	NA	NA	12
Algebra I	3D	99	1	100	80	8	NA	NA	NA	12
Algebra I	4	85	14	99	35	24	15	10	6	11
Algebra I	5	85	14	99	44	13	10	13	7	13
Algebra I	6A	99	1	100	48	40	NA	NA	NA	11
Algebra I	6B	100	0	100	67	22	NA	NA	NA	11
Algebra I	6C	100	0	100	78	10	NA	NA	NA	11
Algebra I	6D	99	1	100	82	6	NA	NA	NA	11
Biology	1	79	19	98	34	33	16	9	NA	9
Biology	2	80	20	100	28	28	20	12	NA	11
Biology	3	90	10	100	22	39	19	6	NA	14
Biology	4	88	12	100	32	26	19	13	NA	9
Biology	5	93	7	100	41	21	13	12	NA	13
Biology	6	85	15	100	28	27	21	10	NA	14
Literature	1	80	19	99	6	26	38	19	NA	11
Literature	2	80	20	100	9	29	35	15	NA	12
Literature	3	80	19	99	13	27	34	12	NA	14
Literature	4	82	18	100	7	36	34	10	NA	12
Literature	5	79	21	100	12	30	31	12	NA	14
Literature	6	79	21	100	10	33	30	11	NA	16

*Note.* Some of the Algebra I CR items were scored by part. For example, 1A in the second column means part A of item 1. B/NS in the last column represents blank/non-scorable. NA means not applicable.

**Table 18–7. Inter-Rater Agreement and Percentage Awarded for Each Score Point for CR Items: Summer**

Content Area	Item	Inter-Rater Agreement % Exact	Inter-Rater Agreement % Adjacent	% Exact + Adjacent Agreement	% Awarded for Score Point 0	% Awarded for Score Point 1	% Awarded for Score Point 2	% Awarded for Score Point 3	% Awarded for Score Point 4	% Awarded for Score Point B/NS
Algebra I	1A	100	0	100	15	83	NA	NA	NA	3
Algebra I	1B	97	3	100	42	56	NA	NA	NA	3
Algebra I	1C	100	0	100	85	12	NA	NA	NA	3
Algebra I	1D	100	0	100	84	14	NA	NA	NA	3
Algebra I	2	89	11	100	28	44	11	9	4	5
Algebra I	3A	100	0	100	56	41	NA	NA	NA	2
Algebra I	3B	100	0	100	59	38	NA	NA	NA	2
Algebra I	3C	100	0	100	89	8	NA	NA	NA	2
Algebra I	3D	100	0	100	77	20	NA	NA	NA	2
Algebra I	4A	100	0	100	26	73	NA	NA	NA	1
Algebra I	4B	100	0	100	77	22	NA	NA	NA	1
Algebra I	4C	100	0	100	89	10	NA	NA	NA	1
Algebra I	4D	100	0	100	96	3	NA	NA	NA	1
Algebra I	5	91	9	100	27	49	12	6	2	4
Algebra I	6A	100	0	100	65	32	NA	NA	NA	3
Algebra I	6B	100	0	100	89	8	NA	NA	NA	3
Algebra I	6C	100	0	100	88	8	NA	NA	NA	3
Algebra I	6D	100	0	100	89	8	NA	NA	NA	3
Biology	1	93	7	100	11	24	41	21	NA	3
Biology	2	100	0	100	53	20	11	10	NA	6
Biology	3	100	0	100	29	27	26	12	NA	5
Biology	4	85	15	100	39	29	19	10	NA	4
Biology	5	93	7	100	37	31	18	9	NA	5
Biology	6	96	4	100	13	22	35	26	NA	3
Literature	1	84	16	100	7	42	43	1	NA	8
Literature	2	84	16	100	10	46	34	1	NA	8
Literature	3	81	19	100	10	51	25	1	NA	12
Literature	4	89	11	100	13	45	26	4	NA	11
Literature	5	84	16	100	15	43	25	3	NA	14
Literature	6	91	9	100	18	38	25	2	NA	16

*Note.* Some of the Algebra I CR items were scored by part. For example, 1A in the second column means part A of item 1. B/NS in the last column represents blank/non-scorable. NA means not applicable.

## CHAPTER NINETEEN: VALIDITY

---

As defined in the *Standards for Educational and Psychological Testing*, validity is “the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 2014, p.11). The *Standards* provide a framework for describing the sources of evidence that should be considered when evaluating validity. These sources include evidence based on test content, response processes, the internal structure of the test, the relationships between test scores and other variables, and the consequences of testing. In addition, when Rasch models are used to analyze assessment data, validity considerations related to those processes should also be explored.

The validity process involves the collection of a variety of evidence to support the proposed test score interpretations and uses. The entire technical report describes the technical aspects of the Keystone Exams in support of their score interpretations and uses. Each of the previous chapters contributes important components that pertain to score validation: test development, test administration, test scoring, item analysis, Rasch calibration, scaling, equating, score reporting, and reliability. This chapter summarizes and synthesizes the evidence based on the framework of the *Standards*. The purposes and intended uses of the Keystone Exams are reviewed first, and then each type of validity evidence is addressed in turn.

### PURPOSES AND INTENDED USES OF THE KEYSTONE EXAMS

The *Standards* emphasize that validity pertains to how test scores are used. To help contextualize the evidence that are presented in this chapter, the purposes of the Keystone Exams will be reviewed first. The Keystone Exams, which began in 2010–2011 for Algebra I, Biology, and Literature, are one component of Pennsylvania’s system of high school graduation requirements. Students take the exams upon completing specific courses. The Keystone Exams results help schools and districts guide students toward meeting state standards. Students who do not score Proficient or above on a Keystone Exam module may choose to complete a project-based assessment for that module, provided that they meet the requirements detailed below.

- The student has taken the course.
- The student was unsuccessful in achieving a score of Proficient or Advanced on the Keystone Exam after at least two attempts.
- The student has met the district’s attendance requirements for the course.
- The student has participated in a satisfactory manner in supplemental instructional services.

### EVIDENCE BASED ON TEST CONTENT

Test content validity evidence for the Keystone Exams rests greatly on establishing a link between each piece of the assessment (i.e., the items) and what students should know and be able to do as prescribed by the Keystone Exams Assessment Anchors and Eligible Content. The Keystone Exams are intended to measure the knowledge and skills described in the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature.

Lane (1999) suggests taking the following steps to support the content validity of an assessment. In the case of Keystone Exams, one should

- Evaluate the degree to which the test specifications represent and align with the knowledge and skills described in the Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature.
- Evaluate the alignment between the Keystone Exams items and test specifications to ensure representativeness.
- Evaluate the extent to which the curriculum aligns with the Assessment Anchors and Eligible Content.

- Conduct content reviews of the Keystone Exams items using a panel of content experts to see whether items measure the intended construct or are the sources of construct-irrelevant variance.
- Conduct fairness reviews of the items to avoid issues related to a specific subpopulation.
- Evaluate procedures for administration and scoring such as the appropriateness of instructions to examinees, time limit for the assessment, and training of raters.
- Submit operational tests to third-party independent reviews.

Chapters Two through Eight of this report present a considerable amount of evidence related to test content. As described in these chapters, all the items were developed and aligned with the Keystone Exams Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature. After development, items underwent multiple rounds of content and bias reviews. After being field-tested, items were reviewed with respect to their statistical properties. Items selected for the operational assessment had to pass content, psychometric, and PDE reviews. Tests were administered according to standardized procedures with allowable accommodations.

Some of the efforts made to ensure content validity are summarized below.

- DRC used Webb’s (1999) Depth of Knowledge (DOK) model to ensure the Keystone Exams items aligned with the Assessment Anchors and Eligible Content and the Academic Content Standards in terms of both content and cognitive levels.
- DRC established detailed test and item/passage development specifications and ensured the items were sufficient in number and adequately distributed across content, levels of cognitive complexity, and levels of difficulty.
- DRC selected qualified item writers and provided training to help ensure they wrote high-quality items.
- All newly developed items were first reviewed by content specialists and editors at DRC to make sure they measured the intended Assessment Anchors and Eligible Content for Algebra I, Biology, and Literature. Appropriateness for the intended students was also considered, as well as DOK, graphics, grammar/punctuation, language demand, and distractor reasonableness.
- Prior to field testing, the test items were submitted to content committees (composed of Pennsylvania educators) for review with respect to the following categories:
  - Overall quality and clarity
  - Anchor, Eligible Content, and/or standard alignment
  - Grade-level appropriateness
  - Difficulty level
  - Domain of knowledge (DOK)
  - Appropriate sources of challenge (e.g., unintended content and skills)
  - Correct answer
  - Quality of distractors
  - Graphics
  - Appropriate language demand
  - Freedom from bias
- The items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. This committee reviewed items for issues related to diversity, gender, and other pertinent factors.

- Items passing all the prior hurdles were tried out as embedded field-test items in the operational test. Several statistical analyses were conducted on the field-test data including classical item analyses, distractor analyses, and differential item functioning (DIF). Items were again carefully reviewed by DRC staff and a committee of Pennsylvania teachers with respect to their statistical characteristics. DIF was used to detect test items that might bias test scores for particular groups. Empirical investigation of DIF strengthens the validity evidence related to score interpretations for students in particular groups by eliminating potential sources of construct-irrelevant variance.
- The Keystone Exams were administered according to standardized procedures with allowable accommodations. Students were given ample time to complete the tests (i.e., there were no speediness issues).
- As described in Chapter Eight, the raters for constructed-response (CR) items were carefully recruited and well trained. Their scoring was monitored throughout the scoring session to ensure that an acceptable level of scoring accuracy was maintained.

## **EVIDENCE BASED ON RESPONSE PROCESS**

Response-process evidence is used to examine the extent to which the cognitive skills and processes employed by students match those identified in the test developer’s defined construct domains for all students and for each subgroup. Think-aloud procedures or cognitive labs can be used to collect this type of evidence. In addition, when an assessment includes CR items, an examination of the extent to which the raters interpret and apply the scoring criteria accurately when assigning scores to students’ responses on CR items also adds response-process validity evidence.

For the operational Keystone Exams offered in winter, spring, and summer, no cognitive lab studies were conducted to collect the response-process evidence. Rather, for all the Keystone Exams, well-organized scorer training and subsequent monitoring of rating accuracy helped ensure that raters strictly followed the scoring criteria and that no features unrelated to the rubric significantly affected their scoring.

## **EVIDENCE BASED ON INTERNAL STRUCTURE**

As described in the *Standards* (2014), internal-structure evidence refers to the degree to which the relationships between test items and test components conform to the construct on which the proposed test interpretations are based. For each Keystone Exam, one total test score as well as module scores were reported (see Chapter Sixteen for more information about the Keystone Exams scores). Several dimensionality studies were conducted in order to provide internal-structure evidence relating to the use of both types of scores.

## **ITEM-TEST CORRELATIONS**

Item-test correlations are provided in Appendix J and discussed in Chapter Eleven. All item correlations for operational items were greater than 0, indicating a positive association with the total score.

## **DIFFERENTIAL ITEM FUNCTIONING (DIF)**

DIF analyses with respect to gender, ethnicity, and administration mode address construct-irrelevant variance, which represents an important threat to the validity of achievement tests. As noted in Chapter Five, newly field-test items were screened and reviewed for DIF. Items identified with C-level DIF must be reviewed and approved by teacher committees in order to be used operationally (for scoring purposes). After operational use, DIF analyses were conducted on the operational items again to investigate whether the DIF code changed between field testing and operational use. Table 19–1 shows the number of items in which the DIF codes changed between field testing and operational testing, either from A/B to C-level DIF, or vice versa. There were no items in which the bias code changed from A/B to C, and few items in which the bias code changed from C to A/B. Items with a changed DIF code tended to be in Keystone Literature. There were insufficient N counts in the summer administration for all subjects, so DIF could not be run.



**Table 19–1. Summary of Bias Code Change from Field Test to Operational Test**

Administration	Content Area	Change	Male/Female	White/Black	PPT/CBT
Winter	Algebra I	AIB->C	0	0	0
Winter	Algebra I	C->AIB	0	0	0
Winter	Biology	AIB->C	0	0	0
Winter	Biology	C->AIB	0	0	0
Winter	Literature	AIB->C	0	0	0
Winter	Literature	C->AIB	1	0	0
Spring	Algebra I	AIB->C	0	0	0
Spring	Algebra I	C->AIB	0	0	0
Spring	Biology	AIB->C	0	0	0
Spring	Biology	C->AIB	0	0	0
Spring	Literature	AIB->C	0	0	0
Spring	Literature	C->AIB	1	0	0
Summer	Algebra I	AIB->C	-	-	-
Summer	Algebra I	C->AIB	-	-	-
Summer	Biology	AIB->C	-	-	-
Summer	Biology	C->AIB	-	-	-
Summer	Literature	AIB->C	-	-	-
Summer	Literature	C->AIB	-	-	-

*Note.* PPT represents the paper-and-pencil-based test, and CBT represents the computer-based test. A dash (-) indicates that there were insufficient N counts to run DIF analyses.

## MODULE CORRELATIONS

Correlations and disattenuated correlations among module scores for the Keystone Exams are presented in Table 19-2 and 19-3, respectively. Values were derived from the Keystone final data files (see Chapter Nine). These data can also provide information on score dimensionality that is part of internal-structure evidence. All Keystone Exams have two modules, which represent unique, non-overlapping topics. The intercorrelations between the modules within the content areas were positive and ranged from 0.82 to 0.88 and 0.79 to 0.82 for spring and winter, respectively. The intercorrelations between modules in different content areas ranged from 0.49 to 0.68 and 0.50 to 0.67 for spring and winter, respectively. Correlations for summer tend to be lower than correlations for spring and winter, likely influenced by the smaller volume of testers and different population of testers who take Keystone in the summer.

**Table 19–2. Correlations among Algebra I, Biology, and Literature Modules**

Administration	Content Area	Module	Algebra I Module 1	Algebra I Module 2	Biology Module 1	Biology Module 2	Literature Module 1	Literature Module 2
Winter	Algebra I	Module 1	-					
Winter	Algebra I	Module 2	0.79	-				
Winter	Biology	Module 1	0.52	0.52	-			
Winter	Biology	Module 2	0.53	0.53	0.80	-		
Winter	Literature	Module 1	0.53	0.50	0.64	0.65	-	
Winter	Literature	Module 2	0.56	0.53	0.66	0.67	0.82	-
Spring	Algebra I	Module 1	-					
Spring	Algebra I	Module 2	0.84	-				
Spring	Biology	Module 1	0.65	0.65	-			
Spring	Biology	Module 2	0.67	0.68	0.88	-		
Spring	Literature	Module 1	0.52	0.49	0.65	0.67	-	
Spring	Literature	Module 2	0.55	0.53	0.70	0.72	0.82	-
Summer	Algebra I	Module 1	-					
Summer	Algebra I	Module 2	0.77	-				
Summer	Biology	Module 1	0.42	0.48	-			
Summer	Biology	Module 2	0.23	0.26	0.83	-		
Summer	Literature	Module 1	0.21	0.35	0.33	0.28	-	
Summer	Literature	Module 2	0.32	0.61	0.44	0.43	0.70	-

The correlations in Table 19–2 are based on the observed module scores. These observed-score correlations are weakened by existing measurement error contained within each module score. As a result, disattenuated correlations could provide an estimate of the relationships among modules if there were no measurement error. (An important caveat is explained further below.) The disattenuated correlation coefficients  $R_{12}$  can be computed by using the formula (Spearman, 1904; Spearman, 1910) below:

$$R_{12} = \frac{r_{12}}{\sqrt{r_{11}r_{22}}},$$

where  $r_{12}$  is the observed correlation, and  $r_{11}$  and  $r_{22}$  are the reliabilities for Module 1 and Module 2. Disattenuated correlations very near 1.00 suggest that the same or very similar constructs are being measured. Values somewhat less than 1.00 suggest that different modules are measuring slightly different aspects of the same construct. Values markedly less than 1.00 suggest the modules reflect different constructs.

Table 19–3 shows the corresponding disattenuated correlations for each Keystone Exam. Given that none of these modules had reliabilities of 1 (see Chapter Eighteen), the disattenuated module correlations are higher than their observed score counterparts.

**Table 19–3. Disattenuated Correlations among Algebra I, Biology, and Literature Modules**

Administration	Content Area	Module	Algebra I Module 1	Algebra I Module 2	Biology Module 1	Biology Module 2	Literature Module 1	Literature Module 2
Winter	Algebra I	Module 1	-					
Winter	Algebra I	Module 2	1.00	-				
Winter	Biology	Module 1	0.65	0.63	-			
Winter	Biology	Module 2	0.67	0.66	0.97	-		
Winter	Literature	Module 1	0.66	0.61	0.76	0.79	-	
Winter	Literature	Module 2	0.69	0.65	0.78	0.81	0.98	-
Spring	Algebra I	Module 1	-					
Spring	Algebra I	Module 2	1.00	-				
Spring	Biology	Module 1	0.75	0.75	-			
Spring	Biology	Module 2	0.77	0.78	0.99	-		
Spring	Literature	Module 1	0.62	0.58	0.75	0.78	-	
Spring	Literature	Module 2	0.65	0.63	0.81	0.84	0.98	-
Summer	Algebra I	Module 1	-					
Summer	Algebra I	Module 2	1.01	-				
Summer	Biology	Module 1	0.52	0.60	-			
Summer	Biology	Module 2	0.28	0.33	0.99	-		
Summer	Literature	Module 1	0.28	0.48	0.42	0.36	-	
Summer	Literature	Module 2	0.41	0.79	0.54	0.53	0.93	-

Similar to the findings from the correlations, the within-content area disattenuated correlations were higher for winter and spring than summer administrations. The higher disattenuated correlations suggest that the within-content area modules might be measuring essentially the same construct. This, in turn, suggests that the within content module scores might not provide unique information about the strengths or weaknesses of many of the students.

On a fairly consistent basis, the disattenuated correlations among the modules within each content area are higher than the correlations among modules across different content areas. In general, within-content area module correlations were greater than 0.97 for winter and spring and greater than 0.93 for summer, while across-content area module disattenuated correlations ranged from 0.28 to 0.84.

It should be noted that some caution is needed when interpreting the disattenuated results because the reliabilities used to calculate the disattenuated correlations are subject to both upward and downward biases. Consequently, some of the values in the table above may be higher or lower than they should be, depending on which bias prevails for any given pair of module scores. When the reliabilities are lower than they should be, the disattenuated correlations will be inflated and, in some instances, can appear higher than the theoretical correlation maximum value of 1.00.

Overall, the internal structure evidence presented supports the assumption that related elements of each of the Keystone Exams are correlated in the intended manner. The modules *within* each content area have stronger relationships than the *across* content area modules, which further supports reporting a total score for each different content area.

The module scores present more of a mixed message. Since the modules in each content area were designed to measure distinct components of the content area, it is reasonable to expect that the inter-content module correlations should be positive and strong but, ideally, not extremely high. However, the disattenuated correlations imply that some modules are essentially measuring the same constructs for most of the students. Consequently, there may be less support for providing results for some module scores beyond the total score. While there is content rationale underlying the creation of the module scores, the empirical correlations illustrate that caution is required when using the module scores to identify a student’s strengths and weaknesses. Certainly, instructional programs should not be based on module score information alone, but rather in conjunction with other sources of evidence available (e.g., teacher observations, other exam performance).

## EVIDENCE BASED ON RELATIONSHIPS WITH OTHER VARIABLES

As described in the *Standards*, “Evidence based on relationships with other variables provides evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations” (AERA, APA, & NCME, 2014, p. 16). This category of evidence refers to external structure evidence and has been classified as three types of evidence: *convergent*, *discriminant*, and *criterion-related*. Convergent evidence is provided by relationships among students’ performances on different assessments intended to measure a similar construct. Discriminant evidence is provided by relationships among students’ performances on different tests intended to measure different constructs. Criterion-related evidence, either predictive or concurrent, is provided by relationships between students’ test scores and their performances on a criterion measure (Cronbach, 1971; Messick, 1989).

The correlations among students’ test scores on different Keystone Exams including Algebra I, Biology, and Literature are shown in Table 19–4 to provide some discriminant validity evidence. In this table, both the observed correlations and the disattenuated correlations (in the parentheses) are reported. Each Keystone Exam measures different constructs, so the correlations among them were not expected to be extremely high. The values in this table are consistent with this expectation. As can be seen, the correlations among the Keystone Exams ranged from 0.43 to 0.75 and the disattenuated correlations ranged from 0.48 to 0.81.

**Table 19–4. Correlations Among Student Performance**

Administration	Content Area	Algebra I	Biology
Winter	Biology	0.63(0.71)	
Winter	Literature	0.61(0.68)	0.74(0.81)
Spring	Biology	0.74(0.80)	
Spring	Literature	0.60(0.66)	0.75(0.81)
Summer	Biology	0.43(0.48)	
Summer	Literature	0.43(0.50)	0.47(0.53)

External evidence for the Keystone Exams is examined by using students’ scores on the 2023 Pennsylvania System of School Assessment (PSSA) as the external criteria. The final 2023 Algebra I, Biology and Literature data files were merged with the PSSA mathematics, science, and ELA data by students’ PA secure IDs. Then the correlations between students’ scores on the Keystone Exams and on the PSSA were calculated as one component of external evidence. This analysis was attempted for all Keystone administrations. However, only enough students were obtained for the spring administration. Table 19–5 summarizes the sample sizes and correlations by grade and content area after the file merging of the spring Keystone Exams and the PSSA.

**Table 19–5. Correlation between Spring Keystone Exam Scores and PSSA Scores by Grade**

Keystone Content Area/ PSSA Subject	Grade 7 N	Grade 7 Correlation	Grade 8 N	Grade 8 Correlation
Algebra I/Mathematics	6372	0.78	28809	0.85
Biology/Science	NA	NA	245	0.74
Literature/ELA	NA	NA	5	-

*Note.* Correlations were not reported for sample sizes less than 10 students.

The correlations within the same content area ranged from 0.74 to 0.85. These results suggest the Keystone Exams measured something similar but not identical to the corresponding PSSA tests. The results also provide external evidence in support of the Keystone Exams as a valid measure of students' achievement.

The collection of external evidence relating to the Keystone Exams is an ongoing process once the data are collected in the future. Other criterion-related evidence can be evaluated by the relationships between the Keystone Exams and criterion variables such as the Scholastic Aptitude Test (SAT), the American College Testing (ACT), or students' Grade Point Average (GPA) in their first college course.

## EVIDENCE BASED ON CONSEQUENCES OF TESTS

Based on the *Standards* (AERA, APA, & NCME, 2014), evidence of the consequences of implementing an assessment program is an additional source of validity information. Both positive and negative (intended and unintended) consequences of score-based inferences must be investigated to fully evaluate the pool of validity evidence.

Lane and Stone (2002) summarized the general *intended* consequences for state assessments and accountability programs:

- Student, teacher, and administrator motivation and effort
- Curriculum and instruction practices (including content and strategies)
- Improved learning for all students
- Content and format of classroom assessments
- Professional development support
- Use and nature of test-preparation activities
- Student, teacher, administrator, and public awareness and beliefs about the assessment, criteria for judging performance, and the use of assessment results

Evidence for the improvement of student learning can be seen by examining the change in the percentage of students who scored Proficient or Advanced from the inception of the Keystone program in 2011 to the most recent administrations. Table 19–6 provides the percentages of students who scored Proficient or Advanced by administration and content. Values are derived from the first-time test-takers during each spring administration and represent administration-specific performance level classification.

**Table 19–6. Percentages of Students earning Proficient or Advanced Performance Level Classifications**

Administration	Algebra I Number	Algebra I Percent	Biology Number	Biology Percent	Literature Number	Literature Percent
Spring 2011	94,697	38.6	46,979	35.7	42,808	49.9
Spring 2013	157,811	47.6	134,995	47.2	117,830	63.1
Spring 2014	124,954	51.5	119,274	52.9	113,477	60.9
Spring 2015	121,255	50.1	115,936	58.0	114,387	67.2
Spring 2016	120,104	50.6	116,345	56.8	112,361	65.0
Spring 2017	118,704	51.7	115,473	56.6	110,966	64.5
Spring 2018	117,345	50.6	114,704	57.2	109,387	64.7
Spring 2019	117,677	48.4	113,834	54.7	110,030	63.3
Spring 2020						
Spring 2021	103,296	37.1	98,419	45.6	96,409	57.7
Spring 2022	120,398	36.6	112,130	47.6	107,815	60.1
Spring 2023	120,395	36.6	113,460	45.8	108,505	58.3

*Note.* Spring 2020 data is not available due to school closures and cancellation of Keystone testing in Spring 2020.

Lane and Stone (2002) also summarized the possible unintended outcomes:

- Narrowing of curriculum and instruction to focus only on the specific standards assessed and ignoring the broader construct reflected in the specified standards
- Use of test-preparation materials that are closely linked to the assessment without making changes to instruction
- Use of unethical test-preparation materials or administration procedures
- Differential performance gains for subgroups of students
- Inappropriate or unfair uses of test scores, such as questionable practices in reassignment of teachers or principals
- For some students, decreased confidence and motivation to learn and to perform well on the assessment because of past experiences with assessments

As noted above, one important piece of consequential evidence pertains to the use of assessment results. As shown in Chapter Sixteen, there are several different types of scores and score reports used for the Keystone Exams. The extent to which various groups of users (e.g., students, teachers) interpret these scores and reports appropriately affects the validity of subsequent uses of these results. Chapter Sixteen is intended to provide accurate and clear test score and report information with the hope that this will help users avoid unintended uses and interpretations of the Keystone Exams results. Nevertheless, evidence pertaining to other consequences of the Keystone Exams needs continued research.

## **EVIDENCE RELATED TO THE USE OF RASCH MODEL**

Since the Rasch model is the basis of all calibration, scaling, and equating analyses associated with the Keystone Exams, the validity of the inferences from these results depends on the degree to which the assumptions of the model are met, as well as the fit between the model and the test data. As discussed in Chapter Twelve, the underlying assumptions of Rasch models were essentially met for all the Keystone Exams data, indicating the appropriateness of using the Rasch models to analyze the Keystone Exams data.

## **VALIDITY EVIDENCE SUMMARY**

Validity evidence related to test content was reviewed earlier in this chapter. The early chapters of this technical report show that a strong link can be established between each item on the Keystone Exam and its associated Eligible Content. Details regarding how the operational Keystone Exams were assembled to reflect the state content standards and detailed information regarding educator reviews (including content, bias, and sensitivity reviews) are presented in Chapter Six.

Module score intercorrelations were also presented in this chapter. In general, within-content area module scores (e.g., Algebra I Module 1 and Algebra Module 2) were correlated more highly than across-content module scores (e.g., Literature Module 1 and Algebra Module 1). Consequently, this provides some favorable evidence regarding the internal and external relationships between the test components.

Moreover, validity of score inferences is bolstered when test scores are consistent. Here, the reliabilities of the total test scores presented in Chapter Eighteen were high.

As reported in Chapter Five, DIF with respect to gender, ethnicity, and test administration mode helps address construct-irrelevant variance, which represents an important threat to the validity of inferences made from achievement test scores. As noted in that chapter, field-test items are screened and reviewed for DIF. Only items approved by data review committees are eligible for operational use.

## CHAPTER TWENTY: SPECIAL STUDY ON TEST ADMINISTRATION MODE

The Keystone Exams were offered in both paper-and-pencil test (PPT) and online computer-based test (CBT) formats. Test administration mode may have an impact on students' performance, especially when the students are more familiar with one mode over the other. Data Recognition Corporation (DRC) conducted an elaborate test administration mode study on the fall 2010 field tests and presented the study results to the Pennsylvania Department of Education (PDE) and Technical Advisory Committee (TAC). In this report, the mode effects are reinvestigated using operational testing results. However, given that schools choose the test administration format, the results from this study should be viewed with caution.

### SUMMARY OF STUDENTS' DEMOGRAPHIC DISTRIBUTIONS

Table 20-1 provides the breakdown by mode for the final sample. As one can see, substantially more student took the tests on paper than on a computer.

**Table 20–1. Distribution by Mode**

Keystone Exam	Paper/Pencil Test (PPT)	Computer-Based Test (CBT)
Algebra I (Number)	115819	47141
Algebra I (Percent)	71.1	28.9
Biology (Number)	89472	48314
Biology (Percent)	64.9	35.1
Literature (Number)	80884	44647
Literature (Percent)	64.4	35.6

Table 20–2 provides the demographic distributions by test administration mode for the final sample. Note that, for all content areas, the percentage of white students who took CBTs was consistently higher than the percentage of white students who took PPTs. In contrast, the percentage of black and Hispanic students who took PPTs was consistently higher than the percentage of those students who took CBTs. Additionally, some test accommodations are available only with CBTs. Thus, equivalence of samples across the two modes may not be assumed. In other words, a direct comparison of summary statistics based on raw total scores is not advisable.

**Table 20–2. Students' Demographic Distribution by Mode**

Demographic or Educational Characteristic	Algebra I PPT Percentage	Algebra I CBT Percentage	Biology PPT Percentage	Biology CBT Percentage	Literature PPT Percentage	Literature CBT Percentage
Female	49.0	48.6	49.4	48.2	48.5	47.7
Male	51.0	51.4	50.6	51.8	51.5	52.3
African American	17.6	12.9	17.1	11.6	17.1	10.9
American Indian	0.2	0.2	0.2	0.2	0.2	0.2
Asian	4.4	4.4	4.3	4.8	4.4	4.5
Hispanic	15.7	13.8	14.7	12.7	13.7	13.7
Multi Racial	4.6	4.3	4.5	4.1	4.3	4.0
Native Hawaiian/ Pacific Islander	0.1	0.1	0.1	0.1	0.1	0.1
White	57.5	64.3	59.2	66.7	60.2	66.5
Economically Disadvantaged	49.2	45.6	48.8	43.1	47.8	42.7
Total number of students	115819	47141	89472	48314	80884	44647

## MODE DIF SUMMARY OF OPERATIONAL ITEMS

Mode DIF can be investigated at the item level using differential item functioning (DIF) analysis. Each year a small number of field test items show C/severe mode DIF. These items were presented to the Pennsylvania data review committee under both test administration modes for review to determine if C level mode DIF can be explained by the presentation format of items. Subsequently, during construction of operational forms, all items were carefully selected by DRC’s staff in test development and psychometric services departments. Special attention was given to avoid using items with C level mode DIF. Table 20–2 presents the number of 2023 operational items with different levels of mode DIF based on previous field test data (banked values) and based on 2023 operational data. At the time of forms construction, no items showed C/severe mode DIF. Additionally, none of these items showed C/severe mode DIF based on 2023 operational data.

**Table 20–3. Mode DIF Summary of Spring Operational Items Only**

Data Source	DIF Category	Algebra I Number of Items	Biology Number of Items	Literature Number of Items
Banked Value	A-	18	23	21
Banked Value	A+	22	30	18
Banked Value	B-	1	1	0
Banked Value	B+	1	0	1
Banked Value	C-	0	0	0
Banked Value	C+	0	0	0
Operational Value	A-	18	24	19
Operational Value	A+	24	30	19
Operational Value	B-	0	0	0
Operational Value	B+	0	0	2
Operational Value	C-	0	0	0
Operational Value	C+	0	0	0

*Note.* A “+” sign indicates the item favors the CBT group (focal); a “-” sign indicates the item favors the PPT group (reference).

## RAW-TO-SCALED-SCORE COMPARISON

To test for mode effect at the test level, raw-to-scaled tables were generated based on the testing results of students who took PPT and students who took CBT compared to the pre-equated solution, used for scoring. Tables 20–4 to 20–6 present the raw-to-scaled score tables for these three groups. Given there are no operational items that showed C mode DIF, it is not a surprise that scaled scores and standard errors are very close across testing modes. For Biology and Literature, all raw scores were classified in the same proficiency category across modes. However, for Algebra the proficiency cut-score (1500) between the CBT and pre-equated solution differed by 1 raw score point. The difference in performance level classification is due to a small difference in the associated scaled scores where the SEM is 16. The small scaled score differences across modes that are within a fraction of SEM do not suggest mode effects.



**Table 20–4. Raw-to-Scaled-Score Comparison for Algebra I**

Raw Score	Scaled Score Pre-Equated	SEM Pre- Equated	Scaled Score PPT	SEM PPT	Scaled Score CBT	SEM CBT
0	1217	92	1217	92	1215	92
1	1278	51	1279	51	1276	51
2	1315	37	1315	37	1313	37
3	1337	30	1337	30	1335	30
4	1353	27	1353	27	1351	27
5	1366	24	1366	24	1364	24
6	1377	22	1377	22	1375	23
7	1386	21	1387	21	1385	21
8	1395	20	1395	20	1393	20
9	1402	19	1403	19	1401	19
10	1409	18	1410	18	1408	18
11	1416	18	1416	18	1415	18
12	1422	17	1422	17	1421	17
13	1428	17	1428	17	1427	17
14	1433	16	1434	16	1432	16
15	1439	16	1439	16	1437	16
16	1444	16	1444	16	1443	16
17	1449	15	1449	15	1447	16
18	1453	15	1453	15	1452	15
19	1458	15	1458	15	1457	15
20	1462	15	1462	15	1461	15
21	1467	15	1467	15	1466	15
22	1471	15	1471	14	1470	15
23	1475	14	1475	14	1474	14
24	1479	14	1479	14	1478	14
25	1483	14	1483	14	1483	14
26	1487	14	1487	14	1487	14
27	1491	14	1491	14	1491	14
28	1495	14	1495	14	1495	14
29	1499	14	1498	14	1498	14
30	1503	14	1502	14	1502	14
31	1507	14	1506	14	1506	14
32	1511	14	1510	14	1510	14
33	1515	14	1514	14	1514	14
34	1519	14	1518	14	1518	14
35	1523	14	1522	14	1522	14
36	1527	14	1526	14	1526	14
37	1531	14	1530	14	1530	14
38	1535	14	1534	14	1534	14
39	1539	15	1538	15	1538	15
40	1543	15	1542	15	1543	15
41	1548	15	1547	15	1547	15

**Table 20–4 (continued). Raw-to-Scaled-Score Comparison for Algebra I**

<b>Raw Score</b>	<b>Scaled Score Pre-Equated</b>	<b>SEM Pre- Equated</b>	<b>Scaled Score PPT</b>	<b>SEM PPT</b>	<b>Scaled Score CBT</b>	<b>SEM CBT</b>
42	1552	15	1551	15	1552	15
43	1557	15	1556	16	1556	15
44	1562	16	1561	16	1561	16
45	1567	16	1566	16	1566	16
46	1572	16	1571	16	1571	16
47	1578	17	1577	17	1577	17
48	1583	17	1583	17	1583	17
49	1589	18	1589	18	1589	18
50	1596	18	1596	18	1596	18
51	1603	19	1603	19	1603	19
52	1611	20	1610	20	1610	20
53	1619	21	1619	21	1619	21
54	1629	23	1628	23	1628	23
55	1640	24	1639	24	1639	24
56	1653	27	1652	27	1652	27
57	1669	30	1668	30	1669	30
58	1691	37	1691	37	1691	37
59	1728	51	1727	51	1727	51
60	1789	92	1789	92	1789	92

**Table 20–5. Raw-to-Scaled-Score Comparison for Biology**

<b>Raw Score</b>	<b>Scaled Score All</b>	<b>SEM All</b>	<b>Scaled Score PPT</b>	<b>SEM PPT</b>	<b>Scaled Score CBT</b>	<b>SEM CBT</b>
0	1220	92	1220	92	1220	92
1	1281	51	1281	51	1281	51
2	1317	36	1317	36	1317	36
3	1338	30	1338	30	1338	30
4	1354	26	1354	26	1354	26
5	1366	24	1366	24	1366	24
6	1376	22	1376	22	1376	22
7	1385	20	1385	20	1385	20
8	1393	19	1393	19	1393	19
9	1400	18	1400	18	1400	18
10	1406	18	1406	18	1406	18
11	1412	17	1412	17	1412	17
12	1417	16	1417	16	1417	16
13	1422	16	1422	16	1422	16
14	1427	15	1427	15	1427	15
15	1432	15	1432	15	1432	15
16	1436	15	1436	15	1436	15
17	1441	14	1441	14	1441	14
18	1445	14	1445	14	1445	14
19	1449	14	1449	14	1449	14
20	1452	14	1452	14	1452	14
21	1456	14	1456	14	1456	14
22	1460	13	1460	13	1460	13
23	1463	13	1463	13	1463	13
24	1467	13	1467	13	1467	13
25	1470	13	1470	13	1470	13
26	1473	13	1473	13	1473	13
27	1477	13	1477	13	1477	13
28	1480	13	1480	13	1480	13
29	1483	13	1483	13	1483	13
30	1486	13	1486	13	1486	13
31	1489	12	1489	12	1489	12
32	1492	12	1492	12	1492	12
33	1496	12	1496	12	1496	12
34	1499	12	1499	12	1499	12
35	1502	12	1502	12	1502	12
36	1505	12	1505	12	1505	12
37	1508	12	1508	12	1508	12
38	1511	12	1511	12	1511	12
39	1514	12	1514	12	1514	12
40	1517	13	1517	13	1517	13
41	1520	13	1520	13	1520	13

**Table 20–5 (continued). Raw-to-Scaled-Score Comparison for Biology**

<b>Raw Score</b>	<b>Scaled Score All</b>	<b>SEM All</b>	<b>Scaled Score PPT</b>	<b>SEM PPT</b>	<b>Scaled Score CBT</b>	<b>SEM CBT</b>
42	1523	13	1523	13	1523	13
43	1527	13	1527	13	1527	13
44	1530	13	1530	13	1530	13
45	1533	13	1533	13	1533	13
46	1537	13	1537	13	1537	13
47	1540	13	1540	13	1540	13
48	1544	14	1544	14	1544	14
49	1548	14	1548	14	1548	14
50	1552	14	1552	14	1552	14
51	1556	14	1556	14	1556	14
52	1560	15	1560	15	1560	15
53	1565	15	1565	15	1565	15
54	1570	16	1570	16	1570	16
55	1575	16	1575	16	1575	16
56	1580	17	1580	17	1580	17
57	1586	18	1586	18	1586	18
58	1593	19	1593	19	1593	19
59	1600	20	1600	20	1600	20
60	1608	21	1608	21	1608	21
61	1618	23	1618	23	1618	23
62	1630	26	1630	26	1630	26
63	1645	29	1645	29	1645	29
64	1666	36	1666	36	1666	36
65	1701	50	1701	50	1701	50
66	1762	91	1762	91	1762	91

**Table 20–6. Raw-to-Scaled-Score Comparison for Literature**

<b>Raw Score</b>	<b>Scaled Score All</b>	<b>SEM All</b>	<b>Scaled Score PPT</b>	<b>SEM PPT</b>	<b>Scaled Score CBT</b>	<b>SEM CBT</b>
0	1202	92	1202	92	1200	92
1	1263	51	1263	51	1262	51
2	1299	37	1299	37	1298	37
3	1321	30	1321	30	1320	30
4	1338	27	1337	27	1336	27
5	1351	24	1350	24	1349	24
6	1361	23	1361	23	1360	23
7	1371	21	1371	21	1370	21
8	1380	20	1379	20	1378	20
9	1387	19	1387	19	1386	19
10	1395	19	1394	19	1393	19
11	1401	18	1401	18	1400	18
12	1408	18	1407	18	1406	18
13	1414	17	1413	17	1412	17
14	1419	17	1419	17	1418	17
15	1425	17	1425	17	1423	16
16	1430	16	1430	16	1429	16
17	1435	16	1435	16	1434	16
18	1441	16	1440	16	1439	16
19	1446	16	1445	16	1444	16
20	1450	16	1450	16	1449	16
21	1455	16	1455	16	1454	15
22	1460	15	1460	15	1458	15
23	1465	15	1465	15	1463	15
24	1470	15	1469	15	1468	15
25	1474	15	1474	15	1473	15
26	1479	15	1479	15	1477	15
27	1484	15	1484	15	1482	15
28	1489	16	1488	16	1487	15
29	1494	16	1493	16	1492	16
30	1498	16	1498	16	1496	16
31	1503	16	1503	16	1501	16
32	1508	16	1508	16	1506	16
33	1514	16	1513	16	1512	16
34	1519	16	1519	16	1517	16
35	1524	17	1524	17	1522	17
36	1530	17	1530	17	1528	17
37	1536	17	1536	17	1534	17
38	1542	18	1542	18	1540	18
39	1548	18	1548	18	1546	18
40	1555	18	1555	19	1553	18
41	1562	19	1562	19	1560	19

**Table 20–6 (continued). Raw-to-Scaled-Score Comparison for Literature**

Raw Score	Scaled Score All	SEM All	Scaled Score PPT	SEM PPT	Scaled Score CBT	SEM CBT
42	1569	20	1569	20	1567	20
43	1577	20	1577	20	1575	20
44	1586	21	1586	21	1584	21
45	1595	22	1596	22	1593	22
46	1606	24	1606	24	1604	24
47	1618	25	1618	26	1616	26
48	1632	28	1633	28	1630	28
49	1650	32	1650	32	1648	32
50	1673	38	1674	38	1672	38
51	1711	52	1712	52	1710	52
52	1774	92	1775	92	1773	92

### MODE DIF SUMMARY OF FIELD-TEST ITEMS

The mode DIF analysis was also conducted on the field-test items embedded in the spring 2023 operational forms. Table 20–7 presents the mode DIF summary of all the embedded field-test items. As can be seen, no Keystone field-test items showed C-level mode DIF. Most items showed A/negligible mode DIF and a few items showed B/moderate DIF.

**Table 20–7. Mode DIF Summary of Field-Test Items**

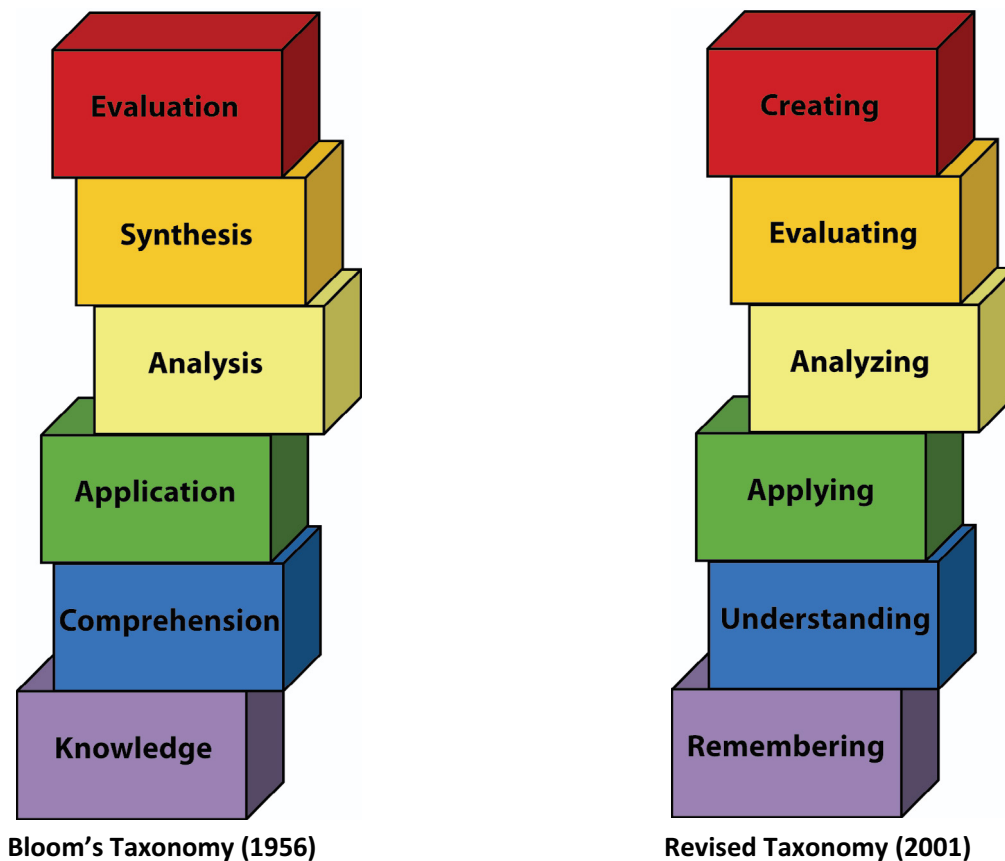
Category	Algebra I Number of Items	Biology Number of Items	Literature Number of Items
A-	118	42	125
A+	170	66	202
B-	0	0	2
B+	0	0	7
C-	0	0	0
C+	0	0	0

## APPENDIX A: UNDERSTANDING DEPTH OF KNOWLEDGE AND COGNITIVE COMPLEXITY

---

One of the steps in the item review process involves Pennsylvania educators' review of items for cognitive complexity or the nature of thinking. One model for classifying thinking into cognitive levels of complexity is Bloom's Taxonomy. Bloom's Taxonomy was first presented in 1956 through the publication, *The Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain*. This taxonomy identifies six levels within the cognitive domain, from the simple recall or recognition of facts, at the lowest level, through increasingly more complex levels, to the highest level which is classified as evaluation.

During the late 1990s, the original Bloom's Taxonomy was revised (Anderson and Krathwohl, 2001). In the 2001 version of Bloom's Taxonomy, the names of the six major cognitive process categories or levels were revised to indicate action (verbs) rather than non-action (nouns) as noted in the graphic below.



More recently, Webb's Depth-of-Knowledge Levels have also been used in the review of items for cognitive demand. Webb's Depth of Knowledge was created by Norman Webb from the Wisconsin Center for Education Research. Webb's definition of depth of knowledge is the degree or complexity of knowledge that the content curriculum standards and expectations require. Therefore, when reviewing items for depth of knowledge, the item is reviewed to determine whether or not it is as demanding cognitively as what the actual content curriculum standard expects. In the case of the Pennsylvania Keystone items, the item meets the criterion if the depth of knowledge of the item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

Webb's Depth of Knowledge includes four levels, from the lowest (basic recall) to the highest (extended thinking). Verb examples that represent each level in Webb's Depth of Knowledge can be found in the information that follows. However, verbs alone do not describe the depth of knowledge. Rather, depth of knowledge also focuses upon how well the students need to know the content before they can respond to a given item.

Because Bloom's Taxonomy (1956) is very familiar to many teachers, information comparing Bloom's Taxonomy and Webb's Depth of Knowledge is provided to Pennsylvania educators during the review of the Keystone items. The comparison serves as a "bridge" for teachers to understand Webb's Depth of Knowledge as compared to Bloom's Taxonomy.

## **ALGEBRA I DEPTH OF KNOWLEDGE**

### **DEPTH OF KNOWLEDGE GUIDELINES FOR REVIEW OF ALGEBRA I, ALGEBRA II, AND GEOMETRY ITEMS**

Committees of Pennsylvania educators review each Keystone Exam item, not only to determine whether or not the item measures what it is intended to measure, but also to determine whether or not the item aligns with the cognitive level or depth of knowledge of the Assessment Anchor as defined by the Eligible Content. The information below provides a definition of the four depth-of-knowledge levels. The charts at the end of the section also provide a comparison between Bloom's Taxonomy and Webb's Depth of Knowledge for mathematics (Algebra I, Algebra II, and Geometry). Included are examples of verbs (i.e., the action). Using this information as well as the charts, Pennsylvania educators are asked to determine the depth of knowledge of each item and to verify that the depth of knowledge of each item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

### **DEFINITIONS OF WEBB'S DEPTH OF KNOWLEDGE**

*Level 1 (Recall)* requires the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics, a one-step, well-defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify Level 1 include "identify," "recall," "recognize," "use," and "measure." Verbs such as "describe" and "explain" could be classified at different levels, depending on what is to be described and explained.

*Level 2 (Skill/Concept)* requires the engagement of some mental processing beyond a habitual response. A Level 2 item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply more than one step. For example, to compare data requires first identifying characteristics of objects or phenomena and then grouping or ordering the objects. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different levels depending on the object of the action. For example, interpreting information from a simple graph, or reading information from the graph, are also at Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is at Level 3. Level 2 activities are not limited only to number skills, but may involve visualization skills and probability skills. Other Level 2 activities include noticing or describing non-trivial patterns; explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

*Level 3 (Strategic Thinking)* requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is at Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be at Level 3. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and deciding which concepts to apply in order to solve a complex problem.

*Level 4 (Extended Thinking)* requires complex reasoning, planning, developing, and thinking most likely over an extended period of time. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2. However, if the student is to conduct a river study that requires



taking into consideration a number of variables, this would be a Level 4. At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and have to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Level 4 activities include designing *and* conducting experiments and projects; developing and proving conjectures; making connections between a finding and related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs.

Note: Multiple-choice and constructed-response items can be written at a depth-of-knowledge Level 4; however, to design an item in this format is difficult, as it would require research, investigation, and application, often over an extended period of time (e.g., performance-based tasks; portfolios; research studies/projects).

(Webb, N. 1997, 1999, 2002, 2005, 2006)

**Table A–1. Bloom’s Taxonomy – Algebra I**

Categories (1956)	Definition	Examples of Action Words*
Knowledge	Student remembers, or recalls appropriate previously learned information.	define; identify; name; select; state; order; (involves a one-step problem)
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	convert; estimate; explain; express; factor; generalize; give example; identify; indicate; locate; picture; (involves two or more steps)
Application	Student selects, transfers, and uses data and principles to complete a task or problem with minimum directions.	apply; choose; compute; employ; interpret; graph; modify; operate; plot; practice; solve; use; (involves three or more steps)
Analysis	Student distinguishes, classifies, and relates assumptions, hypotheses, evidence, or structure of a statement or question.	compare; contrast; correlate; differentiate; discriminate; examine; infer; maximize; minimize; prioritize; subdivide; test
Synthesis	Student originates, integrates, and combines ideas into a product, plan, or proposal that is new to him or her.	arrange; collect; construct; design; develop; formulate; organize; set up; prepare; plan; propose; create experiment and record data
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	appraise; assess; defend estimate; evaluate; judge; predict; rate; validate; verify

**Table A–2. Webb’s Depth of Knowledge – Algebra I**

Categories	Definition	Example of Action Words*
Recall	Student recalls facts, information, procedures, or definitions.	define; identify; name; select; state; order; one step
Basic Application of Skill/ Concept	Student uses information, conceptual knowledge, and procedures.	apply; choose; compute; employ; interpret; graph; modify; operate; plot; practice; solve; use; two or more steps
Strategic Thinking	Student uses reasoning and develops a plan or sequence of steps; process has some complexity.	compare; contrast; correlate; differentiate; discriminate; examine; infer; maximize; minimize; prioritize; subdivide; test
Extended Thinking	Student conducts an investigation, needs time to think and process multiple conditions of the problem or task. (The item/task generally requires several days or weeks to complete.)	arrange; collect; construct; design; develop; formulate; organize; set up; prepare; plan; propose; create experiment and record data

\*Some action words (verbs) can be classified at different depth-of-knowledge levels depending on the context of the item and the complexity of the action.

## BIOLOGY DEPTH OF KNOWLEDGE

---

Note: “Knowledge” can refer both to content knowledge and knowledge of *scientific processes*. This meaning of knowledge is consistent with the National Science Education Standards (NSES), which terms “Science as Inquiry” as its first Content Standard.

---

Committees of Pennsylvania educators review each Keystone Exam item, not only to determine whether or not the item measures what it is intended to measure, but also to determine whether or not the item aligns with the cognitive level or depth of knowledge of the Assessment Anchor as defined by the Eligible Content. The information below provides a definition of the four depth-of-knowledge levels. The charts at the end of the section also provide a comparison between Bloom’s Taxonomy and Webb’s Depth of Knowledge for biology. Included are examples of verbs (i.e., the action). Using this information as well as the charts, Pennsylvania educators are asked to determine the depth of knowledge of each item and to verify that the depth of knowledge of each item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

### DEFINITIONS OF WEBB’S DEPTH OF KNOWLEDGE

*Level 1 (Recall)* requires the recall of information, such as a fact, definition, term, or a simple procedure, as well as performance of a simple science process or procedure. Level 1 only requires students to demonstrate a rote response, use a well-known formula, follow a set procedure (like a recipe), or perform a clearly defined series of steps. A “simple” procedure is well defined and typically involves only one step. Verbs such as “identify,” “recall,” “recognize,” “use,” “calculate,” and “measure” generally represent cognitive work at the recall level. Simple word problems that can be directly translated into and solved by a formula are considered Level 1. Verbs such as “describe” and “explain” could be classified at different depth-of-knowledge levels, depending on the complexity of what is to be described and explained.

A student answering a Level 1 item either knows the answer or does not: that is, the item does not need to be “figured out” or “solved.” In other words, if the knowledge necessary to answer an item automatically provides the answer to it, then the item is at Level 1. If the knowledge needed to answer the item is not automatically provided in the stem, the item is at least at Level 2. Some examples that represent but do not constitute all Level 1 performance are as follows:

- Recall or recognize a fact, term, or property.
- Represent in words or diagrams a scientific concept or relationship.
- Provide or recognize a standard scientific representation for simple phenomenon.
- Perform a routine procedure, such as measuring length.

*Level 2 (Skills and Concepts)* requires the engagement of some mental processing beyond recalling. The content knowledge or process involved is **more complex** than in Level 1. Items require students to make some decisions as to how to approach the question or problem. Keywords that generally distinguish a Level 2 item include “classify,” “organize,” “estimate,” “make observations,” “collect and display data,” and “compare data.” These actions imply **more than one step**. For example, to compare data requires first identifying characteristics of the objects or phenomena and then grouping or ordering the objects. Level 2 activities include making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts. Some action verbs, such as “explain,” “describe,” or “interpret,” could be classified at different depth-of-knowledge levels, depending on the complexity of the action. For example, interpreting information from a simple graph, which requires reading information from the graph, is a Level 2. An item that requires interpretation from a complex graph, such as making decisions regarding features of the graph that need to be considered and how information from the graph can be aggregated, is at Level 3. Some examples that represent but do not constitute all of Level 2 performance are as follows:

- Specify and explain the relationship between facts, terms, properties, or variables.
- Describe and explain examples and non-examples of science concepts.
- **Select a procedure according to specified criteria and perform it.**

- Formulate a routine problem, given data and conditions.
- Organize, represent, and interpret data.

*Level 3 (Strategic Thinking)* requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. The cognitive demands at Level 3 are complex and abstract. The complexity does not result only from the fact that there could be multiple answers, a possibility for both Levels 1 and 2, but because the multi-step task requires more demanding reasoning. In most instances, requiring students to explain their thinking is at Level 3; requiring a very simple explanation or a word or two should be at Level 2. An activity that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Experimental designs in Level 3 typically involve more than one dependent variable. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve non-routine problems. Some examples that represent but do not constitute all Level 3 performance are as follows:

- Identify research questions and design investigations for a scientific problem.
- Solve non-routine problems.
- Develop a scientific model for a complex situation.
- Form conclusions from experimental data.

*Level 4 (Extended Thinking)* requires high cognitive demands and complexity. Students are required to make several connections—relate ideas within the content area or among content areas—and have to select or devise one approach among many alternatives to solve the problem. Many on-demand assessment instruments will not include any assessment activities that could be classified as Level 4. However, standards, goals, and objectives can be stated in such a way as to expect students to perform extended thinking. “Develop generalizations of the results obtained and the strategies used and apply them to new problem situations,” is an example of a grade 8 objective that is a Level 4. Many, but not all, performance assessments and open-ended assessment activities requiring significant thought will be Level 4.

Level 4 involves complex reasoning, experimental design and planning, and probably will require an extended period of time either for the science investigation required by an objective, or for carrying out the multiple steps of an assessment item. However, the extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student is asked to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2 activity. However, if the student conducts a river study that requires taking into consideration a number of variables, this would be a Level 4. Some examples that represent but do not constitute all Level 4 performance are as follows:

- Based on data provided from a complex experiment that is novel to the student, deduct the fundamental relationship between several controlled variables.
- Conduct an investigation, from specifying a problem to designing and carrying out an experiment, to analyzing its data and forming conclusions.

---

Note: Multiple-choice and constructed-response items can be written at a depth-of-knowledge Level 4; however, to design an item in this format is difficult, as it would require research, investigation, and application, often over an extended period of time (e.g. performance-based tasks, portfolios, research studies/projects).

---

**(Webb, N. 1997, 1999, 2002, 2005, 2006)**

**Table A-3. Bloom's Taxonomy – Biology**

Categories (1956)	Definition	Examples of Action Words*
Knowledge	Student remembers or recalls appropriate previously learned information.	identify; recall; observe; recognize; use; calculate; measure; order
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	explain; interpret; describe; classify; identify; recognize; predict
Application	Student selects, transfers, and uses data and principles to complete a task or problem with minimum directions.	apply; classify; experiment; interpret; use; order; calculate
Analysis	Student distinguishes, classifies, and relates assumptions, hypotheses, evidence, or structure of a statement or question.	analyze; order; explain; classify; arrange; compare; contrast; infer; calculate; categorize; examine; experiment; question; test
Synthesis	Student originates, integrates, and combines ideas into a product, plan, or proposal that is new to him or her.	combine; arrange; rearrange; modify; invent; design; construct; organize; predict; infer; conclude; create; experiment and record data
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	evaluate; measure; explain; compare; summarize; predict; test decide; rate; conclude

**Table A-4. Webb's Depth of Knowledge – Biology**

Categories	Definition	Examples of Action Words*
Recall	Student recalls facts, information, procedures, or definitions.	identify; recall; observe; recognize; use; calculate; measure; order
Basic Application of Skill/ Concept	Student uses information, conceptual knowledge, and procedures.	explain; interpret; describe; classify; identify; order; recognize; predict; apply; use; calculate; organize; estimate; observe; collect; and display data
Strategic Thinking	Student uses reasoning and develops a plan or sequence of steps; process has some complexity.	analyze; order; explain; classify; arrange; compare; contrast; infer; interpret; calculate; categorize; examine; experiment; question; predict; evaluate; test
Extended Thinking	Student conducts an investigation, needs time to think and process multiple conditions of the problem or task. (The item/task generally requires several days or weeks to complete.)	combine; arrange; rearrange; propose; evaluate; modify; invent; design; construct; organize; predict; infer; conclude; evaluate; create; experiment and record data

\*Some action words (verbs) can be classified at different depth-of-knowledge levels depending on the context of the item and the complexity of the action.

## LITERATURE DEPTH OF KNOWLEDGE

---

Note: *The levels are based on Valencia and Wixson (2000, pp. 909–935).*

---

Committees of Pennsylvania educators review each Keystone Exam item, not only to determine whether or not the item measures what it is intended to measure, but also to determine whether or not the item aligns with the cognitive level or depth of knowledge of the Assessment Anchor as defined by the Eligible Content. The information below provides a definition of the four depth-of-knowledge levels. The charts at the end of the section also provide a comparison between Bloom’s Taxonomy and Webb’s Depth of Knowledge for literature. Included are examples of verbs (i.e., the action). Using this information as well as the charts, Pennsylvania educators are asked to determine the depth of knowledge of each item and to verify that the depth of knowledge of each item is in alignment with the depth of knowledge of the Assessment Anchor as defined by the Eligible Content.

### DEFINITIONS OF WEBB’S DEPTH OF KNOWLEDGE

*Level 1* requires students to receive or recite facts or to use simple skills or abilities. Oral reading that does not include analysis of the text, as well as basic comprehension of a text, is included. Items require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Some examples that represent but do not constitute all Level 1 performance are as follows:

- Support ideas by reference to verbatim or only slightly paraphrased details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage.

*Level 2* requires the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis of inference is required. Some important concepts are covered, but not in a complex way. Content curriculum standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item’s paraphrasing of both the question and the answer. Some examples that represent but do not constitute all Level 2 performance are as follows:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a selection.
- Identify and summarize the major events in a narrative.

*Level 3* requires deeper knowledge. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Content curriculum standards and items (Assessment Anchors as defined by the Eligible Content) at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students’ application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent but do not constitute all Level 3 performance are as follows:

- Explain or recognize how the author’s purpose affects the interpretation of a selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

*Level 4* requires higher-order thinking and deep knowledge. The content curriculum standard or item at this level will probably require an extended activity, with extended time provided for completing it. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require the application of significant conceptual understanding and higher-order thinking. Students take information from

at least one passage of a text and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts. Some examples that represent but do not constitute all Level 4 performance are as follows:

- Analyze and synthesize information from more than one source.
- Examine and explain alternative perspectives across a variety of sources.
- Describe and illustrate how common themes are found across texts from different cultures.

Note: Multiple-choice and constructed-response items can be written at a depth-of-knowledge Level 4; however, to design an item in this format is difficult, as it would require research, investigation, and application, often over an extended period of time (e.g. performance-based tasks, portfolios, research studies/projects).

(Webb, N. 2005; Valencia and Wixson, 2000)

**Table A-5. Bloom’s Taxonomy – Literature**

Categories (1956)	Definition	Examples of Action Words*
Knowledge	Student remembers or recalls appropriate previously learned information.	define; identify; name; recall; recognize; select; tell
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	describe; distinguish; explain; identify; indicate; interpret; locate; recognize; restate; summarize
Application	Student selects, transfers, and uses data and principles to complete a task or problem with minimum directions.	apply; choose; demonstrate; determine; interpret; inform; select; show; use
Analysis	Student distinguishes, classifies, and relates assumptions, hypotheses, evidence, or structure of a statement or question.	analyze; characterize; compare; contrast; discriminate; distinguish; explain; infer
Synthesis	Student originates, integrates, and combines ideas into a product, plan, or proposal that is new to him or her.	compose; create; develop; formulate; generalize; organize
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	assess; conclude; convince; defend; evaluate; explain; justify; predict; prove; support

**Table A-6. Webb’s Depth of Knowledge – Literature**

Categories	Definition	Examples of Action Words*
Recall	Student recalls facts, information, procedures, or definitions.	define; identify; locate; name; recall; recognize; sequence; tell
Basic Application of Skill/Concept	Student uses information, conceptual knowledge, and procedures.	apply; compare; comprehend; identify; describe; determine; infer; interpret; predict; summarize; use
Strategic Thinking	Student uses reasoning and develops a plan or sequence of steps; process has some complexity.	analyze; cite evidence; compare; contrast; draw conclusions; explain; generalize; infer; interpret; evaluate; recognize; summarize; support
Extended Thinking	Student conducts an investigation, needs time to think and process multiple conditions of the problem or task. (The item/task generally requires several days or weeks to complete.)	describe and illustrate; evaluate; examine and explain; analyze; synthesize

\*Some action words (verbs) can be classified at different depth-of-knowledge levels depending on the context of the item and the complexity of the action.

## APPENDIX B: GENERAL SCORING GUIDELINES

---

### ALGEBRA I

#### 4 Points

- The response demonstrates a *thorough* understanding of the mathematical concepts and procedures required by the task.
- The response provides correct answer(s) with clear and complete mathematical procedures shown and a correct explanation, as required by the task. Response may contain a minor “blemish” or omission in work or explanation that does not detract from demonstrating a thorough understanding.

#### 3 Points

- The response demonstrates a *general* understanding of the mathematical concepts and procedures required by the task.
- The response and explanation (as required by the task) are mostly complete and correct. The response may have minor errors or omissions that do not detract from demonstrating a general understanding.

#### 2 Points

- The response demonstrates a *partial* understanding of the mathematical concepts and procedures required by the task.
- The response is somewhat correct with partial understanding of the required mathematical concepts and/or procedures demonstrated and/or explained. The response may contain some work that is incomplete or unclear.

#### 1 Point

- The response demonstrates a *minimal* understanding of the mathematical concepts and procedures required by the task.

#### 0 Points

- The response has no correct answer and *insufficient* evidence to demonstrate any understanding of the mathematical concepts and procedures required by the task.

Special Categories within zero reported separately:

<b>Blank</b>	Blank, entirely erased, entirely crossed out, or consists entirely of whitespace
<b>Refusal</b>	Refusal to respond to the task
<b>Off Task</b>	Makes no reference to the item but is not an intentional refusal
<b>Foreign Language</b>	Written entirely in a language other than English
<b>Illegible</b>	Illegible or incoherent

COPYRIGHT© PA DEPARTMENT OF EDUCATION. DO NOT DUPLICATE.

## BIOLOGY

### 3 Points

- The response demonstrates a *thorough* understanding of the scientific content, concepts, and/or procedures required by the task(s).
- The response provides a clear, complete, and correct response as required by the task(s). The response may contain a minor blemish or omission in work or explanation that does not detract from demonstrating a thorough understanding.

### 2 Points

- The response demonstrates a *partial* understanding of the scientific content, concepts, and/or procedures required by the task(s).
- The response is somewhat correct with partial understanding of the required scientific content, concepts, and/or procedures demonstrated and/or explained. The response may contain some work that is incomplete or unclear.

### 1 Point

- The response demonstrates a *minimal* understanding of the scientific content, concepts, and/or procedures required by the task(s).
- The response is somewhat correct with minimal understanding of the required scientific content, concepts, and/or procedures demonstrated and/or explained. The response may contain some work that is incomplete or unclear.

### 0 Points

- The response provides *insufficient* evidence to demonstrate any understanding of the scientific content, concepts, and/or procedures as required by the task(s).
- The response may show only information copied or rephrased from the question or insufficient correct information to receive a score of 1.

Special Categories within zero reported separately:

**Blank** .....Blank, entirely erased, entirely crossed out, or consists entirely of whitespace

**Refusal** .....Refusal to respond to the task

**Off Task** .....Makes no reference to the item but is not an intentional refusal

**Foreign Language** .....Written entirely in a language other than English

**Illegible** .....Illegible or incoherent

COPYRIGHT© PA DEPARTMENT OF EDUCATION. DO NOT DUPLICATE.



## LITERATURE

### 3 Points

- The response provides a clear, complete, and accurate answer to the task.
- The response provides relevant and specific information from the passage.

### 2 Points

- The response provides a partial answer to the task.
- The response provides limited information from the passage and may include inaccuracies.

### 1 Point

- The response provides a minimal answer to the task.
- The response provides little or no information from the passage and may include inaccuracies.

OR

- The response relates minimally to the task.

### 0 Points

- The response is totally incorrect or irrelevant or contains insufficient information to demonstrate comprehension.

Special Categories within zero reported separately:

**Blank** .....Blank, entirely erased, entirely crossed out, or consists entirely of white space

**Refusal** .....Refusal to respond to the task

**Off Task** .....Makes no reference to the item but is not an intentional refusal

**Foreign Language** .....Written entirely in a language other than English

**Illegible** .....Illegible or incoherent

COPYRIGHT© PA DEPARTMENT OF EDUCATION. DO NOT DUPLICATE.

## APPENDIX C: ITEM AND TEST DEVELOPMENT PROCESS FOR THE KEYSTONE EXAMS

Table C–1. Item and Test Development Process for the Keystone Exams

Step	Description
1. Review Guiding Documentation	Each year item and test development specialists meet internally to review all guiding documentation related to the Keystone Exams. Documentation reviewed includes the test design blueprints, the Keystone Assessment Anchors and Eligible Content, the test item specifications, the test style specifications (style guide), and all test content descriptions.
2. Meet with PDE to Confirm Understanding of Program	The goal of the meeting each year is to ensure that item and test development teams have a clear understanding of PDE’s vision for test development. A successful development cycle requires a clear understanding of Pennsylvania’s content-area test specifications and of any unique interpretations of the Keystone Assessment Anchors (if any).
3. Create Preliminary Test Item Development Plan	Item and test development specialists generate a preliminary development plan which includes an overview of the program, the internal and external (PDE) review and approval processes, a projected schedule for development of test items—including the number of test items to be developed for review by PDE and subsequent review by the committees of Pennsylvania educators. Item and test development specialists also generate strategies for securing passages and developing passage-based items, etc.
4. Meet with PDE to Finalize Test Item Development Plan	Over the course of the meeting, item and test development specialists verify all steps in the development process including timelines and schedules for test item/test development.
5. Analyze Item Bank	Existing test items in the current Keystone Exams Item Bank are reviewed for technical psychometric quality as well as for their match to the Assessment Anchors. During this phase, test development specialists also make a tally of the test items by Assessment Anchor—including test development specialists’ best thinking regarding the number of usable test items in the existing item bank. A tally is also made of the number of usable passages, as well as other stimulus prompts in the bank, including science scenarios.
6. Refine Test Item Development Plan to Include Writers and Subcontractors	Item and test development specialists identify the writers who will write the test items (test development specialists or other professional item writers, subcontractors, etc.), the estimated number of writers needed, the qualifications of writers, and the approximate number of test items to be submitted by each source.
7. Train Item Writers	Item and test development specialists train item writers, as needed. Item writers who have written for the Keystone Exams in the past receive updated information, as needed.
8. Write and Review Items	Test items are written by item writers after training is complete, and feedback is provided by the item and test development specialists to item writers on a regular basis. As test items are written, they are reviewed and edited in a series of internal reviews. Item and test development specialists review and edit items to include, but not limited to, the following: match to Assessment Anchor/Eligible Content, relevance to purpose, accuracy of content, item difficulty, interest level, depth of knowledge and cognitive complexity, adherence to the principles of Universal Design, and freedom from issues of bias/fairness/sensitivity. At the same time, the process of procuring permissions also begins, including securing permissions for passages, art, etc.
9. Enter Test Items into Database	Upon acceptance from item writers, test items are entered into the item management system, IDEAS (Item Development and Educational Assessment System). Item data stored in the system database includes, but is not limited to, the following: readability, cognitive level, estimated level of difficulty, alignment to assessment anchors, and correlation to stimulus passages.
10. Prepare Item Set for Sample Item Review by PDE	Item and test development specialists prepare a subset of the items for review by PDE.
11. PDE Conducts Sample Item Review	After a subset of the items is submitted to PDE for review, PDE reviews the items and provides feedback to item and test development teams via a conference call. Items are revised per PDE feedback.

**Table C–1 (continued). Item and Test Development Process for the Keystone Exams**

Step	Description
12. Continue to Write and Review Items	The remaining items are written, and feedback is provided by the item and test development specialists to item writers on a regular basis. Items are entered into the item management system, IDEAS (Item Development and Educational Assessment System) (See step 8 and step 9).
13. Review Items Prior to Test Item Review and Validation Sessions	Prior to New Item Content Review, all items are submitted to PDE for review. Item and test development specialists incorporate all PDE feedback, and PDE-requested edits to items are made.
14. Prepare for Test Item Review Sessions (the New Item Content Review and the Bias, Fairness, and Sensitivity Review)	Item and test development specialists prepare all items and stimulus passages for review by the New Item Content Review Committee (consisting of Pennsylvania educators) and by the separate Bias, Fairness, and Sensitivity Committee (consisting of a panel of experts). Item and test development specialists also prepare training materials needed for training committee members to review items for content or for bias, fairness, and sensitivity issues. All training materials and other ancillary materials (e.g. agendas, presentations, etc.) are also developed and then submitted to PDE for review and approval. Invitations are also sent to Pennsylvania educators and national experts from PDE-approved committee lists.
15. Conduct Test Item Review Sessions (the New Item Content Review and the Bias, Fairness, and Sensitivity Review)	Committees of Pennsylvania educators and national experts review items in two meetings: one addressing item content and quality, the other addressing bias, fairness, and sensitivity. PDE, with support from item and test development specialists, presents training on how to review new test items for content considerations or bias/fairness/sensitivity issues. At the New Item Content Review, suggested edits to test items are made and/or replacement test items are written during the actual item review so that both the committee and the PDE are able to observe changes to the test items and approve the test items during the committee review process. At the Bias, Fairness, and Sensitivity Review, experts in bias, fairness, and sensitivity review all test items and passages and come to a consensus about any issues that are noted. At both meetings the results are carefully documented.
16. Conduct Item Review Resolution and Cleanup	Following the conclusion of the New Item Content Review Committee meetings, PDE re-examines the consensus changes suggested by the committee members during the New Item Content Review Committee meetings. DRC item and test development specialists then record all of PDE’s follow-up decisions and changes. During this cleanup process, PDE either accepts the changes as requested by the committee, or PDE rejects the decision of the committee. If a committee decision is rejected, PDE provides an alternate decision for DRC to implement. During this cleanup process, PDE also interprets the report from the Bias, Fairness, and Sensitivity Committee meetings and subsequently applies changes to test items and passages. DRC item and test development specialists then apply the changes to the test items and passages per PDE’s decisions.
17. Submit Field-Test Items for Final Sign-Off	PDE-approved changes are applied to the items, non-permissioned passages, prompts, etc. (Changes reflect PDE’s arbitration of the committee decisions.) Once all revisions to the items, non-permissioned passage text, and/or the art used by test items and passages are completed, the test items are submitted to PDE for final review and sign-off. (Changes requested to permissioned passages are sought from the publisher of record, and, if approved by the copyright holders, changes are implemented.) [PDE’s approval process for field-test items generally occurs simultaneously with PDE’s approval of the core test forms. See step 25.]
	To follow the path for new field-test items, skip to step 22, or to follow the chronological test development path, continue with step 18.
18. Review Results of the Field Test	Following the administration of a field-test form and the subsequent rangefinding and field-test scoring processes for field-test items, performance data for all field-test items are analyzed by DRC psychometricians and test development specialists. Test item performance data that meet certain triggering criteria are flagged for additional reviews by test development specialists. Flagged field-test items with extreme performance data are considered psychometrically unusable and are removed from future operational consideration. Field-test items with marginal performance data are prepared for the Field-Test Item Data Review meeting.

**Table C–1 (continued). Item and Test Development Process for the Keystone Exams**

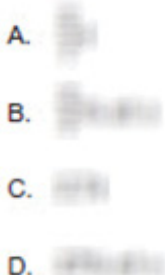




Step	Description
19. Prepare for Field-Test Item Data Review	Test development specialists prepare all items and stimulus passages for review by the Field-Test Item Data Review Committee (which consists of Pennsylvania educators). Psychometricians also prepare training materials needed for training committee members to review items for their performance. All training materials and other ancillary materials (e.g. agendas, presentations, etc.) are submitted to PDE for review and approval. Invitations are also sent to Pennsylvania educators from PDE-approved committee lists.
20. Conduct Field-Test Item Data Review	Committees of Pennsylvania educators review the performance data of flagged field-test items. Psychometricians present training on how to review field-test items based on their performance data. At the Item Data Review, committee members examine the performance of the items and determine whether the field-test item is technically sound and appropriate for use on an operational Keystone Exams test. Since test items cannot be modified at the Field-Test Item Data Review, the committee can either accept an item as is or the committee can reject the item.
21. Conduct Field-Test Item Data Review Reconciliation	Following the conclusion of the Field-Test Item Data Review Committee meetings, PDE re-examines the consensus decisions (accept or reject) suggested by the committee members during the Field-Test Item Data Review Committee meetings. Test development specialists record all of PDE’s follow-up decisions and changes. During this cleanup process, PDE either accepts the decisions of the data review committee, or PDE rejects the decisions of the data review committee. If a committee decision is not accepted, PDE provides an alternate decision for test development specialists to implement. All PDE-approved changes to the test items status (accepted or rejected) are incorporated into the Item Development and Educational Assessment System, IDEAS.
22. Select Items to Fill Core, Field-Test, and Equating Block Positions in Core and Field-Test Forms	After the PDE-approved changes to the new field-test items is completed AND the results of the prior field test have been finalized following data review, test development specialists collaborate with psychometricians to follow the Test Design Blueprints and build requirements to make the initial selection of items for core and field-test positions for all test forms. In later administrations, core-to-core linking items will also be selected during this step.
23. Review Core and Equating Block Selections	After test content and psychometric requirements have been achieved for core, the core items are provided to PDE for review and approval. Any changes to the content of the core requested by PDE are balanced with psychometric requirements until all core positions are approved by PDE, test development specialists, and psychometricians.
24. Construct Test Forms	Items, passages, and test components are assembled into forms using the form construction and typesetting function of DRC’s Item Development and Educational Assessment System, IDEAS. Forms are reviewed internally for style and formatting requirements.
25. Review Typeset Forms	After forms are constructed in IDEAS, draft hard copies of the forms are produced and presented to PDE for review and approval. Any changes to the content of the core requested by PDE are balanced with psychometric requirements until all core positions are approved by PDE, test development specialists, and psychometricians. PDE also re-reviews all field-test items appearing in the test forms. DRC applies changes to the field-test items as required.
26. Print Test Forms	Following PDE’s approval of the test forms, DRC completes a series of final proofing of all test forms. Final forms (along with ancillary materials) are then approved for printing.
27. Assemble Documentation of Test Materials	Metadata for each test item and form is documented and proofed, including: grade, form, session/section, item sequence, reporting category, Assessment Anchor, Descriptor, Eligible Content, number of points, item type, number of answer options, item usage, stimulus ID, etc.
28. Prepare Online Forms	Following approval of the print forms, all online forms are prepared. Forms are rendered in form sets, and items and forms are compared for continuity with the print forms as well as to ensure that all tools and features are functioning as expected.
	To follow the path for new field-test items, return to step 18.

# APPENDIX D: ITEM AND DATA REVIEW CARD EXAMPLES

## ITEM REVIEW CARD EXAMPLE

<p>Standard: Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.</p>	<p><b>PA Keystone Item Card</b></p>
<p>1. <i>[Blurred text]</i></p> <p>A. <i>[Blurred text]</i></p> <p>B. <i>[Blurred text]</i></p> <p>C. <i>[Blurred text]</i></p> <p>D. <i>[Blurred text]</i></p>	
	Item ID
	<i>[Blurred]</i>
	Content Area
	Science
	Course
	Biology
	Scenario ID
	<i>[Blurred]</i>
	Scenario Title
	<i>[Blurred]</i>
	Grade
	HS
	KAACS Standards
	BIO.A.2.3.2
	Item Type
	Multiple Choice
	Points
	1
	Depth of Knowledge
	2
	Est Difficulty
	Medium
	Key
	<i>[Blurred]</i>

# DATA REVIEW CARD EXAMPLE

<p>Standard: Compare and/or order any real numbers (rational and irrational may be mixed).</p>	<p><b>PA Keystone Data Card</b></p>
<p>1. </p>	<p>Item ID</p> <p>Content Area</p> <p>Mathematics</p> <p>Course</p> <p>Algebra I</p> <p>Passage ID</p> <p>Passage Title</p> <p>Grade</p> <p>HS</p> <p>Standards</p> <p>KAACS: A1.1.1.1.1</p> <p>Item Type</p> <p>Multiple Choice</p> <p>Points</p> <p>1</p> <p>Calculator</p> <p>Yes</p> <p>Depth of Knowledge</p> <p>2</p> <p>Est Difficulty</p> <p>High</p> <p>Key</p> <p>Focus</p>
<p>A: </p>	
<p>B: </p>	
<p>C: </p>	
<p>D: </p>	

PA3 - Data Card continued

Administration

Form Name	Use Function	Rptg Flag	Seq	Period	Year	Session	Calc	Mode/Ext	Grade
						1	Yes		HS

Traditional Statistics

N	P-Val	Mean	Item Total Corr
	0.34		0.10

Fit Statistics

Outfit t	Infit t	Outfit MnSq	Infit MnSq	Chi-sq	Deg Free	Item Fit	Fit
9.9	9.9	1.28	1.18				

IRT Statistics

Label	Final	Final S.E.	Preliminary	Preliminary S.E.
Location	1.39	0.02		

Distractor/Step Specific

Label	Proportion	Corr	Avg Meas	Threshold
A*	0.34	0.10		
B	0.24	0.11		
C	0.25	-0.22		
D	0.17	0.01		
MULTS	0.00			
OMITS	0.00			

DIF Analysis

Category	Bias Code	Num Value	N - Ref	N - Focal
MALEFEMALE	A-	-0.13	4709	4550
PAPERONLINE	A+	0.15	8242	1029
WHITEBLACK	A-	-0.23	6812	1245
WHITEHISPANIC	A-	-0.16	6812	726





The purpose of this form is to provide guidelines to the item review process in terms of item characteristics that are essential in building a fair and balanced assessment. Use these guidelines in conjunction with the Item Rating Sheet when recording your feedback on individual items.

Content Alignment		Options
Standards, Anchors, Eligible Content	Does the content of the item align with the Standard/Anchor/Eligible Content? Each item was written to assess a particular Standard/Anchor/Eligible Content statement which is indicated on the individual Item Card. Consider the degree to which the item is, in fact, aligned with the indicated eligible content. In making this judgment, it is important to consider whether the <b>content</b> is aligned (e.g., do the eligible content and the item both deal with fractions) and whether the required <b>performance</b> is aligned (e.g., if the eligible content calls for a comparison to be made, is this reflected in the item).	<b>HIGHER</b> —Aligns to the higher level of the EC <b>LOWER</b> —Aligns to the lower level of the EC <b>NONE</b> —No alignment with EC

Rigor Level Alignment		Options
Grade	Is the item grade-level appropriate? Is the content consistent with the experiences of a student at the grade level assessed? Is the challenge level appropriate for the grade?	<b>ABOVE</b> Grade Level <b>AT</b> Grade Level <b>BELOW</b> Grade Level
Difficulty	Do you agree with the item’s difficulty rating? Item Difficulty is indicated as Low, Medium, and High. Is your rating in agreement with the difficulty rating on the Item Form?	<b>HIGH</b> <b>MEDIUM</b> <b>LOW</b>
Depth of Knowledge	Depth of Knowledge is based on the alignment work of Norman Webb. Rate each item based on the cognitive demand, using the following levels: <ol style="list-style-type: none"> <li>1. Recall – <b>Recall</b> of a fact, information, or procedure.</li> <li>2. Basic Application of Skill or Concept – <b>Use</b> of information, conceptual knowledge, procedures, two or more steps, etc.</li> <li>3. Strategic Thinking – Requires reasoning, developing a plan or sequence of steps; has some complexity; more than one possible answer.</li> <li>4. Extended Thinking – Requires an investigation, time to think and process multiple conditions of the problem or task, and more than 10 minutes to do non-routine manipulations. (This level is generally not assessed in on-demand assessments.)</li> </ol>	<b>4</b> = Extended Thinking <b>3</b> = Strategic Thinking <b>2</b> = Basic Application <b>1</b> = Recall

Rigor Level Alignment		Options
Source of Challenge	Is the source of challenge appropriately targeted to the content? The hardest part of the item (i.e., source of challenge) should be the content that is targeted. For example, in mathematics, the mathematics should be the major source of challenge rather than the wording or graphic. Students should not give an incorrect answer to a mathematics item because the reading level is too high or a graphic is flawed. Conversely, students should not give correct answers for reasons such as prior knowledge that make the answer to the question obvious (e.g., if the question asks which country has the largest population and students are to read a graph that includes China, there is no need to read the graph to answer the question).	Y = Yes N = No

Technical Design		Options
Correct Answer	Is there one clear, correct answer option? There should be no other answer that “could” be correct. CAUTION: This does not mean that “good” distractors are unfair.	Y = Yes N = No
Distractors	Are distractors fair and appropriate? Distractors that are appropriate offer students reasonable choices that can be arrived at by making common errors. There should be no distractors that make no sense at all. It should be possible to examine each option and to reason how a student with some deficiency in knowledge or skill could choose it. The distractors should be formatted according to acceptable standards of test construction (e.g., a phrase that is common to each distractor should be in the stem). Are the graphics clear and accurate?	Y = Yes N = No N/A = OE items do not have distractors
Graphics		Y = Yes N = No N/A = No graphic

Universal Design		Options
Language Demand	Is language clear, well-formatted, and precise? Does the item use correct terminology for the content area? In order for all students to enter into the questions of the assessment, they must be able to understand them. If the items are formatted poorly, use unnecessarily complex words or phrases, or use figures or layouts that are difficult to understand, some students will give incorrect answers due to these factors rather than the content that is being assessed.	Y = Yes N = No
Bias	Is the item free of bias? All students will not be able to enter into the assessment if bias considerations are not resolved. Does the item contain clear bias problems? A <i>thorough, independent bias review</i> (separate from this meeting) <i>will be completed for all items</i> .	Y = Yes N = No

Status		Options
Acceptance Status	This is an overall judgment about the item. Based on the consensus of the committee, indicate whether the item was approved without revision to the content of the item or whether the item was accepted by the committee after revision of the content of the item. If there is a dissenting view (opposed to the committee consensus), record a	—Approved as is —Accepted with suggested revisions

## NOTES:

- If you leave a box blank on the Item Rating Sheet, it will be recorded to indicate that you did not have any specific feedback for that item or issue.
- If you object to the consensus of the committee, please note this on the item rating sheet and then record a brief explanation of the dissenting view on the back of the Item Rating Sheet.
- **Do NOT remove any items from the item binder at any time.**
- You must sign your item rating sheet.

# APPENDIX F: TALLY SHEETS

ALGEBRA I WINTER 2022/2023													
Keystone Exam										Algebra I			
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items				Points				
					Number of Core Items				Core Points				
					MC	SCR	ECR	Total	MC	SCR	ECR	Total	
A1.1: Operations and Linear Equations & Inequalities	1			<b>Operations with Real Numbers and Expressions</b>		1		1			4		4
	1	1	1	Compare and/or order any real numbers.	1			1	1				1
	1	1	2	Simplify square roots.									
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	1			1	1				1
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1			1	1				1
	1	4	1	Use estimation to solve problems.									
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1			1	1				1
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1			1	1				1
	1	5	3	Simplify/reduce a rational algebraic expression.	1			1	1				1
	Total For Assessment Anchor A1.1.1					6	1		7	6	4		10
	2			<b>Linear Equations</b>			1	1				4	4
	2	1	1	Write, solve, and/or apply a linear equation.	1			1	1				1
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	2			2	2				2
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1			1	1				1
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1			1	1				1
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1			1	1				1
	Total For Assessment Anchor A1.1.2					6		1	7	6		4	10
	3			<b>Linear Inequalities</b>		1		1		4			4
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line.	1			1	1				1
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	1			1	1				1
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1			1	1				1
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	2			2	2				2
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1			1	1				1
	Total For Assessment Anchor A1.1.3					6	1		7	6	4		10
	Total For Reporting Category A1.1					18	2	1	21	18	8	4	30

**Keystone Exam**

**Algebra I**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items				Points				
					Number of Core Items				Core Points				
					MC	SCR	ECR	Total	MC	SCR	ECR	Total	
A1.2: Linear Functions & Data Organizations	1			<b>Functions</b>		1		1		4		4	
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	1			1	1			1	
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1			1	1			1	
	1	1	3	Identify the domain or range of a relation.	1			1	1			1	
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	2			2	2			2	
	1	2	2	Translate from one representation of a linear function to another.	1			1	1			1	
	<b>Total For Assessment Anchor A1.2.1</b>					<b>6</b>	<b>1</b>		<b>7</b>	<b>6</b>	<b>4</b>		<b>10</b>
	2			<b>Coordinate Geometry</b>			1	1			4		4
	2	1	1	Identify, describe, and/or use constant rates of change.	1			1	1				1
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1			1	1				1
	2	1	3	Write or identify a linear equation when given...	1			1	1				1
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1			1	1				1
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	2			2	2				2
	<b>Total For Assessment Anchor A1.2.2</b>					<b>6</b>		<b>1</b>	<b>7</b>	<b>6</b>		<b>4</b>	<b>10</b>
	3			<b>Data Analysis</b>		1		1		4			4
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	1			1	1				1
	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	2			2	2				2
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1			1	1				1
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	1			1	1				1
	3	3	1	Find probabilities for compound events.	1			1	1				1
	<b>Total For Assessment Anchor A1.2.3</b>					<b>6</b>	<b>1</b>		<b>7</b>	<b>6</b>	<b>4</b>		<b>10</b>
<b>Total For Reporting Category A1.2</b>					<b>18</b>	<b>2</b>	<b>1</b>	<b>21</b>	<b>18</b>	<b>8</b>	<b>4</b>	<b>30</b>	

ALGEBRA I SPRING 2023

Keystone Exam

Algebra I

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items				Points				
					Number of Core Items				Core Points				
					MC	SCR	ECR	Total	MC	SCR	ECR	Total	
A1.1: Operations and Linear Equations & Inequalities	1			<b>Operations with Real Numbers and Expressions</b>		1		1		4		4	
	1	1	1	Compare and/or order any real numbers.	1			1	1			1	
	1	1	2	Simplify square roots.	1			1	1			1	
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.	1			1	1			1	
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1			1	1			1	
	1	4	1	Use estimation to solve problems.									
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1			1	1			1	
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.									
	1	5	3	Simplify/reduce a rational algebraic expression.	1			1	1			1	
	Total For Assessment Anchor A1.1.1					6	1		7	6	4		10
	2			<b>Linear Equations</b>		1		1		4		4	
	2	1	1	Write, solve, and/or apply a linear equation.	1			1	1			1	
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	2			2	2			2	
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1			1	1			1	
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	Total For Assessment Anchor A1.1.2					6	1		7	6	4		10
	3			<b>Linear Inequalities</b>			1	1			4	4	
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	1			1	1			1	
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	2			2	2			2	
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1			1	1			1	
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	Total For Assessment Anchor A1.1.3					6		1	7	6		4	10
	Total For Reporting Category A1.1					18	2	1	21	18	8	4	30

**Keystone Exam**

**Algebra I**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items				Points				
					Number of Core Items				Core Points				
					MC	SCR	ECR	Total	MC	SCR	ECR	Total	
A1.2: Linear Functions & Data Organizations	1			<b>Functions</b>			1	1			4	4	
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	1			1	1			1	
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1			1	1			1	
	1	1	3	Identify the domain or range of a relation.	1			1	1			1	
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	2			2	2			2	
	1	2	2	Translate from one representation of a linear function to another.	1			1	1			1	
	<b>Total For Assessment Anchor A1.2.1</b>					<b>6</b>		<b>1</b>	<b>7</b>	<b>6</b>		<b>4</b>	<b>10</b>
	2			<b>Coordinate Geometry</b>			1	1			4	4	
	2	1	1	Identify, describe, and/or use constant rates of change.	1			1	1			1	
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1			1	1			1	
	2	1	3	Write or identify a linear equation when given...	1			1	1			1	
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1			1	1			1	
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	2			2	2			2	
	<b>Total For Assessment Anchor A1.2.2</b>					<b>6</b>		<b>1</b>	<b>7</b>	<b>6</b>		<b>4</b>	<b>10</b>
	3			<b>Data Analysis</b>		1		1		4		4	
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	1			1	1			1	
	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	2			2	2			2	
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1			1	1			1	
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	1			1	1			1	
	3	3	1	Find probabilities for compound events.	1			1	1			1	
	<b>Total For Assessment Anchor A1.2.3</b>					<b>6</b>	<b>1</b>		<b>7</b>	<b>6</b>	<b>4</b>		<b>10</b>
	<b>Total For Reporting Category A1.2</b>					<b>18</b>	<b>1</b>	<b>2</b>	<b>21</b>	<b>18</b>	<b>4</b>	<b>8</b>	<b>30</b>

## Keystone Exam

## Algebra I

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items				Points				
					Number of Core Items				Core Points				
					MC	SCR	ECR	Total	MC	SCR	ECR	Total	
A1.1: Operations and Linear Equations & Inequalities	1			<b>Operations with Real Numbers and Expressions</b>		1		1		4		4	
	1	1	1	Compare and/or order any real numbers.	1			1	1			1	
	1	1	2	Simplify square roots.	1			1	1			1	
	1	2	1	Find the Greatest Common Factor (GCF) and/or the Least Common Multiple (LCM) for sets of monomials.									
	1	3	1	Simplify/evaluate expressions involving properties/laws of exponents, roots, and/or absolute values to solve problems.	1			1	1			1	
	1	4	1	Use estimation to solve problems.	1			1	1			1	
	1	5	1	Add, subtract, and/or multiply polynomial expressions (express answers in simplest form).	1			1	1			1	
	1	5	2	Factor algebraic expressions, including difference of squares and trinomials.	1			1	1			1	
	1	5	3	Simplify/reduce a rational algebraic expression.									
	Total For Assessment Anchor A1.1.1					6	1		7	6	4		10
	2			<b>Linear Equations</b>		1		1		4		4	
	2	1	1	Write, solve, and/or apply a linear equation.	2			2	2			2	
	2	1	2	Use and/or identify an algebraic property to justify any step in an equation-solving process.	1			1	1			1	
	2	1	3	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	2	2	1	Write and/or solve a system of linear equations (including problem situations) using graphing, substitution, and/or elimination.	1			1	1			1	
	2	2	2	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	Total For Assessment Anchor A1.1.2					6	1		7	6	4		10
	3			<b>Linear Inequalities</b>			1	1			4	4	
	3	1	1	Write or solve compound inequalities and/or graph their solution sets on a number line .	2			2	2			2	
	3	1	2	Identify or graph the solution set to a linear inequality on a number line.	1			1	1			1	
	3	1	3	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	3	2	1	Write and/or solve a system of linear inequalities using graphing.	1			1	1			1	
	3	2	2	Interpret solutions to problems in the context of the problem situation.	1			1	1			1	
	Total For Assessment Anchor A1.1.3					6		1	7	6		4	10
Total For Reporting Category A1.1					18	2	1	21	18	8	4	30	



Keystone Exam					Algebra I								
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items				Points				
					Number of Core Items				Core Points				
					MC	SCR	ECR	Total	MC	SCR	ECR	Total	
A1.2: Linear Functions & Data Organizations	1			<b>Functions</b>		1		1		4		4	
	1	1	1	Analyze a set of data for the existence of a pattern and represent the pattern algebraically and/or graphically.	1			1	1			1	
	1	1	2	Determine whether a relation is a function, given a set of points or a graph.	1			1	1			1	
	1	1	3	Identify the domain or range of a relation.	2			2	2			2	
	1	2	1	Create, interpret, and/or use the equation, graph, or table of a linear function.	1			1	1			1	
	1	2	2	Translate from one representation of a linear function to another.	1			1	1			1	
	<b>Total For Assessment Anchor A1.2.1</b>					<b>6</b>	<b>1</b>		<b>7</b>	<b>6</b>	<b>4</b>		<b>10</b>
	2			<b>Coordinate Geometry</b>			1	1			4		4
	2	1	1	Identify, describe, and/or use constant rates of change.	1			1	1				1
	2	1	2	Apply the concept of linear rate of change (slope) to solve problems.	1			1	1				1
	2	1	3	Write or identify a linear equation when given...	2			2	2				2
	2	1	4	Determine the slope and/or y-intercept represented by a linear equation or graph.	1			1	1				1
	2	2	1	Draw, identify, find, and/or write an equation for a line of best fit for a scatter plot.	1			1	1				1
	<b>Total For Assessment Anchor A1.2.2</b>					<b>6</b>		<b>1</b>	<b>7</b>	<b>6</b>		<b>4</b>	<b>10</b>
	3			<b>Data Analysis</b>		1		1		4			4
	3	1	1	Calculate and/or interpret the range, quartiles, and interquartile range of data.	2			2	2				2
	3	2	1	Estimate or calculate to make predictions based on a circle, line, bar graph, measures of central tendency, or other representations.	1			1	1				1
	3	2	2	Analyze data, make predictions, and/or answer questions based on displayed data.	1			1	1				1
	3	2	3	Make predictions using the equations or graphs of best-fit lines of scatter plots.	1			1	1				1
	3	3	1	Find probabilities for compound events.	1			1	1				1
	<b>Total For Assessment Anchor A1.2.3</b>					<b>6</b>	<b>1</b>		<b>7</b>	<b>6</b>	<b>4</b>		<b>10</b>
	<b>Total For Reporting Category A1.2</b>					<b>18</b>	<b>2</b>	<b>1</b>	<b>21</b>	<b>18</b>	<b>8</b>	<b>4</b>	<b>30</b>

Keystone Exam				Biology							
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
BIO.A: Basic Biological Principles	1			<b>Basic Biological Principles</b>							
	1	1	1	Describe the characteristics of life shared by all prokaryotic and eukaryotic organisms.	1	1	2	1	3	4	
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	2		2	2		2	
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.	2		2	2		2	
	Total For Assessment Anchor BIO.A.1					5	1	6	5	3	8
	2			<b>The Chemical Basis for Life</b>							
	2	1	1	Describe the unique properties of water and how these properties support life on Earth.	2		2	2		2	
	2	2	1	Explain how carbon is uniquely suited to form biological macromolecules.	1		1	1		1	
	2	2	2	Describe how biological macromolecules form from monomers.	1		1	1		1	
	2	2	3	Compare the structure and function of carbohydrates, lipids, proteins, and nucleic acids in organisms.	1		1	1		1	
	2	3	1	Describe the role of an enzyme as a catalyst in regulating a specific biochemical reaction.	2		2	2		2	
	2	3	2	Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.	2		2	2		2	
	Total For Assessment Anchor BIO.A.2					9		9	9		9
	3			<b>Bioenergetics</b>							
	3	1	1	Describe the fundamental roles of plastids (e.g., chloroplasts) and mitochondria in energy transformations.	1	1	2	1	3	4	
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	2		2	2		2	
	3	2	2	Describe the role of ATP in biochemical reactions.	1		1	1		1	
	Total For Assessment Anchor BIO.A.3					4	1	5	4	3	7
	4			<b>Homeostasis and Transport</b>							
	4	1	1	Describe how the structure of the plasma membrane allows it to function as a regulatory structure and/or protective barrier for a cell.	1	1	2	1	3	4	
	4	1	2	Compare the mechanisms that transport materials across the plasma membrane.	1		1	1		1	
	4	1	3	Describe how membrane-bound cellular organelles.	2		2	2		2	
	4	2	1	Explain how organisms maintain homeostasis.	2		2	2		2	
	Total For Assessment Anchor BIO.A.4					6	1	7	6	3	9
	Total For Reporting Category BIO.A					24	3	27	24	9	33

**Keystone Exam**

**Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
BIO.B: Cell Growth and Reproduction	1			<b>Cell Growth and Reproduction</b>							
	1	1	1	Describe the events that occur during the cell cycle: interphase, nuclear division.	1		1	1		1	
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	2		2	2		2	
	1	2	1	Describe how the process of DNA replication results in the transmission and/or conservation of genetic information.		1	1		3	3	
	1	2	2	Explain the functional relationships between DNA, genes, alleles, and chromosomes and their roles in inheritance.	1		1	1		1	
	Total For Assessment Anchor BIO.B.1					4	1	5	4	3	7
	2			<b>Genetics</b>							
	2	1	1	Describe and/or predict observed patterns of inheritance.	1		1	1		1	
	2	1	2	Describe processes that can alter composition or number of chromosomes.	1		1	1		1	
	2	2	1	Describe how the processes of transcription and translation are similar in all organisms.	1		1	1		1	
	2	2	2	Describe the role of ribosomes, endoplasmic reticulum, Golgi apparatus, and the nucleus in the production of specific types of proteins.	1		1	1		1	
	2	3	1	Describe how genetic mutations alter the DNA sequence and may or may not affect phenotype.	1	1	2	1	3	4	
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	1		1	1		1	
	Total For Assessment Anchor BIO.B.2					6	1	7	6	3	9
	3			<b>Theory of Evolution</b>							
	3	1	1	Explain how natural selection can impact allele frequencies of a population.	1		1	1		1	
	3	1	2	Describe the factors that can contribute to the development of new species.	1		1	1		1	
	3	1	3	Explain how genetic mutations may result in genotypic and phenotypic variations within a population.	1		1	1		1	
	3	2	1	Interpret evidence supporting the theory of evolution.	2		2	2		2	
	3	3	1	Distinguish between the scientific terms: hypothesis, inference, law, theory, principle, fact, and observation.	1		1	1		1	
	Total For Assessment Anchor BIO.B.3					6		6	6		6
	4			<b>Ecology</b>							
	4	1	1	Describe the levels of ecological organization.	1		1	1		1	
	4	1	2	Describe characteristic biotic and abiotic components of aquatic and terrestrial ecosystems.	1		1	1		1	
	4	2	1	Describe how energy flows through an ecosystem.	1	1	2	1	3	4	
	4	2	2	Describe biotic interactions in an ecosystem.	1		1	1		1	
	4	2	3	Describe how matter recycles through an ecosystem.	1		1	1		1	
	4	2	4	Describe how ecosystems change in response to natural and human disturbances.	2		2	2		2	
	4	2	5	Describe the effects of limiting factors on population dynamics and potential species extinction.	1		1	1		1	
	Total For Assessment Anchor BIO.B.4					8	1	9	8	3	11
	Total For Assessment Anchor BIO.B					24	3	27	24	9	33

**BIOLOGY SPRING 2023**

<b>Keystone Exam</b>				<b>Biology</b>							
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
<b>BIO.A: Basic Biological Principles</b>	1			<b>Basic Biological Principles</b>							
	1	1	1	Describe the characteristics of life shared by all prokaryotic and eukaryotic organisms.	2		2	2		2	
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	2		2	2		2	
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.		1	1		3	3	
	Total For Assessment Anchor BIO.A.1					4	1	5	4	3	7
	2			<b>The Chemical Basis for Life</b>							
	2	1	1	Describe the unique properties of water and how these properties support life on Earth.	1	1	2	1	3	4	
	2	2	1	Explain how carbon is uniquely suited to form biological macromolecules.	1		1	1		1	
	2	2	2	Describe how biological macromolecules form from monomers.	1		1	1		1	
	2	2	3	Compare the structure and function of carbohydrates, lipids, proteins, and nucleic acids in organisms.	2		2	2		2	
	2	3	1	Describe the role of an enzyme as a catalyst in regulating a specific biochemical reaction.	1		1	1		1	
	2	3	2	Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.	1		1	1		1	
	Total For Assessment Anchor BIO.A.2					7	1	8	7	3	10
	3			<b>Bioenergetics</b>							
	3	1	1	Describe the fundamental roles of plastids (e.g., chloroplasts) and mitochondria in energy transformations.	3		3	3		3	
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	2		2	2		2	
	3	2	2	Describe the role of ATP in biochemical reactions.	1	1	2	1	3	4	
	Total For Assessment Anchor BIO.A.3					6	1	7	6	3	9
	4			<b>Homeostasis and Transport</b>							
	4	1	1	Describe how the structure of the plasma membrane allows it to function as a regulatory structure and/or protective barrier for a cell.	1		1	1		1	
	4	1	2	Compare the mechanisms that transport materials across the plasma membrane.	2		2	2		2	
	4	1	3	Describe how membrane-bound cellular organelles.	2		2	2		2	
	4	2	1	Explain how organisms maintain homeostasis.	2		2	2		2	
	Total For Assessment Anchor BIO.A.4					7		7	7		7
	Total For Reporting Category BIO.A					24	3	27	24	9	33

**Keystone Exam**

**Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
<b>BIO.B: Cell Growth and Reproduction</b>	1			<b>Cell Growth and Reproduction</b>							
	1	1	1	Describe the events that occur during the cell cycle: interphase, nuclear division.	1	1	2	1	3	4	
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	2		2	2		2	
	1	2	1	Describe how the process of DNA replication results in the transmission and/or conservation of genetic information.	1		1	1		1	
	1	2	2	Explain the functional relationships between DNA, genes, alleles, and chromosomes and their roles in inheritance.	1		1	1		1	
	Total For Assessment Anchor BIO.B.1					5	1	6	5	3	8
	2				<b>Genetics</b>						
	2	1	1		Describe and/or predict observed patterns of inheritance.	1		1	1		1
	2	1	2		Describe processes that can alter composition or number of chromosomes.	1		1	1		1
	2	2	1		Describe how the processes of transcription and translation are similar in all organisms.	1		1	1		1
	2	2	2		Describe the role of ribosomes, endoplasmic reticulum, Golgi apparatus, and the nucleus in the production of specific types of proteins.	2		2	2		2
	2	3	1		Describe how genetic mutations alter the DNA sequence and may or may not affect phenotype.	1		1	1		1
	2	4	1		Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.						
	Total For Assessment Anchor BIO.B.2					6		6	6		6
	3				<b>Theory of Evolution</b>						
	3	1	1		Explain how natural selection can impact allele frequencies of a population.		1	1		3	3
	3	1	2		Describe the factors that can contribute to the development of new species.	2		2	2		2
	3	1	3		Explain how genetic mutations may result in genotypic and phenotypic variations within a population.	1		1	1		1
	3	2	1		Interpret evidence supporting the theory of evolution.	1		1	1		1
	3	3	1		Distinguish between the scientific terms: hypothesis, inference, law, theory, principle, fact, and observation.	1		1	1		1
	Total For Assessment Anchor BIO.B.3					5	1	6	5	3	8
	4				<b>Ecology</b>						
	4	1	1		Describe the levels of ecological organization.	2		2	2		2
	4	1	2		Describe characteristic biotic and abiotic components of aquatic and terrestrial ecosystems.	2		2	2		2
	4	2	1		Describe how energy flows through an ecosystem.	1	1	2	1	3	4
	4	2	2		Describe biotic interactions in an ecosystem.	1		1	1		1
	4	2	3		Describe how matter recycles through an ecosystem.						
	4	2	4		Describe how ecosystems change in response to natural and human disturbances.	1		1	1		1
	4	2	5		Describe the effects of limiting factors on population dynamics and potential species extinction.	1		1	1		1
	Total For Assessment Anchor BIO.B.4					8	1	9	8	3	11
	Total For Assessment Anchor BIO.B					24	3	27	24	9	33

Keystone Exam				Biology							
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
BIO.A: Basic Biological Principles	1			<b>Basic Biological Principles</b>							
	1	1	1	Describe the characteristics of life shared by all prokaryotic and eukaryotic organisms.	2		2	2		2	
	1	2	1	Compare cellular structures and their functions in prokaryotic and eukaryotic cells.	2		2	2		2	
	1	2	2	Describe and interpret relationships between structure and function at various levels of biological organization.	2		2	2		2	
	Total For Assessment Anchor BIO.A.1					6		6	6		6
	2				<b>The Chemical Basis for Life</b>						
	2	1	1	Describe the unique properties of water and how these properties support life on Earth.	2		2	2		2	
	2	2	1	Explain how carbon is uniquely suited to form biological macromolecules.	1		1	1		1	
	2	2	2	Describe how biological macromolecules form from monomers.	1		1	1		1	
	2	2	3	Compare the structure and function of carbohydrates, lipids, proteins, and nucleic acids in organisms.	1		1	1		1	
	2	3	1	Describe the role of an enzyme as a catalyst in regulating a specific biochemical reaction.	1		1	1		1	
	2	3	2	Explain how factors such as pH, temperature, and concentration levels can affect enzyme function.	2		2	2		2	
	Total For Assessment Anchor BIO.A.2					8		8	8		8
	3				<b>Bioenergetics</b>						
	3	1	1	Describe the fundamental roles of plastids (e.g., chloroplasts) and mitochondria in energy transformations.	1	1	2	1	3	4	
	3	2	1	Compare the basic transformation of energy during photosynthesis and cellular respiration.	2		2	2		2	
	3	2	2	Describe the role of ATP in biochemical reactions.	2		2	2		2	
	Total For Assessment Anchor BIO.A.3					5	1	6	5	3	8
	4				<b>Homeostasis and Transport</b>						
	4	1	1	Describe how the structure of the plasma membrane allows it to function as a regulatory structure and/or protective barrier for a cell.	2		2	2		2	
	4	1	2	Compare the mechanisms that transport materials across the plasma membrane.	1		1	1		1	
	4	1	3	Describe how membrane-bound cellular organelles.	2	1	3	2	3	5	
	4	2	1	Explain how organisms maintain homeostasis.		1	1		3	3	
	Total For Assessment Anchor BIO.A.4					5	2	7	5	6	11
	Total For Reporting Category BIO.A					24	3	27	24	9	33

**Keystone Exam**

**Biology**

Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
BIO.B: Cell Growth and Reproduction	1			<b>Cell Growth and Reproduction</b>							
	1	1	1	Describe the events that occur during the cell cycle: interphase, nuclear division.	2		2	2		2	
	1	1	2	Compare the processes and outcomes of mitotic and meiotic nuclear divisions.	1		1	1		1	
	1	2	1	Describe how the process of DNA replication results in the transmission and/or conservation of genetic information.	2		2	2		2	
	1	2	2	Explain the functional relationships between DNA, genes, alleles, and chromosomes and their roles in inheritance.	1		1	1		1	
	Total For Assessment Anchor BIO.B.1					6		6	6		6
	2			<b>Genetics</b>							
	2	1	1	Describe and/or predict observed patterns of inheritance.	2		2	2		2	
	2	1	2	Describe processes that can alter composition or number of chromosomes.	1		1	1		1	
	2	2	1	Describe how the processes of transcription and translation are similar in all organisms.	1		1	1		1	
	2	2	2	Describe the role of ribosomes, endoplasmic reticulum, Golgi apparatus, and the nucleus in the production of specific types of proteins.	1		1	1		1	
	2	3	1	Describe how genetic mutations alter the DNA sequence and may or may not affect phenotype.	1		1	1		1	
	2	4	1	Explain how genetic engineering has impacted the fields of medicine, forensics, and agriculture.	1		1	1		1	
	Total For Assessment Anchor BIO.B.2					7		7	7		7
	3			<b>Theory of Evolution</b>							
	3	1	1	Explain how natural selection can impact allele frequencies of a population.	1	1	2	1	3	4	
	3	1	2	Describe the factors that can contribute to the development of new species.	1		1	1		1	
	3	1	3	Explain how genetic mutations may result in genotypic and phenotypic variations within a population.	2		2	2		2	
	3	2	1	Interpret evidence supporting the theory of evolution.	1		1	1		1	
	3	3	1	Distinguish between the scientific terms: hypothesis, inference, law, theory, principle, fact, and observation.							
	Total For Assessment Anchor BIO.B.3					5	1	6	5	3	8
	4			<b>Ecology</b>							
	4	1	1	Describe the levels of ecological organization.	2		2	2		2	
	4	1	2	Describe characteristic biotic and abiotic components of aquatic and terrestrial ecosystems.	1		1	1		1	
	4	2	1	Describe how energy flows through an ecosystem.	1		1	1		1	
	4	2	2	Describe biotic interactions in an ecosystem.							
	4	2	3	Describe how matter recycles through an ecosystem.	1	1	2	1	3	4	
	4	2	4	Describe how ecosystems change in response to natural and human disturbances.		1	1		3	3	
	4	2	5	Describe the effects of limiting factors on population dynamics and potential species extinction.	1		1	1		1	
	Total For Assessment Anchor BIO.B.4					6	2	8	6	6	12
	Total For Assessment Anchor BIO.B					24	3	27	24	9	33

LITERATURE WINTER 2022/2023

Keystone Exam					Literature						
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
	1			<b>Reading for Meaning—Fiction</b>							
	1	1	1	Identify and/or analyze the author's intended purpose of a text.							
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	1		1	1		1	
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	2	1	3	2	3	5	
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.							
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	1		1	1		1	
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.							
	1	2	4	Draw conclusions about connotations of words.							
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.							
	1	3	2	Summarize the key details and events of a fictional text, in part or as a whole.	1		1	1		1	
	Total For Assessment Anchor L.F.1				5	1	6	5	3	8	
	2			<b>Analyzing and Interpreting Literature—Fiction</b>							
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.							
	2	1	2	Cite evidence from a text to support generalizations.	1		1	1		1	
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	1		1	1		1	
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	1		1	1		1	
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.							
	2	2	4	Compare and evaluate the characteristics that distinguish narrative, poetry, and drama.							
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of fiction:	1	1	2	1	3	4	
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of fiction:	1		1	1		1	
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of fiction:	1		1	1		1	
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of fiction:	1		1	1		1	
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of fiction:		1	1		3	3	
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of fiction:	1		1	1		1	
	2	4	1	Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.	1		1	1		1	
	2	5	1	Identify, explain, interpret, describe, and/or analyze the effects of personification, simile, metaphor, hyperbole, satire, foreshadowing, flashback, imagery, allegory, symbolism, dialect, allusion, and irony in a text.	3		3	3		3	
	2	5	2	Identify, explain, and analyze the structure of poems and sound devices.							
	2	5	3	Identify and analyze how stage directions, monologue, dialogue, soliloquy, and dialect support dramatic script.							
	Total For Assessment Anchor L.F.2				12	2	14	12	6	18	
	Total For Reporting Category L.F				17	3	20	17	9	26	



Keystone Exam				Literature							
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
	1			<b>Reading for Meaning—Nonfiction</b>							
	1	1	1	Identify and/or analyze the author's intended purpose of a text.							
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	1		1	1		1	
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of nonfiction to effectively communicate an idea or concept.							
	1	1	4	Explain how an author's use of key words or phrases in text informs and influences the reader.	1		1	1		1	
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.							
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	1		1	1		1	
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	1		1	1		1	
	1	2	4	Draw conclusions about connotations of words.	1		1	1		1	
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	1		1	1		1	
	1	3	2	Summarize the key details and events of a nonfictional text, in part or as a whole.							
	1	3	3	Analyze the interrelationships of ideas and events in text to determine how one idea or event may interact and influence another.		1	1		3	3	
	Total For Assessment Anchor L.N.1				6	1	7	6	3	9	
	2			<b>Data Analysis</b>							
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1		1	1		1	
	2	1	2	Cite evidence from a text to support generalizations.	1		1	1		1	
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	1		1	1		1	
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	1		1	1		1	
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.							
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:		1	1		3	3	
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of nonfiction:							
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of nonfiction:							
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of nonfiction:	1		1	1		1	
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of nonfiction:	1		1	1		1	
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:							
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.	1		1	1		1	
	2	4	2	Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions.							
	2	4	3	Explain, interpret, and/or analyze the effect of text organization, including headings, graphics, and charts.							
	2	4	4	Make connections between a text and the content of graphics and charts.							
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.							
	2	5	1	Differentiate between fact and opinion.	1		1	1		1	
	2	5	2	Explain, interpret, describe, and/or analyze the use of facts and opinions in a text.	1		1	1		1	
	2	5	3	Distinguish essential from nonessential information.							
	2	5	4	Identify, explain, and/or interpret bias and propaganda techniques in nonfictional text.	1		1	1		1	
	2	5	5	Explain, describe, and/or analyze the effectiveness of bias (explicit and implicit) and propaganda techniques in nonfictional text.	1		1	1		1	
	2	5	6	Explain, interpret, describe, and/or analyze the author's defense of a claim to make a point or construct an argument in nonfictional text.		1	1		3	3	
	Total For Assessment Anchor L.N.2				11	2	13	11	6	17	
	Total For Reporting Category L.N				17	3	20	17	9	26	

LITERATURE SPRING 2023

Keystone Exam						Literature					
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
	1			<b>Reading for Meaning—Fiction</b>							
	1	1	1	Identify and/or analyze the author's intended purpose of a text.							
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	1		1	1		1	
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	1	1	2	1	3	4	
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	1		1	1		1	
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	1		1	1		1	
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	1		1	1		1	
	1	2	4	Draw conclusions about connotations of words.	1		1	1		1	
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	1		1	1		1	
	1	3	2	Summarize the key details and events of a fictional text, in part or as a whole.	1		1	1		1	
	Total For Assessment Anchor L.F.1				8	1	9	8	3	11	
	2			<b>Analyzing and Interpreting Literature—Fiction</b>							
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1		1	1		1	
	2	1	2	Cite evidence from a text to support generalizations.		1	1		3	3	
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.							
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.	1		1	1		1	
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.							
	2	2	4	Compare and evaluate the characteristics that distinguish narrative, poetry, and drama.							
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of fiction:	1		1	1		1	
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of fiction:							
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of fiction:	2		2	2		2	
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of fiction:	2		2	2		2	
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of fiction:							
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of fiction:	1		1	1		1	
	2	4	1	Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.							
	2	5	1	Identify, explain, interpret, describe, and/or analyze the effects of personification, simile, metaphor, hyperbole, satire, foreshadowing, flashback, imagery, allegory, symbolism, dialect, allusion, and irony in a text.	1	1	2	1	3	4	
	2	5	2	Identify, explain, and analyze the structure of poems and sound devices.							
	2	5	3	Identify and analyze how stage directions, monologue, dialogue, soliloquy, and dialect support dramatic script.							
	Total For Assessment Anchor L.F.2				9	2	11	9	6	15	
Total For Reporting Category L.F					17	3	20	17	9	26	

Keystone Exam				Literature							
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
	1			<b>Reading for Meaning—Nonfiction</b>							
	1	1	1	Identify and/or analyze the author's intended purpose of a text.							
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.	1	1	2	1	3	4	
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of nonfiction to effectively communicate an idea or concept.	1		1	1		1	
	1	1	4	Explain how an author's use of key words or phrases in text informs and influences the reader.							
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.							
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.	1		1	1		1	
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	1		1	1		1	
	1	2	4	Draw conclusions about connotations of words.	1		1	1		1	
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.							
	1	3	2	Summarize the key details and events of a nonfictional text, in part or as a whole.							
	1	3	3	Analyze the interrelationships of ideas and events in text to determine how one idea or event may interact and influence another.	1		1	1		1	
	Total For Assessment Anchor L.N.1				6	1	7	6	3	9	
	2			<b>Data Analysis</b>							
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1		1	1		1	
	2	1	2	Cite evidence from a text to support generalizations.	1		1	1		1	
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.							
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.							
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.							
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:	1		1	1		1	
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of nonfiction:	1		1	1		1	
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of nonfiction:							
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of nonfiction:		1	1		3	3	
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of nonfiction:	2		2	2		2	
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:	2		2	2		2	
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.							
	2	4	2	Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions.							
	2	4	3	Explain, interpret, and/or analyze the effect of text organization, including headings, graphics, and charts.							
	2	4	4	Make connections between a text and the content of graphics and charts.							
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.							
	2	5	1	Differentiate between fact and opinion.							
	2	5	2	Explain, interpret, describe, and/or analyze the use of facts and opinions in a text.							
	2	5	3	Distinguish essential from nonessential information.							
	2	5	4	Identify, explain, and/or interpret bias and propaganda techniques in nonfictional text.	1		1	1		1	
	2	5	5	Explain, describe, and/or analyze the effectiveness of bias (explicit and implicit) and propaganda techniques in nonfictional text.							
	2	5	6	Explain, interpret, describe, and/or analyze the author's defense of a claim to make a point or construct an argument in nonfictional text.	2	1	3	2	3	5	
	Total For Assessment Anchor L.N.2				11	2	13	11	6	17	
	Total For Reporting Category L.N				17	3	20	17	9	26	

LITERATURE SUMMER 2023

Keystone Exam				Literature							
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
	1			<b>Reading for Meaning—Fiction</b>							
	1	1	1	Identify and/or analyze the author's intended purpose of a text.							
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.		1	1		3	3	
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of fiction to effectively communicate an idea or concept.	1		1	1			1
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.	1		1	1			1
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.							
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.							
	1	2	4	Draw conclusions about connotations of words.	1		1	1			1
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.	2		2	2			2
	1	3	2	Summarize the key details and events of a fictional text, in part or as a whole.							
	Total For Assessment Anchor L.F.1				5	1	6	5	3	8	
	2			<b>Analyzing and Interpreting Literature—Fiction</b>							
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1		1	1			1
	2	1	2	Cite evidence from a text to support generalizations.	1		1	1			1
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.							
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.							
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.							
	2	2	4	Compare and evaluate the characteristics that distinguish narrative, poetry, and drama.							
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of fiction:	1	2	3	1	6	7	
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of fiction:	1		1	1			1
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of fiction:	1		1	1			1
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of fiction:	2		2	2			2
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of fiction:	1		1	1			1
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of fiction:	1		1	1			1
	2	4	1	Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.							
	2	5	1	Identify, explain, interpret, describe, and/or analyze the effects of personification, simile, metaphor, hyperbole, satire, foreshadowing, flashback, imagery, allegory, symbolism, dialect, allusion, and irony in a text.	3		3	3			3
	2	5	2	Identify, explain, and analyze the structure of poems and sound devices.							
	2	5	3	Identify and analyze how stage directions, monologue, dialogue, soliloquy, and dialect support dramatic script.							
	Total For Assessment Anchor L.F.2				12	2	14	12	6	18	
	Total For Reporting Category L.F				17	3	20	17	9	26	

Keystone Exam					Literature						
Reporting Category	Assessment Anchor	Descriptor (Sub-anchor)	Eligible Content	Focus	Items			Points			
					Number of Core Items			Core Points			
					MC	CR	Total	MC	CR	Total	
	1			<b>Reading for Meaning—Nonfiction</b>							
	1	1	1	Identify and/or analyze the author's intended purpose of a text.							
	1	1	2	Explain, describe, and/or analyze examples of a text that support the author's intended purpose.							
	1	1	3	Analyze, interpret, and evaluate how authors use techniques and elements of nonfiction to effectively communicate an idea or concept.	1	1	2	1	3	4	
	1	1	4	Explain how an author's use of key words or phrases in text informs and influences the reader.	1		1	1		1	
	1	2	1	Identify and/or apply a synonym or antonym of a word used in a text.							
	1	2	2	Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word with an affix from a text.							
	1	2	3	Use context clues to determine or clarify the meaning of unfamiliar, multiple-meaning, or ambiguous words.	2		2	2		2	
	1	2	4	Draw conclusions about connotations of words.	1		1	1		1	
	1	3	1	Identify and/or explain stated or implied main ideas and relevant supporting details from a text.							
	1	3	2	Summarize the key details and events of a nonfictional text, in part or as a whole.							
	1	3	3	Analyze the interrelationships of ideas and events in text to determine how one idea or event may interact and influence another.	1		1	1		1	
	Total For Assessment Anchor L.N.1				6	1	7	6	3	9	
	2			<b>Data Analysis</b>							
	2	1	1	Make inferences and/or draw conclusions based on analysis of a text.	1		1	1		1	
	2	1	2	Cite evidence from a text to support generalizations.							
	2	2	1	Analyze how literary form relates to and/or influences meaning of a text.	2		2	2		2	
	2	2	2	Compare and evaluate the characteristics that distinguish fiction from literary nonfiction.							
	2	2	3	Explain, interpret, compare, describe, analyze, and/or evaluate connections between texts.							
	2	3	1	Explain, interpret, compare, describe, analyze, and/or evaluate character in a variety of nonfiction:							
	2	3	2	Explain, interpret, compare, describe, analyze, and/or evaluate setting in a variety of nonfiction:							
	2	3	3	Explain, interpret, compare, describe, analyze, and/or evaluate plot in a variety of nonfiction:							
	2	3	4	Explain, interpret, compare, describe, analyze, and/or evaluate theme in a variety of nonfiction:							
	2	3	5	Explain, interpret, compare, describe, analyze, and/or evaluate tone, style, and/or mood in a variety of nonfiction:							
	2	3	6	Explain, interpret, compare, describe, analyze, and/or evaluate point of view in a variety of nonfiction:	1		1	1		1	
	2	4	1	Identify, analyze, and evaluate the structure and format of complex informational texts.	1	1	2	1	3	4	
	2	4	2	Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions.							
	2	4	3	Explain, interpret, and/or analyze the effect of text organization, including headings, graphics, and charts.							
	2	4	4	Make connections between a text and the content of graphics and charts.							
	2	4	5	Analyze and evaluate how graphics and charts clarify, simplify, and organize complex informational texts.							
	2	5	1	Differentiate between fact and opinion.							
	2	5	2	Explain, interpret, describe, and/or analyze the use of facts and opinions in a text.	1		1	1		1	
	2	5	3	Distinguish essential from nonessential information.							
	2	5	4	Identify, explain, and/or interpret bias and propaganda techniques in nonfictional text.	1		1	1		1	
	2	5	5	Explain, describe, and/or analyze the effectiveness of bias (explicit and implicit) and propaganda techniques in nonfictional text.	1		1	1		1	
	2	5	6	Explain, interpret, describe, and/or analyze the author's defense of a claim to make a point or construct an argument in nonfictional text.	3	1	4	3	3	6	
	Total For Assessment Anchor L.N.2				11	2	13	11	6	17	
	Total For Reporting Category L.N				17	3	20	17	9	26	

## APPENDIX G: KEYSTONE EXAMS MODULE LAYOUT PLANS

**Table G–1A. Winter 2022/2023, Spring 2023 and Summer 2023 Algebra I Keystone Exams Section Layout Plan**

Module	Number of MC	Estimated MC Item Breakdown	Number of CR	Estimated CR Item Breakdown	Testing Time	Administration Time
1	23	18—Operational (Core) Items; 5—Embedded Field-Test Items	4	3—Operational (Core) Items; 1— Embedded Field-Test Items	75	85–90
2	23	18—Operational (Core) Items; 5—Embedded Field-Test Items	4	3—Operational (Core) Items; 1— Embedded Field-Test Items	75	85–90

**Table G–1B. Winter 2022/2023, Spring 2023 and Summer 2023 Biology Keystone Exams Section Layout Plan**

Module	Number of MC	Estimated MC Item Breakdown	Number of CR	Estimated CR Item Breakdown	Testing Time	Administration Time
1	32	24—Operational (Core) Items; 8—Embedded Field-Test Items	4	3—Operational (Core) Items; 1— Embedded Field-Test Items	72	82–87
2	32	24—Operational (Core) Items; 8—Embedded Field-Test Items	4	3—Operational (Core) Items; 1— Embedded Field-Test Items	72	82–87

**Table G–1L. Winter 2022/2023, Spring 2023 and Summer 2023 Literature Keystone Exams Section Layout Plan**

Module	Number of MC	Estimated MC Item Breakdown	Number of CR	Estimated CR Item Breakdown	Testing Time	Administration Time
1	23	17—Operational (Core) Items; 6—Embedded Field-Test Items	4	3—Operational (Core) Items; 1— Embedded Field-Test Items	73	83–88
2	23	17—Operational (Core) Items; 6—Embedded Field-Test Items	4	3—Operational (Core) Items; 1— Embedded Field-Test Items	73	83–88

## APPENDIX H: MEAN RAW SCORES BY FORM

---

Tables in this appendix present the mean raw scores for all forms by content area and is based on the final data for the spring administration (see Chapter Nine). In addition, accommodated forms (i.e., Spanish, Braille, large print, VSL) were removed to best represent comparisons among similar groups of test-takers.

**Table H-1. Mean Raw Scores by Form**

<b>Column Heading</b>	<b>Definition</b>
Form	Form
<i>N</i>	Number of students
<i>L</i>	Length (Number of Items)
<i>Pts</i>	Points possible
<i>Min</i>	Minimum
<i>Max</i>	Maximum
<i>Mean</i>	Mean
<i>Med</i>	Median
<i>SD</i>	Standard deviation

## ALGEBRA I: SPRING

Table H-2. Algebra I Mean Raw Scores by Form Table

Form	N	L	Pts	Min	Max	Mean	Med	SD
All	117351	42	60	0	60	26.0	24	12.9
1	4921	42	60	1	60	25.9	24	12.9
2	4912	42	60	1	59	25.9	24	12.8
3	4845	42	60	2	60	25.9	24	12.8
4	4880	42	60	3	59	26.0	24	12.9
5	4881	42	60	1	59	25.8	23	12.9
6	4746	42	60	2	59	26.0	24	12.9
7	4885	42	60	2	60	25.8	24	12.8
8	4912	42	60	2	59	25.9	24	12.9
9	4905	42	60	1	59	26.0	24	12.9
10	4900	42	60	2	60	26.0	24	13.1
11	4897	42	60	1	60	25.9	24	12.7
12	4910	42	60	1	59	26.0	24	12.8
13	4892	42	60	2	60	26.0	24	12.7
14	4953	42	60	4	60	26.0	24	12.9
15	4879	42	60	1	60	26.4	25	12.8
16	4929	42	60	2	60	26.0	24	12.9
17	4866	42	60	2	60	25.7	24	12.8
18	4881	42	60	2	59	26.0	24	12.9
19	4878	42	60	2	60	26.0	24	13.0
20	4855	42	60	1	60	26.1	24	12.8
21	4923	42	60	3	59	25.9	24	12.9
22	4877	42	60	0	60	26.0	23	12.9
23	4914	42	60	2	59	26.0	24	12.8
24	4910	42	60	3	60	26.0	24	13.0



## BIOLOGY: SPRING

Table H-3. Biology Mean Raw Scores by Form Table

Form	N	L	Pts	Min	Max	Mean	Med	SD
All	110271	54	66	1	66	34.0	33	15.1
1	18420	54	66	3	66	34.2	33	15.1
2	18475	54	66	3	66	34.1	33	15.0
3	18382	54	66	3	66	33.9	33	15.1
4	18276	54	66	1	66	33.9	33	15.1
5	18302	54	66	3	66	34.0	33	15.0
6	18416	54	66	4	66	33.9	33	15.2

## LITERATURE: SPRING

Table H-4. Literature Mean Raw Scores by Form Table

Form	N	L	Pts	Min	Max	Mean	Med	SD
All	108453	40	52	1	52	31.3	33	10.3
1	4520	40	52	3	52	31.4	33	10.5
2	4510	40	52	4	51	31.0	33	10.2
3	4552	40	52	2	52	31.4	33	10.4
4	4574	40	52	3	52	31.4	33	10.3
5	4526	40	52	1	52	31.4	33	10.2
6	4551	40	52	3	51	31.3	33	10.2
7	4524	40	52	3	52	31.6	33	10.2
8	4537	40	52	3	52	31.2	33	10.4
9	4526	40	52	2	52	31.1	33	10.1
10	4496	40	52	4	52	31.3	33	10.4
11	4508	40	52	2	52	31.3	33	10.3
12	4513	40	52	2	51	30.9	32	10.0
13	4509	40	52	4	51	31.4	33	10.3
14	4500	40	52	3	52	31.3	33	10.2
15	4439	40	52	3	52	31.5	33	10.2
16	4548	40	52	3	52	31.4	33	10.2
17	4498	40	52	3	52	31.6	34	10.3
18	4520	40	52	2	52	31.0	33	10.4
19	4474	40	52	3	52	31.1	33	10.5
20	4505	40	52	2	52	31.1	33	10.4
21	4584	40	52	2	52	30.9	32	10.1
22	4485	40	52	4	52	31.8	33	10.2
23	4493	40	52	2	52	31.5	33	10.5
24	4561	40	52	3	52	31.3	33	10.3

## APPENDIX I: DEMOGRAPHIC AND ACCOMMODATION DATA

### WINTER

**Table I-1. Students Assessed on the Winter Keystone: Algebra I**

Description	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	1	2	3	274	6,088	18,168	13,310	1,677	39,523
Total number of CBT processed (Number)	1	1	18	318	3,974	8,183	5,687	572	18,754
Total number of tests processed (Number)	2	3	21	592	10,062	26,351	18,997	2,249	58,277
Total number of tests processed with a score (Number)	2	3	21	588	9,475	24,762	17,271	1,934	54,056
Total number of tests processed with a score (Percent)	100	100	100	99.3	94.2	94	90.9	86	92.8
Total number of tests processed without a score (Number)	0	0	0	4	587	1,589	1,726	315	4,221
Total number of tests processed without a score (Percent)	0	0	0	.7	5.8	6	9.1	14	7.2

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table I-2. Students Assessed on the Winter Keystone: Biology**

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	0	0	1,649	11,224	12,410	1,790	27,073
Total number of CBT processed (Number)	0	0	1,695	6,551	6,703	904	15,853
Total number of tests processed (Number)	0	0	3,344	17,775	19,113	2,694	42,926
Total number of tests processed with a score (Number)	0	0	3,072	16,788	17,229	2,327	39,416
Total number of tests processed with a score (Percent)	0	0	91.9	94.4	90.1	86.4	91.8
Total number of tests processed without a score (Number)	0	0	272	987	1,884	367	3,510
Total number of tests processed without a score (Percent)	0	0	8.1	5.6	9.9	13.6	8.2

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–3. Students Assessed on the Winter Keystone: Literature**

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	3	0	493	8,405	12,762	2,304	23,967
Total number of CBT processed (Number)	0	0	221	5,303	7,514	868	13,906
Total number of tests processed (Number)	3	0	714	13,708	20,276	3,172	37,873
Total number of tests processed with a score (Number)	1	0	559	13,151	18,103	2,713	34,527
Total number of tests processed with a score (Percent)	33.3	0	78.3	95.9	89.3	85.5	91.2
Total number of tests processed without a score (Number)	2	0	155	557	2,173	459	3,346
Total number of tests processed without a score (Percent)	66.7	0	21.7	4.1	10.7	14.5	8.8

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–4. Counts of Students without Scores on the Winter Keystone: Algebra I**

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	0	1	295	544	588	87	1,515
Extended absence from school (Percent)	0	0	0	25	50.3	34.2	34.1	27.6	35.9
Non-attempt (Number)	0	0	0	1	178	700	606	161	1,646
Non-attempt (Percent)	0	0	0	25	30.3	44.1	35.1	51.1	39
Medical emergency (Number)	0	0	0	0	8	20	21	1	50
Medical emergency (Percent)	0	0	0	0	1.4	1.3	1.2	.3	1.2
Parental request - Chapter 4 (Number)	0	0	0	1	6	33	66	10	116
Parental request - Chapter 4 (Percent)	0	0	0	25	1	2.1	3.8	3.2	2.7
Parental request - Other reasons (Number)	0	0	0	1	11	71	126	11	220
Parental request - Other reasons (Percent)	0	0	0	25	1.9	4.5	7.3	3.5	5.2
Other reasons (Number)	0	0	0	0	89	221	319	45	674
Other reasons (Percent)	0	0	0	0	15.2	13.9	18.5	14.3	16
Total not assessed (Number)	0	0	0	4	587	1,589	1,726	315	4,221

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table I-5. Counts of Students without Scores on the Winter Keystone: Biology**

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	138	409	553	77	1,177
Extended absence from school (Percent)	0	0	50.7	41.4	29.4	21	33.5
Non-attempt (Number)	0	0	66	336	717	188	1,307
Non-attempt (Percent)	0	0	24.3	34	38.1	51.2	37.2
Medical emergency (Number)	0	0	2	22	30	1	55
Medical emergency (Percent)	0	0	.7	2.2	1.6	.3	1.6
Parental request - Chapter 4 (Number)	0	0	0	29	74	13	116
Parental request - Chapter 4 (Percent)	0	0	0	2.9	3.9	3.5	3.3
Parental request - Other reasons (Number)	0	0	6	37	146	25	214
Parental request - Other reasons (Percent)	0	0	2.2	3.7	7.7	6.8	6.1
Other reasons (Number)	0	0	60	154	364	63	641
Other reasons (Percent)	0	0	22.1	15.6	19.3	17.2	18.3
Total not assessed (Number)	0	0	272	987	1,884	367	3,510

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I-6. Counts of Students without Scores on the Winter Keystone: Literature**

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	112	284	670	113	1,179
Extended absence from school (Percent)	0	0	72.3	51	30.8	24.6	35.2
Non-attempt (Number)	2	0	21	146	835	229	1,233
Non-attempt (Percent)	100	0	13.5	26.2	38.4	49.9	36.8
EL in first year in U.S. schools (Number)	0	0	0	0	2	1	3
EL in first year in U.S. schools (Percent)	0	0	0	0	.1	.2	.1
Medical emergency (Number)	0	0	0	21	28	3	52
Medical emergency (Percent)	0	0	0	3.8	1.3	.7	1.6
Parental request - Chapter 4 (Number)	0	0	0	11	71	14	96
Parental request - Chapter 4 (Percent)	0	0	0	2	3.3	3.1	2.9
Parental request - Other reasons (Number)	0	0	2	25	177	27	231
Parental request - Other reasons (Percent)	0	0	1.3	4.5	8.1	5.9	6.9
Other reasons (Number)	0	0	20	70	390	72	552
Other reasons (Percent)	0	0	12.9	12.6	17.9	15.7	16.5
Total not assessed (Number)	2	0	155	557	2,173	459	3,346

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–7. Demographic Characteristics of Students taking the Winter Keystone: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	0	1	13	291	4,978	12,320	8,362	854	26,819
Female (Percent)	0	33.3	61.9	49.5	52.5	49.8	48.4	44.2	49.6
Male (Number)	2	2	8	297	4,496	12,441	8,908	1,080	27,234
Male (Percent)	100	66.7	38.1	50.5	47.5	50.2	51.6	55.8	50.4
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	0	17	47	33	6	103
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	0	.2	.2	.2	.3	.2
Asian (not Hispanic) (Number)	1	3	10	71	351	729	449	81	1,695
Asian (not Hispanic) (Percent)	50	100	47.6	12.1	3.7	2.9	2.6	4.2	3.1
Black or African American (not Hispanic) (Number)	0	0	0	30	1,164	4,552	3,751	512	10,009
Black or African American (not Hispanic) (Percent)	0	0	0	5.1	12.3	18.4	21.7	26.5	18.5
Hispanic (any race) (Number)	0	0	0	77	1,235	3,862	3,134	499	8,807
Hispanic (any race) (Percent)	0	0	0	13.1	13	15.6	18.1	25.8	16.3
Multi-Racial (not Hispanic) (Number)	0	0	2	26	375	1,066	781	93	2,343
Multi-Racial (not Hispanic) (Percent)	0	0	9.5	4.4	4	4.3	4.5	4.8	4.3
White (not Hispanic) (Number)	1	0	9	384	6,329	14,481	9,105	739	31,048
White (not Hispanic) (Percent)	50	0	42.9	65.3	66.8	58.5	52.7	38.2	57.4
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	4	24	16	4	48
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	0	.1	.1	.2	.1
IEP (not gifted) (Number)	0	0	0	13	768	4,143	4,200	549	9,673
IEP (not gifted) (Percent)	0	0	0	2.2	8.1	16.7	24.3	28.4	17.9
Student exited IEP in last 2 years (Number)	0	0	0	0	99	223	140	9	471
Student exited IEP in last 2 years (Percent)	0	0	0	0	1	.9	.8	.5	.9
Title I (Number)	0	0	1	73	1,983	7,931	6,348	994	17,330
Title I (Percent)	0	0	4.8	12.4	20.9	32	36.8	51.4	32.1
Title III served (Number)	0	0	0	2	186	1,105	1,054	239	2,586
Title III served (Percent)	0	0	0	.3	2	4.5	6.1	12.4	4.8
Title III not served (Number)	0	0	0	43	980	2,521	2,086	196	5,826
Title III not served (Percent)	0	0	0	7.3	10.3	10.2	12.1	10.1	10.8
Migrant student (Number)	0	0	0	0	6	19	10	4	39
Migrant student (Percent)	0	0	0	0	.1	.1	.1	.2	.1
EL enrolled first year (Number)	0	0	0	0	1	2	10	5	18
EL enrolled first year (Percent)	0	0	0	0	0	0	.1	.3	0
EL enrolled not first year (Number)	0	0	0	2	198	1,169	1,136	242	2,747
EL enrolled not first year (Percent)	0	0	0	.3	2.1	4.7	6.6	12.5	5.1
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	1	0	30	72	46	5	154
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	4.8	0	.3	.3	.3	.3	.3

**Table I-7 (continued). Demographic Characteristics of Students taking the Winter Keystone: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	0	17	38	31	2	88
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	0	.2	.2	.2	.1	.2
Former EL no longer monitored (Number)	0	0	0	10	130	491	326	40	997
Former EL no longer monitored (Percent)	0	0	0	1.7	1.4	2	1.9	2.1	1.8
LIFE first year (Number)	0	0	0	0	0	0	0	0	0
LIFE first year (Percent)	0	0	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	0	0	4	9	15	6	34
LIFE not first year (Percent)	0	0	0	0	0	0	.1	.3	.1
Former EL exited and in 3rd year of monitoring (Number)	0	0	0	5	13	44	45	4	111
Former EL exited and in 3rd year of monitoring (Percent)	0	0	0	.9	.1	.2	.3	.2	.2
Former EL exited and in 4th year of monitoring (Number)	0	0	0	10	63	57	51	3	184
Former EL exited and in 4th year of monitoring (Percent)	0	0	0	1.7	.7	.2	.3	.2	.3
Foreign exchange student (Number)	0	0	0	0	0	26	3	2	31
Foreign exchange student (Percent)	0	0	0	0	0	.1	0	.1	.1
Economically disadvantaged (Number)	1	0	0	144	3,608	12,659	9,968	1,265	27,645
Economically disadvantaged (Percent)	50	0	0	24.5	38.1	51.1	57.7	65.4	51.1
Historically Underperforming Subgroup (Number)	1	0	0	151	4,068	14,685	11,882	1,523	32,310
Historically Underperforming Subgroup (Percent)	50	0	0	25.7	42.9	59.3	68.8	78.7	59.8
Enrollment in school of residence after Oct 1 (Number)	0	0	0	4	77	208	228	32	549
Enrollment in school of residence after Oct 1 (Percent)	0	0	0	.7	.8	.8	1.3	1.7	1
Enrollment in district of residence after Oct 1 (Number)	0	0	0	2	63	166	181	24	436
Enrollment in district of residence after Oct 1 (Percent)	0	0	0	.3	.7	.7	1	1.2	.8
Enrollment as PA resident after Oct 1 (Number)	0	0	0	1	39	91	82	10	223
Enrollment as PA resident after Oct 1 (Percent)	0	0	0	.2	.4	.4	.5	.5	.4
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	3	4	24	5,128	5,097	3,640	614	14,510
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	100	19	4.1	54.1	20.6	21.1	31.7	26.8
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	1	3	15	905	1,814	1,670	296	4,704

**Table I-7 (continued). Demographic Characteristics of Students taking the Winter Keystone: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	33.3	14.3	2.6	9.6	7.3	9.7	15.3	8.7
Military family (Number)	0	0	0	1	54	127	73	8	263
Military family (Percent)	0	0	0	.2	.6	.5	.4	.4	.5
Homeless (Number)	0	0	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0	0	0
Foster (Number)	0	0	0	0	22	98	104	22	246
Foster (Percent)	0	0	0	0	.2	.4	.6	1.1	.5
Home schooled (Number)	0	0	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	5	27	24	17	73
Court/agency placed (Percent)	0	0	0	0	.1	.1	.1	.9	.1
Number of assessed students (Number)	2	3	21	588	9,475	24,762	17,271	1,934	54,056

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade



**Table I–8. Demographic Characteristics of Students taking the Winter Keystone: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	0	0	1,494	8,230	8,288	1,028	19,040
Female (Percent)	0	0	48.6	49	48.1	44.2	48.3
Male (Number)	0	0	1,578	8,558	8,940	1,299	20,375
Male (Percent)	0	0	51.4	51	51.9	55.8	51.7
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	7	29	30	6	72
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.2	.2	.2	.3	.2
Asian (not Hispanic) (Number)	0	0	127	545	497	94	1,263
Asian (not Hispanic) (Percent)	0	0	4.1	3.2	2.9	4	3.2
Black or African American (not Hispanic) (Number)	0	0	666	2,614	3,450	750	7,480
Black or African American (not Hispanic) (Percent)	0	0	21.7	15.6	20	32.2	19
Hispanic (any race) (Number)	0	0	533	2,198	2,795	520	6,046
Hispanic (any race) (Percent)	0	0	17.4	13.1	16.2	22.3	15.3
Multi-Racial (not Hispanic) (Number)	0	0	132	731	819	103	1,785
Multi-Racial (not Hispanic) (Percent)	0	0	4.3	4.4	4.8	4.4	4.5
White (not Hispanic) (Number)	0	0	1,603	10,647	9,617	848	22,715
White (not Hispanic) (Percent)	0	0	52.2	63.4	55.8	36.4	57.6
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	4	24	19	6	53
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.1	.1	.1	.3	.1
IEP (not gifted) (Number)	0	0	450	2,972	4,267	535	8,224
IEP (not gifted) (Percent)	0	0	14.6	17.7	24.8	23	20.9
Student exited IEP in last 2 years (Number)	0	0	20	123	114	16	273
Student exited IEP in last 2 years (Percent)	0	0	.7	.7	.7	.7	.7
Title I (Number)	0	0	650	4,099	5,944	1,282	11,975
Title I (Percent)	0	0	21.2	24.4	34.5	55.1	30.4
Title III served (Number)	0	0	93	531	946	264	1,834
Title III served (Percent)	0	0	3	3.2	5.5	11.3	4.7
Title III not served (Number)	0	0	309	1,816	2,130	191	4,446
Title III not served (Percent)	0	0	10.1	10.8	12.4	8.2	11.3
Migrant student (Number)	0	0	3	6	16	4	29
Migrant student (Percent)	0	0	.1	0	.1	.2	.1
EL enrolled first year (Number)	0	0	1	1	11	6	19
EL enrolled first year (Percent)	0	0	0	0	.1	.3	0
EL enrolled not first year (Number)	0	0	101	585	1,038	263	1,987
EL enrolled not first year (Percent)	0	0	3.3	3.5	6	11.3	5
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	7	50	55	6	118
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	.2	.3	.3	.3	.3
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	1	28	29	2	60
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	.2	.2	.1	.2

**Table I–8 (continued). Demographic Characteristics of Students taking the Winter Keystone: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Former EL no longer monitored (Number)	0	0	67	330	307	40	744
Former EL no longer monitored (Percent)	0	0	2.2	2	1.8	1.7	1.9
LIFE first year (Number)	0	0	0	0	0	0	0
LIFE first year (Percent)	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	1	15	17	7	40
LIFE not first year (Percent)	0	0	0	.1	.1	.3	.1
Former EL exited and in 3rd year of monitoring (Number)	0	0	6	35	50	2	93
Former EL exited and in 3rd year of monitoring (Percent)	0	0	.2	.2	.3	.1	.2
Former EL exited and in 4th year of monitoring (Number)	0	0	17	37	58	5	117
Former EL exited and in 4th year of monitoring (Percent)	0	0	.6	.2	.3	.2	.3
Foreign exchange student (Number)	0	0	0	1	25	1	27
Foreign exchange student (Percent)	0	0	0	0	.1	0	.1
Economically disadvantaged (Number)	0	0	1,479	7,610	9,767	1,535	20,391
Economically disadvantaged (Percent)	0	0	48.1	45.3	56.7	66	51.7
Historically Underperforming Subgroup (Number)	0	0	1,669	9,100	11,653	1,785	24,207
Historically Underperforming Subgroup (Percent)	0	0	54.3	54.2	67.6	76.7	61.4
Enrollment in school of residence after Oct 1 (Number)	0	0	44	139	232	42	457
Enrollment in school of residence after Oct 1 (Percent)	0	0	1.4	.8	1.3	1.8	1.2
Enrollment in district of residence after Oct 1 (Number)	0	0	38	114	182	36	370
Enrollment in district of residence after Oct 1 (Percent)	0	0	1.2	.7	1.1	1.5	.9
Enrollment as PA resident after Oct 1 (Number)	0	0	26	63	82	17	188
Enrollment as PA resident after Oct 1 (Percent)	0	0	.8	.4	.5	.7	.5
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	1,796	2,964	3,204	733	8,697
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	58.5	17.7	18.6	31.5	22.1
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	415	1,086	1,512	302	3,315
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	13.5	6.5	8.8	13	8.4
Military family (Number)	0	0	12	45	57	10	124
Military family (Percent)	0	0	.4	.3	.3	.4	.3
Homeless (Number)	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0
Foster (Number)	0	0	8	66	99	16	189
Foster (Percent)	0	0	.3	.4	.6	.7	.5
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	2	20	46	17	85
Court/agency placed (Percent)	0	0	.1	.1	.3	.7	.2
Number of assessed students (Number)	0	0	3,072	16,788	17,229	2,327	39,416

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–9. Demographic Characteristics of Students taking the Winter Keystone: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	0	0	244	6,396	8,073	1,139	15,852
Female (Percent)	0	0	43.6	48.6	44.6	42	45.9
Male (Number)	0	0	315	6,755	10,030	1,574	18,674
Male (Percent)	0	0	56.4	51.4	55.4	58	54.1
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	3	24	30	7	64
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	.5	.2	.2	.3	.2
Asian (not Hispanic) (Number)	0	0	9	498	501	110	1,118
Asian (not Hispanic) (Percent)	0	0	1.6	3.8	2.8	4.1	3.2
Black or African American (not Hispanic) (Number)	0	0	130	1,865	3,564	798	6,357
Black or African American (not Hispanic) (Percent)	0	0	23.3	14.2	19.7	29.4	18.4
Hispanic (any race) (Number)	0	0	112	1,892	2,935	581	5,520
Hispanic (any race) (Percent)	0	0	20	14.4	16.2	21.4	16
Multi-Racial (not Hispanic) (Number)	0	0	27	573	815	126	1,541
Multi-Racial (not Hispanic) (Percent)	0	0	4.8	4.4	4.5	4.6	4.5
White (not Hispanic) (Number)	0	0	277	8,283	10,246	1,086	19,892
White (not Hispanic) (Percent)	0	0	49.6	63	56.6	40	57.6
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	1	16	12	5	34
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	.2	.1	.1	.2	.1
IEP (not gifted) (Number)	0	0	101	2,029	4,952	647	7,729
IEP (not gifted) (Percent)	0	0	18.1	15.4	27.4	23.8	22.4
Student exited IEP in last 2 years (Number)	0	0	5	98	143	17	263
Student exited IEP in last 2 years (Percent)	0	0	.9	.7	.8	.6	.8
Title I (Number)	0	0	200	2,435	5,885	1,384	9,904
Title I (Percent)	0	0	35.8	18.5	32.5	51	28.7
Title III served (Number)	0	0	15	275	1,002	304	1,596
Title III served (Percent)	0	0	2.7	2.1	5.5	11.2	4.6
Title III not served (Number)	0	0	65	1,215	2,016	255	3,551
Title III not served (Percent)	0	0	11.6	9.2	11.1	9.4	10.3
Migrant student (Number)	0	0	0	7	22	7	36
Migrant student (Percent)	0	0	0	.1	.1	.3	.1
EL enrolled first year (Number)	0	0	0	0	2	6	8
EL enrolled first year (Percent)	0	0	0	0	0	.2	0
EL enrolled not first year (Number)	0	0	16	305	1,106	308	1,735
EL enrolled not first year (Percent)	0	0	2.9	2.3	6.1	11.4	5
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	35	39	12	86
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	.3	.2	.4	.2
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	17	37	3	57
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	.1	.2	.1	.2

**Table I–9 (continued). Demographic Characteristics of Students taking the Winter Keystone: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Former EL no longer monitored (Number)	0	0	10	296	312	54	672
Former EL no longer monitored (Percent)	0	0	1.8	2.3	1.7	2	1.9
LIFE first year (Number)	0	0	0	0	0	0	0
LIFE first year (Percent)	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	0	4	14	5	23
LIFE not first year (Percent)	0	0	0	0	.1	.2	.1
Former EL exited and in 3rd year of monitoring (Number)	0	0	1	33	44	4	82
Former EL exited and in 3rd year of monitoring (Percent)	0	0	.2	.3	.2	.1	.2
Former EL exited and in 4th year of monitoring (Number)	0	0	3	22	53	4	82
Former EL exited and in 4th year of monitoring (Percent)	0	0	.5	.2	.3	.1	.2
Foreign exchange student (Number)	0	0	0	3	34	1	38
Foreign exchange student (Percent)	0	0	0	0	.2	0	.1
Economically disadvantaged (Number)	0	0	368	5,653	10,267	1,701	17,989
Economically disadvantaged (Percent)	0	0	65.8	43	56.7	62.7	52.1
Historically Underperforming Subgroup (Number)	0	0	393	6,583	12,427	2,013	21,416
Historically Underperforming Subgroup (Percent)	0	0	70.3	50.1	68.6	74.2	62
Enrollment in school of residence after Oct 1 (Number)	0	0	13	131	221	47	412
Enrollment in school of residence after Oct 1 (Percent)	0	0	2.3	1	1.2	1.7	1.2
Enrollment in district of residence after Oct 1 (Number)	0	0	11	114	173	33	331
Enrollment in district of residence after Oct 1 (Percent)	0	0	2	.9	1	1.2	1
Enrollment as PA resident after Oct 1 (Number)	0	0	3	65	84	17	169
Enrollment as PA resident after Oct 1 (Percent)	0	0	.5	.5	.5	.6	.5
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	123	1,868	3,335	930	6,256
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	22	14.2	18.4	34.3	18.1
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	87	876	1,659	324	2,946
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	15.6	6.7	9.2	11.9	8.5
Military family (Number)	0	0	2	68	62	10	142
Military family (Percent)	0	0	.4	.5	.3	.4	.4
Homeless (Number)	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0
Foster (Number)	0	0	5	44	111	23	183
Foster (Percent)	0	0	.9	.3	.6	.8	.5
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	8	29	35	18	90
Court/agency placed (Percent)	0	0	1.4	.2	.2	.7	.3
Number of assessed students (Number)	1	0	559	13,151	18,103	2,713	34,527

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–10. Incidence of Presentation Accommodations Received on the Winter Keystone: Algebra I**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	3	N/A	3
Braille format (Percent)	0	N/A	0
Large print format (Number)	11	N/A	11
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	179	350	529
Some test items/questions read aloud (Percent)	.5	2	1
All test items/questions read aloud (Number)	161	393	554
All test items/questions read aloud (Percent)	.4	2.2	1
Test items/questions signed (Number)	0	1	1
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for EL student (Number)	2	12	14
Test items/questions interpreted for EL student (Percent)	0	.1	0
Amplification device (Number)	5	5	10
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	3	3
Magnification device (Percent)	0	0	0
Color overlay (Number)	1	N/A	1
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	8	9	17
Other (per Accommodations Guidelines) (Percent)	0	.1	0
Spanish version (Number)	368	N/A	368
Spanish version (Percent)	1	N/A	.7
Audio (Number)	N/A	598	598
Audio (Percent)	N/A	3.4	1.1
Color Chooser (Number)	N/A	11	11
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	11	11
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast (Number)	N/A	2	2
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	36,213	17,843	54,056

**Table I–11. Incidence of Presentation Accommodations Received on the Winter Keystone: Biology**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	6	N/A	6
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	119	334	453
Some test items/questions read aloud (Percent)	.5	2.2	1.1
All test items/questions read aloud (Number)	130	392	522
All test items/questions read aloud (Percent)	.5	2.6	1.3
Test items/questions signed (Number)	0	2	2
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for EL student (Number)	0	4	4
Test items/questions interpreted for EL student (Percent)	0	0	0
Amplification device (Number)	1	6	7
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	4	4
Magnification device (Percent)	0	0	0
Color overlay (Number)	1	N/A	1
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	6	12	18
Other (per Accommodations Guidelines) (Percent)	0	.1	0
Spanish version (Number)	271	N/A	271
Spanish version (Percent)	1.1	N/A	.7
Audio (Number)	N/A	663	663
Audio (Percent)	N/A	4.4	1.7
Color Chooser (Number)	N/A	18	18
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	18	18
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast (Number)	N/A	7	7
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	24,415	15,001	39,416

**Table I–12. Incidence of Presentation Accommodations Received on the Winter Keystone: Literature**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	9	N/A	9
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	2	N/A	2
Computer Assistive Technology (Percent)	0	N/A	0
Amplification device (Number)	3	4	7
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	2	2
Magnification device (Percent)	0	0	0
Color overlay (Number)	1	N/A	1
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	5	9	14
Other (per Accommodations Guidelines) (Percent)	0	.1	0
Color Chooser (Number)	N/A	17	17
Color Chooser (Percent)	N/A	.1	0
Contrasting Text Chooser (Number)	N/A	17	17
Contrasting Text Chooser (Percent)	N/A	.1	0
Reverse Contrast (Number)	N/A	5	5
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Number of assessed students (Number)	21,365	13,162	34,527

**Table I–13. Incidence of Response Accommodations Received on the Winter Keystone: Algebra I**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	8	0	8
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	8	2	10
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	13	0	13
Test administrator transcribed student responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	5	0	5
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Number)	0	1	1
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	3	N/A	3
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	2	N/A	2
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	1	0	1
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for EL student (Number)	75	50	125
Translation dictionary for EL student (Percent)	.2	.3	.2
Other (per Accommodations Guidelines) (Number)	6	1	7
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	36,213	17,843	54,056



**Table I–14. Incidence of Response Accommodations Received on the Winter Keystone: Biology**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	4	0	4
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	8	2	10
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	7	0	7
Test administrator transcribed student responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	5	0	5
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Number)	1	0	1
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	3	N/A	3
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	0	N/A	0
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	1	0	1
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for EL student (Number)	53	55	108
Translation dictionary for EL student (Percent)	.2	.4	.3
Other (per Accommodations Guidelines) (Number)	5	0	5
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	24,415	15,001	39,416

**Table I–15. Incidence of Response Accommodations Received on the Winter Keystone: Literature**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	4	0	4
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	12	2	14
Test administrator scribed open-ended responses at student's direction (Percent)	.1	0	0
Test administrator transcribed student responses (Number)	30	1	31
Test administrator transcribed student responses (Percent)	.1	0	.1
Keyboard, word processor, or computer (Number)	19	N/A	19
Keyboard, word processor, or computer (Percent)	.1	N/A	.1
Braille/Notetaker (Number)	0	N/A	0
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	1	N/A	1
Computer Assistive Technology (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	2	1	3
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	21,365	13,162	34,527

**Table I–16. Incidence of Setting Accommodations Received on the Winter Keystone: Algebra I**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	38	6	44
One-on-one setting (Percent)	.1	0	.1
Small group setting (Number)	3,611	1,789	5,400
Small group setting (Percent)	10	10	10
Other (per Accommodations Guidelines) (Number)	9	33	42
Other (per Accommodations Guidelines) (Percent)	0	.2	.1
Number of assessed students (Number)	36,213	17,843	54,056

**Table I–17. Incidence of Setting Accommodations Received on the Winter Keystone: Biology**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	1	0	1
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	35	10	45
One-on-one setting (Percent)	.1	.1	.1
Small group setting (Number)	2,619	1,667	4,286
Small group setting (Percent)	10.7	11.1	10.9
Other (per Accommodations Guidelines) (Number)	17	23	40
Other (per Accommodations Guidelines) (Percent)	.1	.2	.1
Number of assessed students (Number)	24,415	15,001	39,416

**Table I–18. Incidence of Setting Accommodations Received on the Winter Keystone: Literature**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	27	16	43
One-on-one setting (Percent)	.1	.1	.1
Small group setting (Number)	2,621	1,694	4,315
Small group setting (Percent)	12.3	12.9	12.5
Other (per Accommodations Guidelines) (Number)	20	33	53
Other (per Accommodations Guidelines) (Percent)	.1	.3	.2
Number of assessed students (Number)	21,365	13,162	34,527

**Table I–19. Incidence of Timing Accommodations Received on the Winter Keystone: Algebra I**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	2,499	1,063	3,562
Extended time (Percent)	6.9	6	6.6
Frequent breaks (Number)	188	337	525
Frequent breaks (Percent)	.5	1.9	1
Changed test schedule (Number)	97	7	104
Changed test schedule (Percent)	.3	0	.2
Other (per Accommodations Guidelines) (Number)	20	1	21
Other (per Accommodations Guidelines) (Percent)	.1	0	0
Number of assessed students (Number)	36,213	17,843	54,056

**Table I–20. Incidence of Timing Accommodations Received on the Winter Keystone: Biology**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	784	712	1,496
Extended time (Percent)	3.2	4.7	3.8
Frequent breaks (Number)	135	343	478
Frequent breaks (Percent)	.6	2.3	1.2
Changed test schedule (Number)	86	3	89
Changed test schedule (Percent)	.4	0	.2
Other (per Accommodations Guidelines) (Number)	1	3	4
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	24,415	15,001	39,416

**Table I–21. Incidence of Timing Accommodations Received on the Winter Keystone: Literature**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	2,310	1,088	3,398
Extended time (Percent)	10.8	8.3	9.8
Frequent breaks (Number)	167	343	510
Frequent breaks (Percent)	.8	2.6	1.5
Changed test schedule (Number)	46	10	56
Changed test schedule (Percent)	.2	.1	.2
Other (per Accommodations Guidelines) (Number)	9	3	12
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	21,365	13,162	34,527

**Table I–22. Accommodation Rate for Non-IEP and IEP Students on the Winter Keystone Exams: Algebra I**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	29,834	14,549	44,383
Non-Accommodated (Number)	27,415	14,089	41,504
Non-Accommodated (Percent)	91.9	96.8	93.5
Accommodated (Number)	2,419	460	2,879
Accommodated (Percent)	8.1	3.2	6.5
IEP Students (Number)	6,379	3,294	9,673
Non-Accommodated (Number)	3,079	1,418	4,497
Non-Accommodated (Percent)	48.3	43	46.5
Accommodated (Number)	3,300	1,876	5,176
Accommodated (Percent)	51.7	57	53.5

**Table I–23. Accommodation Rate for Non-IEP and IEP Students on the Winter Keystone Exams: Biology**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	19,524	11,668	31,192
Non-Accommodated (Number)	18,669	11,429	30,098
Non-Accommodated (Percent)	95.6	98	96.5
Accommodated (Number)	855	239	1,094
Accommodated (Percent)	4.4	2	3.5
IEP Students (Number)	4,891	3,333	8,224
Non-Accommodated (Number)	2,452	1,509	3,961
Non-Accommodated (Percent)	50.1	45.3	48.2
Accommodated (Number)	2,439	1,824	4,263
Accommodated (Percent)	49.9	54.7	51.8

**Table I–24. Accommodation Rate for Non-IEP and IEP Students on the Winter Keystone Exams: Literature**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	16,698	10,100	26,798
Non-Accommodated (Number)	14,659	9,678	24,337
Non-Accommodated (Percent)	87.8	95.8	90.8
Accommodated (Number)	2,039	422	2,461
Accommodated (Percent)	12.2	4.2	9.2
IEP Students (Number)	4,667	3,062	7,729
Non-Accommodated (Number)	2,182	1,389	3,571
Non-Accommodated (Percent)	46.8	45.4	46.2
Accommodated (Number)	2,485	1,673	4,158
Accommodated (Percent)	53.2	54.6	53.8

**Table I–25. Incidence of IEP and EL Students Receiving Accommodations on the Winter Keystone: Algebra I**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Some test items/questions read aloud (Number)	3	3	8	165
PPT - Some test items/questions read aloud (Percent)	1.5	.1	0	2.7
PPT - All test items/questions read aloud (Number)	4	2	2	153
PPT - All test items/questions read aloud (Percent)	2	.1	0	2.5
PPT - Small group setting (Number)	58	145	367	3,041
PPT - Small group setting (Percent)	29.3	7.2	1.3	49.2
PPT - Extended time (Number)	18	156	1,752	573
PPT - Extended time (Percent)	9.1	7.8	6.3	9.3
PPT - Frequent breaks (Number)	2	11	10	165
PPT - Frequent breaks (Percent)	1	.5	0	2.7
PPT - Number assessed (Number)	198	2,005	27,829	6,181
CBT - Some test items/questions read aloud (Number)	7	8	8	327
CBT - Some test items/questions read aloud (Percent)	9	1.7	.1	10.2
CBT - All test items/questions read aloud (Number)	9	2	9	373
CBT - All test items/questions read aloud (Percent)	11.5	.4	.1	11.6
CBT - Small group setting (Number)	27	18	102	1,642
CBT - Small group setting (Percent)	34.6	3.7	.7	51.1
CBT - Extended time (Number)	7	12	311	733
CBT - Extended time (Percent)	9	2.5	2.2	22.8
CBT - Frequent breaks (Number)	4	0	20	313
CBT - Frequent breaks (Percent)	5.1	0	.1	9.7
CBT - Number assessed (Number)	78	484	14,065	3,216
Total - Some test items/questions read aloud (Number)	10	11	16	492
Total - Some test items/questions read aloud (Percent)	3.6	.4	0	5.2
Total - All test items/questions read aloud (Number)	13	4	11	526
Total - All test items/questions read aloud (Percent)	4.7	.2	0	5.6
Total - Small group setting (Number)	85	163	469	4,683
Total - Small group setting (Percent)	30.8	6.5	1.1	49.8
Total - Extended time (Number)	25	168	2,063	1,306
Total - Extended time (Percent)	9.1	6.7	4.9	13.9
Total - Frequent breaks (Number)	6	11	30	478
Total - Frequent breaks (Percent)	2.2	.4	.1	5.1
Total - Number assessed (Number)	276	2,489	41,894	9,397

**Table I–26. Incidence of IEP and EL Students Receiving Accommodations on the Winter Keystone: Biology**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Some test items/questions read aloud (Number)	3	2	1	113
PPT - Some test items/questions read aloud (Percent)	2.3	.2	0	2.4
PPT - All test items/questions read aloud (Number)	6	1	3	120
PPT - All test items/questions read aloud (Percent)	4.6	.1	0	2.5
PPT - Small group setting (Number)	51	96	220	2,252
PPT - Small group setting (Percent)	39.2	7.4	1.2	47.3
PPT - Extended time (Number)	8	56	417	303
PPT - Extended time (Percent)	6.2	4.3	2.3	6.4
PPT - Frequent breaks (Number)	4	6	10	115
PPT - Frequent breaks (Percent)	3.1	.5	.1	2.4
PPT - Number assessed (Number)	130	1,303	18,221	4,761
CBT - Some test items/questions read aloud (Number)	12	5	12	305
CBT - Some test items/questions read aloud (Percent)	13.2	1	.1	9.4
CBT - All test items/questions read aloud (Number)	9	0	11	372
CBT - All test items/questions read aloud (Percent)	9.9	0	.1	11.5
CBT - Small group setting (Number)	23	15	99	1,530
CBT - Small group setting (Percent)	25.3	3.1	.9	47.2
CBT - Extended time (Number)	9	5	91	607
CBT - Extended time (Percent)	9.9	1	.8	18.7
CBT - Frequent breaks (Number)	6	0	26	311
CBT - Frequent breaks (Percent)	6.6	0	.2	9.6
CBT - Number assessed (Number)	91	482	11,186	3,242
Total - Some test items/questions read aloud (Number)	15	7	13	418
Total - Some test items/questions read aloud (Percent)	6.8	.4	0	5.2
Total - All test items/questions read aloud (Number)	15	1	14	492
Total - All test items/questions read aloud (Percent)	6.8	.1	0	6.1
Total - Small group setting (Number)	74	111	319	3,782
Total - Small group setting (Percent)	33.5	6.2	1.1	47.3
Total - Extended time (Number)	17	61	508	910
Total - Extended time (Percent)	7.7	3.4	1.7	11.4
Total - Frequent breaks (Number)	10	6	36	426
Total - Frequent breaks (Percent)	4.5	.3	.1	5.3
Total - Number assessed (Number)	221	1,785	29,407	8,003

**Table I–27. Incidence of IEP and EL Students Receiving Accommodations on the Winter Keystone: Literature**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Small group setting (Number)	54	70	237	2,260
PPT - Small group setting (Percent)	39.7	6.4	1.5	49.9
PPT - Extended time (Number)	16	90	1,664	540
PPT - Extended time (Percent)	11.8	8.2	10.7	11.9
PPT - Frequent breaks (Number)	1	4	19	143
PPT - Frequent breaks (Percent)	.7	.4	.1	3.2
PPT - Number assessed (Number)	136	1,091	15,607	4,531
CBT - Small group setting (Number)	29	25	102	1,538
CBT - Small group setting (Percent)	38.7	5.7	1.1	51.5
CBT - Extended time (Number)	14	11	324	739
CBT - Extended time (Percent)	18.7	2.5	3.4	24.7
CBT - Frequent breaks (Number)	7	0	26	310
CBT - Frequent breaks (Percent)	9.3	0	.3	10.4
CBT - Number assessed (Number)	75	441	9,659	2,987
Total - Small group setting (Number)	83	95	339	3,798
Total - Small group setting (Percent)	39.3	6.2	1.3	50.5
Total - Extended time (Number)	30	101	1,988	1,279
Total - Extended time (Percent)	14.2	6.6	7.9	17
Total - Frequent breaks (Number)	8	4	45	453
Total - Frequent breaks (Percent)	3.8	.3	.2	6
Total - Number assessed (Number)	211	1,532	25,266	7,518



## SUMMER

**Table I–28. Students Assessed on the Summer Keystone: Algebra I**

Description	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	0	0	4	27	34	11	7	0	83
Total number of CBT processed (Number)	0	0	8	35	68	58	25	3	197
Total number of tests processed (Number)	0	0	12	62	102	69	32	3	280
Total number of tests processed with a score (Number)	0	0	12	62	94	66	27	3	264
Total number of tests processed with a score (Percent)	0	0	100	100	92.2	95.7	84.4	100	94.3
Total number of tests processed without a score (Number)	0	0	0	0	8	3	5	0	16
Total number of tests processed without a score (Percent)	0	0	0	0	7.8	4.3	15.6	0	5.7

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table I–29. Students Assessed on the Summer Keystone: Biology**

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	0	21	24	35	7	0	87
Total number of CBT processed (Number)	0	0	57	52	26	4	139
Total number of tests processed (Number)	0	21	81	87	33	4	226
Total number of tests processed with a score (Number)	0	21	74	80	31	4	210
Total number of tests processed with a score (Percent)	0	100	91.4	92	93.9	100	92.9
Total number of tests processed without a score (Number)	0	0	7	7	2	0	16
Total number of tests processed without a score (Percent)	0	0	8.6	8	6.1	0	7.1

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–30. Students Assessed on the Summer Keystone: Literature**

Description	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Total number of PPT processed (Number)	0	0	15	22	8	2	47
Total number of CBT processed (Number)	0	0	1	37	34	1	73
Total number of tests processed (Number)	0	0	16	59	42	3	120
Total number of tests processed with a score (Number)	0	0	16	55	37	3	111
Total number of tests processed with a score (Percent)	0	0	100	93.2	88.1	100	92.5
Total number of tests processed without a score (Number)	0	0	0	4	5	0	9
Total number of tests processed without a score (Percent)	0	0	0	6.8	11.9	0	7.5

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–31. Counts of Students without Scores on the Summer Keystone: Algebra I**

Reason for Non-Assessment	Other*	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	0	0	0	0	1	0	1
Extended absence from school (Percent)	0	0	0	0	0	0	20	0	6.3
Non-attempt (Number)	0	0	0	0	7	2	3	0	12
Non-attempt (Percent)	0	0	0	0	87.5	66.7	60	0	75
Medical emergency (Number)	0	0	0	0	0	0	0	0	0
Medical emergency (Percent)	0	0	0	0	0	0	0	0	0
Parental request - Chapter 4 (Number)	0	0	0	0	0	0	0	0	0
Parental request - Chapter 4 (Percent)	0	0	0	0	0	0	0	0	0
Parental request - Other reasons (Number)	0	0	0	0	0	0	0	0	0
Parental request - Other reasons (Percent)	0	0	0	0	0	0	0	0	0
Other reasons (Number)	0	0	0	0	1	1	1	0	3
Other reasons (Percent)	0	0	0	0	12.5	33.3	20	0	18.8
Total not assessed (Number)	0	0	0	0	8	3	5	0	16

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table I–32. Counts of Students without Scores on the Summer Keystone: Biology**

Reason for Non-Assessment	Other*	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12	Total
Extended absence from school (Number)	0	0	0	0	2	0	2
Extended absence from school (Percent)	0	0	0	0	100	0	12.5
Non-attempt (Number)	0	0	7	5	0	0	12
Non-attempt (Percent)	0	0	100	71.4	0	0	75
Medical emergency (Number)	0	0	0	0	0	0	0
Medical emergency (Percent)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Number)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Percent)	0	0	0	0	0	0	0
Parental request - Other reasons (Number)	0	0	0	0	0	0	0
Parental request - Other reasons (Percent)	0	0	0	0	0	0	0
Other reasons (Number)	0	0	0	2	0	0	2
Other reasons (Percent)	0	0	0	28.6	0	0	12.5
Total not assessed (Number)	0	0	7	7	2	0	16

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–33. Counts of Students without Scores on the Summer Keystone: Literature**

<b>Reason for Non-Assessment</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Extended absence from school (Number)	0	0	0	0	2	0	2
Extended absence from school (Percent)	0	0	0	0	40	0	22.2
Non-attempt (Number)	0	0	0	3	2	0	5
Non-attempt (Percent)	0	0	0	75	40	0	55.6
EL in first year in U.S. schools (Number)	0	0	0	0	0	0	0
EL in first year in U.S. schools (Percent)	0	0	0	0	0	0	0
Medical emergency (Number)	0	0	0	0	0	0	0
Medical emergency (Percent)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Number)	0	0	0	0	0	0	0
Parental request - Chapter 4 (Percent)	0	0	0	0	0	0	0
Parental request - Other reasons (Number)	0	0	0	0	0	0	0
Parental request - Other reasons (Percent)	0	0	0	0	0	0	0
Other reasons (Number)	0	0	0	1	1	0	2
Other reasons (Percent)	0	0	0	25	20	0	22.2
Total not assessed (Number)	0	0	0	4	5	0	9

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–34. Demographic Characteristics of Students taking the Summer Keystone: Algebra I**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	0	0	7	38	48	34	14	1	142
Female (Percent)	0	0	58.3	61.3	51.1	51.5	51.9	33.3	53.8
Male (Number)	0	0	5	23	46	32	13	2	121
Male (Percent)	0	0	41.7	37.1	48.9	48.5	48.1	66.7	45.8
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	0	1	0	0	0	1
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	0	1.1	0	0	0	.4
Asian (not Hispanic) (Number)	0	0	6	3	4	1	3	0	17
Asian (not Hispanic) (Percent)	0	0	50	4.8	4.3	1.5	11.1	0	6.4
Black or African American (not Hispanic) (Number)	0	0	0	3	9	7	0	0	19
Black or African American (not Hispanic) (Percent)	0	0	0	4.8	9.6	10.6	0	0	7.2
Hispanic (any race) (Number)	0	0	1	6	8	5	2	0	22
Hispanic (any race) (Percent)	0	0	8.3	9.7	8.5	7.6	7.4	0	8.3
Multi-Racial (not Hispanic) (Number)	0	0	0	1	3	4	0	1	9
Multi-Racial (not Hispanic) (Percent)	0	0	0	1.6	3.2	6.1	0	33.3	3.4
White (not Hispanic) (Number)	0	0	5	48	69	49	22	2	195
White (not Hispanic) (Percent)	0	0	41.7	77.4	73.4	74.2	81.5	66.7	73.9
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	0	0	0	0	0
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	0	0	0	0	0
IEP (not gifted) (Number)	0	0	1	3	9	7	2	0	22
IEP (not gifted) (Percent)	0	0	8.3	4.8	9.6	10.6	7.4	0	8.3
Student exited IEP in last 2 years (Number)	0	0	0	2	1	0	0	0	3
Student exited IEP in last 2 years (Percent)	0	0	0	3.2	1.1	0	0	0	1.1
Title I (Number)	0	0	0	3	9	10	2	0	24
Title I (Percent)	0	0	0	4.8	9.6	15.2	7.4	0	9.1
Title III served (Number)	0	0	0	0	0	1	0	0	1
Title III served (Percent)	0	0	0	0	0	1.5	0	0	.4
Title III not served (Number)	0	0	0	4	27	8	8	0	47
Title III not served (Percent)	0	0	0	6.5	28.7	12.1	29.6	0	17.8
Migrant student (Number)	0	0	0	0	0	0	0	0	0
Migrant student (Percent)	0	0	0	0	0	0	0	0	0
EL enrolled first year (Number)	0	0	0	0	0	0	0	0	0
EL enrolled first year (Percent)	0	0	0	0	0	0	0	0	0
EL enrolled not first year (Number)	0	0	0	1	0	1	0	0	2
EL enrolled not first year (Percent)	0	0	0	1.6	0	1.5	0	0	.8
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	0	0	0	0	0	0
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	0	0	0	0	0	0

**Table I–34 (continued). Demographic Characteristics of Students taking the Summer Keystone: Algebra**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	0	0	0	0	0	0
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	0	0	0	0	0	0
Former EL no longer monitored (Number)	0	0	2	0	2	0	0	0	4
Former EL no longer monitored (Percent)	0	0	16.7	0	2.1	0	0	0	1.5
LIFE first year (Number)	0	0	0	0	0	0	0	0	0
LIFE first year (Percent)	0	0	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	0	0	0	0	0	0	0
LIFE not first year (Percent)	0	0	0	0	0	0	0	0	0
Former EL exited and in 3rd year of monitoring (Number)	0	0	0	0	1	0	0	0	1
Former EL exited and in 3rd year of monitoring (Percent)	0	0	0	0	1.1	0	0	0	.4
Former EL exited and in 4th year of monitoring (Number)	0	0	0	1	0	0	0	0	1
Former EL exited and in 4th year of monitoring (Percent)	0	0	0	1.6	0	0	0	0	.4
Foreign exchange student (Number)	0	0	0	0	0	0	0	0	0
Foreign exchange student (Percent)	0	0	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	0	4	8	39	21	12	1	85
Economically disadvantaged (Percent)	0	0	33.3	12.9	41.5	31.8	44.4	33.3	32.2
Historically Underperforming Subgroup (Number)	0	0	5	11	44	25	13	1	99
Historically Underperforming Subgroup (Percent)	0	0	41.7	17.7	46.8	37.9	48.1	33.3	37.5
Enrollment in school of residence after Oct 1 (Number)	0	0	0	0	1	1	0	0	2
Enrollment in school of residence after Oct 1 (Percent)	0	0	0	0	1.1	1.5	0	0	.8
Enrollment in district of residence after Oct 1 (Number)	0	0	0	0	1	0	0	0	1
Enrollment in district of residence after Oct 1 (Percent)	0	0	0	0	1.1	0	0	0	.4
Enrollment as PA resident after Oct 1 (Number)	0	0	0	0	1	0	0	0	1
Enrollment as PA resident after Oct 1 (Percent)	0	0	0	0	1.1	0	0	0	.4
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	0	1	45	1	1	0	48
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	0	1.6	47.9	1.5	3.7	0	18.2
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	0	1	6	1	1	0	9
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	0	1.6	6.4	1.5	3.7	0	3.4
Military family (Number)	0	0	0	1	0	0	0	0	1
Military family (Percent)	0	0	0	1.6	0	0	0	0	.4
Homeless (Number)	0	0	0	0	0	0	0	0	0

**Table I–34 (continued). Demographic Characteristics of Students taking the Summer Keystone: Algebra**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.6</b>	<b>Gr.7</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Homeless (Percent)	0	0	0	0	0	0	0	0	0
Foster (Number)	0	0	0	0	1	0	0	0	1
Foster (Percent)	0	0	0	0	1.1	0	0	0	.4
Home schooled (Number)	0	0	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	0	0	0	0	0
Court/agency placed (Percent)	0	0	0	0	0	0	0	0	0
Number of assessed students (Number)	0	0	12	62	94	66	27	3	264

\*Other combines students coded as (1) below Grade 6, (2) ungraded, or (3) without a coded grade

**Table I–35. Demographic Characteristics of Students taking the Summer Keystone: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	0	10	38	44	14	0	106
Female (Percent)	0	47.6	51.4	55	45.2	0	50.5
Male (Number)	0	11	36	36	17	4	104
Male (Percent)	0	52.4	48.6	45	54.8	100	49.5
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	1	1	0	0	2
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	1.4	1.3	0	0	1
Asian (not Hispanic) (Number)	0	8	7	0	2	0	17
Asian (not Hispanic) (Percent)	0	38.1	9.5	0	6.5	0	8.1
Black or African American (not Hispanic) (Number)	0	1	4	14	3	0	22
Black or African American (not Hispanic) (Percent)	0	4.8	5.4	17.5	9.7	0	10.5
Hispanic (any race) (Number)	0	2	4	7	1	0	14
Hispanic (any race) (Percent)	0	9.5	5.4	8.8	3.2	0	6.7
Multi-Racial (not Hispanic) (Number)	0	3	4	5	2	0	14
Multi-Racial (not Hispanic) (Percent)	0	14.3	5.4	6.3	6.5	0	6.7
White (not Hispanic) (Number)	0	7	54	53	23	4	141
White (not Hispanic) (Percent)	0	33.3	73	66.3	74.2	100	67.1
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	0	0	0
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	0	0	0
IEP (not gifted) (Number)	0	2	13	13	5	0	33
IEP (not gifted) (Percent)	0	9.5	17.6	16.3	16.1	0	15.7
Student exited IEP in last 2 years (Number)	0	0	2	1	0	0	3
Student exited IEP in last 2 years (Percent)	0	0	2.7	1.3	0	0	1.4
Title I (Number)	0	0	5	18	1	0	24
Title I (Percent)	0	0	6.8	22.5	3.2	0	11.4
Title III served (Number)	0	0	2	3	1	0	6
Title III served (Percent)	0	0	2.7	3.8	3.2	0	2.9
Title III not served (Number)	0	0	27	12	7	0	46
Title III not served (Percent)	0	0	36.5	15	22.6	0	21.9
Migrant student (Number)	0	0	0	0	0	0	0
Migrant student (Percent)	0	0	0	0	0	0	0
EL enrolled first year (Number)	0	0	0	0	0	0	0
EL enrolled first year (Percent)	0	0	0	0	0	0	0
EL enrolled not first year (Number)	0	0	2	3	1	0	6
EL enrolled not first year (Percent)	0	0	2.7	3.8	3.2	0	2.9
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	0	0	0	0
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	0	0	0	0
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	0	0	0	0

**Table I–35 (continued). Demographic Characteristics of Students taking the Summer Keystone: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	0	0	0	0
Former EL no longer monitored (Number)	0	1	1	1	0	0	3
Former EL no longer monitored (Percent)	0	4.8	1.4	1.3	0	0	1.4
LIFE first year (Number)	0	0	0	0	0	0	0
LIFE first year (Percent)	0	0	0	0	0	0	0
LIFE not first year (Number)	0	0	0	0	0	0	0
LIFE not first year (Percent)	0	0	0	0	0	0	0
Former EL exited and in 3rd year of monitoring (Number)	0	0	1	0	0	0	1
Former EL exited and in 3rd year of monitoring (Percent)	0	0	1.4	0	0	0	.5
Former EL exited and in 4th year of monitoring (Number)	0	0	0	0	0	0	0
Former EL exited and in 4th year of monitoring (Percent)	0	0	0	0	0	0	0
Foreign exchange student (Number)	0	0	0	0	0	0	0
Foreign exchange student (Percent)	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	1	27	34	12	0	74
Economically disadvantaged (Percent)	0	4.8	36.5	42.5	38.7	0	35.2
Historically Underperforming Subgroup (Number)	0	2	37	42	15	0	96
Historically Underperforming Subgroup (Percent)	0	9.5	50	52.5	48.4	0	45.7
Enrollment in school of residence after Oct 1 (Number)	0	0	0	3	1	0	4
Enrollment in school of residence after Oct 1 (Percent)	0	0	0	3.8	3.2	0	1.9
Enrollment in district of residence after Oct 1 (Number)	0	0	0	2	0	0	2
Enrollment in district of residence after Oct 1 (Percent)	0	0	0	2.5	0	0	1
Enrollment as PA resident after Oct 1 (Number)	0	0	0	1	0	0	1
Enrollment as PA resident after Oct 1 (Percent)	0	0	0	1.3	0	0	.5
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	38	3	2	0	43
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	51.4	3.8	6.5	0	20.5
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	2	1	1	0	4
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	2.7	1.3	3.2	0	1.9
Military family (Number)	0	0	0	0	0	0	0
Military family (Percent)	0	0	0	0	0	0	0
Homeless (Number)	0	0	0	0	0	0	0
Homeless (Percent)	0	0	0	0	0	0	0
Foster (Number)	0	0	0	0	0	0	0
Foster (Percent)	0	0	0	0	0	0	0



**Table I–35 (continued). Demographic Characteristics of Students taking the Summer Keystone: Biology**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	0	0	0
Court/agency placed (Percent)	0	0	0	0	0	0	0
Number of assessed students (Number)	0	21	74	80	31	4	210

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–36. Demographic Characteristics of Students taking the Summer Keystone: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Female (Number)	0	0	4	29	20	2	55
Female (Percent)	0	0	25	52.7	54.1	66.7	49.5
Male (Number)	0	0	12	26	17	1	56
Male (Percent)	0	0	75	47.3	45.9	33.3	50.5
American Indian/Alaskan Native (not Hispanic) (Number)	0	0	0	1	0	0	1
American Indian/Alaskan Native (not Hispanic) (Percent)	0	0	0	1.8	0	0	.9
Asian (not Hispanic) (Number)	0	0	1	1	0	0	2
Asian (not Hispanic) (Percent)	0	0	6.3	1.8	0	0	1.8
Black or African American (not Hispanic) (Number)	0	0	1	11	7	1	20
Black or African American (not Hispanic) (Percent)	0	0	6.3	20	18.9	33.3	18
Hispanic (any race) (Number)	0	0	1	8	6	1	16
Hispanic (any race) (Percent)	0	0	6.3	14.5	16.2	33.3	14.4
Multi-Racial (not Hispanic) (Number)	0	0	0	3	1	0	4
Multi-Racial (not Hispanic) (Percent)	0	0	0	5.5	2.7	0	3.6
White (not Hispanic) (Number)	0	0	13	31	23	1	68
White (not Hispanic) (Percent)	0	0	81.3	56.4	62.2	33.3	61.3
Native Hawaiian or Other Pacific Islander (not Hispanic) (Number)	0	0	0	0	0	0	0
Native Hawaiian or Other Pacific Islander (not Hispanic) (Percent)	0	0	0	0	0	0	0
IEP (not gifted) (Number)	0	0	7	13	9	1	30
IEP (not gifted) (Percent)	0	0	43.8	23.6	24.3	33.3	27
Student exited IEP in last 2 years (Number)	0	0	0	0	0	0	0
Student exited IEP in last 2 years (Percent)	0	0	0	0	0	0	0
Title I (Number)	0	0	0	7	1	0	8
Title I (Percent)	0	0	0	12.7	2.7	0	7.2
Title III served (Number)	0	0	0	2	2	0	4
Title III served (Percent)	0	0	0	3.6	5.4	0	3.6
Title III not served (Number)	0	0	1	7	7	0	15
Title III not served (Percent)	0	0	6.3	12.7	18.9	0	13.5
Migrant student (Number)	0	0	0	0	0	0	0
Migrant student (Percent)	0	0	0	0	0	0	0
EL enrolled first year (Number)	0	0	0	0	1	0	1
EL enrolled first year (Percent)	0	0	0	0	2.7	0	.9
EL enrolled not first year (Number)	0	0	0	2	0	0	2
EL enrolled not first year (Percent)	0	0	0	3.6	0	0	1.8
Exited ESL/bilingual program and in first year of monitoring (Number)	0	0	0	0	0	0	0
Exited ESL/bilingual program and in first year of monitoring (Percent)	0	0	0	0	0	0	0

**Table I–36 (continued). Demographic Characteristics of Students taking the Summer Keystone: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Exited ESL/bilingual program and in 2nd year of monitoring (Number)	0	0	0	0	0	0	0
Exited ESL/bilingual program and in 2nd year of monitoring (Percent)	0	0	0	0	0	0	0
Former EL no longer monitored (Number)	0	0	0	0	0	0	0
Former EL no longer monitored (Percent)	0	0	0	0	0	0	0
LIFE first year (Number)	0	0	0	0	1	0	1
LIFE first year (Percent)	0	0	0	0	2.7	0	.9
LIFE not first year (Number)	0	0	0	0	0	0	0
LIFE not first year (Percent)	0	0	0	0	0	0	0
Former EL exited and in 3rd year of monitoring (Number)	0	0	0	0	0	0	0
Former EL exited and in 3rd year of monitoring (Percent)	0	0	0	0	0	0	0
Former EL exited and in 4th year of monitoring (Number)	0	0	0	0	0	0	0
Former EL exited and in 4th year of monitoring (Percent)	0	0	0	0	0	0	0
Foreign exchange student (Number)	0	0	0	0	0	0	0
Foreign exchange student (Percent)	0	0	0	0	0	0	0
Economically disadvantaged (Number)	0	0	6	22	16	2	46
Economically disadvantaged (Percent)	0	0	37.5	40	43.2	66.7	41.4
Historically Underperforming Subgroup (Number)	0	0	11	31	21	2	65
Historically Underperforming Subgroup (Percent)	0	0	68.8	56.4	56.8	66.7	58.6
Enrollment in school of residence after Oct 1 (Number)	0	0	0	1	2	0	3
Enrollment in school of residence after Oct 1 (Percent)	0	0	0	1.8	5.4	0	2.7
Enrollment in district of residence after Oct 1 (Number)	0	0	0	1	2	0	3
Enrollment in district of residence after Oct 1 (Percent)	0	0	0	1.8	5.4	0	2.7
Enrollment as PA resident after Oct 1 (Number)	0	0	0	1	2	0	3
Enrollment as PA resident after Oct 1 (Percent)	0	0	0	1.8	5.4	0	2.7
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	0	3	2	0	5
Enrollment in school of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	0	5.5	5.4	0	4.5
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Number)	0	0	0	1	1	0	2
Enrollment in district of residence after previous Oct 1 but on/before current Oct 1 (Percent)	0	0	0	1.8	2.7	0	1.8
Military family (Number)	0	0	0	0	0	0	0
Military family (Percent)	0	0	0	0	0	0	0
Homeless (Number)	0	0	0	0	0	0	0

**Table I–36 (continued). Demographic Characteristics of Students taking the Summer Keystone: Literature**

<b>Demographic or Educational Characteristic</b>	<b>Other*</b>	<b>Gr.8</b>	<b>Gr.9</b>	<b>Gr.10</b>	<b>Gr.11</b>	<b>Gr.12</b>	<b>Total</b>
Homeless (Percent)	0	0	0	0	0	0	0
Foster (Number)	0	0	0	0	0	0	0
Foster (Percent)	0	0	0	0	0	0	0
Home schooled (Number)	0	0	0	0	0	0	0
Home schooled (Percent)	0	0	0	0	0	0	0
Court/agency placed (Number)	0	0	0	0	0	0	0
Court/agency placed (Percent)	0	0	0	0	0	0	0
Number of assessed students (Number)	0	0	16	55	37	3	111

\*Other combines students coded as (1) below Grade 8, (2) ungraded, or (3) without a coded grade

**Table I–37. Incidence of Presentation Accommodations Received on the Summer Keystone: Algebra I**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	0	N/A	0
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	0	1	1
Some test items/questions read aloud (Percent)	0	.5	.4
All test items/questions read aloud (Number)	0	0	0
All test items/questions read aloud (Percent)	0	0	0
Test items/questions signed (Number)	0	0	0
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for EL student (Number)	0	0	0
Test items/questions interpreted for EL student (Percent)	0	0	0
Amplification device (Number)	0	0	0
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	0	0
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Spanish version (Number)	0	N/A	0
Spanish version (Percent)	0	N/A	0
Audio (Number)	N/A	1	1
Audio (Percent)	N/A	.5	.4
Color Chooser (Number)	N/A	0	0
Color Chooser (Percent)	N/A	0	0
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Reverse Contrast (Number)	N/A	0	0
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	78	186	264

**Table I–38. Incidence of Presentation Accommodations Received on the Summer Keystone: Biology**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	0	N/A	0
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Some test items/questions read aloud (Number)	0	1	1
Some test items/questions read aloud (Percent)	0	.8	.5
All test items/questions read aloud (Number)	0	0	0
All test items/questions read aloud (Percent)	0	0	0
Test items/questions signed (Number)	0	0	0
Test items/questions signed (Percent)	0	0	0
Test items/questions interpreted for EL student (Number)	0	0	0
Test items/questions interpreted for EL student (Percent)	0	0	0
Amplification device (Number)	0	0	0
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	0	0
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Spanish version (Number)	0	N/A	0
Spanish version (Percent)	0	N/A	0
Audio (Number)	N/A	1	1
Audio (Percent)	N/A	.8	.5
Color Chooser (Number)	N/A	0	0
Color Chooser (Percent)	N/A	0	0
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Reverse Contrast (Number)	N/A	0	0
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Video Sign Language (Number)	N/A	0	0
Video Sign Language (Percent)	N/A	0	0
Number of assessed students (Number)	83	127	210

**Table I–39. Incidence of Presentation Accommodations Received on the Summer Keystone: Literature**

Type of Presentation Accommodation	PPT	CBT	Total
Braille format (Number)	0	N/A	0
Braille format (Percent)	0	N/A	0
Large print format (Number)	0	N/A	0
Large print format (Percent)	0	N/A	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Amplification device (Number)	0	0	0
Amplification device (Percent)	0	0	0
Magnification device (Number)	0	0	0
Magnification device (Percent)	0	0	0
Color overlay (Number)	0	N/A	0
Color overlay (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Color Chooser (Number)	N/A	0	0
Color Chooser (Percent)	N/A	0	0
Contrasting Text Chooser (Number)	N/A	0	0
Contrasting Text Chooser (Percent)	N/A	0	0
Reverse Contrast (Number)	N/A	0	0
Reverse Contrast (Percent)	N/A	0	0
Refreshable Braille (Number)	N/A	0	0
Refreshable Braille (Percent)	N/A	0	0
Number of assessed students (Number)	42	69	111

**Table I–40. Incidence of Response Accommodations Received on the Summer Keystone: Algebra I**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	0	0	0
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	0	0	0
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	0	0	0
Test administrator transcribed student responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	0	N/A	0
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	0	N/A	0
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for EL student (Number)	0	0	0
Translation dictionary for EL student (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	78	186	264



**Table I–41. Incidence of Response Accommodations Received on the Summer Keystone: Biology**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	0	0	0
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	0	0	0
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	0	0	0
Test administrator transcribed student responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed student's signed responses (Percent)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Number)	0	0	0
Qualified interpreter translated, transcribed, and/or scribed EL student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	0	N/A	0
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	0	N/A	0
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Translation dictionary for EL student (Number)	0	0	0
Translation dictionary for EL student (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	83	127	210

**Table I–42. Incidence of Response Accommodations Received on the Summer Keystone: Literature**

Type of Response Accommodation	PPT	CBT	Total
Test administrator marked multiple-choice responses at student's direction (Number)	0	0	0
Test administrator marked multiple-choice responses at student's direction (Percent)	0	0	0
Test administrator scribed open-ended responses at student's direction (Number)	0	0	0
Test administrator scribed open-ended responses at student's direction (Percent)	0	0	0
Test administrator transcribed student responses (Number)	0	0	0
Test administrator transcribed student responses (Percent)	0	0	0
Keyboard, word processor, or computer (Number)	0	N/A	0
Keyboard, word processor, or computer (Percent)	0	N/A	0
Braille/Notetaker (Number)	0	N/A	0
Braille/Notetaker (Percent)	0	N/A	0
Augmentative communication device (Number)	0	0	0
Augmentative communication device (Percent)	0	0	0
Computer Assistive Technology (Number)	0	N/A	0
Computer Assistive Technology (Percent)	0	N/A	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	42	69	111

**Table I–43. Incidence of Setting Accommodations Received on the Summer Keystone: Algebra I**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	0	0	0
One-on-one setting (Percent)	0	0	0
Small group setting (Number)	3	2	5
Small group setting (Percent)	3.8	1.1	1.9
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	78	186	264

**Table I–44. Incidence of Setting Accommodations Received on the Summer Keystone: Biology**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	0	0	0
One-on-one setting (Percent)	0	0	0
Small group setting (Number)	7	2	9
Small group setting (Percent)	8.4	1.6	4.3
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	83	127	210

**Table I–45. Incidence of Setting Accommodations Received on the Summer Keystone: Literature**

Type of Setting Accommodation	PPT	CBT	Total
Hospital/home setting (Number)	0	0	0
Hospital/home setting (Percent)	0	0	0
One-on-one setting (Number)	0	0	0
One-on-one setting (Percent)	0	0	0
Small group setting (Number)	10	3	13
Small group setting (Percent)	23.8	4.3	11.7
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	42	69	111

**Table I–46. Incidence of Timing Accommodations Received on the Summer Keystone: Algebra I**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	15	0	15
Extended time (Percent)	19.2	0	5.7
Frequent breaks (Number)	0	0	0
Frequent breaks (Percent)	0	0	0
Changed test schedule (Number)	0	0	0
Changed test schedule (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	78	186	264

**Table I–47. Incidence of Timing Accommodations Received on the Summer Keystone: Biology**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	2	0	2
Extended time (Percent)	2.4	0	1
Frequent breaks (Number)	0	0	0
Frequent breaks (Percent)	0	0	0
Changed test schedule (Number)	0	0	0
Changed test schedule (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	83	127	210

**Table I–48. Incidence of Timing Accommodations Received on the Summer Keystone: Literature**

Type of Timing Accommodation	PPT	CBT	Total
Extended time (Number)	7	1	8
Extended time (Percent)	16.7	1.4	7.2
Frequent breaks (Number)	0	2	2
Frequent breaks (Percent)	0	2.9	1.8
Changed test schedule (Number)	0	0	0
Changed test schedule (Percent)	0	0	0
Other (per Accommodations Guidelines) (Number)	0	0	0
Other (per Accommodations Guidelines) (Percent)	0	0	0
Number of assessed students (Number)	42	69	111

**Table I–49. Accommodation Rate for Non-IEP and IEP Students on the Summer Keystone Exams: Algebra I**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	71	171	242
Non-Accommodated (Number)	58	170	228
Non-Accommodated (Percent)	81.7	99.4	94.2
Accommodated (Number)	13	1	14
Accommodated (Percent)	18.3	.6	5.8
IEP Students (Number)	7	15	22
Non-Accommodated (Number)	3	14	17
Non-Accommodated (Percent)	42.9	93.3	77.3
Accommodated (Number)	4	1	5
Accommodated (Percent)	57.1	6.7	22.7

**Table I–50. Accommodation Rate for Non-IEP and IEP Students on the Summer Keystone Exams: Biology**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	65	112	177
Non-Accommodated (Number)	65	111	176
Non-Accommodated (Percent)	100	99.1	99.4
Accommodated (Number)	0	1	1
Accommodated (Percent)	0	.9	.6
IEP Students (Number)	18	15	33
Non-Accommodated (Number)	11	14	25
Non-Accommodated (Percent)	61.1	93.3	75.8
Accommodated (Number)	7	1	8
Accommodated (Percent)	38.9	6.7	24.2

**Table I–51. Accommodation Rate for Non-IEP and IEP Students on the Summer Keystone Exams: Literature**

<b>Student Subgroup Tested</b>	<b>PPT</b>	<b>CBT</b>	<b>Total</b>
Non-IEP Students (Number)	29	52	81
Non-Accommodated (Number)	27	52	79
Non-Accommodated (Percent)	93.1	100	97.5
Accommodated (Number)	2	0	2
Accommodated (Percent)	6.9	0	2.5
IEP Students (Number)	13	17	30
Non-Accommodated (Number)	4	14	18
Non-Accommodated (Percent)	30.8	82.4	60
Accommodated (Number)	9	3	12
Accommodated (Percent)	69.2	17.6	40

**Table I–52. Incidence of IEP and EL Students Receiving Accommodations on the Summer Keystone: Algebra I**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Some test items/questions read aloud (Number)	0	0	0	0
PPT - Some test items/questions read aloud (Percent)	0	0	0	0
PPT - All test items/questions read aloud (Number)	0	0	0	0
PPT - All test items/questions read aloud (Percent)	0	0	0	0
PPT - Small group setting (Number)	0	0	0	3
PPT - Small group setting (Percent)	0	0	0	42.9
PPT - Extended time (Number)	0	0	13	2
PPT - Extended time (Percent)	0	0	18.6	28.6
PPT - Frequent breaks (Number)	0	0	0	0
PPT - Frequent breaks (Percent)	0	0	0	0
PPT - Number assessed (Number)	0	1	70	7
CBT - Some test items/questions read aloud (Number)	0	0	0	1
CBT - Some test items/questions read aloud (Percent)	0	0	0	6.7
CBT - All test items/questions read aloud (Number)	0	0	0	0
CBT - All test items/questions read aloud (Percent)	0	0	0	0
CBT - Small group setting (Number)	0	0	1	1
CBT - Small group setting (Percent)	0	0	.6	6.7
CBT - Extended time (Number)	0	0	0	0
CBT - Extended time (Percent)	0	0	0	0
CBT - Frequent breaks (Number)	0	0	0	0
CBT - Frequent breaks (Percent)	0	0	0	0
CBT - Number assessed (Number)	0	1	170	15
Total - Some test items/questions read aloud (Number)	0	0	0	1
Total - Some test items/questions read aloud (Percent)	0	0	0	4.5
Total - All test items/questions read aloud (Number)	0	0	0	0
Total - All test items/questions read aloud (Percent)	0	0	0	0
Total - Small group setting (Number)	0	0	1	4
Total - Small group setting (Percent)	0	0	.4	18.2
Total - Extended time (Number)	0	0	13	2
Total - Extended time (Percent)	0	0	5.4	9.1
Total - Frequent breaks (Number)	0	0	0	0
Total - Frequent breaks (Percent)	0	0	0	0
Total - Number assessed (Number)	0	2	240	22

**Table I–53. Incidence of IEP and EL Students Receiving Accommodations on the Summer Keystone: Biology**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Some test items/questions read aloud (Number)	0	0	0	0
PPT - Some test items/questions read aloud (Percent)	0	0	0	0
PPT - All test items/questions read aloud (Number)	0	0	0	0
PPT - All test items/questions read aloud (Percent)	0	0	0	0
PPT - Small group setting (Number)	0	0	0	7
PPT - Small group setting (Percent)	0	0	0	38.9
PPT - Extended time (Number)	0	0	0	2
PPT - Extended time (Percent)	0	0	0	11.1
PPT - Frequent breaks (Number)	0	0	0	0
PPT - Frequent breaks (Percent)	0	0	0	0
PPT - Number assessed (Number)	0	4	61	18
CBT - Some test items/questions read aloud (Number)	0	0	0	1
CBT - Some test items/questions read aloud (Percent)	0	0	0	6.7
CBT - All test items/questions read aloud (Number)	0	0	0	0
CBT - All test items/questions read aloud (Percent)	0	0	0	0
CBT - Small group setting (Number)	0	0	1	1
CBT - Small group setting (Percent)	0	0	.9	6.7
CBT - Extended time (Number)	0	0	0	0
CBT - Extended time (Percent)	0	0	0	0
CBT - Frequent breaks (Number)	0	0	0	0
CBT - Frequent breaks (Percent)	0	0	0	0
CBT - Number assessed (Number)	0	2	110	15
Total - Some test items/questions read aloud (Number)	0	0	0	1
Total - Some test items/questions read aloud (Percent)	0	0	0	3
Total - All test items/questions read aloud (Number)	0	0	0	0
Total - All test items/questions read aloud (Percent)	0	0	0	0
Total - Small group setting (Number)	0	0	1	8
Total - Small group setting (Percent)	0	0	.6	24.2
Total - Extended time (Number)	0	0	0	2
Total - Extended time (Percent)	0	0	0	6.1
Total - Frequent breaks (Number)	0	0	0	0
Total - Frequent breaks (Percent)	0	0	0	0
Total - Number assessed (Number)	0	6	171	33

**Table I–54. Incidence of IEP and EL Students Receiving Accommodations on the Summer Keystone: Literature**

<b>Accommodation Received by Administration Mode</b>	<b>Both IEP and EL</b>	<b>EL and non-IEP</b>	<b>General Education (non-IEP or EL)</b>	<b>IEP and non-EL</b>
PPT - Small group setting (Number)	0	0	1	9
PPT - Small group setting (Percent)	0	0	3.7	69.2
PPT - Extended time (Number)	0	0	1	6
PPT - Extended time (Percent)	0	0	3.7	46.2
PPT - Frequent breaks (Number)	0	0	0	0
PPT - Frequent breaks (Percent)	0	0	0	0
PPT - Number assessed (Number)	0	2	27	13
CBT - Small group setting (Number)	0	0	0	3
CBT - Small group setting (Percent)	0	0	0	17.6
CBT - Extended time (Number)	0	0	0	1
CBT - Extended time (Percent)	0	0	0	5.9
CBT - Frequent breaks (Number)	0	0	0	2
CBT - Frequent breaks (Percent)	0	0	0	11.8
CBT - Number assessed (Number)	0	1	51	17
Total - Small group setting (Number)	0	0	1	12
Total - Small group setting (Percent)	0	0	1.3	40
Total - Extended time (Number)	0	0	1	7
Total - Extended time (Percent)	0	0	1.3	23.3
Total - Frequent breaks (Number)	0	0	0	2
Total - Frequent breaks (Percent)	0	0	0	6.7
Total - Number assessed (Number)	0	3	78	30



## APPENDIX J: ITEM STATISTICS

Appendix J presents item statistics for each item type (multiple-choice and constructed-response) by content area and administration (winter, spring, and summer). Form values of 0 represent items that appeared on all forms. It is important to note that only the spring forms have embedded field-test items (see Chapter Five for additional information).

**Table J–1. Item Statistics**

Column Heading	Definition
Ref	Reference line number
PubID	Item ID
Form	Test form
<i>N</i>	Number of students
PVal	P-Value
P()	Proportion selecting given response ('-' represents omitted responses and '*' represents multiple responses)
ITCorr	Item-total correlation
Corr()	Correlation of options/points and total test score
Meas	Rasch item difficulty measure estimate
SEM	Standard error of Rasch item difficulty measure estimate
z-Infit	z infit statistic
MS-Infit	Mean square infit statistic
z-Outfit	z outfit statistic
MS-Outfit	Mean square outfit statistic
M/F	Male/Female DIF code
W/B	White/Black DIF code
W/H	White/Hispanic DIF code
O/P	Online computer-based test/paper-pencil-based test DIF code

## MULTIPLE-CHOICE ITEMS

Table J–2. Algebra I Multiple-Choice Item Statistics: Winter

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	274135	0	54824	0.72	0.72	0.09	0.12	0.07	0.00	0.00	0.38	0.38	-0.23	-0.23	-0.12	-1.10	0.01	-9.90	0.85	-9.90	0.79
2	278794	0	54824	0.48	0.18	0.16	0.18	0.48	0.00	0.00	0.30	-0.10	-0.23	-0.06	0.30	0.12	0.01	9.90	1.07	9.90	1.09
3	303331	0	54824	0.41	0.12	0.14	0.33	0.41	0.00	0.00	0.34	-0.20	-0.21	-0.06	0.34	0.21	0.01	-0.40	1.00	3.54	1.02
4	324099	0	54824	0.36	0.08	0.24	0.36	0.32	0.01	0.00	0.16	-0.10	0.05	0.16	-0.13	1.05	0.01	9.90	1.43	9.90	1.67
5	324137	0	54824	0.44	0.19	0.44	0.28	0.09	0.01	0.00	0.40	-0.09	0.40	-0.24	-0.17	0.19	0.01	-8.69	0.97	-6.80	0.97
6	336720	0	54824	0.41	0.41	0.14	0.29	0.15	0.01	0.00	0.33	0.33	-0.18	-0.15	-0.06	0.46	0.01	9.90	1.08	9.90	1.11
7	336778	0	54824	0.29	0.20	0.24	0.26	0.29	0.01	0.00	0.31	-0.16	-0.16	-0.01	0.31	0.83	0.01	4.16	1.02	9.18	1.07
8	336787	0	54824	0.24	0.24	0.21	0.33	0.21	0.01	0.00	0.22	0.22	-0.09	-0.05	-0.07	0.95	0.01	5.68	1.03	9.90	1.12
9	399594	0	54824	0.31	0.24	0.31	0.20	0.23	0.01	0.00	0.22	-0.08	0.22	-0.11	-0.03	1.22	0.01	9.90	1.36	9.90	1.55
10	400466	0	54824	0.25	0.25	0.20	0.40	0.14	0.00	0.00	0.30	0.30	-0.08	-0.14	-0.08	1.22	0.01	9.90	1.07	9.90	1.21
11	416431	0	54824	0.59	0.59	0.19	0.12	0.10	0.00	0.00	0.40	0.40	-0.26	-0.20	-0.08	-0.29	0.01	-9.90	0.92	-9.90	0.92
12	416616	0	54824	0.45	0.14	0.22	0.45	0.18	0.01	0.00	0.26	-0.14	-0.12	0.26	-0.06	0.41	0.01	9.90	1.17	9.90	1.22
13	417142	0	54824	0.40	0.16	0.40	0.33	0.11	0.00	0.00	0.23	-0.09	0.23	-0.11	-0.08	0.63	0.01	9.90	1.21	9.90	1.33
14	418254	0	54824	0.48	0.27	0.10	0.14	0.48	0.01	0.00	0.42	-0.22	-0.20	-0.13	0.42	-0.65	0.01	7.83	1.03	3.73	1.02
15	418759	0	54824	0.44	0.21	0.18	0.44	0.16	0.01	0.00	0.36	-0.15	-0.17	0.36	-0.12	-0.12	0.01	-1.05	1.00	-2.92	0.99
16	418784	0	54824	0.26	0.50	0.10	0.26	0.14	0.00	0.00	0.13	0.02	-0.13	0.13	-0.08	0.90	0.01	9.90	1.13	9.90	1.28
17	418787	0	54824	0.45	0.17	0.10	0.45	0.28	0.00	0.00	0.23	-0.10	-0.13	0.23	-0.08	0.46	0.01	9.90	1.22	9.90	1.28
18	418894	0	54824	0.52	0.08	0.17	0.22	0.52	0.00	0.00	0.35	-0.17	-0.14	-0.18	0.35	-0.47	0.01	1.95	1.01	-0.02	1.00
19	418896	0	54824	0.31	0.33	0.31	0.12	0.24	0.00	0.00	0.20	-0.05	0.20	-0.09	-0.09	1.01	0.01	9.90	1.23	9.90	1.40
20	419060	0	54824	0.32	0.21	0.20	0.32	0.25	0.01	0.00	0.24	-0.08	-0.17	0.24	-0.01	0.83	0.01	9.90	1.14	9.90	1.25
21	419201	0	54824	0.23	0.23	0.14	0.26	0.36	0.00	0.00	0.24	0.24	-0.10	-0.04	-0.09	1.08	0.01	6.00	1.03	9.90	1.15
22	419206	0	54824	0.42	0.42	0.23	0.19	0.15	0.01	0.00	0.29	0.29	-0.10	-0.15	-0.09	0.29	0.01	9.90	1.07	9.90	1.09
23	419207	0	54824	0.37	0.23	0.26	0.37	0.13	0.00	0.00	0.34	-0.16	-0.07	0.34	-0.18	0.16	0.01	-3.85	0.99	-3.74	0.98
24	420444	0	54824	0.24	0.19	0.23	0.33	0.24	0.01	0.00	0.23	-0.11	-0.08	-0.03	0.23	1.38	0.01	9.90	1.18	9.90	1.36
25	495142	0	54824	0.28	0.20	0.35	0.28	0.18	0.00	0.00	0.26	-0.13	0.00	0.26	-0.16	1.02	0.01	9.90	1.10	9.90	1.21
26	496393	0	54824	0.55	0.10	0.55	0.16	0.19	0.01	0.00	0.38	-0.15	0.38	-0.20	-0.16	-0.57	0.01	-5.55	0.98	-4.79	0.97
27	496415	0	54824	0.64	0.09	0.13	0.13	0.64	0.01	0.00	0.39	-0.18	-0.19	-0.19	0.39	-1.01	0.01	-9.90	0.96	-5.76	0.96
28	496419	0	54824	0.78	0.12	0.78	0.05	0.05	0.00	0.00	0.32	-0.17	0.32	-0.18	-0.16	-1.82	0.01	5.29	1.03	-0.66	0.99
29	499992	0	54824	0.78	0.07	0.78	0.10	0.04	0.00	0.00	0.37	-0.19	0.37	-0.21	-0.17	-1.58	0.01	-9.90	0.86	-9.90	0.78
30	502443	0	54824	0.35	0.36	0.35	0.21	0.08	0.00	0.00	0.15	0.04	0.15	-0.12	-0.14	1.09	0.01	9.90	1.44	9.90	1.69
31	577978	0	54824	0.43	0.15	0.18	0.43	0.23	0.01	0.00	0.31	-0.11	-0.23	0.31	-0.05	0.23	0.01	9.90	1.05	9.90	1.07

**Table J-2 (continued). Algebra I Multiple-Choice Item Statistics: Winter**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
32	582566	0	54824	0.50	0.50	0.28	0.16	0.05	0.01	0.00	0.42	0.42	-0.25	-0.14	-0.17	-0.46	0.01	-9.90	0.96	-9.90	0.94
33	583819	0	54824	0.55	0.55	0.14	0.21	0.10	0.00	0.00	0.39	0.39	-0.17	-0.19	-0.18	-0.30	0.01	-9.90	0.94	-9.90	0.93
34	583862	0	54824	0.67	0.67	0.06	0.07	0.20	0.01	0.00	0.41	0.41	-0.20	-0.21	-0.21	-0.94	0.01	-9.90	0.89	-9.90	0.84
35	584009	0	54824	0.39	0.38	0.10	0.13	0.39	0.00	0.00	0.51	-0.20	-0.18	-0.27	0.51	-0.17	0.01	-9.90	0.87	-9.90	0.85
36	584269	0	54824	0.34	0.28	0.34	0.22	0.15	0.01	0.00	0.22	0.06	0.22	-0.16	-0.16	0.59	0.01	9.90	1.12	9.90	1.18

**Table J-3. Biology Multiple-Choice Item Statistics: Winter**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	271319	0	40077	0.59	0.10	0.59	0.15	0.16	0.01	0.00	0.36	-0.20	0.36	-0.23	-0.08	-0.57	0.01	-9.90	0.94	-8.67	0.95
2	274151	0	40077	0.41	0.14	0.25	0.19	0.41	0.00	0.00	0.33	-0.11	-0.12	-0.17	0.33	0.44	0.01	9.90	1.10	9.90	1.12
3	277840	0	40077	0.39	0.39	0.10	0.11	0.40	0.01	0.00	0.35	0.35	-0.20	-0.15	-0.12	0.10	0.01	-6.07	0.98	-3.23	0.98
4	278852	0	40077	0.35	0.35	0.25	0.35	0.05	0.00	0.00	0.50	-0.24	-0.22	0.50	-0.14	-0.26	0.01	-9.90	0.89	-9.90	0.86
5	278863	0	40077	0.37	0.24	0.18	0.20	0.37	0.01	0.00	0.43	-0.14	-0.17	-0.20	0.43	0.11	0.01	-9.90	0.90	-9.90	0.88
6	278923	0	40077	0.48	0.09	0.25	0.48	0.18	0.00	0.00	0.34	-0.17	-0.17	0.34	-0.13	-0.27	0.01	-1.12	1.00	-2.35	0.99
7	302086	0	40077	0.39	0.22	0.22	0.16	0.39	0.01	0.00	0.26	-0.10	-0.10	-0.10	0.26	0.08	0.01	9.90	1.05	9.90	1.07
8	303245	0	40077	0.33	0.15	0.36	0.15	0.33	0.01	0.00	0.35	-0.14	-0.15	-0.11	0.35	0.77	0.01	9.90	1.07	9.90	1.10
9	313499	0	40077	0.36	0.11	0.29	0.36	0.22	0.01	0.00	0.20	-0.09	-0.04	0.20	-0.11	0.65	0.01	9.90	1.22	9.90	1.30
10	314633	0	40077	0.42	0.18	0.18	0.20	0.42	0.01	0.00	0.36	-0.12	-0.15	-0.16	0.36	0.12	0.01	-1.77	0.99	-2.17	0.99
11	321591	0	40077	0.37	0.37	0.19	0.29	0.15	0.00	0.00	0.32	0.32	-0.09	-0.18	-0.11	0.36	0.01	4.91	1.02	5.66	1.03
12	337698	0	40077	0.44	0.24	0.44	0.23	0.10	0.00	0.00	0.30	-0.09	0.30	-0.14	-0.16	0.16	0.01	9.90	1.07	9.90	1.06
13	339676	0	40077	0.45	0.45	0.11	0.23	0.20	0.01	0.00	0.33	0.33	-0.15	-0.15	-0.11	0.13	0.01	9.25	1.04	7.64	1.04
14	340998	0	40077	0.54	0.15	0.14	0.54	0.17	0.00	0.00	0.40	-0.15	-0.21	0.40	-0.20	-0.51	0.01	-9.90	0.93	-9.90	0.91
15	341002	0	40077	0.49	0.20	0.17	0.49	0.14	0.01	0.00	0.40	-0.15	-0.19	0.40	-0.18	-0.46	0.01	-9.90	0.96	-9.90	0.93
16	342282	0	40077	0.65	0.09	0.14	0.12	0.65	0.00	0.00	0.44	-0.20	-0.21	-0.24	0.44	-0.96	0.01	-9.90	0.87	-9.90	0.80
17	342301	0	40077	0.37	0.26	0.37	0.18	0.17	0.01	0.00	0.27	-0.01	0.27	-0.18	-0.13	0.41	0.01	9.90	1.09	9.90	1.11
18	408528	0	40077	0.34	0.20	0.34	0.15	0.31	0.00	0.00	0.27	-0.07	0.27	-0.04	-0.18	0.78	0.01	9.90	1.18	9.90	1.26
19	408688	0	40077	0.39	0.39	0.18	0.21	0.21	0.00	0.00	0.28	0.28	-0.12	-0.10	-0.11	0.22	0.01	9.90	1.06	9.90	1.07
20	409189	0	40077	0.39	0.22	0.39	0.23	0.16	0.00	0.00	0.28	-0.04	0.28	-0.15	-0.15	0.60	0.01	9.90	1.17	9.90	1.22
21	409857	0	40077	0.49	0.49	0.17	0.21	0.12	0.01	0.00	0.39	0.39	-0.18	-0.20	-0.12	-0.25	0.01	-9.90	0.95	-9.90	0.93
22	410515	0	40077	0.48	0.09	0.13	0.30	0.48	0.01	0.00	0.26	-0.18	-0.17	-0.03	0.26	0.31	0.01	9.90	1.19	9.90	1.23
23	410545	0	40077	0.33	0.15	0.33	0.14	0.38	0.01	0.00	0.23	-0.05	0.23	-0.23	-0.01	0.48	0.01	9.90	1.09	9.90	1.13
24	410549	0	40077	0.39	0.22	0.22	0.39	0.16	0.01	0.00	0.24	-0.07	-0.13	0.24	-0.07	0.28	0.01	9.90	1.11	9.90	1.13
25	410620	0	40077	0.39	0.20	0.39	0.25	0.15	0.01	0.00	0.29	-0.09	0.29	-0.15	-0.10	0.38	0.01	9.90	1.09	9.90	1.10
26	468433	0	40077	0.39	0.39	0.31	0.15	0.14	0.00	0.00	0.32	0.32	-0.11	-0.17	-0.12	0.51	0.01	9.90	1.09	9.90	1.12
27	468441	0	40077	0.48	0.07	0.21	0.24	0.48	0.00	0.00	0.36	-0.20	-0.22	-0.09	0.36	0.08	0.01	3.62	1.01	2.01	1.01
28	471915	0	40077	0.57	0.17	0.57	0.15	0.11	0.01	0.00	0.39	-0.15	0.39	-0.21	-0.18	-0.63	0.01	-9.90	0.94	-9.90	0.91
29	471925	0	40077	0.41	0.14	0.11	0.41	0.33	0.01	0.00	0.29	-0.08	-0.20	0.29	-0.10	0.22	0.01	9.90	1.06	9.90	1.08
30	479411	0	40077	0.30	0.36	0.17	0.17	0.30	0.00	0.00	0.24	-0.07	-0.05	-0.15	0.24	1.19	0.01	9.90	1.36	9.90	1.52
31	480273	0	40077	0.40	0.10	0.16	0.40	0.33	0.00	0.00	0.09	-0.14	-0.09	0.09	0.07	0.69	0.01	9.90	1.43	9.90	1.55
32	480848	0	40077	0.46	0.16	0.18	0.46	0.19	0.00	0.00	0.37	-0.15	-0.19	0.37	-0.15	-0.01	0.01	-5.33	0.98	-5.93	0.97
33	480853	0	40077	0.41	0.22	0.41	0.21	0.15	0.00	0.00	0.31	-0.02	0.31	-0.22	-0.14	0.29	0.01	9.90	1.07	9.90	1.08

**Table J-3 (continued). Biology Multiple-Choice Item Statistics: Winter**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	492415	0	40077	0.28	0.24	0.24	0.23	0.28	0.01	0.00	0.24	0.03	-0.08	-0.19	0.24	1.21	0.01	9.90	1.31	9.90	1.43
35	496405	0	40077	0.36	0.36	0.24	0.13	0.26	0.01	0.00	0.23	0.23	0.01	-0.23	-0.07	0.53	0.01	9.90	1.16	9.90	1.21
36	566216	0	40077	0.40	0.15	0.22	0.23	0.40	0.00	0.00	0.44	-0.18	-0.15	-0.21	0.44	-0.11	0.01	-9.90	0.91	-9.90	0.88
37	569369	0	40077	0.48	0.20	0.16	0.48	0.15	0.01	0.00	0.37	-0.14	-0.15	0.37	-0.19	-0.28	0.01	-9.90	0.96	-7.29	0.96
38	574709	0	40077	0.51	0.14	0.20	0.51	0.15	0.01	0.00	0.36	-0.20	-0.16	0.36	-0.12	-0.34	0.01	-9.90	0.97	-6.37	0.96
39	574711	0	40077	0.44	0.44	0.20	0.14	0.21	0.01	0.00	0.36	0.36	-0.16	-0.17	-0.11	0.20	0.01	5.13	1.02	6.40	1.04
40	574722	0	40077	0.42	0.22	0.42	0.23	0.13	0.00	0.00	0.38	-0.21	0.38	-0.13	-0.12	-0.06	0.01	-9.90	0.96	-9.90	0.94
41	575117	0	40077	0.42	0.13	0.42	0.24	0.21	0.00	0.00	0.31	-0.15	0.31	-0.16	-0.09	0.24	0.01	9.90	1.06	9.90	1.07
42	575126	0	40077	0.50	0.13	0.16	0.21	0.50	0.01	0.00	0.38	-0.18	-0.21	-0.11	0.38	-0.35	0.01	-9.90	0.95	-9.11	0.95
43	575128	0	40077	0.29	0.29	0.29	0.19	0.23	0.00	0.00	0.26	0.26	-0.13	-0.05	-0.09	1.01	0.01	9.90	1.17	9.90	1.25
44	579367	0	40077	0.54	0.54	0.18	0.16	0.12	0.00	0.00	0.42	0.42	-0.18	-0.19	-0.21	-0.71	0.01	-9.90	0.95	-9.90	0.92
45	579370	0	40077	0.40	0.40	0.19	0.26	0.14	0.00	0.00	0.39	0.39	-0.22	-0.14	-0.12	0.23	0.01	-8.12	0.97	-7.05	0.96
46	580505	0	40077	0.40	0.40	0.22	0.20	0.17	0.01	0.00	0.39	0.39	-0.08	-0.21	-0.18	-0.25	0.01	-9.90	0.96	-9.90	0.95
47	582371	0	40077	0.49	0.49	0.15	0.15	0.20	0.00	0.00	0.40	0.40	-0.19	-0.18	-0.16	-0.21	0.01	-9.90	0.94	-9.90	0.92
48	582769	0	40077	0.37	0.35	0.09	0.19	0.37	0.00	0.00	0.41	-0.15	-0.13	-0.22	0.41	-0.11	0.01	-9.90	0.93	-9.90	0.92

**Table J-4. Literature Multiple-Choice Item Statistics: Winter**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	500529	0	35222	0.29	0.36	0.06	0.29	0.29	0.00	0.00	0.22	-0.07	-0.27	0.22	0.00	1.47	0.01	9.90	1.20	9.90	1.50
2	500530	0	35222	0.42	0.42	0.07	0.11	0.39	0.00	0.00	0.22	0.22	-0.25	-0.17	0.01	0.78	0.01	9.90	1.25	9.90	1.37
3	500532	0	35222	0.60	0.08	0.09	0.23	0.60	0.00	0.00	0.36	-0.20	-0.24	-0.13	0.36	0.13	0.01	9.58	1.05	9.90	1.10
4	500533	0	35222	0.46	0.46	0.31	0.12	0.10	0.00	0.00	0.37	0.37	-0.10	-0.24	-0.19	0.54	0.01	9.90	1.06	9.90	1.13
5	500536	0	35222	0.54	0.22	0.09	0.15	0.54	0.00	0.00	0.46	-0.16	-0.26	-0.24	0.46	0.28	0.01	-8.69	0.96	-5.91	0.96
6	500537	0	35222	0.66	0.13	0.66	0.12	0.08	0.00	0.00	0.46	-0.29	0.46	-0.16	-0.24	0.03	0.01	-9.90	0.91	-9.90	0.90
7	500538	0	35222	0.42	0.10	0.34	0.42	0.13	0.00	0.00	0.28	-0.19	-0.04	0.28	-0.17	1.02	0.01	9.90	1.25	9.90	1.42
8	500539	0	35222	0.70	0.04	0.70	0.06	0.20	0.00	0.00	0.44	-0.23	0.44	-0.23	-0.26	-0.73	0.01	-9.90	0.92	-9.90	0.86
9	500540	0	35222	0.67	0.67	0.17	0.07	0.09	0.00	0.00	0.47	0.47	-0.26	-0.28	-0.19	-0.79	0.01	-5.75	0.97	-8.13	0.90
10	500541	0	35222	0.49	0.08	0.49	0.23	0.20	0.00	0.00	0.35	-0.25	0.35	-0.13	-0.13	0.57	0.01	9.90	1.11	9.90	1.15
11	502738	0	35222	0.51	0.19	0.07	0.22	0.51	0.01	0.00	0.42	-0.21	-0.23	-0.15	0.42	-0.02	0.01	7.41	1.04	6.50	1.05
12	502741	0	35222	0.40	0.40	0.12	0.15	0.33	0.01	0.00	0.30	0.30	-0.26	-0.19	0.03	1.18	0.01	9.90	1.23	9.90	1.45
13	502743	0	35222	0.67	0.13	0.11	0.09	0.67	0.01	0.00	0.52	-0.23	-0.27	-0.28	0.52	-1.01	0.01	3.91	1.03	-6.24	0.92
14	502745	0	35222	0.37	0.09	0.37	0.22	0.31	0.01	0.00	0.24	-0.18	0.24	-0.18	0.04	1.20	0.01	9.90	1.27	9.90	1.50
15	502747	0	35222	0.43	0.24	0.19	0.14	0.43	0.01	0.00	0.42	-0.17	-0.12	-0.23	0.42	0.15	0.01	9.90	1.05	8.31	1.07
16	502748	0	35222	0.48	0.14	0.30	0.48	0.07	0.01	0.00	0.46	-0.20	-0.20	0.46	-0.24	0.11	0.01	-4.92	0.98	-3.73	0.97
17	502750	0	35222	0.61	0.61	0.10	0.22	0.07	0.01	0.00	0.46	0.46	-0.27	-0.19	-0.25	-0.54	0.01	1.95	1.01	-1.61	0.98
18	515760	0	35222	0.40	0.21	0.40	0.10	0.28	0.01	0.00	0.22	-0.14	0.22	-0.24	0.06	1.00	0.01	9.90	1.29	9.90	1.45
19	515761	0	35222	0.58	0.19	0.09	0.58	0.13	0.01	0.00	0.45	-0.21	-0.23	0.45	-0.20	-0.11	0.01	-7.71	0.96	-9.90	0.92
20	515762	0	35222	0.56	0.56	0.18	0.16	0.10	0.01	0.00	0.37	0.37	-0.19	-0.20	-0.10	0.15	0.01	9.90	1.05	9.90	1.10
21	515763	0	35222	0.37	0.37	0.36	0.15	0.12	0.01	0.00	0.20	0.20	-0.08	-0.19	0.05	1.44	0.01	9.90	1.43	9.90	1.75
22	515765	0	35222	0.53	0.09	0.17	0.21	0.53	0.01	0.00	0.45	-0.22	-0.20	-0.20	0.45	0.20	0.01	-6.87	0.97	-5.33	0.96
23	515766	0	35222	0.56	0.16	0.56	0.19	0.08	0.01	0.00	0.42	-0.19	0.42	-0.20	-0.21	0.03	0.01	-1.74	0.99	-1.71	0.99
24	515767	0	35222	0.38	0.25	0.15	0.22	0.38	0.01	0.00	0.34	-0.05	-0.22	-0.14	0.34	0.73	0.01	9.90	1.07	9.90	1.13
25	515770	0	35222	0.41	0.11	0.29	0.19	0.41	0.01	0.00	0.34	-0.20	-0.10	-0.14	0.34	1.05	0.01	9.90	1.15	9.90	1.28
26	528447	0	35222	0.83	0.08	0.83	0.05	0.04	0.00	0.00	0.40	-0.24	0.40	-0.24	-0.16	-1.41	0.01	-9.90	0.82	-9.90	0.69
27	528469	0	35222	0.40	0.19	0.26	0.40	0.15	0.00	0.00	0.21	-0.14	-0.01	0.21	-0.12	1.25	0.01	9.90	1.39	9.90	1.69
28	528470	0	35222	0.52	0.15	0.13	0.20	0.52	0.00	0.00	0.24	-0.11	-0.16	-0.06	0.24	0.79	0.01	9.90	1.33	9.90	1.47
29	528472	0	35222	0.69	0.69	0.09	0.16	0.06	0.00	0.00	0.44	0.44	-0.23	-0.22	-0.22	-0.71	0.01	-9.90	0.94	-7.17	0.92
30	528474	0	35222	0.58	0.16	0.58	0.12	0.13	0.00	0.00	0.43	-0.14	0.43	-0.22	-0.25	0.23	0.01	-1.40	0.99	-4.08	0.97
31	528475	0	35222	0.47	0.47	0.19	0.13	0.21	0.00	0.00	0.36	0.36	-0.12	-0.22	-0.15	0.41	0.01	9.90	1.07	9.90	1.09
32	528477	0	35222	0.59	0.59	0.11	0.23	0.06	0.00	0.00	0.34	0.34	-0.25	-0.12	-0.15	0.16	0.01	9.90	1.08	9.90	1.12
33	528478	0	35222	0.61	0.06	0.15	0.61	0.18	0.00	0.00	0.43	-0.25	-0.24	0.43	-0.17	-0.33	0.01	-3.64	0.98	-2.44	0.98

**Table J-4 (continued). Literature Multiple-Choice Item Statistics: Winter**

Ref	PubID	Form	<i>N</i>	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	528479	0	35222	0.75	0.08	0.08	0.09	0.75	0.00	0.00	0.49	-0.25	-0.27	-0.25	0.49	-0.81	0.01	-9.90	0.78	-9.90	0.70

**Table J-5. Algebra I Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	619193	0	118915	0.40	0.34	0.40	0.16	0.09	0.00	0.00	0.35	-0.05	0.35	-0.29	-0.13					0.79	0.01	9.90	1.16	9.90	1.22
2	629867	0	118915	0.47	0.21	0.16	0.15	0.47	0.01	0.00	0.51	-0.22	-0.20	-0.24	0.51					0.36	0.01	-9.90	0.95	-7.44	0.95
3	629873	0	118915	0.49	0.17	0.49	0.19	0.15	0.01	0.00	0.35	-0.14	0.35	-0.15	-0.16					0.47	0.01	9.90	1.17	9.90	1.21
4	633236	0	118915	0.52	0.09	0.26	0.13	0.52	0.01	0.00	0.52	-0.25	-0.23	-0.23	0.52					0.02	0.01	-6.22	0.97	-4.59	0.96
5	633815	0	118915	0.58	0.13	0.16	0.58	0.13	0.00	0.00	0.45	-0.20	-0.24	0.45	-0.19					0.09	0.01	0.26	1.00	-2.39	0.98
6	634334	0	118915	0.51	0.51	0.20	0.16	0.13	0.01	0.00	0.40	0.40	-0.20	-0.20	-0.12					0.46	0.01	9.90	1.10	9.90	1.10
7	651332	0	118915	0.71	0.71	0.06	0.15	0.07	0.00	0.00	0.43	0.43	-0.22	-0.26	-0.19					-0.84	0.01	-3.87	0.98	-2.16	0.97
8	663500	0	118915	0.56	0.10	0.12	0.22	0.56	0.01	0.00	0.58	-0.25	-0.30	-0.27	0.58					-0.16	0.01	-9.90	0.88	-9.90	0.83
9	664234	0	118915	0.67	0.11	0.67	0.10	0.12	0.01	0.00	0.54	-0.26	0.54	-0.22	-0.31					-0.43	0.01	-9.90	0.84	-9.90	0.76
10	665911	0	118915	0.85	0.03	0.05	0.85	0.06	0.00	0.00	0.38	-0.17	-0.22	0.38	-0.22					-1.62	0.02	-9.90	0.85	-9.90	0.73
11	666864	0	118915	0.57	0.09	0.16	0.57	0.17	0.00	0.00	0.47	-0.24	-0.22	0.47	-0.21					0.15	0.01	-1.67	0.99	-5.77	0.96
12	667247	0	118915	0.32	0.25	0.25	0.17	0.32	0.00	0.00	0.45	-0.14	-0.21	-0.15	0.45					1.26	0.01	-0.32	1.00	6.06	1.06
13	667528	0	118915	0.53	0.14	0.15	0.53	0.17	0.00	0.00	0.44	-0.24	-0.12	0.44	-0.24					0.18	0.01	6.12	1.03	3.35	1.03
14	705083	0	118915	0.56	0.10	0.18	0.16	0.56	0.00	0.00	0.60	-0.27	-0.30	-0.27	0.60					0.10	0.01	-9.90	0.82	-9.90	0.75
15	418272	0	118915	0.68	0.13	0.68	0.08	0.10	0.01	0.00	0.42	-0.17	0.42	-0.26	-0.19					-0.39	0.01	-5.41	0.97	-5.73	0.94
16	420439	0	118915	0.39	0.39	0.20	0.25	0.15	0.01	0.00	0.41	0.41	-0.14	-0.16	-0.19					1.07	0.01	9.90	1.12	9.90	1.20
17	492843	0	118915	0.53	0.31	0.53	0.09	0.06	0.01	0.00	0.30	-0.06	0.30	-0.25	-0.18					0.25	0.01	9.90	1.22	9.90	1.30
18	492849	0	118915	0.59	0.16	0.59	0.20	0.05	0.01	0.00	0.50	-0.20	0.50	-0.32	-0.17					0.02	0.01	-9.90	0.94	-9.90	0.91
19	493934	0	118915	0.38	0.09	0.29	0.23	0.38	0.00	0.00	0.44	-0.10	-0.20	-0.21	0.44					1.13	0.01	9.90	1.08	9.90	1.12
20	495371	0	118915	0.35	0.20	0.22	0.35	0.22	0.01	0.00	0.37	-0.16	-0.06	0.37	-0.19					1.34	0.01	9.90	1.16	9.90	1.32
21	495373	0	118915	0.47	0.17	0.25	0.47	0.11	0.00	0.00	0.46	-0.11	-0.25	0.46	-0.25					0.93	0.01	9.90	1.09	9.90	1.17
22	496200	0	118915	0.58	0.09	0.11	0.58	0.21	0.01	0.00	0.47	-0.24	-0.19	0.47	-0.23					-0.05	0.01	-4.95	0.98	-3.62	0.97
23	496412	0	118915	0.42	0.31	0.15	0.42	0.12	0.01	0.00	0.38	-0.07	-0.23	0.38	-0.20					0.87	0.01	9.90	1.13	9.90	1.20
24	499993	0	118915	0.37	0.37	0.27	0.22	0.13	0.00	0.00	0.29	0.29	-0.08	-0.07	-0.21					0.75	0.01	9.90	1.21	9.90	1.28
25	502444	0	118915	0.52	0.52	0.14	0.17	0.17	0.00	0.00	0.49	0.49	-0.26	-0.20	-0.20					0.22	0.01	-5.10	0.98	-4.24	0.97
26	506316	0	118915	0.59	0.59	0.22	0.13	0.06	0.00	0.00	0.40	0.40	-0.15	-0.28	-0.19					-0.06	0.01	9.90	1.06	9.01	1.08
27	566676	0	118915	0.57	0.12	0.57	0.20	0.11	0.00	0.00	0.47	-0.17	0.47	-0.29	-0.20					-0.03	0.01	-5.96	0.97	-2.78	0.98
28	569278	0	118915	0.59	0.59	0.17	0.15	0.08	0.01	0.00	0.54	0.54	-0.27	-0.26	-0.24					-0.38	0.01	-9.90	0.93	-8.10	0.92
29	569280	0	118915	0.43	0.14	0.21	0.23	0.43	0.01	0.00	0.38	-0.14	-0.12	-0.20	0.38					0.91	0.01	9.90	1.14	9.90	1.21
30	574491	0	118915	0.45	0.16	0.17	0.45	0.22	0.00	0.00	0.32	-0.22	-0.17	0.32	-0.04					0.67	0.01	9.90	1.19	9.90	1.27
31	583822	0	118915	0.35	0.22	0.26	0.18	0.35	0.00	0.00	0.40	-0.22	-0.09	-0.15	0.40					1.22	0.01	9.90	1.09	9.90	1.17
32	584008	0	118915	0.78	0.06	0.78	0.09	0.07	0.00	0.00	0.39	-0.23	0.39	-0.22	-0.18					-0.83	0.01	-9.90	0.85	-9.90	0.80
33	584024	0	118915	0.38	0.21	0.38	0.16	0.25	0.00	0.00	0.33	-0.09	0.33	-0.11	-0.20					1.06	0.01	9.90	1.19	9.90	1.32



Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	584275	0	118915	0.43	0.20	0.24	0.12	0.43	0.01	0.00	0.47	-0.12	-0.25	-0.22	0.47					0.61	0.01	-0.74	1.00	0.10	1.00
35	585774	0	118915	0.40	0.12	0.10	0.38	0.40	0.01	0.00	0.36	-0.14	-0.14	-0.17	0.36					1.03	0.01	9.90	1.17	9.90	1.25
36	585779	0	118915	0.59	0.15	0.19	0.59	0.08	0.01	0.00	0.42	-0.16	-0.22	0.42	-0.22					0.07	0.01	8.51	1.04	4.61	1.04
37	618113	1	6524	0.55	0.21	0.09	0.55	0.15	0.00	0.00	0.27	-0.05	-0.19	0.27	-0.17	A+	A-	A+	A+	-0.17	0.06	6.57	1.16	4.58	1.20
38	708516	1	6524	0.34	0.26	0.24	0.15	0.34	0.00	0.00	0.37	-0.08	-0.21	-0.13	0.37	A+	A-	A-	A+	0.95	0.06	2.02	1.06	2.35	1.10
39	719043	1	6524	0.55	0.19	0.55	0.19	0.07	0.00	0.00	0.34	-0.13	0.34	-0.23	-0.10	A+	A-	A+	A-	-0.17	0.06	5.42	1.13	2.91	1.13
40	720103	1	6524	0.59	0.18	0.59	0.10	0.12	0.00	0.00	0.27	-0.20	0.27	-0.20	0.03	A+	A-	A-	A+	-0.35	0.06	8.40	1.21	8.77	1.45
41	720138	1	6524	0.56	0.19	0.15	0.56	0.09	0.00	0.00	0.37	-0.13	-0.21	0.37	-0.18	A+	A-	A+	A+	-0.21	0.06	3.10	1.07	3.74	1.17
42	720222	1	6524	0.86	0.04	0.04	0.05	0.86	0.00	0.00	0.36	-0.18	-0.20	-0.20	0.36	A+	B-	A-	A+	-2.07	0.08	-3.01	0.85	-3.48	0.65
43	720641	1	6524	0.35	0.19	0.35	0.23	0.22	0.00	0.00	0.25	-0.12	0.25	-0.06	-0.10	A+	A-	A+	A+	0.86	0.06	5.35	1.15	5.65	1.24
44	720646	1	6524	0.56	0.56	0.12	0.18	0.13	0.00	0.00	0.49	0.49	-0.22	-0.27	-0.20	A+	A-	A-	A-	-0.24	0.06	-4.84	0.89	-4.15	0.83
45	721668	1	6524	0.72	0.72	0.10	0.12	0.06	0.00	0.00	0.48	0.48	-0.26	-0.27	-0.21	A+	A-	A+	A-	-1.08	0.06	-6.76	0.81	-5.15	0.69
46	721802	1	6524	0.32	0.21	0.23	0.24	0.32	0.01	0.00	0.35	-0.12	-0.09	-0.18	0.35	A+	A-	A+	A-	1.07	0.06	2.48	1.07	3.70	1.17
47	613759	2	4910	0.44	0.15	0.25	0.15	0.44	0.00	0.00	0.44	-0.16	-0.24	-0.15	0.44	A+	A-	A+	A+	0.67	0.06	0.64	1.02	1.09	1.04
48	618127	2	4910	0.48	0.48	0.12	0.07	0.32	0.00	0.00	0.49	0.49	-0.24	-0.21	-0.24	A-	B-	B-	A-	0.45	0.06	-2.98	0.93	-3.00	0.89
49	624120	2	4910	0.40	0.23	0.40	0.24	0.13	0.00	0.00	0.35	-0.21	0.35	-0.02	-0.20	A+	A-	A-	A+	0.91	0.06	3.84	1.11	3.66	1.15
50	629871	2	4910	0.35	0.20	0.30	0.15	0.35	0.00	0.00	0.46	-0.26	-0.09	-0.20	0.46	A+	A-	A-	A+	1.18	0.06	-1.05	0.97	-0.03	1.00
51	661908	2	4910	0.58	0.14	0.13	0.58	0.14	0.00	0.00	0.44	-0.23	-0.24	0.44	-0.15	A+	A+	A+	A+	-0.04	0.06	-1.43	0.97	-0.77	0.97
52	717294	2	4910	0.44	0.15	0.44	0.19	0.21	0.00	0.00	0.33	-0.15	0.33	-0.17	-0.10	A+	A+	A+	A+	0.66	0.06	6.45	1.17	5.45	1.21
53	718918	2	4910	0.50	0.14	0.50	0.10	0.26	0.00	0.00	0.38	-0.16	0.38	-0.28	-0.11	A+	A+	A-	A+	0.38	0.06	3.19	1.08	3.91	1.15
54	719045	2	4910	0.32	0.26	0.21	0.20	0.32	0.00	0.00	0.56	-0.23	-0.28	-0.10	0.56	A-	A-	A-	A-	1.37	0.06	-5.28	0.85	-2.57	0.88
55	720093	2	4910	0.37	0.13	0.22	0.37	0.28	0.00	0.00	0.41	-0.10	-0.24	0.41	-0.14	A+	A+	A-	A-	1.07	0.06	1.02	1.03	2.14	1.09
56	593283	2	4910	0.64	0.08	0.16	0.12	0.64	0.00	0.00	0.53	-0.26	-0.28	-0.25	0.53	A-	A+	A-	A-	-0.34	0.06	-7.12	0.83	-5.87	0.73
57	618120	3	4840	0.57	0.57	0.24	0.08	0.11	0.00	0.00	0.48	0.48	-0.27	-0.19	-0.22	A-	A-	A-	A+	0.02	0.06	-1.99	0.95	-1.00	0.96
58	618747	3	4840	0.54	0.15	0.25	0.54	0.05	0.00	0.00	0.44	-0.29	-0.17	0.44	-0.19	A+	A+	A-	A+	0.15	0.06	-0.29	0.99	0.20	1.01
59	717291	3	4840	0.63	0.07	0.13	0.17	0.63	0.00	0.00	0.34	-0.18	-0.11	-0.21	0.34	A+	A-	A+	A-	-0.29	0.06	3.78	1.10	4.17	1.21
60	717564	3	4840	0.12	0.12	0.10	0.29	0.48	0.01	0.00	0.20	0.20	-0.14	-0.10	0.04	A-	A-	A+	A-	2.86	0.09	2.50	1.15	9.90	2.62
61	719036	3	4840	0.55	0.10	0.14	0.21	0.55	0.00	0.00	0.36	-0.22	-0.20	-0.11	0.36	A+	A+	A+	A+	0.12	0.06	2.61	1.06	1.31	1.05
62	720098	3	4840	0.58	0.12	0.58	0.12	0.17	0.00	0.00	0.44	-0.22	0.44	-0.29	-0.13	A+	A+	A+	A+	-0.04	0.06	-1.82	0.96	-1.64	0.93
63	721807	3	4840	0.38	0.09	0.38	0.35	0.18	0.00	0.00	0.48	-0.07	0.48	-0.25	-0.23	A-	A+	A+	A+	1.01	0.06	-1.42	0.96	0.46	1.02
64	722114	3	4840	0.76	0.11	0.76	0.07	0.06	0.01	0.00	0.44	-0.24	0.44	-0.25	-0.21	A+	B-	A-	A+	-1.04	0.07	-3.90	0.88	-3.13	0.79
65	579827	3	4840	0.54	0.09	0.54	0.26	0.10	0.00	0.00	0.43	-0.19	0.43	-0.25	-0.15	A-	A-	A-	A+	0.15	0.06	0.46	1.01	-0.22	0.99
66	584401	3	4840	0.39	0.37	0.15	0.39	0.08	0.00	0.00	0.39	-0.12	-0.23	0.39	-0.17	A-	A-	A+	A+	0.93	0.06	1.93	1.05	2.25	1.09

Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
67	618739	4	4878	0.51	0.08	0.20	0.51	0.21	0.00	0.00	0.29	-0.10	-0.10	0.29	-0.19	A-	A-	A-	A+	0.33	0.06	6.08	1.15	5.79	1.22
68	623227	4	4878	0.48	0.16	0.17	0.18	0.48	0.00	0.00	0.42	-0.26	-0.10	-0.18	0.42	A+	A+	A+	A-	0.48	0.06	1.58	1.04	0.56	1.02
69	624556	4	4878	0.43	0.43	0.18	0.23	0.16	0.00	0.00	0.37	0.37	-0.21	-0.16	-0.08	A+	A-	A+	A+	0.72	0.06	4.19	1.11	4.18	1.16
70	651334	4	4878	0.48	0.25	0.14	0.48	0.13	0.00	0.00	0.28	-0.04	-0.19	0.28	-0.17	A-	A+	A+	A+	0.49	0.06	7.07	1.18	6.22	1.24
71	665413	4	4878	0.44	0.15	0.20	0.22	0.44	0.00	0.00	0.42	-0.24	-0.13	-0.17	0.42	A-	A-	A-	A+	0.70	0.06	1.91	1.05	2.20	1.08
72	709466	4	4878	0.48	0.48	0.17	0.13	0.22	0.00	0.00	0.51	0.51	-0.26	-0.24	-0.18	A+	A-	A+	A+	0.48	0.06	-3.57	0.92	-3.47	0.88
73	717730	4	4878	0.42	0.35	0.42	0.17	0.06	0.00	0.00	0.37	-0.14	0.37	-0.22	-0.12	A-	A-	A-	A-	0.80	0.06	3.96	1.11	4.19	1.16
74	718921	4	4878	0.32	0.09	0.32	0.49	0.09	0.00	0.00	0.33	-0.18	0.33	-0.12	-0.14	A-	A-	A-	A-	1.32	0.06	1.71	1.05	3.78	1.18
75	722118	4	4878	0.40	0.17	0.27	0.15	0.40	0.00	0.00	0.44	-0.14	-0.17	-0.24	0.44	A-	A-	A-	A+	0.88	0.06	-0.17	1.00	0.78	1.03
76	583348	4	4878	0.69	0.69	0.10	0.14	0.06	0.00	0.00	0.45	0.45	-0.24	-0.23	-0.23	A+	A-	A-	A-	-0.61	0.06	-2.61	0.93	-2.17	0.88
77	615106	5	4878	0.33	0.26	0.19	0.33	0.22	0.00	0.00	0.23	-0.09	-0.11	0.23	-0.06	A+	A-	A-	A+	1.28	0.06	6.29	1.20	7.19	1.36
78	621484	5	4878	0.34	0.28	0.30	0.34	0.07	0.00	0.00	0.25	-0.11	-0.06	0.25	-0.16	A+	A+	A+	A+	1.21	0.06	7.29	1.22	8.28	1.41
79	623231	5	4878	0.27	0.27	0.30	0.29	0.14	0.00	0.00	0.39	0.39	-0.03	-0.20	-0.18	C-	A-	A-	A-	1.63	0.07	-0.07	1.00	1.27	1.07
80	665942	5	4878	0.35	0.12	0.23	0.29	0.35	0.00	0.00	0.32	-0.12	-0.23	-0.02	0.32	A+	A+	A-	A-	1.16	0.06	4.37	1.13	4.23	1.19
81	667534	5	4878	0.37	0.14	0.29	0.19	0.37	0.00	0.00	0.21	-0.11	-0.07	-0.06	0.21	A-	A-	A-	A+	1.06	0.06	9.90	1.32	8.93	1.41
82	709463	5	4878	0.55	0.16	0.55	0.23	0.05	0.00	0.00	0.48	-0.27	0.48	-0.25	-0.14	A-	A-	A-	A+	0.08	0.06	-3.65	0.92	-3.02	0.88
83	720449	5	4878	0.54	0.12	0.54	0.24	0.10	0.00	0.00	0.51	-0.16	0.51	-0.30	-0.25	A-	A-	A-	A+	0.16	0.06	-3.80	0.91	-2.93	0.89
84	720631	5	4878	0.66	0.15	0.66	0.10	0.09	0.00	0.00	0.51	-0.32	0.51	-0.19	-0.24	A+	A-	A-	A+	-0.45	0.06	-5.45	0.87	-4.63	0.78
85	720635	5	4878	0.40	0.40	0.23	0.17	0.19	0.00	0.00	0.50	0.50	-0.11	-0.32	-0.19	A+	A-	A-	A-	0.87	0.06	-2.38	0.94	-1.71	0.94
86	722116	5	4878	0.42	0.26	0.15	0.16	0.42	0.01	0.00	0.49	-0.17	-0.22	-0.23	0.49	A+	A+	A+	A-	0.75	0.06	-1.24	0.97	-0.88	0.97
87	618116	6	4742	0.70	0.10	0.10	0.70	0.10	0.01	0.00	0.45	-0.22	-0.25	0.45	-0.21	A+	A+	A-	A+	-0.68	0.06	-2.49	0.93	-2.47	0.86
88	622276	6	4742	0.66	0.08	0.16	0.10	0.66	0.00	0.00	0.61	-0.25	-0.38	-0.27	0.61	A+	A+	A-	A-	-0.47	0.06	-9.90	0.75	-7.36	0.66
89	633232	6	4742	0.36	0.25	0.36	0.14	0.24	0.01	0.00	0.45	-0.18	0.45	-0.09	-0.24	B-	A+	A+	A+	1.13	0.06	-1.20	0.97	0.19	1.01
90	651415	6	4742	0.58	0.16	0.20	0.58	0.05	0.01	0.00	0.53	-0.27	-0.29	0.53	-0.20	A+	A-	A+	A+	-0.04	0.06	-5.45	0.87	-4.85	0.81
91	664759	6	4742	0.66	0.11	0.13	0.66	0.11	0.00	0.00	0.36	-0.29	-0.12	0.36	-0.12	A-	A-	A-	A-	-0.45	0.06	2.12	1.06	1.76	1.09
92	665943	6	4742	0.49	0.09	0.49	0.09	0.32	0.01	0.00	0.40	-0.18	0.40	-0.26	-0.16	A+	A+	A-	A+	0.43	0.06	3.68	1.10	3.12	1.12
93	717285	6	4742	0.32	0.28	0.18	0.21	0.32	0.01	0.00	0.35	-0.14	-0.16	-0.10	0.35	A+	A-	A-	A-	1.33	0.06	3.50	1.11	3.75	1.19
94	717286	6	4742	0.64	0.21	0.64	0.08	0.07	0.01	0.00	0.46	-0.22	0.46	-0.27	-0.23	A+	A+	A+	A+	-0.34	0.06	-4.07	0.90	-1.63	0.92
95	717562	6	4742	0.60	0.60	0.21	0.09	0.11	0.00	0.00	0.39	0.39	-0.11	-0.25	-0.24	A+	A+	A+	A-	-0.13	0.06	1.86	1.05	1.72	1.08
96	593282	6	4742	0.62	0.62	0.14	0.16	0.07	0.01	0.00	0.47	0.47	-0.21	-0.25	-0.24	A+	A-	A-	A+	-0.27	0.06	-3.05	0.93	-3.11	0.86
97	613765	7	4883	0.50	0.32	0.50	0.10	0.08	0.00	0.00	0.13	-0.03	0.13	-0.10	-0.07	A+	A-	A+	A-	0.38	0.06	9.90	1.36	9.90	1.54
98	621925	7	4883	0.36	0.37	0.17	0.36	0.09	0.00	0.00	0.30	0.08	-0.31	0.30	-0.23	A-	A+	A+	A+	1.08	0.06	4.25	1.12	4.62	1.20
99	623498	7	4883	0.78	0.78	0.07	0.06	0.09	0.00	0.00	0.43	0.43	-0.20	-0.23	-0.24	A+	A-	A+	A+	-1.15	0.07	-2.68	0.91	-3.10	0.78

**Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
100	634331	7	4883	0.17	0.46	0.27	0.17	0.10	0.00	0.00	0.20	-0.06	-0.11	0.20	0.02	A-	A-	A-	A+	2.39	0.08	2.35	1.11	7.60	1.78
101	665171	7	4883	0.72	0.10	0.11	0.72	0.06	0.00	0.00	0.39	-0.26	-0.15	0.39	-0.21	A-	A-	A-	A+	-0.80	0.06	-2.13	0.94	-1.48	0.91
102	665417	7	4883	0.45	0.14	0.18	0.22	0.45	0.01	0.00	0.28	-0.12	-0.05	-0.18	0.28	A+	A+	A+	A-	0.60	0.06	6.19	1.16	5.10	1.20
103	717560	7	4883	0.44	0.21	0.18	0.17	0.44	0.00	0.00	0.43	-0.11	-0.29	-0.15	0.43	A-	A-	A-	A-	0.68	0.06	1.48	1.04	1.29	1.05
104	717739	7	4883	0.51	0.14	0.51	0.15	0.20	0.00	0.00	0.40	-0.19	0.40	-0.25	-0.10	A-	A-	A-	A+	0.30	0.06	1.87	1.04	1.95	1.08
105	720642	7	4883	0.30	0.26	0.20	0.24	0.30	0.00	0.00	0.35	-0.03	-0.10	-0.25	0.35	A-	A-	A+	A-	1.46	0.06	3.33	1.11	3.68	1.19
106	575146	7	4883	0.58	0.58	0.11	0.20	0.10	0.01	0.00	0.50	0.50	-0.24	-0.26	-0.22	A+	A-	A-	A+	-0.07	0.06	-3.78	0.92	-2.74	0.88
107	618963	8	4912	0.51	0.51	0.11	0.25	0.13	0.00	0.00	0.47	0.47	-0.27	-0.13	-0.26	A-	A-	A+	A-	0.33	0.06	-0.61	0.99	-0.59	0.98
108	633056	8	4912	0.37	0.14	0.12	0.36	0.37	0.00	0.00	0.32	-0.21	-0.20	-0.03	0.32	A-	A-	A+	A-	1.03	0.06	5.04	1.15	5.72	1.25
109	633817	8	4912	0.48	0.16	0.22	0.48	0.14	0.00	0.00	0.39	-0.08	-0.26	0.39	-0.16	A+	A+	A+	A+	0.44	0.06	1.78	1.04	1.77	1.06
110	717561	8	4912	0.44	0.11	0.44	0.14	0.30	0.00	0.00	0.42	-0.24	0.42	-0.29	-0.06	A+	A+	A+	A+	0.65	0.06	1.82	1.05	1.74	1.06
111	717738	8	4912	0.65	0.07	0.10	0.18	0.65	0.00	0.00	0.34	-0.20	-0.22	-0.12	0.34	A-	A-	A-	A-	-0.41	0.06	2.85	1.07	3.03	1.16
112	718762	8	4912	0.66	0.13	0.66	0.11	0.09	0.01	0.00	0.47	-0.22	0.47	-0.27	-0.20	A-	A-	A-	A+	-0.48	0.06	-3.94	0.91	-3.00	0.85
113	719038	8	4912	0.72	0.12	0.07	0.09	0.72	0.00	0.00	0.54	-0.29	-0.25	-0.28	0.54	A+	A-	A-	A+	-0.81	0.06	-7.66	0.80	-5.98	0.67
114	720439	8	4912	0.67	0.67	0.12	0.10	0.10	0.00	0.00	0.50	0.50	-0.26	-0.22	-0.26	A-	A+	A-	A+	-0.52	0.06	-4.86	0.88	-3.88	0.81
115	722110	8	4912	0.51	0.39	0.51	0.06	0.04	0.00	0.00	0.48	-0.28	0.48	-0.24	-0.21	A-	B-	B-	A+	0.30	0.06	-1.38	0.97	-1.23	0.96
116	585780	8	4912	0.24	0.31	0.30	0.24	0.15	0.00	0.00	0.23	0.09	-0.22	0.23	-0.10	A-	A+	A-	A+	1.85	0.07	5.22	1.20	8.61	1.65
117	614780	9	4905	0.64	0.13	0.15	0.64	0.08	0.00	0.00	0.47	-0.20	-0.27	0.47	-0.22	A-	A-	A-	A-	-0.33	0.06	-2.50	0.94	-2.10	0.90
118	614789	9	4905	0.48	0.11	0.20	0.21	0.48	0.00	0.00	0.47	-0.24	-0.22	-0.17	0.47	A+	A-	A+	A+	0.49	0.06	-1.67	0.96	-1.21	0.96
119	618129	9	4905	0.54	0.12	0.17	0.17	0.54	0.00	0.00	0.49	-0.29	-0.19	-0.21	0.49	A+	A-	A-	A-	0.18	0.06	-3.11	0.93	-2.23	0.91
120	668408	9	4905	0.39	0.09	0.14	0.38	0.39	0.00	0.00	0.45	-0.25	-0.18	-0.17	0.45	A+	A+	A+	A+	0.98	0.06	0.65	1.02	1.83	1.07
121	709467	9	4905	0.33	0.10	0.33	0.15	0.41	0.00	0.00	0.46	-0.15	0.46	-0.27	-0.15	A-	A+	A+	A+	1.27	0.06	-1.72	0.95	-0.11	0.99
122	718919	9	4905	0.50	0.14	0.50	0.15	0.20	0.00	0.00	0.29	-0.18	0.29	-0.24	0.02	A+	A+	A+	A+	0.36	0.06	8.16	1.21	6.31	1.26
123	720217	9	4905	0.47	0.47	0.15	0.25	0.12	0.00	0.00	0.48	0.48	-0.16	-0.22	-0.26	A-	A-	A-	A+	0.51	0.06	-1.46	0.96	-1.26	0.95
124	720437	9	4905	0.47	0.15	0.29	0.47	0.08	0.00	0.00	0.42	-0.27	-0.13	0.42	-0.19	A-	A+	A+	A+	0.53	0.06	1.59	1.04	1.16	1.04
125	720446	9	4905	0.39	0.34	0.39	0.17	0.10	0.00	0.00	0.13	0.02	0.13	-0.11	-0.11	A+	A+	A+	A+	0.97	0.06	9.90	1.39	9.90	1.54
126	721805	9	4905	0.51	0.51	0.15	0.23	0.10	0.00	0.00	0.38	0.38	-0.21	-0.12	-0.20	A+	A+	A+	A+	0.31	0.06	2.41	1.06	3.00	1.12
127	621623	10	4898	0.30	0.18	0.30	0.18	0.34	0.00	0.00	0.21	-0.10	0.21	-0.11	-0.03	A+	A+	A+	A+	1.46	0.06	8.03	1.26	9.82	1.55
128	621929	10	4898	0.35	0.35	0.17	0.08	0.39	0.00	0.00	0.12	0.12	-0.33	-0.27	0.30	A-	A-	A-	A-	1.18	0.06	9.90	1.37	9.56	1.46
129	623223	10	4898	0.48	0.11	0.15	0.25	0.48	0.00	0.00	0.49	-0.23	-0.29	-0.15	0.49	A-	A-	A+	A-	0.46	0.06	-2.72	0.94	-2.52	0.91
130	624278	10	4898	0.68	0.09	0.14	0.68	0.08	0.00	0.00	0.49	-0.26	-0.29	0.49	-0.17	A-	B-	A-	A+	-0.59	0.06	-5.10	0.87	0.03	1.00
131	629874	10	4898	0.56	0.16	0.56	0.12	0.15	0.00	0.00	0.38	-0.20	0.38	-0.17	-0.16	A+	A-	A-	A+	0.06	0.06	2.60	1.06	1.98	1.08
132	633803	10	4898	0.45	0.13	0.45	0.31	0.11	0.00	0.00	0.32	-0.21	0.32	-0.05	-0.22	A+	A+	A+	A-	0.64	0.06	6.29	1.16	8.57	1.35

Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
133	717729	10	4898	0.46	0.11	0.23	0.46	0.20	0.00	0.00	0.30	-0.21	0.01	0.30	-0.23	A-	A-	A-	A+	0.59	0.06	6.39	1.16	4.86	1.19
134	720443	10	4898	0.45	0.14	0.20	0.21	0.45	0.00	0.00	0.47	-0.22	-0.17	-0.21	0.47	A+	A-	A+	A+	0.65	0.06	-1.78	0.96	-0.78	0.97
135	722310	10	4898	0.43	0.22	0.26	0.43	0.08	0.00	0.00	0.46	-0.29	-0.22	0.46	-0.04	A+	A-	A-	A-	0.73	0.06	-0.81	0.98	-0.48	0.98
136	584135	10	4898	0.70	0.70	0.10	0.14	0.07	0.00	0.00	0.46	0.46	-0.21	-0.26	-0.23	A-	A-	A+	A+	-0.66	0.06	-3.63	0.90	-2.64	0.85
137	614785	11	4896	0.50	0.08	0.23	0.19	0.50	0.00	0.00	0.33	-0.19	-0.06	-0.22	0.33	A+	A+	A+	A-	0.37	0.06	2.92	1.07	2.14	1.07
138	618109	11	4896	0.46	0.37	0.46	0.09	0.08	0.00	0.00	0.37	-0.13	0.37	-0.23	-0.19	A+	A+	A+	A+	0.59	0.06	3.42	1.08	3.50	1.12
139	623490	11	4896	0.28	0.43	0.18	0.28	0.10	0.01	0.00	0.14	0.12	-0.17	0.14	-0.18	A-	A-	A-	A+	1.57	0.06	7.16	1.24	8.02	1.45
140	633230	11	4896	0.68	0.68	0.10	0.17	0.05	0.00	0.00	0.39	0.39	-0.22	-0.19	-0.20	A+	A-	A-	A+	-0.58	0.06	-0.19	1.00	-0.19	0.99
141	664750	11	4896	0.44	0.20	0.44	0.20	0.15	0.00	0.00	0.32	-0.07	0.32	-0.17	-0.17	A+	A+	A-	A+	0.67	0.06	6.05	1.15	6.12	1.21
142	665988	11	4896	0.62	0.10	0.62	0.19	0.08	0.00	0.00	0.49	-0.23	0.49	-0.26	-0.24	B-	B-	A-	A+	-0.26	0.06	-5.14	0.88	-4.19	0.84
143	718925	11	4896	0.27	0.51	0.09	0.13	0.27	0.00	0.00	0.20	0.15	-0.26	-0.26	0.20	A-	A+	A-	A-	1.64	0.06	6.19	1.21	6.80	1.39
144	720444	11	4896	0.58	0.58	0.10	0.27	0.05	0.00	0.00	0.34	0.34	-0.24	-0.10	-0.20	A+	A+	A+	A+	-0.01	0.06	3.60	1.08	4.43	1.17
145	720637	11	4896	0.42	0.15	0.24	0.42	0.19	0.00	0.00	0.34	-0.06	-0.15	0.34	-0.21	A+	A-	A+	A+	0.76	0.06	4.02	1.10	3.63	1.12
146	720768	11	4896	0.28	0.11	0.28	0.18	0.42	0.00	0.00	0.08	-0.15	0.08	-0.22	0.20	A-	A-	A+	A-	1.58	0.06	8.65	1.30	9.90	1.62
147	614790	12	4910	0.58	0.25	0.58	0.09	0.08	0.01	0.00	0.45	-0.18	0.45	-0.24	-0.26	A-	A-	A-	A+	-0.02	0.06	0.02	1.00	0.78	1.03
148	618748	12	4910	0.61	0.09	0.61	0.09	0.21	0.00	0.00	0.43	-0.18	0.43	-0.25	-0.21	A+	A-	A+	A+	-0.16	0.06	0.11	1.00	0.20	1.01
149	624125	12	4910	0.45	0.35	0.45	0.12	0.08	0.00	0.00	0.32	0.01	0.32	-0.31	-0.23	A-	A-	A+	A+	0.62	0.06	4.94	1.12	4.28	1.15
150	624281	12	4910	0.52	0.11	0.17	0.52	0.20	0.01	0.00	0.45	-0.07	-0.28	0.45	-0.24	A-	A+	A+	A+	0.27	0.06	-1.56	0.96	-1.89	0.94
151	708513	12	4910	0.23	0.27	0.24	0.26	0.23	0.00	0.00	0.20	-0.22	0.08	-0.04	0.20	A-	A-	A-	A-	1.93	0.07	5.64	1.22	6.59	1.48
152	719899	12	4910	0.43	0.19	0.29	0.09	0.43	0.00	0.00	0.42	-0.21	-0.09	-0.28	0.42	A-	A-	A-	A+	0.76	0.06	-0.16	1.00	1.02	1.04
153	721672	12	4910	0.37	0.13	0.13	0.37	0.36	0.00	0.00	0.21	-0.21	-0.16	0.21	0.06	A+	A-	A+	A+	1.04	0.06	8.46	1.24	9.38	1.40
154	722192	12	4910	0.42	0.20	0.42	0.24	0.14	0.00	0.00	0.34	-0.07	0.34	-0.20	-0.16	A-	A-	A-	A+	0.79	0.06	2.77	1.07	2.39	1.09
155	569304	12	4910	0.40	0.40	0.25	0.18	0.17	0.00	0.00	0.43	0.43	-0.01	-0.27	-0.26	A-	A+	A+	A-	0.90	0.06	-1.15	0.97	-0.40	0.99
156	585776	12	4910	0.67	0.67	0.15	0.07	0.10	0.00	0.00	0.47	0.47	-0.24	-0.22	-0.25	A-	A-	A-	A+	-0.53	0.06	-2.96	0.93	-1.57	0.92
157	618960	13	4890	0.29	0.29	0.45	0.16	0.10	0.00	0.00	0.17	0.17	0.01	-0.13	-0.12	A-	A-	A-	A-	1.55	0.06	6.84	1.23	8.54	1.53
158	665419	13	4890	0.71	0.16	0.71	0.07	0.05	0.00	0.00	0.47	-0.30	0.47	-0.21	-0.20	B-	A-	A-	A+	-0.73	0.06	-4.23	0.89	-3.69	0.80
159	666491	13	4890	0.47	0.47	0.23	0.17	0.12	0.00	0.00	0.44	0.44	-0.15	-0.19	-0.25	A-	A-	A-	A+	0.54	0.06	1.72	1.04	2.12	1.08
160	666878	13	4890	0.51	0.15	0.51	0.16	0.17	0.00	0.00	0.47	-0.13	0.47	-0.27	-0.23	A+	A-	A+	A-	0.32	0.06	-0.80	0.98	-0.62	0.98
161	708519	13	4890	0.42	0.42	0.13	0.23	0.22	0.00	0.00	0.43	0.43	-0.13	-0.13	-0.27	A-	A-	A+	A-	0.80	0.06	-1.09	0.97	-1.07	0.96
162	717731	13	4890	0.64	0.07	0.24	0.64	0.05	0.00	0.00	0.44	-0.15	-0.32	0.44	-0.17	A-	A-	A-	A+	-0.35	0.06	-2.62	0.93	-2.39	0.89
163	718764	13	4890	0.48	0.12	0.10	0.29	0.48	0.00	0.00	0.42	-0.30	-0.23	-0.09	0.42	A-	A-	A+	A-	0.46	0.06	0.57	1.01	0.71	1.02
164	720097	13	4890	0.24	0.26	0.25	0.24	0.24	0.00	0.00	0.11	-0.05	-0.08	0.02	0.11	A-	A+	A+	A-	1.84	0.07	7.83	1.30	9.90	1.87
165	721799	13	4890	0.70	0.70	0.14	0.09	0.06	0.00	0.00	0.52	0.52	-0.32	-0.24	-0.23	A+	A-	A-	A-	-0.68	0.06	-6.98	0.82	-5.08	0.74

Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
166	583346	13	4890	0.67	0.11	0.12	0.67	0.10	0.00	0.00	0.51	-0.23	-0.28	0.51	-0.24	A+	A-	A-	A+	-0.52	0.06	-4.28	0.89	-4.48	0.79
167	618738	14	4951	0.62	0.22	0.08	0.07	0.62	0.00	0.00	0.41	-0.20	-0.20	-0.22	0.41	A-	B-	A-	A+	-0.25	0.06	-0.17	1.00	1.51	1.07
168	663713	14	4951	0.59	0.08	0.10	0.23	0.59	0.00	0.00	0.40	-0.15	-0.23	-0.20	0.40	A-	A-	A-	A+	-0.09	0.06	1.92	1.05	4.78	1.22
169	717559	14	4951	0.22	0.12	0.18	0.22	0.48	0.00	0.00	0.11	-0.16	0.02	0.11	0.00	A+	A-	A-	A-	2.01	0.07	7.00	1.30	9.90	1.92
170	720767	14	4951	0.70	0.14	0.70	0.09	0.08	0.00	0.00	0.41	-0.24	0.41	-0.22	-0.16	A+	A+	A-	A+	-0.68	0.06	-0.66	0.98	0.14	1.01
171	721667	14	4951	0.58	0.28	0.08	0.58	0.07	0.00	0.00	0.36	-0.16	-0.23	0.36	-0.18	A-	A-	A-	A-	-0.02	0.06	1.85	1.04	1.00	1.04
172	722112	14	4951	0.62	0.62	0.16	0.14	0.08	0.00	0.00	0.50	0.50	-0.28	-0.22	-0.23	A+	A-	A+	A-	-0.23	0.06	-5.53	0.87	-4.25	0.82
173	574881	14	4951	0.43	0.18	0.43	0.22	0.16	0.00	0.00	0.60	-0.22	0.60	-0.24	-0.30	A-	A+	A+	A+	0.71	0.06	-7.58	0.82	-5.78	0.80
174	579833	14	4951	0.68	0.11	0.05	0.68	0.16	0.01	0.00	0.49	-0.28	-0.21	0.49	-0.25	B-	A-	A-	A+	-0.56	0.06	-6.82	0.84	-5.64	0.73
175	584002	14	4951	0.23	0.23	0.60	0.07	0.09	0.00	0.00	0.30	0.30	-0.11	-0.18	-0.07	B-	A-	A+	A+	1.93	0.07	2.34	1.09	6.86	1.52
176	584068	14	4951	0.26	0.19	0.13	0.42	0.26	0.00	0.00	0.26	-0.18	-0.17	0.03	0.26	A+	A+	A-	A-	1.72	0.07	4.73	1.17	8.04	1.55
177	618114	15	4877	0.40	0.30	0.13	0.16	0.40	0.00	0.00	0.47	-0.21	-0.12	-0.25	0.47	B-	B-	A-	A-	0.95	0.06	-1.17	0.97	-0.01	1.00
178	624283	15	4877	0.76	0.76	0.08	0.11	0.05	0.00	0.00	0.48	0.48	-0.27	-0.24	-0.25	A+	A-	A+	A+	-0.99	0.07	-5.62	0.83	-5.43	0.66
179	633228	15	4877	0.48	0.17	0.16	0.18	0.48	0.00	0.00	0.33	-0.08	-0.13	-0.21	0.33	A-	A-	A-	A+	0.50	0.06	5.94	1.16	5.65	1.22
180	664235	15	4877	0.59	0.11	0.21	0.59	0.09	0.00	0.00	0.43	-0.14	-0.27	0.43	-0.20	A+	A+	A-	A+	-0.06	0.06	0.56	1.01	-0.07	1.00
181	665912	15	4877	0.76	0.11	0.76	0.07	0.06	0.00	0.00	0.41	-0.20	0.41	-0.23	-0.22	A-	A+	A-	A+	-1.00	0.07	-3.78	0.88	-2.69	0.82
182	665913	15	4877	0.21	0.10	0.21	0.13	0.55	0.00	0.00	0.05	-0.26	0.05	0.07	0.07	A-	A+	A+	A+	2.08	0.07	8.19	1.34	9.90	2.00
183	718387	15	4877	0.65	0.65	0.12	0.17	0.06	0.00	0.00	0.42	0.42	-0.20	-0.22	-0.20	A-	A+	A-	A+	-0.35	0.06	-0.77	0.98	-1.17	0.94
184	719897	15	4877	0.77	0.05	0.06	0.77	0.11	0.00	0.00	0.45	-0.18	-0.25	0.45	-0.27	A+	A-	A-	A+	-1.10	0.07	-4.76	0.85	-3.88	0.73
185	720639	15	4877	0.43	0.43	0.10	0.37	0.10	0.00	0.00	0.51	0.51	-0.27	-0.28	-0.11	A-	A-	A-	A+	0.78	0.06	-2.40	0.94	-1.25	0.95
186	569310	15	4877	0.76	0.08	0.76	0.07	0.09	0.00	0.00	0.37	-0.23	0.37	-0.21	-0.14	A+	A-	A+	A+	-1.01	0.07	-1.60	0.95	0.86	1.06
187	618126	16	4928	0.33	0.38	0.33	0.13	0.16	0.01	0.00	0.17	0.08	0.17	-0.18	-0.16	A+	A+	A+	A-	1.29	0.06	9.22	1.29	9.27	1.46
188	624123	16	4928	0.58	0.17	0.58	0.12	0.12	0.00	0.00	0.41	-0.13	0.41	-0.28	-0.19	A-	A+	A+	A+	-0.04	0.06	0.20	1.00	0.31	1.01
189	633816	16	4928	0.55	0.55	0.22	0.13	0.10	0.00	0.00	0.52	0.52	-0.30	-0.23	-0.19	A-	A-	A+	A+	0.14	0.06	-2.99	0.93	-3.39	0.87
190	651330	16	4928	0.35	0.14	0.25	0.25	0.35	0.00	0.00	0.19	-0.13	0.01	-0.10	0.19	A+	A-	A+	A-	1.15	0.06	9.90	1.31	9.28	1.43
191	661905	16	4928	0.59	0.59	0.27	0.08	0.06	0.00	0.00	0.32	0.32	-0.10	-0.24	-0.19	A+	A-	A-	A+	-0.09	0.06	5.27	1.13	4.47	1.20
192	663268	16	4928	0.29	0.05	0.07	0.58	0.29	0.00	0.00	0.33	-0.16	-0.20	-0.13	0.33	A-	A-	A-	A-	1.50	0.06	1.64	1.05	2.56	1.13
193	667243	16	4928	0.70	0.11	0.10	0.70	0.09	0.00	0.00	0.43	-0.23	-0.22	0.43	-0.20	A+	A-	A+	A+	-0.65	0.06	-2.70	0.93	-2.13	0.88
194	719898	16	4928	0.61	0.17	0.15	0.61	0.07	0.01	0.00	0.50	-0.23	-0.29	0.50	-0.21	A-	A+	A-	A+	-0.17	0.06	-3.04	0.93	-2.37	0.90
195	720769	16	4928	0.44	0.27	0.44	0.18	0.10	0.00	0.00	0.35	-0.02	0.35	-0.24	-0.22	A-	A+	A+	A+	0.70	0.06	4.63	1.12	3.97	1.15
196	593802	16	4928	0.46	0.46	0.22	0.15	0.16	0.00	0.00	0.30	0.30	-0.07	-0.18	-0.14	A+	A+	A+	A-	0.56	0.06	6.58	1.17	6.81	1.26
197	621624	17	4864	0.30	0.14	0.30	0.25	0.30	0.00	0.00	0.35	-0.04	-0.20	-0.12	0.35	B-	A-	A-	A-	1.42	0.06	4.16	1.13	4.95	1.26
198	622278	17	4864	0.39	0.11	0.36	0.14	0.39	0.00	0.00	0.34	-0.21	-0.04	-0.23	0.34	A-	A-	A-	A-	0.93	0.06	2.97	1.08	3.13	1.13

Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
199	622279	17	4864	0.65	0.08	0.05	0.65	0.22	0.00	0.00	0.53	-0.16	-0.23	0.53	-0.37	A-	A-	A-	A-	-0.43	0.06	-5.55	0.87	-3.69	0.82
200	633227	17	4864	0.60	0.14	0.60	0.12	0.14	0.00	0.00	0.47	-0.21	0.47	-0.25	-0.23	A-	A-	A-	A+	-0.19	0.06	-2.92	0.93	-2.97	0.87
201	634332	17	4864	0.40	0.08	0.40	0.37	0.14	0.00	0.00	0.33	-0.21	0.33	-0.06	-0.20	A-	A+	A+	A+	0.87	0.06	6.09	1.17	5.33	1.21
202	664752	17	4864	0.73	0.04	0.08	0.15	0.73	0.00	0.00	0.54	-0.22	-0.28	-0.33	0.54	A+	A-	A-	A+	-0.87	0.06	-7.42	0.80	-6.41	0.64
203	717289	17	4864	0.44	0.44	0.26	0.15	0.15	0.00	0.00	0.40	0.40	-0.14	-0.16	-0.22	A-	A-	A-	A+	0.64	0.06	2.80	1.07	4.00	1.15
204	717735	17	4864	0.63	0.63	0.14	0.13	0.10	0.00	0.00	0.32	0.32	-0.14	-0.16	-0.16	A+	A-	A+	A-	-0.31	0.06	4.22	1.11	4.15	1.21
205	718917	17	4864	0.55	0.22	0.12	0.55	0.10	0.00	0.00	0.39	-0.16	-0.25	0.39	-0.14	A-	A-	A-	A+	0.09	0.06	1.54	1.04	1.48	1.06
206	720764	17	4864	0.44	0.16	0.20	0.20	0.44	0.00	0.00	0.43	-0.24	-0.17	-0.13	0.43	A+	A-	A+	A-	0.68	0.06	-0.95	0.98	-0.93	0.97
207	618742	18	4880	0.41	0.41	0.34	0.11	0.13	0.00	0.00	0.42	0.42	-0.08	-0.25	-0.25	A-	A+	A+	A+	0.84	0.06	0.98	1.03	1.91	1.07
208	618746	18	4880	0.47	0.21	0.15	0.16	0.47	0.00	0.00	0.51	-0.15	-0.28	-0.24	0.51	B+	A+	A+	A-	0.52	0.06	-2.02	0.95	-2.29	0.92
209	624285	18	4880	0.70	0.10	0.13	0.70	0.07	0.00	0.00	0.51	-0.24	-0.30	0.51	-0.22	A+	A-	A-	A+	-0.68	0.06	-5.10	0.87	-4.82	0.74
210	635791	18	4880	0.52	0.14	0.52	0.21	0.12	0.00	0.00	0.31	-0.15	0.31	-0.18	-0.09	A-	A-	A+	A+	0.25	0.06	7.93	1.20	6.53	1.27
211	708517	18	4880	0.49	0.07	0.20	0.24	0.49	0.00	0.00	0.43	-0.22	-0.13	-0.24	0.43	A+	A+	A+	A+	0.45	0.06	1.52	1.04	1.28	1.05
212	717563	18	4880	0.28	0.10	0.53	0.28	0.09	0.00	0.00	0.31	-0.13	-0.07	0.31	-0.21	A+	A+	A+	A+	1.61	0.07	1.69	1.06	6.14	1.38
213	719892	18	4880	0.47	0.17	0.19	0.17	0.47	0.00	0.00	0.47	-0.27	-0.23	-0.11	0.47	A+	A-	A+	A-	0.55	0.06	-1.82	0.96	-1.30	0.95
214	720218	18	4880	0.53	0.08	0.53	0.09	0.30	0.00	0.00	0.21	-0.22	0.21	-0.26	0.08	A-	A+	A-	A+	0.23	0.06	9.72	1.25	8.31	1.35
215	720440	18	4880	0.64	0.09	0.16	0.11	0.64	0.00	0.00	0.47	-0.26	-0.21	-0.23	0.47	A+	A-	A-	A+	-0.36	0.06	-3.78	0.91	-2.33	0.89
216	720632	18	4880	0.74	0.74	0.07	0.10	0.09	0.00	0.00	0.49	0.49	-0.23	-0.27	-0.26	A+	A-	A-	A-	-0.94	0.06	-4.22	0.88	-4.15	0.74
217	617830	19	4877	0.46	0.14	0.25	0.46	0.16	0.00	0.00	0.46	-0.22	-0.21	0.46	-0.16	A+	A-	A-	A+	0.59	0.06	-0.30	0.99	0.42	1.02
218	618741	19	4877	0.30	0.42	0.30	0.12	0.16	0.00	0.00	0.27	0.16	0.27	-0.29	-0.28	A+	A+	A+	A+	1.50	0.06	5.42	1.18	7.06	1.41
219	622273	19	4877	0.55	0.05	0.23	0.15	0.55	0.00	0.00	0.57	-0.18	-0.36	-0.24	0.57	A-	A-	B-	A-	0.10	0.06	-7.40	0.83	-5.99	0.78
220	624121	19	4877	0.50	0.50	0.13	0.23	0.13	0.01	0.00	0.51	0.51	-0.29	-0.15	-0.26	A-	A-	A-	A+	0.37	0.06	-3.34	0.92	-2.38	0.91
221	633804	19	4877	0.55	0.55	0.16	0.20	0.09	0.00	0.00	0.40	0.40	-0.20	-0.17	-0.20	A+	A+	A+	A-	0.14	0.06	1.61	1.04	1.69	1.07
222	651328	19	4877	0.66	0.08	0.14	0.12	0.66	0.00	0.00	0.54	-0.28	-0.28	-0.24	0.54	A+	A+	A-	A+	-0.45	0.06	-7.24	0.83	-5.41	0.74
223	664104	19	4877	0.46	0.24	0.46	0.13	0.17	0.00	0.00	0.22	-0.04	0.22	-0.13	-0.11	A-	A+	A-	A-	0.60	0.06	8.93	1.24	7.51	1.30
224	665949	19	4877	0.59	0.19	0.59	0.12	0.09	0.01	0.00	0.50	-0.25	0.50	-0.22	-0.23	A+	A-	A-	A+	-0.09	0.06	-4.94	0.89	-3.91	0.84
225	720438	19	4877	0.40	0.20	0.40	0.27	0.13	0.00	0.00	0.27	-0.14	0.27	-0.06	-0.14	A+	A-	A+	A+	0.91	0.06	7.87	1.22	7.74	1.33
226	582654	19	4877	0.46	0.22	0.20	0.46	0.12	0.00	0.00	0.37	-0.25	-0.09	0.37	-0.13	A+	A-	A-	A+	0.58	0.06	3.11	1.08	2.69	1.10
227	629877	20	4854	0.76	0.05	0.76	0.13	0.06	0.00	0.00	0.36	-0.21	0.36	-0.16	-0.23	A+	A-	A-	A+	-1.03	0.07	-1.74	0.95	-2.26	0.85
228	633809	20	4854	0.46	0.46	0.13	0.27	0.13	0.00	0.00	0.41	0.41	-0.15	-0.19	-0.20	A-	A-	A+	A+	0.57	0.06	2.62	1.07	2.27	1.08
229	635934	20	4854	0.36	0.36	0.17	0.15	0.31	0.00	0.00	0.41	0.41	-0.14	-0.22	-0.13	A+	A-	A-	A-	1.12	0.06	1.18	1.03	1.74	1.07
230	665916	20	4854	0.59	0.11	0.08	0.59	0.22	0.00	0.00	0.45	-0.25	-0.17	0.45	-0.23	B-	A-	A-	A+	-0.07	0.06	-1.28	0.97	-0.66	0.97
231	718384	20	4854	0.37	0.25	0.13	0.37	0.24	0.00	0.00	0.13	0.02	-0.26	0.13	0.05	A+	A+	A-	A+	1.07	0.06	9.90	1.35	9.90	1.51

Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
232	718758	20	4854	0.52	0.18	0.52	0.13	0.17	0.00	0.00	0.40	-0.17	0.40	-0.23	-0.14	A+	A-	A-	A-	0.28	0.06	0.43	1.01	0.05	1.00
233	718763	20	4854	0.48	0.11	0.17	0.48	0.24	0.00	0.00	0.45	-0.20	-0.21	0.45	-0.19	A-	A-	A-	A+	0.50	0.06	-0.26	0.99	-0.63	0.98
234	718767	20	4854	0.40	0.23	0.19	0.18	0.40	0.00	0.00	0.25	-0.01	-0.11	-0.20	0.25	A+	A+	A+	A-	0.91	0.06	9.26	1.27	8.50	1.36
235	721808	20	4854	0.64	0.64	0.14	0.14	0.08	0.00	0.00	0.57	0.57	-0.31	-0.30	-0.22	A+	A-	A-	A-	-0.32	0.06	-8.12	0.82	-6.91	0.71
236	569276	20	4854	0.73	0.08	0.12	0.73	0.07	0.00	0.00	0.46	-0.19	-0.33	0.46	-0.18	A+	A-	A+	A-	-0.83	0.06	-3.36	0.91	-3.11	0.82
237	629865	21	4921	0.54	0.54	0.21	0.16	0.08	0.00	0.00	0.39	0.39	-0.13	-0.22	-0.20	A-	A+	A+	A+	0.15	0.06	1.64	1.04	1.01	1.04
238	633808	21	4921	0.45	0.22	0.15	0.45	0.18	0.00	0.00	0.23	-0.11	-0.23	0.23	0.04	A-	A-	A+	A+	0.62	0.06	8.45	1.23	7.72	1.30
239	665982	21	4921	0.38	0.15	0.38	0.24	0.23	0.00	0.00	0.29	-0.14	0.29	-0.11	-0.10	A-	A-	A+	A+	0.98	0.06	6.72	1.19	6.82	1.29
240	709472	21	4921	0.56	0.16	0.16	0.12	0.56	0.00	0.00	0.45	-0.15	-0.24	-0.25	0.45	A-	A+	A+	A-	0.04	0.06	-1.81	0.96	-1.07	0.96
241	719893	21	4921	0.43	0.10	0.10	0.37	0.43	0.00	0.00	0.39	-0.22	-0.20	-0.14	0.39	A+	A-	A+	A+	0.73	0.06	0.62	1.02	0.97	1.04
242	720643	21	4921	0.42	0.42	0.46	0.08	0.04	0.00	0.00	0.35	0.35	-0.12	-0.27	-0.20	A-	A-	A-	A+	0.81	0.06	4.84	1.13	4.38	1.17
243	720649	21	4921	0.49	0.49	0.14	0.21	0.16	0.00	0.00	0.50	0.50	-0.21	-0.25	-0.20	A+	B-	A-	A-	0.40	0.06	-4.26	0.90	-2.80	0.90
244	721671	21	4921	0.36	0.37	0.21	0.36	0.06	0.00	0.00	0.47	-0.19	-0.24	0.47	-0.14	A+	A+	A-	A+	1.11	0.06	-2.26	0.94	1.22	1.05
245	566672	21	4921	0.45	0.12	0.26	0.45	0.16	0.01	0.00	0.30	-0.24	-0.05	0.30	-0.12	A-	A-	A-	A+	0.62	0.06	7.28	1.19	7.05	1.27
246	583825	21	4921	0.50	0.18	0.50	0.14	0.18	0.00	0.00	0.31	-0.17	0.31	-0.17	-0.07	A+	A-	A+	A-	0.38	0.06	6.05	1.15	5.96	1.23
247	615107	22	4876	0.21	0.21	0.58	0.14	0.07	0.00	0.00	0.26	0.26	0.03	-0.21	-0.18	A+	A+	A-	A-	2.07	0.07	1.81	1.07	5.41	1.43
248	622283	22	4876	0.36	0.35	0.10	0.19	0.36	0.00	0.00	0.25	-0.16	-0.21	0.06	0.25	A-	A-	A+	A+	1.15	0.06	6.81	1.20	6.18	1.28
249	624288	22	4876	0.64	0.05	0.13	0.18	0.64	0.00	0.00	0.56	-0.19	-0.33	-0.30	0.56	A+	A-	A-	A-	-0.36	0.06	-8.60	0.80	-6.90	0.70
250	633057	22	4876	0.34	0.19	0.34	0.19	0.28	0.00	0.00	0.27	-0.15	0.27	-0.11	-0.05	A-	A-	A-	A+	1.26	0.06	6.21	1.19	6.60	1.32
251	664756	22	4876	0.63	0.11	0.63	0.17	0.08	0.00	0.00	0.45	-0.21	0.45	-0.21	-0.25	A-	A-	A-	A+	-0.30	0.06	-3.13	0.93	-1.10	0.95
252	717287	22	4876	0.80	0.80	0.07	0.09	0.04	0.00	0.00	0.44	0.44	-0.23	-0.26	-0.20	A+	A-	A+	A+	-1.29	0.07	-4.65	0.84	-3.80	0.72
253	718385	22	4876	0.58	0.19	0.12	0.11	0.58	0.00	0.00	0.53	-0.30	-0.23	-0.22	0.53	A+	A+	A+	A+	-0.05	0.06	-6.79	0.85	-5.98	0.77
254	718920	22	4876	0.41	0.14	0.29	0.15	0.41	0.00	0.00	0.30	-0.17	0.00	-0.24	0.30	A+	A-	A+	A-	0.85	0.06	7.16	1.20	6.25	1.25
255	721803	22	4876	0.46	0.20	0.18	0.46	0.15	0.00	0.00	0.33	-0.04	-0.23	0.33	-0.16	A+	A-	A+	A+	0.56	0.06	4.91	1.12	4.32	1.16
256	569294	22	4876	0.50	0.31	0.12	0.50	0.06	0.00	0.00	0.55	-0.35	-0.22	0.55	-0.15	B-	A-	A-	A+	0.39	0.06	-6.21	0.86	-5.17	0.83
257	651422	23	4913	0.45	0.45	0.19	0.18	0.18	0.01	0.00	0.55	0.55	-0.24	-0.20	-0.26	A-	A+	A+	A+	0.64	0.06	-3.62	0.91	-3.02	0.89
258	665910	23	4913	0.76	0.05	0.10	0.76	0.08	0.00	0.00	0.47	-0.22	-0.27	0.47	-0.23	A+	A-	A-	A-	-1.09	0.07	-3.68	0.89	-3.74	0.75
259	708521	23	4913	0.50	0.15	0.25	0.50	0.09	0.00	0.00	0.47	-0.26	-0.24	0.47	-0.11	A-	A-	A-	A+	0.39	0.06	-2.37	0.94	-2.13	0.93
260	718382	23	4913	0.50	0.23	0.14	0.50	0.13	0.00	0.00	0.40	-0.19	-0.20	0.40	-0.15	A+	A-	A+	A-	0.37	0.06	1.09	1.03	1.05	1.04
261	720220	23	4913	0.48	0.14	0.19	0.19	0.48	0.00	0.00	0.41	-0.21	-0.11	-0.22	0.41	A+	A+	A+	A+	0.50	0.06	0.31	1.01	-0.26	0.99
262	720634	23	4913	0.76	0.04	0.08	0.12	0.76	0.00	0.00	0.47	-0.20	-0.28	-0.27	0.47	A-	A-	A-	A+	-1.04	0.07	-5.82	0.83	-4.57	0.71
263	720773	23	4913	0.33	0.33	0.26	0.22	0.18	0.00	0.00	0.42	0.42	-0.20	-0.13	-0.14	A-	A-	A-	A-	1.28	0.06	0.16	1.00	1.49	1.07
264	722184	23	4913	0.42	0.42	0.18	0.21	0.18	0.00	0.00	0.47	0.47	-0.26	-0.21	-0.11	A+	A+	A-	A-	0.76	0.06	0.50	1.01	1.45	1.05

**Table J-5 (continued). Algebra I Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
265	572795	23	4913	0.52	0.18	0.52	0.19	0.11	0.00	0.00	0.32	-0.12	0.32	-0.12	-0.20	A-	A+	A+	A+	0.29	0.06	6.22	1.16	5.32	1.21
266	583826	23	4913	0.38	0.10	0.38	0.28	0.25	0.00	0.00	0.49	-0.21	0.49	-0.23	-0.15	C-	A-	A-	A-	1.03	0.06	-2.20	0.94	-0.98	0.96
267	623226	24	4908	0.66	0.04	0.66	0.05	0.24	0.00	0.00	0.44	-0.19	0.44	-0.17	-0.30	A+	B-	A+	A+	-0.46	0.06	-2.19	0.94	-2.32	0.88
268	633060	24	4908	0.61	0.17	0.61	0.08	0.14	0.00	0.00	0.41	-0.19	0.41	-0.22	-0.19	A-	A-	A+	A+	-0.18	0.06	1.31	1.03	-0.68	0.97
269	633814	24	4908	0.66	0.21	0.66	0.08	0.05	0.00	0.00	0.30	-0.16	0.30	-0.22	-0.07	A-	A-	A-	A+	-0.44	0.06	4.71	1.13	5.33	1.31
270	717737	24	4908	0.70	0.70	0.12	0.12	0.06	0.00	0.00	0.51	0.51	-0.30	-0.24	-0.25	A-	A-	A+	A+	-0.71	0.06	-6.74	0.82	-4.74	0.74
271	718381	24	4908	0.71	0.10	0.10	0.71	0.08	0.00	0.00	0.45	-0.20	-0.24	0.45	-0.25	A+	A-	A-	A+	-0.78	0.06	-4.43	0.88	-2.96	0.82
272	719039	24	4908	0.52	0.10	0.22	0.15	0.52	0.00	0.00	0.46	-0.25	-0.16	-0.24	0.46	A+	A+	A+	A-	0.26	0.06	-2.69	0.94	-1.48	0.94
273	719901	24	4908	0.41	0.20	0.11	0.27	0.41	0.00	0.00	0.39	0.03	-0.25	-0.28	0.39	A-	A+	A-	A+	0.84	0.06	1.16	1.03	2.18	1.08
274	721666	24	4908	0.68	0.68	0.10	0.14	0.08	0.00	0.00	0.50	0.50	-0.21	-0.31	-0.23	A+	A+	A-	A+	-0.58	0.06	-5.21	0.87	-3.80	0.80
275	721673	24	4908	0.31	0.20	0.32	0.17	0.31	0.00	0.00	0.28	-0.19	0.10	-0.25	0.28	A-	A+	A+	A-	1.44	0.06	5.57	1.18	6.14	1.33
276	577983	24	4908	0.51	0.13	0.21	0.51	0.15	0.00	0.00	0.39	-0.26	-0.22	0.39	-0.05	A-	A-	A+	A-	0.32	0.06	2.93	1.07	2.42	1.09



**Table J-6. Biology Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	620408	0	112360	0.45	0.45	0.22	0.19	0.13	0.00	0.00	0.44	0.44	-0.11	-0.26	-0.20					0.67	0.02	3.04	1.03	3.38	1.05
2	621667	0	112360	0.55	0.55	0.10	0.20	0.14	0.01	0.00	0.58	0.58	-0.24	-0.33	-0.23					0.19	0.02	-9.90	0.84	-9.90	0.80
3	634398	0	112360	0.42	0.38	0.42	0.11	0.09	0.00	0.00	0.30	-0.01	0.30	-0.26	-0.21					0.86	0.02	9.90	1.18	9.90	1.24
4	635463	0	112360	0.31	0.33	0.21	0.31	0.15	0.00	0.00	0.38	-0.21	-0.18	0.38	0.00					0.94	0.02	0.10	1.00	4.27	1.06
5	653079	0	112360	0.66	0.17	0.66	0.08	0.08	0.01	0.00	0.51	-0.25	0.51	-0.30	-0.22					-0.43	0.02	-9.90	0.90	-9.21	0.84
6	654031	0	112360	0.57	0.15	0.11	0.57	0.15	0.01	0.00	0.48	-0.21	-0.28	0.48	-0.18					0.11	0.02	-3.47	0.97	-4.75	0.93
7	655221	0	112360	0.65	0.14	0.12	0.65	0.09	0.00	0.00	0.40	-0.21	-0.15	0.40	-0.23					-0.12	0.02	0.24	1.00	0.23	1.00
8	656089	0	112360	0.72	0.72	0.11	0.10	0.06	0.00	0.00	0.55	0.55	-0.25	-0.33	-0.26					-0.85	0.03	-9.90	0.85	-9.90	0.74
9	656227	0	112360	0.65	0.04	0.11	0.65	0.19	0.00	0.00	0.46	-0.13	-0.26	0.46	-0.28					-0.17	0.02	-8.67	0.92	-8.52	0.87
10	656235	0	112360	0.54	0.15	0.14	0.16	0.54	0.00	0.00	0.46	-0.16	-0.20	-0.28	0.46					0.35	0.02	-1.93	0.98	-2.74	0.96
11	658152	0	112360	0.62	0.13	0.62	0.12	0.13	0.01	0.00	0.47	-0.17	0.47	-0.28	-0.22					-0.26	0.02	-5.35	0.95	-3.71	0.94
12	658158	0	112360	0.44	0.19	0.11	0.44	0.25	0.00	0.00	0.44	-0.21	-0.17	0.44	-0.18					0.71	0.02	2.33	1.02	3.39	1.05
13	658181	0	112360	0.39	0.39	0.37	0.15	0.09	0.00	0.00	0.44	0.44	-0.09	-0.27	-0.24					0.51	0.02	-0.60	0.99	-1.40	0.98
14	659519	0	112360	0.42	0.15	0.15	0.28	0.42	0.00	0.00	0.45	-0.22	-0.20	-0.16	0.45					0.90	0.02	1.41	1.01	2.43	1.04
15	659536	0	112360	0.59	0.11	0.14	0.59	0.15	0.00	0.00	0.49	-0.23	-0.22	0.49	-0.24					0.04	0.02	-6.55	0.94	-6.35	0.91
16	659662	0	112360	0.58	0.17	0.58	0.14	0.11	0.00	0.00	0.47	-0.19	0.47	-0.25	-0.23					0.08	0.02	-3.38	0.97	-4.54	0.94
17	659665	0	112360	0.56	0.19	0.09	0.15	0.56	0.01	0.00	0.57	-0.34	-0.23	-0.21	0.57					0.38	0.02	-9.90	0.87	-9.90	0.84
18	659693	0	112360	0.55	0.13	0.11	0.55	0.20	0.00	0.00	0.36	-0.15	-0.23	0.36	-0.14					0.15	0.02	7.70	1.07	7.59	1.11
19	659699	0	112360	0.70	0.07	0.09	0.70	0.14	0.00	0.00	0.48	-0.25	-0.26	0.48	-0.23					-0.55	0.02	-9.90	0.86	-9.90	0.80
20	659700	0	112360	0.56	0.18	0.56	0.13	0.12	0.00	0.00	0.53	-0.21	0.53	-0.28	-0.25					0.22	0.02	-9.90	0.90	-9.35	0.88
21	660378	0	112360	0.64	0.10	0.15	0.11	0.64	0.00	0.00	0.40	-0.20	-0.17	-0.21	0.40					-0.37	0.02	2.29	1.02	3.62	1.07
22	662922	0	112360	0.70	0.70	0.10	0.11	0.09	0.00	0.00	0.50	0.50	-0.29	-0.28	-0.19					-0.49	0.02	-9.90	0.84	-9.90	0.76
23	662923	0	112360	0.45	0.45	0.28	0.17	0.10	0.00	0.00	0.42	0.42	-0.11	-0.22	-0.25					0.66	0.02	5.07	1.05	6.13	1.08
24	663988	0	112360	0.61	0.15	0.61	0.15	0.09	0.00	0.00	0.49	-0.24	0.49	-0.23	-0.24					0.02	0.02	-8.03	0.93	-5.88	0.92
25	665499	0	112360	0.64	0.08	0.08	0.64	0.20	0.00	0.00	0.50	-0.21	-0.22	0.50	-0.30					-0.07	0.02	-9.90	0.88	-9.90	0.85
26	665975	0	112360	0.74	0.05	0.11	0.10	0.74	0.00	0.00	0.56	-0.23	-0.32	-0.32	0.56					-0.68	0.03	-9.90	0.74	-9.90	0.60
27	665979	0	112360	0.62	0.15	0.62	0.14	0.09	0.00	0.00	0.54	-0.26	0.54	-0.29	-0.24					-0.20	0.02	-9.90	0.87	-9.90	0.84
28	278884	0	112360	0.70	0.06	0.70	0.13	0.11	0.00	0.00	0.36	-0.20	0.36	-0.24	-0.12					-0.36	0.02	-1.55	0.98	-2.41	0.96
29	409045	0	112360	0.49	0.17	0.13	0.21	0.49	0.00	0.00	0.33	-0.13	-0.15	-0.15	0.33					0.55	0.02	9.90	1.16	9.90	1.18
30	409176	0	112360	0.54	0.54	0.19	0.22	0.05	0.00	0.00	0.64	0.64	-0.35	-0.35	-0.15					0.02	0.02	-9.90	0.79	-9.90	0.74
31	411164	0	112360	0.56	0.06	0.12	0.56	0.25	0.00	0.00	0.48	-0.27	-0.26	0.48	-0.19					0.20	0.02	-6.19	0.94	-3.86	0.95
32	413399	0	112360	0.64	0.15	0.64	0.12	0.09	0.00	0.00	0.45	-0.23	0.45	-0.24	-0.19					-0.25	0.02	-4.52	0.96	-5.68	0.91
33	419071	0	112360	0.60	0.09	0.14	0.17	0.60	0.00	0.00	0.57	-0.23	-0.19	-0.38	0.57					-0.33	0.02	-9.90	0.88	-9.90	0.83

**Table J-6 (continued). Biology Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	468404	0	112360	0.52	0.52	0.15	0.18	0.14	0.01	0.00	0.53	0.53	-0.26	-0.19	-0.26					0.39	0.02	-9.55	0.91	-8.75	0.89
35	469028	0	112360	0.41	0.41	0.31	0.12	0.16	0.00	0.00	0.34	0.34	-0.03	-0.22	-0.22					0.65	0.02	9.90	1.11	9.90	1.15
36	469829	0	112360	0.58	0.11	0.58	0.17	0.14	0.00	0.00	0.47	-0.21	0.47	-0.31	-0.14					-0.08	0.02	-2.78	0.97	-1.92	0.97
37	480300	0	112360	0.32	0.10	0.12	0.32	0.45	0.01	0.00	0.38	-0.16	-0.16	0.38	-0.13					1.10	0.02	2.10	1.02	6.21	1.10
38	565880	0	112360	0.55	0.12	0.11	0.22	0.55	0.00	0.00	0.53	-0.29	-0.22	-0.24	0.53					0.38	0.02	-7.98	0.93	-8.10	0.90
39	566500	0	112360	0.68	0.68	0.16	0.08	0.08	0.00	0.00	0.44	0.44	-0.20	-0.24	-0.24					-0.59	0.02	-1.93	0.98	-4.02	0.92
40	571651	0	112360	0.44	0.10	0.30	0.16	0.44	0.00	0.00	0.46	-0.22	-0.15	-0.26	0.46					0.77	0.02	-0.16	1.00	-0.29	1.00
41	571652	0	112360	0.71	0.21	0.04	0.71	0.03	0.00	0.00	0.41	-0.26	-0.20	0.41	-0.19					-0.72	0.03	-3.85	0.96	-3.24	0.93
42	574776	0	112360	0.57	0.57	0.15	0.21	0.07	0.00	0.00	0.45	0.45	-0.21	-0.24	-0.20					-0.03	0.02	-0.26	1.00	-2.81	0.96
43	575008	0	112360	0.52	0.52	0.28	0.13	0.06	0.00	0.00	0.49	0.49	-0.19	-0.31	-0.21					0.31	0.02	-3.34	0.97	-4.41	0.94
44	575114	0	112360	0.58	0.18	0.58	0.13	0.11	0.00	0.00	0.53	-0.24	0.53	-0.29	-0.21					0.14	0.02	-9.90	0.90	-9.90	0.85
45	575116	0	112360	0.52	0.52	0.09	0.30	0.09	0.00	0.00	0.48	0.48	-0.27	-0.21	-0.22					0.29	0.02	-3.01	0.97	-4.23	0.94
46	578634	0	112360	0.47	0.25	0.47	0.17	0.11	0.01	0.00	0.45	-0.07	0.45	-0.28	-0.27					0.65	0.02	2.23	1.02	3.28	1.04
47	578639	0	112360	0.52	0.10	0.15	0.23	0.52	0.01	0.00	0.55	-0.27	-0.28	-0.21	0.55					0.15	0.02	-9.90	0.90	-9.90	0.86
48	582365	0	112360	0.56	0.18	0.14	0.56	0.11	0.00	0.00	0.53	-0.25	-0.28	0.53	-0.20					0.11	0.02	-9.70	0.91	-9.90	0.86
49	621570	1	20518	0.63	0.63	0.08	0.12	0.17	0.00	0.00	0.43	0.43	-0.19	-0.21	-0.24	A+	A-	A+	A+	-0.39	0.06	-1.36	0.97	0.07	1.00
50	621571	1	20518	0.44	0.15	0.44	0.18	0.23	0.00	0.00	0.41	-0.18	0.41	-0.24	-0.11	A-	A+	A+	A-	0.57	0.06	0.74	1.02	0.69	1.02
51	621668	1	20518	0.58	0.58	0.17	0.12	0.13	0.00	0.00	0.43	0.43	-0.20	-0.24	-0.18	A+	A+	A+	A-	-0.11	0.06	-0.58	0.99	-1.31	0.95
52	623010	1	20518	0.45	0.45	0.15	0.27	0.13	0.00	0.00	0.29	0.29	-0.10	-0.08	-0.21	A-	A-	A-	A-	0.51	0.06	6.80	1.17	5.81	1.20
53	623620	1	20518	0.39	0.20	0.39	0.23	0.18	0.00	0.00	0.10	0.03	0.10	-0.19	0.07	A+	A-	A-	A+	0.85	0.06	9.90	1.36	9.90	1.46
54	623624	1	20518	0.48	0.24	0.48	0.09	0.19	0.00	0.00	0.44	-0.13	0.44	-0.21	-0.25	A+	A-	A-	A-	0.40	0.06	-0.60	0.99	-1.15	0.96
55	624176	1	20518	0.34	0.14	0.31	0.34	0.21	0.00	0.00	0.22	-0.15	-0.09	0.22	-0.02	A-	A-	A-	A+	1.13	0.06	5.27	1.15	6.02	1.26
56	624184	1	20518	0.61	0.15	0.16	0.61	0.08	0.00	0.00	0.31	-0.17	-0.07	0.31	-0.22	A-	A-	A-	A+	-0.27	0.06	4.29	1.10	4.03	1.17
57	624369	1	20518	0.48	0.15	0.17	0.20	0.48	0.00	0.00	0.37	-0.20	-0.23	-0.06	0.37	A+	A-	A-	A-	0.41	0.06	3.58	1.08	3.33	1.11
58	632273	1	20518	0.66	0.10	0.09	0.66	0.15	0.00	0.00	0.40	-0.25	-0.21	0.40	-0.14	A+	A+	A-	A+	-0.52	0.06	-1.31	0.97	2.46	1.12
59	634558	1	20518	0.59	0.09	0.59	0.20	0.12	0.00	0.00	0.51	-0.13	0.51	-0.30	-0.28	A-	A+	A-	A-	-0.16	0.06	-5.54	0.88	-4.45	0.84
60	635326	1	20518	0.29	0.12	0.29	0.14	0.46	0.00	0.00	0.24	-0.10	0.24	-0.20	-0.01	A-	A+	A+	A+	1.43	0.06	5.67	1.18	7.12	1.38
61	635455	1	20518	0.86	0.03	0.86	0.06	0.05	0.00	0.00	0.44	-0.20	0.44	-0.27	-0.25	A+	A-	A-	A+	-1.83	0.08	-5.55	0.75	-6.61	0.47
62	635460	1	20518	0.37	0.10	0.28	0.26	0.37	0.00	0.00	0.34	-0.22	-0.02	-0.20	0.34	A-	A-	A-	A+	0.96	0.06	2.96	1.08	3.74	1.14
63	635841	1	20518	0.32	0.40	0.13	0.14	0.32	0.00	0.00	0.29	0.00	-0.13	-0.26	0.29	A+	A-	A-	A-	1.21	0.06	4.32	1.13	5.41	1.24
64	635844	1	20518	0.53	0.21	0.53	0.13	0.13	0.00	0.00	0.40	-0.15	0.40	-0.28	-0.12	A+	A+	A+	A+	0.14	0.06	2.33	1.05	2.01	1.07
65	618333	2	18475	0.53	0.17	0.53	0.20	0.09	0.00	0.00	0.42	-0.17	0.42	-0.24	-0.17	A+	A+	A+	A+	0.22	0.06	0.16	1.00	-0.31	0.99
66	621569	2	18475	0.61	0.07	0.61	0.08	0.24	0.00	0.00	0.40	-0.22	0.40	-0.26	-0.16	A-	A-	A-	A+	-0.19	0.06	0.53	1.01	-0.64	0.97

**Table J-6 (continued). Biology Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
67	621574	2	18475	0.46	0.46	0.08	0.29	0.16	0.00	0.00	0.29	0.29	-0.21	-0.14	-0.05	A-	A-	A+	A-	0.59	0.06	6.04	1.15	5.66	1.19
68	623009	2	18475	0.57	0.17	0.06	0.20	0.57	0.00	0.00	0.54	-0.20	-0.25	-0.33	0.54	A-	A-	A-	A+	0.04	0.06	-6.67	0.86	-6.16	0.80
69	623298	2	18475	0.41	0.13	0.41	0.20	0.25	0.00	0.00	0.34	-0.20	0.34	-0.19	-0.05	A+	A+	A+	A+	0.82	0.06	2.47	1.06	2.67	1.09
70	623630	2	18475	0.39	0.08	0.27	0.26	0.39	0.00	0.00	0.32	-0.21	-0.19	-0.03	0.32	A+	A+	A-	A+	0.94	0.06	6.10	1.16	6.25	1.24
71	624178	2	18475	0.47	0.20	0.17	0.47	0.15	0.00	0.00	0.44	-0.33	-0.07	0.44	-0.16	A-	A-	A-	A+	0.52	0.06	-1.90	0.96	-1.40	0.96
72	624180	2	18475	0.45	0.09	0.25	0.20	0.45	0.00	0.00	0.41	-0.19	-0.02	-0.34	0.41	A-	A-	A-	A-	0.64	0.06	-1.15	0.97	-1.47	0.95
73	630597	2	18475	0.45	0.21	0.45	0.24	0.10	0.00	0.00	0.19	0.08	0.19	-0.18	-0.15	A+	A+	A-	A+	0.62	0.06	9.90	1.32	9.90	1.42
74	630615	2	18475	0.57	0.12	0.57	0.18	0.12	0.00	0.00	0.47	-0.25	0.47	-0.21	-0.21	A-	A-	A+	A+	0.02	0.06	-2.55	0.94	-2.87	0.90
75	630634	2	18475	0.53	0.53	0.21	0.14	0.11	0.00	0.00	0.29	0.29	-0.20	0.05	-0.25	A-	A+	A+	A+	0.23	0.06	5.75	1.13	7.37	1.26
76	634393	2	18475	0.47	0.16	0.20	0.15	0.47	0.00	0.00	0.48	-0.20	-0.23	-0.19	0.48	A-	A+	A+	A+	0.51	0.06	-1.61	0.96	-1.64	0.95
77	634400	2	18475	0.55	0.16	0.17	0.55	0.12	0.00	0.00	0.49	-0.19	-0.25	0.49	-0.23	A+	A-	A-	A+	0.14	0.06	-3.11	0.93	-3.61	0.88
78	634701	2	18475	0.54	0.10	0.11	0.54	0.26	0.00	0.00	0.46	-0.23	-0.24	0.46	-0.20	A-	A+	A+	A-	0.20	0.06	-2.22	0.95	-2.94	0.90
79	635245	2	18475	0.41	0.13	0.41	0.17	0.29	0.00	0.00	0.33	-0.24	0.33	-0.19	-0.01	A+	A-	A-	A+	0.86	0.06	4.28	1.11	4.55	1.16
80	635840	2	18475	0.35	0.37	0.12	0.16	0.35	0.00	0.00	0.28	0.10	-0.19	-0.31	0.28	A+	A-	A+	A+	1.19	0.06	5.54	1.16	5.38	1.23
81	622974	3	18380	0.61	0.08	0.21	0.09	0.61	0.00	0.00	0.52	-0.24	-0.25	-0.29	0.52	A+	A-	A-	A+	-0.19	0.06	-4.99	0.89	-4.64	0.82
82	623296	3	18380	0.51	0.08	0.51	0.28	0.13	0.00	0.00	0.43	-0.03	0.43	-0.23	-0.31	A+	A-	A-	A+	0.31	0.06	-0.47	0.99	-0.33	0.99
83	624166	3	18380	0.59	0.11	0.59	0.13	0.16	0.00	0.00	0.49	-0.24	0.49	-0.33	-0.14	A+	A+	A+	A-	-0.11	0.06	-3.56	0.92	-1.27	0.95
84	625818	3	18380	0.60	0.26	0.06	0.60	0.07	0.00	0.00	0.29	-0.06	-0.25	0.29	-0.19	A-	A+	A+	A-	-0.15	0.06	5.20	1.13	6.39	1.27
85	630582	3	18380	0.55	0.18	0.55	0.12	0.14	0.00	0.00	0.41	-0.18	0.41	-0.22	-0.18	A+	A+	A+	A+	0.10	0.06	-0.02	1.00	-0.67	0.98
86	630619	3	18380	0.46	0.06	0.46	0.22	0.26	0.00	0.00	0.33	-0.23	0.33	-0.20	-0.07	A-	A-	A+	A+	0.56	0.06	5.95	1.15	5.72	1.20
87	634394	3	18380	0.38	0.16	0.38	0.18	0.28	0.00	0.00	0.28	-0.19	0.28	-0.07	-0.08	A-	A+	A-	A+	0.99	0.06	6.30	1.18	7.05	1.29
88	634396	3	18380	0.69	0.69	0.09	0.06	0.16	0.00	0.00	0.44	0.44	-0.26	-0.26	-0.17	A-	A+	A+	A+	-0.60	0.06	-2.28	0.94	-1.91	0.91
89	634399	3	18380	0.26	0.20	0.28	0.26	0.26	0.00	0.00	0.22	0.00	-0.11	-0.11	0.22	A-	A+	A+	A-	1.71	0.07	3.55	1.13	6.88	1.45
90	634549	3	18380	0.44	0.16	0.17	0.44	0.23	0.00	0.00	0.38	-0.26	-0.22	0.38	-0.03	A+	A-	A-	A+	0.66	0.06	2.27	1.06	2.88	1.10
91	634588	3	18380	0.82	0.82	0.08	0.06	0.04	0.00	0.00	0.43	0.43	-0.27	-0.23	-0.18	A+	A-	A-	A+	-1.42	0.07	-3.89	0.86	-4.09	0.70
92	634589	3	18380	0.79	0.09	0.79	0.06	0.06	0.00	0.00	0.38	-0.14	0.38	-0.25	-0.22	A-	A-	A-	A-	-1.21	0.07	-2.70	0.91	-1.40	0.90
93	635457	3	18380	0.35	0.35	0.15	0.34	0.16	0.00	0.00	0.33	0.33	-0.23	-0.05	-0.13	A-	A+	A-	A-	1.14	0.06	3.43	1.10	4.54	1.20
94	635458	3	18380	0.42	0.14	0.42	0.19	0.25	0.00	0.00	0.34	-0.17	0.34	-0.20	-0.06	A+	A+	A-	A+	0.80	0.06	3.63	1.09	4.05	1.15
95	635948	3	18380	0.42	0.42	0.14	0.27	0.17	0.00	0.00	0.40	0.40	-0.10	-0.20	-0.18	A-	A+	A-	A-	0.79	0.06	0.57	1.01	1.14	1.04
96	635949	3	18380	0.53	0.12	0.53	0.12	0.23	0.00	0.00	0.55	-0.20	0.55	-0.26	-0.29	A-	A-	A-	A-	0.24	0.06	-5.25	0.88	-5.15	0.83
97	618334	4	18273	0.48	0.19	0.48	0.22	0.10	0.00	0.00	0.25	-0.06	0.25	-0.07	-0.23	A+	A-	A+	A+	0.46	0.06	9.19	1.22	8.41	1.29
98	624988	4	18273	0.43	0.23	0.05	0.43	0.29	0.00	0.00	0.34	-0.08	-0.25	0.34	-0.17	A-	A-	A-	A-	0.72	0.06	3.50	1.09	2.73	1.09
99	625832	4	18273	0.56	0.56	0.16	0.09	0.19	0.00	0.00	0.52	0.52	-0.27	-0.24	-0.23	A+	A+	A+	A-	0.09	0.06	-4.82	0.90	-4.26	0.86

**Table J-6 (continued). Biology Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
100	626323	4	18273	0.56	0.35	0.04	0.04	0.56	0.00	0.00	0.38	-0.23	-0.20	-0.16	0.38	A-	A+	A+	A+	0.07	0.06	3.16	1.07	1.37	1.05
101	630583	4	18273	0.50	0.18	0.50	0.17	0.14	0.00	0.00	0.44	-0.14	0.44	-0.27	-0.17	A+	A+	A+	A+	0.36	0.06	-0.91	0.98	-1.54	0.95
102	630610	4	18273	0.42	0.20	0.23	0.42	0.14	0.00	0.00	0.25	-0.05	-0.11	0.25	-0.15	A+	A+	A+	A+	0.76	0.06	6.83	1.17	6.24	1.22
103	633581	4	18273	0.47	0.47	0.17	0.25	0.10	0.00	0.00	0.42	0.42	-0.18	-0.15	-0.25	A-	A-	A+	A+	0.51	0.06	-0.71	0.98	-0.38	0.99
104	634404	4	18273	0.61	0.02	0.11	0.61	0.26	0.00	0.00	0.42	-0.13	-0.17	0.42	-0.30	A+	A-	A-	A+	-0.16	0.06	-1.14	0.97	-1.08	0.96
105	634584	4	18273	0.37	0.37	0.35	0.10	0.19	0.00	0.00	0.38	0.38	-0.19	-0.28	-0.02	A-	A+	A-	A-	1.07	0.06	1.16	1.03	2.58	1.10
106	634587	4	18273	0.60	0.17	0.60	0.15	0.07	0.00	0.00	0.25	-0.16	0.25	-0.05	-0.17	A-	A-	A+	A-	-0.16	0.06	7.42	1.18	6.30	1.26
107	634707	4	18273	0.49	0.12	0.49	0.26	0.12	0.00	0.00	0.45	-0.20	0.45	-0.20	-0.20	A+	A+	A+	A+	0.41	0.06	-1.73	0.96	-2.02	0.94
108	635238	4	18273	0.69	0.07	0.10	0.69	0.14	0.00	0.00	0.48	-0.28	-0.24	0.48	-0.22	A+	A-	A-	A+	-0.59	0.06	-3.99	0.90	-4.15	0.81
109	635328	4	18273	0.35	0.28	0.24	0.13	0.35	0.00	0.00	0.51	-0.26	-0.16	-0.16	0.51	A-	A+	A+	A-	1.15	0.06	-2.80	0.93	-1.55	0.94
110	635946	4	18273	0.55	0.25	0.55	0.11	0.09	0.00	0.00	0.44	-0.18	0.44	-0.22	-0.23	A-	A-	A+	A+	0.14	0.06	-0.77	0.98	-1.95	0.94
111	635947	4	18273	0.62	0.05	0.18	0.15	0.62	0.00	0.00	0.50	-0.23	-0.30	-0.21	0.50	A-	A-	A-	A+	-0.22	0.06	-4.91	0.89	-4.49	0.83
112	732477	4	18273	0.21	0.21	0.09	0.24	0.46	0.00	0.00	0.02	0.02	-0.14	-0.30	0.32	B-	A+	A+	A-	2.06	0.07	7.78	1.34	9.90	1.90
113	618099	5	18300	0.74	0.09	0.74	0.03	0.13	0.00	0.00	0.43	-0.28	0.43	-0.17	-0.22	A+	A+	A+	A+	-0.87	0.06	-4.70	0.87	-4.76	0.74
114	621670	5	18300	0.44	0.44	0.20	0.25	0.11	0.00	0.00	0.31	0.31	-0.04	-0.11	-0.28	A-	A+	A+	A-	0.68	0.06	3.80	1.09	3.84	1.13
115	623295	5	18300	0.34	0.15	0.43	0.34	0.08	0.00	0.00	0.28	-0.23	-0.01	0.28	-0.17	A-	A+	A+	A+	1.21	0.06	4.51	1.13	5.47	1.23
116	623300	5	18300	0.41	0.14	0.24	0.41	0.21	0.00	0.00	0.39	-0.19	-0.08	0.39	-0.21	A-	A-	A-	A+	0.84	0.06	0.84	1.02	1.13	1.04
117	624986	5	18300	0.45	0.45	0.16	0.28	0.10	0.00	0.00	0.40	0.40	-0.24	-0.09	-0.22	A-	A-	A-	A-	0.61	0.06	1.25	1.03	0.87	1.03
118	626329	5	18300	0.54	0.14	0.54	0.21	0.11	0.00	0.00	0.28	-0.17	0.28	-0.12	-0.10	A+	A+	A+	A+	0.19	0.06	5.28	1.12	4.68	1.16
119	628190	5	18300	0.33	0.12	0.33	0.14	0.40	0.00	0.00	0.18	-0.21	0.18	-0.20	0.11	A-	A-	A+	A-	1.25	0.06	8.59	1.26	9.50	1.44
120	628195	5	18300	0.26	0.26	0.45	0.14	0.14	0.00	0.00	-0.04	-0.04	0.32	-0.25	-0.16	A+	A-	A-	A-	1.66	0.06	9.90	1.42	9.90	1.72
121	630592	5	18300	0.39	0.20	0.39	0.14	0.27	0.00	0.00	0.46	-0.19	0.46	-0.24	-0.14	A+	A+	A-	A+	0.96	0.06	-3.82	0.91	-2.36	0.92
122	630626	5	18300	0.36	0.16	0.36	0.27	0.21	0.00	0.00	0.38	-0.22	0.38	-0.22	-0.01	A+	A-	A-	A+	1.11	0.06	1.60	1.04	3.10	1.12
123	633583	5	18300	0.74	0.07	0.11	0.08	0.74	0.00	0.00	0.51	-0.26	-0.24	-0.30	0.51	A+	A-	A+	A+	-0.90	0.06	-7.35	0.79	-5.95	0.68
124	634408	5	18300	0.46	0.07	0.21	0.26	0.46	0.00	0.00	0.55	-0.22	-0.34	-0.18	0.55	A+	A-	A+	A+	0.58	0.06	-5.62	0.88	-4.81	0.85
125	634706	5	18300	0.40	0.28	0.17	0.15	0.40	0.00	0.00	0.31	-0.08	-0.14	-0.16	0.31	A-	A-	A-	A+	0.91	0.06	2.36	1.06	3.57	1.13
126	635234	5	18300	0.51	0.14	0.19	0.16	0.51	0.00	0.00	0.56	-0.20	-0.25	-0.30	0.56	A+	A-	A-	A-	0.34	0.06	-6.24	0.87	-5.76	0.83
127	636060	5	18300	0.32	0.15	0.32	0.28	0.25	0.00	0.00	0.08	-0.14	0.08	-0.13	0.17	A+	A+	A-	A-	1.31	0.06	9.90	1.39	9.90	1.59
128	636061	5	18300	0.82	0.04	0.10	0.82	0.04	0.00	0.00	0.38	-0.23	-0.20	0.38	-0.19	A+	A-	A-	A-	-1.41	0.07	-3.37	0.87	-2.48	0.81
129	620406	6	18414	0.56	0.16	0.56	0.14	0.14	0.00	0.00	0.47	-0.18	0.47	-0.25	-0.22	A+	A+	A-	A+	0.09	0.06	-2.07	0.95	-2.21	0.92
130	621671	6	18414	0.61	0.07	0.08	0.61	0.24	0.00	0.00	0.40	-0.28	-0.26	0.40	-0.11	A-	A-	A+	A+	-0.16	0.06	0.14	1.00	-0.22	0.99
131	621673	6	18414	0.62	0.08	0.11	0.62	0.19	0.00	0.00	0.39	-0.24	-0.27	0.39	-0.10	A+	A+	A-	A+	-0.26	0.06	-1.31	0.97	-1.07	0.95
132	621674	6	18414	0.45	0.20	0.25	0.45	0.10	0.00	0.00	0.41	-0.18	-0.13	0.41	-0.25	A+	A-	A+	A+	0.64	0.06	-0.58	0.99	-0.23	0.99

**Table J-6 (continued). Biology Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
133	623618	6	18414	0.53	0.53	0.06	0.14	0.27	0.00	0.00	0.32	0.32	-0.18	-0.24	-0.08	A+	A-	A-	A+	0.23	0.06	5.32	1.12	5.89	1.21
134	623621	6	18414	0.42	0.25	0.15	0.42	0.18	0.00	0.00	0.47	-0.13	-0.23	0.47	-0.22	A+	A+	A+	A-	0.78	0.06	-2.71	0.94	-1.84	0.94
135	628193	6	18414	0.38	0.40	0.11	0.38	0.11	0.00	0.00	0.05	0.21	-0.23	0.05	-0.19	A+	A+	A+	A+	0.96	0.06	9.90	1.39	9.90	1.55
136	628194	6	18414	0.33	0.14	0.33	0.32	0.20	0.00	0.00	-0.02	-0.08	-0.02	0.14	-0.06	A+	A+	A-	A+	1.24	0.06	9.90	1.45	9.90	1.66
137	630574	6	18414	0.39	0.16	0.19	0.26	0.39	0.00	0.00	0.38	-0.10	-0.15	-0.20	0.38	A-	A-	A+	A+	0.94	0.06	3.04	1.08	3.82	1.14
138	630596	6	18414	0.52	0.10	0.16	0.52	0.21	0.00	0.00	0.50	-0.19	-0.24	0.50	-0.24	A+	A+	A+	A-	0.28	0.06	-4.58	0.90	-4.55	0.86
139	630601	6	18414	0.28	0.19	0.28	0.23	0.29	0.00	0.00	0.11	-0.14	0.11	-0.09	0.10	A+	A+	A+	A+	1.55	0.06	6.88	1.23	9.43	1.53
140	630624	6	18414	0.42	0.11	0.17	0.42	0.29	0.00	0.00	0.36	-0.23	-0.21	0.36	-0.05	A-	A-	A-	A+	0.78	0.06	3.59	1.09	3.64	1.13
141	634401	6	18414	0.45	0.08	0.31	0.45	0.16	0.00	0.00	0.41	-0.21	-0.24	0.41	-0.08	A+	A-	A-	A-	0.61	0.06	-0.20	1.00	-0.08	1.00
142	635247	6	18414	0.24	0.24	0.37	0.15	0.24	0.00	0.00	0.19	0.19	0.11	-0.23	-0.11	A+	A-	A-	A-	1.82	0.07	4.85	1.18	8.36	1.56
143	636063	6	18414	0.68	0.08	0.68	0.12	0.11	0.00	0.00	0.53	-0.27	0.53	-0.29	-0.24	A+	A+	A-	A-	-0.58	0.06	-6.60	0.84	-6.02	0.72
144	636064	6	18414	0.49	0.19	0.20	0.49	0.11	0.00	0.00	0.38	-0.11	-0.17	0.38	-0.24	A+	A+	A+	A-	0.43	0.06	0.97	1.02	0.18	1.01

Table J-7. Literature Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	666339	0	108475	0.68	0.08	0.68	0.19	0.05	0.00	0.00	0.43	-0.23	0.43	-0.22	-0.22					0.03	0.01	6.81	1.04	6.72	1.07
2	666345	0	108475	0.54	0.11	0.06	0.54	0.28	0.00	0.00	0.34	-0.23	-0.32	0.34	-0.03					0.95	0.01	9.90	1.14	9.90	1.24
3	666346	0	108475	0.82	0.04	0.05	0.09	0.82	0.00	0.00	0.49	-0.27	-0.28	-0.25	0.49					-0.80	0.01	-9.90	0.84	-9.90	0.82
4	666347	0	108475	0.57	0.57	0.11	0.23	0.08	0.00	0.00	0.44	0.44	-0.29	-0.16	-0.21					0.64	0.01	5.80	1.03	8.56	1.07
5	666348	0	108475	0.40	0.28	0.40	0.08	0.24	0.00	0.00	0.34	-0.14	0.34	-0.16	-0.14					1.28	0.01	9.90	1.09	9.90	1.20
6	666350	0	108475	0.68	0.04	0.17	0.11	0.68	0.00	0.00	0.52	-0.25	-0.31	-0.22	0.52					-0.20	0.01	2.96	1.02	-0.96	0.99
7	666351	0	108475	0.62	0.62	0.10	0.18	0.09	0.00	0.00	0.42	0.42	-0.24	-0.15	-0.24					0.37	0.01	9.90	1.06	9.90	1.12
8	666813	0	108475	0.69	0.04	0.10	0.16	0.69	0.00	0.00	0.55	-0.28	-0.25	-0.33	0.55					-0.09	0.01	-9.90	0.91	-9.90	0.83
9	666814	0	108475	0.78	0.11	0.09	0.78	0.02	0.00	0.00	0.49	-0.39	-0.17	0.49	-0.23					-0.76	0.01	-1.36	0.99	2.40	1.04
10	666816	0	108475	0.77	0.09	0.08	0.05	0.77	0.00	0.00	0.34	-0.14	-0.18	-0.22	0.34					-0.75	0.01	9.90	1.18	9.90	1.69
11	666818	0	108475	0.61	0.04	0.61	0.30	0.05	0.00	0.00	0.30	-0.21	0.30	-0.12	-0.20					0.08	0.01	9.90	1.31	9.90	1.58
12	666819	0	108475	0.74	0.13	0.74	0.10	0.03	0.00	0.00	0.42	-0.13	0.42	-0.31	-0.24					-0.50	0.01	9.90	1.11	9.90	1.34
13	666820	0	108475	0.67	0.67	0.14	0.07	0.11	0.00	0.00	0.49	0.49	-0.24	-0.27	-0.24					-0.19	0.01	8.72	1.06	7.07	1.08
14	666824	0	108475	0.47	0.32	0.08	0.47	0.13	0.00	0.00	0.37	-0.12	-0.25	0.37	-0.16					1.23	0.01	9.90	1.12	9.90	1.22
15	667352	0	108475	0.25	0.25	0.25	0.28	0.21	0.00	0.00	0.22	-0.06	0.22	-0.14	0.00					1.71	0.01	9.90	1.09	9.90	1.33
16	667355	0	108475	0.73	0.73	0.07	0.07	0.13	0.00	0.00	0.46	0.46	-0.29	-0.30	-0.15					-0.27	0.01	-3.03	0.98	-1.12	0.99
17	580157	0	108475	0.53	0.05	0.53	0.08	0.33	0.00	0.00	0.35	-0.29	0.35	-0.23	-0.10					1.23	0.01	9.90	1.15	9.90	1.29
18	580159	0	108475	0.73	0.15	0.05	0.73	0.07	0.00	0.00	0.46	-0.23	-0.28	0.46	-0.23					-0.24	0.01	-3.01	0.98	-4.77	0.94
19	580161	0	108475	0.64	0.64	0.09	0.24	0.02	0.00	0.00	0.41	0.41	-0.12	-0.32	-0.18					0.26	0.01	9.90	1.07	6.25	1.06
20	580162	0	108475	0.64	0.18	0.64	0.14	0.03	0.00	0.00	0.35	-0.16	0.35	-0.18	-0.22					0.30	0.01	9.90	1.14	9.90	1.18
21	580163	0	108475	0.78	0.10	0.05	0.78	0.07	0.00	0.00	0.47	-0.24	-0.28	0.47	-0.25					-0.58	0.01	-9.90	0.90	-8.57	0.88
22	580165	0	108475	0.61	0.61	0.12	0.18	0.09	0.00	0.00	0.40	0.40	-0.10	-0.27	-0.19					0.34	0.01	9.90	1.11	9.90	1.14
23	580166	0	108475	0.83	0.06	0.05	0.83	0.06	0.00	0.00	0.46	-0.28	-0.28	0.46	-0.20					-1.10	0.02	-4.07	0.96	-6.94	0.87
24	580167	0	108475	0.69	0.19	0.04	0.08	0.69	0.00	0.00	0.39	-0.21	-0.27	-0.16	0.39					-0.05	0.01	9.90	1.09	9.90	1.24
25	580168	0	108475	0.50	0.50	0.13	0.31	0.06	0.00	0.00	0.25	0.25	-0.16	-0.07	-0.15					1.04	0.01	9.90	1.26	9.90	1.38
26	580169	0	108475	0.65	0.06	0.65	0.22	0.06	0.00	0.00	0.42	-0.20	0.42	-0.19	-0.29					-0.09	0.01	9.90	1.16	9.90	1.21
27	586928	0	108475	0.49	0.37	0.06	0.49	0.08	0.00	0.00	0.39	-0.16	-0.21	0.39	-0.26					1.10	0.01	9.90	1.08	9.90	1.14
28	586929	0	108475	0.60	0.09	0.03	0.27	0.60	0.00	0.00	0.36	-0.24	-0.23	-0.14	0.36					0.31	0.01	9.90	1.17	9.90	1.29
29	586930	0	108475	0.80	0.80	0.02	0.12	0.06	0.00	0.00	0.38	0.38	-0.22	-0.22	-0.19					-0.84	0.01	6.56	1.06	7.79	1.13
30	586932	0	108475	0.50	0.10	0.05	0.35	0.50	0.00	0.00	0.30	-0.15	-0.20	-0.12	0.30					1.54	0.01	9.90	1.28	9.90	1.52
31	586933	0	108475	0.86	0.86	0.06	0.04	0.03	0.00	0.00	0.48	0.48	-0.30	-0.28	-0.20					-1.27	0.02	-9.90	0.86	-9.90	0.71
32	586936	0	108475	0.58	0.18	0.58	0.12	0.11	0.00	0.00	0.40	-0.10	0.40	-0.26	-0.23					0.79	0.01	9.90	1.07	9.90	1.13
33	586937	0	108475	0.76	0.05	0.13	0.76	0.06	0.00	0.00	0.26	-0.13	-0.17	0.26	-0.11					-0.15	0.01	9.90	1.12	9.90	1.26

Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	586994	0	108475	0.63	0.13	0.13	0.63	0.11	0.00	0.00	0.46	-0.30	-0.13	0.46	-0.25					0.55	0.01	-5.42	0.97	-2.19	0.98
35	709645	1	4569	0.68	0.05	0.09	0.18	0.68	0.00	0.00	0.39	-0.25	-0.20	-0.18	0.39	A-	A-	A-	A-	0.03	0.06	1.54	1.05	2.27	1.11
36	709648	1	4569	0.64	0.07	0.13	0.64	0.16	0.00	0.00	0.22	-0.13	-0.11	0.22	-0.09	A+	A+	A+	A+	0.28	0.06	8.65	1.26	9.90	1.52
37	709652	1	4569	0.77	0.03	0.10	0.77	0.09	0.00	0.00	0.39	-0.29	-0.25	0.39	-0.12	A+	B+	A+	A-	-0.57	0.07	0.12	1.00	-0.15	0.99
38	709653	1	4569	0.57	0.57	0.17	0.15	0.11	0.00	0.00	0.38	0.38	-0.14	-0.19	-0.20	A+	A+	A+	A-	0.65	0.06	1.31	1.03	1.71	1.06
39	709654	1	4569	0.71	0.08	0.16	0.05	0.71	0.00	0.00	0.35	-0.14	-0.13	-0.33	0.35	A-	A-	A-	A-	-0.13	0.06	2.34	1.08	2.45	1.13
40	709741	1	4569	0.68	0.11	0.68	0.17	0.04	0.00	0.00	0.40	-0.21	0.40	-0.21	-0.22	A-	A-	A-	A-	0.05	0.06	1.25	1.04	0.40	1.02
41	709742	1	4569	0.51	0.19	0.18	0.51	0.12	0.00	0.00	0.38	-0.07	-0.21	0.38	-0.25	B-	A-	A-	A+	0.95	0.06	3.10	1.07	3.81	1.14
42	709743	1	4569	0.81	0.06	0.05	0.07	0.81	0.00	0.00	0.48	-0.24	-0.30	-0.23	0.48	B-	C-	C-	B+	-0.86	0.07	-2.54	0.90	-2.58	0.80
43	709747	1	4569	0.62	0.04	0.07	0.26	0.62	0.00	0.00	0.42	-0.32	-0.29	-0.13	0.42	C-	B-	A-	A+	0.34	0.06	1.57	1.04	1.68	1.07
44	709749	1	4569	0.41	0.33	0.41	0.09	0.17	0.00	0.00	0.05	0.26	0.05	-0.29	-0.16	A-	A-	A-	A+	1.47	0.06	9.90	1.42	9.90	1.71
45	709750	1	4569	0.66	0.07	0.23	0.66	0.04	0.00	0.00	0.32	-0.24	-0.09	0.32	-0.26	A+	A+	A-	A+	0.13	0.06	4.43	1.13	2.90	1.14
46	711017	1	4569	0.91	0.02	0.91	0.05	0.02	0.00	0.00	0.40	-0.20	0.40	-0.28	-0.16	A-	A-	B-	A+	-1.91	0.10	-3.67	0.76	-4.30	0.50
47	709643	2	4510	0.46	0.05	0.46	0.35	0.14	0.00	0.00	0.34	-0.26	0.34	-0.13	-0.15	B-	A-	A-	A+	1.17	0.06	2.57	1.06	4.43	1.16
48	709644	2	4510	0.89	0.03	0.03	0.89	0.04	0.00	0.00	0.35	-0.25	-0.17	0.35	-0.17	A-	A+	A-	A+	-1.60	0.09	-0.98	0.94	2.68	1.36
49	709647	2	4510	0.73	0.07	0.12	0.08	0.73	0.00	0.00	0.48	-0.20	-0.32	-0.21	0.48	A+	A-	A-	A-	-0.29	0.07	-2.69	0.91	-1.93	0.89
50	709650	2	4510	0.78	0.11	0.78	0.07	0.04	0.00	0.00	0.33	-0.14	0.33	-0.22	-0.17	A-	A-	A-	A+	-0.65	0.07	3.07	1.12	3.27	1.25
51	709651	2	4510	0.67	0.03	0.67	0.05	0.25	0.00	0.00	0.46	-0.20	0.46	-0.27	-0.29	A-	A+	A+	A-	0.04	0.06	-0.11	1.00	-0.85	0.96
52	709655	2	4510	0.59	0.09	0.18	0.59	0.14	0.00	0.00	0.33	-0.28	-0.13	0.33	-0.09	A+	A-	A-	A-	0.51	0.06	4.28	1.11	5.64	1.23
53	709740	2	4510	0.39	0.39	0.43	0.14	0.04	0.00	0.00	0.23	0.23	-0.02	-0.17	-0.20	A-	A-	A-	A+	1.53	0.06	7.30	1.18	8.96	1.41
54	709745	2	4510	0.80	0.09	0.80	0.06	0.05	0.00	0.00	0.53	-0.34	0.53	-0.32	-0.18	B-	A-	A-	A+	-0.79	0.07	-5.13	0.81	-4.88	0.67
55	709746	2	4510	0.46	0.30	0.14	0.46	0.10	0.00	0.00	0.30	-0.17	-0.11	0.30	-0.11	A-	A-	A-	A+	1.14	0.06	4.69	1.11	4.77	1.18
56	709748	2	4510	0.72	0.72	0.11	0.09	0.07	0.00	0.00	0.51	0.51	-0.29	-0.26	-0.23	A+	A-	A-	A-	-0.26	0.06	-4.28	0.87	-4.06	0.79
57	709751	2	4510	0.70	0.07	0.12	0.10	0.70	0.00	0.00	0.45	-0.25	-0.20	-0.23	0.45	A-	B-	A-	A+	-0.11	0.06	-1.82	0.95	-2.26	0.89
58	709752	2	4510	0.66	0.15	0.12	0.07	0.66	0.00	0.00	0.34	-0.15	-0.18	-0.19	0.34	A+	A-	A-	A+	0.11	0.06	3.24	1.09	4.09	1.20
59	718796	3	4551	0.91	0.04	0.03	0.02	0.91	0.00	0.00	0.44	-0.21	-0.27	-0.24	0.44	A+	A-	A-	A+	-1.87	0.10	-3.83	0.76	-1.31	0.82
60	718797	3	4551	0.84	0.07	0.06	0.84	0.03	0.00	0.00	0.45	-0.25	-0.26	0.45	-0.20	A+	A-	A-	A+	-1.07	0.08	-2.62	0.88	-1.36	0.88
61	718800	3	4551	0.83	0.83	0.04	0.05	0.07	0.00	0.00	0.42	0.42	-0.25	-0.25	-0.19	A-	A-	A-	A-	-1.04	0.08	-2.97	0.87	-0.75	0.93
62	718803	3	4551	0.77	0.15	0.04	0.04	0.77	0.00	0.00	0.39	-0.11	-0.31	-0.29	0.39	A-	A-	A-	A-	-0.55	0.07	-1.12	0.96	4.34	1.33
63	718805	3	4551	0.84	0.04	0.84	0.06	0.05	0.00	0.00	0.55	-0.29	0.55	-0.30	-0.29	A+	A+	A+	A-	-1.13	0.08	-5.59	0.76	-5.84	0.53
64	718806	3	4551	0.65	0.13	0.09	0.65	0.13	0.00	0.00	0.34	-0.24	-0.22	0.34	-0.05	A+	A+	A-	A-	0.20	0.06	3.87	1.11	4.01	1.20
65	722855	3	4551	0.89	0.89	0.02	0.07	0.02	0.00	0.00	0.37	0.37	-0.21	-0.22	-0.21	A-	A-	A+	A+	-1.63	0.09	-0.85	0.95	0.32	1.04
66	722856	3	4551	0.64	0.04	0.14	0.18	0.64	0.00	0.00	0.24	-0.24	-0.10	-0.07	0.24	A-	A-	A-	A+	0.26	0.06	8.00	1.24	7.95	1.40

**Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
67	722858	3	4551	0.76	0.07	0.76	0.05	0.12	0.00	0.00	0.56	-0.30	0.56	-0.31	-0.29	A-	A-	A-	A-	-0.49	0.07	-4.11	0.86	-3.78	0.76
68	722862	3	4551	0.74	0.08	0.74	0.10	0.08	0.00	0.00	0.44	-0.28	0.44	-0.17	-0.24	A+	A+	A+	A+	-0.32	0.07	-1.09	0.96	-1.41	0.91
69	722863	3	4551	0.69	0.69	0.13	0.07	0.11	0.00	0.00	0.48	0.48	-0.26	-0.22	-0.25	A-	A+	A+	A-	-0.01	0.06	-2.60	0.92	-2.48	0.88
70	722864	3	4551	0.61	0.61	0.17	0.05	0.18	0.00	0.00	0.30	0.30	-0.08	-0.27	-0.14	B-	A-	A-	A-	0.45	0.06	7.41	1.21	6.21	1.28
71	718798	4	4573	0.56	0.30	0.56	0.09	0.04	0.00	0.00	0.35	-0.13	0.35	-0.20	-0.27	A-	A+	A-	A-	0.69	0.06	3.36	1.09	4.12	1.17
72	718799	4	4573	0.84	0.84	0.08	0.03	0.05	0.00	0.00	0.39	0.39	-0.15	-0.24	-0.27	A-	A-	A+	A-	-1.06	0.08	-2.11	0.91	-0.80	0.92
73	718802	4	4573	0.77	0.09	0.09	0.77	0.05	0.00	0.00	0.48	-0.29	-0.23	0.48	-0.24	A-	A-	A-	A-	-0.51	0.07	-3.19	0.89	-3.07	0.80
74	718804	4	4573	0.72	0.72	0.05	0.14	0.09	0.00	0.00	0.49	0.49	-0.32	-0.15	-0.33	A+	A-	A+	A-	-0.20	0.07	-2.54	0.92	-1.67	0.90
75	718807	4	4573	0.84	0.02	0.12	0.02	0.84	0.00	0.00	0.35	-0.24	-0.17	-0.24	0.35	A-	A-	A-	A-	-1.05	0.08	-0.22	0.99	1.96	1.20
76	721728	4	4573	0.92	0.02	0.92	0.02	0.03	0.00	0.00	0.37	-0.24	0.37	-0.19	-0.18	A+	A+	A-	A+	-2.02	0.10	-2.22	0.85	-2.70	0.63
77	722852	4	4573	0.92	0.02	0.03	0.04	0.92	0.00	0.00	0.40	-0.21	-0.23	-0.23	0.40	A+	A+	A-	A-	-1.95	0.10	-1.21	0.92	-2.18	0.70
78	722854	4	4573	0.70	0.03	0.70	0.20	0.07	0.00	0.00	0.41	-0.20	0.41	-0.20	-0.28	A-	A+	A-	A+	-0.07	0.06	1.47	1.05	0.79	1.04
79	722857	4	4573	0.78	0.11	0.07	0.78	0.05	0.00	0.00	0.46	-0.25	-0.23	0.46	-0.26	A-	A-	A-	A+	-0.60	0.07	-1.55	0.94	-2.27	0.84
80	722859	4	4573	0.79	0.79	0.05	0.05	0.12	0.00	0.00	0.51	0.51	-0.29	-0.29	-0.26	A+	A+	A+	A-	-0.65	0.07	-2.92	0.89	-3.87	0.73
81	722861	4	4573	0.35	0.39	0.11	0.35	0.14	0.00	0.00	0.24	-0.02	-0.18	0.24	-0.13	A-	A-	A-	A+	1.76	0.06	4.64	1.12	6.96	1.37
82	723242	4	4573	0.69	0.04	0.69	0.21	0.06	0.00	0.00	0.30	-0.16	0.30	-0.12	-0.23	A-	A+	A-	A+	-0.01	0.06	5.32	1.17	4.99	1.29
83	719960	5	4524	0.42	0.15	0.34	0.09	0.42	0.00	0.00	0.28	-0.28	0.06	-0.24	0.28	C-	A-	A-	A+	1.40	0.06	6.47	1.15	8.11	1.34
84	719961	5	4524	0.73	0.20	0.73	0.05	0.02	0.00	0.00	0.36	-0.16	0.36	-0.27	-0.22	A+	A-	A-	A+	-0.26	0.07	1.67	1.06	0.47	1.03
85	719964	5	4524	0.75	0.08	0.12	0.05	0.75	0.00	0.00	0.53	-0.30	-0.24	-0.32	0.53	B-	A-	B-	A-	-0.38	0.07	-3.50	0.88	-2.80	0.84
86	719965	5	4524	0.68	0.09	0.05	0.68	0.18	0.00	0.00	0.45	-0.26	-0.29	0.45	-0.19	C-	A-	A-	A-	0.03	0.06	-1.72	0.95	-1.81	0.92
87	719966	5	4524	0.37	0.17	0.37	0.23	0.24	0.00	0.00	0.15	-0.14	0.15	-0.01	-0.03	A+	A+	A+	A-	1.68	0.06	9.89	1.25	9.90	1.55
88	721742	5	4524	0.50	0.14	0.50	0.14	0.22	0.00	0.00	0.41	-0.14	0.41	-0.15	-0.25	B-	A-	A+	A+	1.00	0.06	-0.88	0.98	0.48	1.02
89	722936	5	4524	0.96	0.02	0.96	0.01	0.01	0.00	0.00	0.33	-0.19	0.33	-0.18	-0.18	A+	A-	B-	A+	-2.74	0.14	-1.35	0.85	-1.61	0.69
90	722939	5	4524	0.86	0.86	0.07	0.05	0.02	0.00	0.00	0.49	0.49	-0.27	-0.32	-0.22	A-	A-	A-	A-	-1.25	0.08	-1.88	0.90	-3.42	0.69
91	722941	5	4524	0.65	0.03	0.65	0.26	0.07	0.00	0.00	0.33	-0.18	0.33	-0.15	-0.25	A-	A-	A-	A+	0.24	0.06	3.95	1.11	3.86	1.17
92	722942	5	4524	0.80	0.80	0.05	0.11	0.04	0.00	0.00	0.41	0.41	-0.21	-0.22	-0.26	A+	A-	A-	A-	-0.72	0.07	-0.60	0.98	-0.68	0.95
93	722943	5	4524	0.74	0.74	0.05	0.13	0.07	0.00	0.00	0.13	0.13	-0.09	0.02	-0.17	A-	A+	A+	A+	-0.36	0.07	9.49	1.37	9.90	1.87
94	722947	5	4524	0.71	0.04	0.71	0.17	0.08	0.00	0.00	0.40	-0.26	0.40	-0.24	-0.17	A+	A-	B-	A+	-0.12	0.06	0.85	1.03	-0.36	0.98
95	719956	6	4549	0.68	0.04	0.02	0.26	0.68	0.00	0.00	0.13	-0.30	-0.24	0.08	0.13	A+	A+	A-	A+	0.05	0.06	9.90	1.32	9.90	1.85
96	719957	6	4549	0.88	0.03	0.03	0.88	0.06	0.00	0.00	0.49	-0.24	-0.28	0.49	-0.28	A-	B-	C-	A+	-1.41	0.09	-4.12	0.78	-5.08	0.55
97	719958	6	4549	0.62	0.14	0.62	0.08	0.16	0.00	0.00	0.20	-0.18	0.20	-0.25	0.09	A-	A+	A+	A-	0.36	0.06	8.79	1.25	8.82	1.38
98	719963	6	4549	0.64	0.64	0.09	0.12	0.14	0.00	0.00	0.46	0.46	-0.29	-0.28	-0.13	A-	A-	A-	A+	0.27	0.06	-0.26	0.99	-0.79	0.97
99	719967	6	4549	0.62	0.18	0.62	0.11	0.08	0.00	0.00	0.38	-0.10	0.38	-0.24	-0.24	A+	A-	A-	A-	0.35	0.06	2.33	1.06	3.08	1.13



**Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
100	719968	6	4549	0.68	0.68	0.14	0.13	0.04	0.00	0.00	0.39	0.39	-0.21	-0.18	-0.24	A+	A+	A+	A-	0.05	0.06	0.95	1.03	2.01	1.10
101	722937	6	4549	0.28	0.05	0.06	0.28	0.60	0.00	0.00	0.07	-0.06	-0.27	0.07	0.10	A+	A-	A+	A+	2.11	0.06	9.90	1.32	9.90	1.91
102	722938	6	4549	0.68	0.21	0.06	0.05	0.68	0.00	0.00	0.30	-0.02	-0.30	-0.28	0.30	A-	A+	A+	A-	0.02	0.06	2.83	1.09	3.96	1.20
103	722944	6	4549	0.72	0.72	0.19	0.03	0.05	0.00	0.00	0.31	0.31	-0.10	-0.16	-0.28	A-	A+	A+	A-	-0.20	0.06	1.98	1.07	2.30	1.13
104	722945	6	4549	0.51	0.06	0.31	0.13	0.51	0.00	0.00	0.18	-0.24	0.11	-0.24	0.18	A-	A-	A+	A+	0.95	0.06	9.90	1.24	9.35	1.34
105	722946	6	4549	0.85	0.02	0.11	0.85	0.02	0.00	0.00	0.47	-0.24	-0.33	0.47	-0.19	A-	A-	A-	A+	-1.16	0.08	-2.29	0.89	-2.46	0.78
106	723042	6	4549	0.60	0.19	0.16	0.60	0.05	0.00	0.00	0.36	-0.18	-0.16	0.36	-0.19	A+	A+	A-	A-	0.50	0.06	3.64	1.09	3.23	1.12
107	717814	7	4524	0.26	0.34	0.33	0.26	0.07	0.00	0.00	0.22	0.04	-0.11	0.22	-0.25	A-	A-	A-	A-	2.29	0.06	4.20	1.14	8.64	1.61
108	717815	7	4524	0.64	0.08	0.24	0.04	0.64	0.00	0.00	0.34	-0.29	-0.10	-0.21	0.34	A-	A-	A-	A-	0.28	0.06	4.12	1.11	3.68	1.16
109	717816	7	4524	0.47	0.22	0.47	0.16	0.14	0.00	0.00	0.39	-0.17	0.39	-0.18	-0.16	A+	A+	A-	A+	1.16	0.06	-0.37	0.99	1.33	1.05
110	717817	7	4524	0.77	0.09	0.77	0.09	0.04	0.00	0.00	0.47	-0.32	0.47	-0.24	-0.16	A+	A-	A-	B-	-0.54	0.07	-4.11	0.86	-2.32	0.85
111	718079	7	4524	0.53	0.08	0.05	0.53	0.34	0.00	0.00	0.33	-0.30	-0.16	0.33	-0.10	A-	A-	A-	A-	0.88	0.06	3.47	1.08	3.52	1.13
112	720579	7	4524	0.45	0.12	0.12	0.31	0.45	0.00	0.00	0.42	-0.23	-0.23	-0.12	0.42	C-	A-	A-	A+	1.28	0.06	-1.37	0.97	0.55	1.02
113	720580	7	4524	0.75	0.08	0.15	0.75	0.02	0.00	0.00	0.42	-0.28	-0.21	0.42	-0.20	A+	A-	A-	A+	-0.38	0.07	-2.86	0.91	-2.99	0.82
114	720581	7	4524	0.62	0.17	0.62	0.09	0.13	0.00	0.00	0.28	-0.02	0.28	-0.26	-0.15	A+	A-	A-	A-	0.40	0.06	6.46	1.17	9.20	1.41
115	720582	7	4524	0.61	0.61	0.17	0.14	0.07	0.00	0.00	0.46	0.46	-0.22	-0.23	-0.23	A-	A-	A-	A+	0.45	0.06	-2.01	0.95	-2.78	0.89
116	720583	7	4524	0.44	0.17	0.28	0.11	0.44	0.00	0.00	0.27	-0.09	-0.22	0.00	0.27	B-	A-	B-	A+	1.33	0.06	7.54	1.18	8.06	1.32
117	720585	7	4524	0.47	0.17	0.08	0.27	0.47	0.00	0.00	0.46	-0.19	-0.34	-0.14	0.46	A-	A-	A+	A-	1.15	0.06	-3.18	0.93	-1.55	0.95
118	721748	7	4524	0.65	0.65	0.21	0.03	0.11	0.00	0.00	0.42	0.42	-0.23	-0.26	-0.20	A-	A-	A-	A+	0.21	0.06	0.49	1.01	0.25	1.01
119	717808	8	4534	0.75	0.75	0.05	0.14	0.06	0.00	0.00	0.35	0.35	-0.25	-0.17	-0.14	A+	A+	A+	A-	-0.40	0.07	-0.45	0.98	0.82	1.05
120	717809	8	4534	0.13	0.12	0.13	0.06	0.69	0.00	0.00	-0.01	-0.29	-0.01	-0.22	0.32	A-	A-	A-	A+	3.22	0.08	3.76	1.21	9.90	2.63
121	717810	8	4534	0.65	0.07	0.09	0.65	0.18	0.00	0.00	0.27	-0.08	-0.23	0.27	-0.10	A+	A+	A+	A-	0.17	0.06	5.45	1.16	4.89	1.23
122	717812	8	4534	0.57	0.57	0.19	0.10	0.14	0.00	0.00	0.38	0.38	-0.12	-0.25	-0.18	A+	A-	A+	A-	0.63	0.06	1.94	1.05	2.37	1.09
123	717813	8	4534	0.75	0.08	0.07	0.09	0.75	0.00	0.00	0.54	-0.33	-0.29	-0.22	0.54	A+	A-	A-	A+	-0.43	0.07	-4.82	0.84	-5.26	0.71
124	717819	8	4534	0.59	0.22	0.59	0.10	0.08	0.00	0.00	0.20	0.01	0.20	-0.17	-0.16	A-	A+	A-	A+	0.50	0.06	8.98	1.24	8.70	1.36
125	720584	8	4534	0.58	0.12	0.20	0.58	0.09	0.00	0.00	0.44	-0.25	-0.18	0.44	-0.20	A-	A-	A-	A+	0.57	0.06	-0.65	0.98	-0.33	0.99
126	720588	8	4534	0.73	0.04	0.14	0.73	0.09	0.00	0.00	0.40	-0.23	-0.21	0.40	-0.20	A-	A-	A-	A+	-0.27	0.06	-0.88	0.97	-1.11	0.94
127	720589	8	4534	0.49	0.05	0.49	0.33	0.13	0.00	0.00	0.21	-0.25	0.21	0.09	-0.27	A-	A-	A-	A-	1.03	0.06	8.44	1.20	7.40	1.27
128	720590	8	4534	0.72	0.72	0.10	0.11	0.08	0.00	0.00	0.36	0.36	-0.26	-0.17	-0.11	A-	A-	A-	A-	-0.20	0.06	2.85	1.09	3.39	1.19
129	720690	8	4534	0.18	0.43	0.10	0.18	0.28	0.00	0.00	-0.08	0.21	-0.20	-0.08	-0.02	B+	A+	A-	A+	2.83	0.07	8.91	1.42	9.90	2.88
130	722265	8	4534	0.24	0.49	0.06	0.24	0.21	0.00	0.00	0.17	0.01	-0.23	0.17	-0.05	A-	A+	A-	A-	2.38	0.07	2.40	1.08	7.68	1.58
131	720158	9	4525	0.55	0.27	0.55	0.03	0.14	0.00	0.00	0.49	-0.33	0.49	-0.26	-0.15	A+	A+	A+	A-	0.69	0.06	-3.18	0.92	-0.47	0.98
132	720163	9	4525	0.53	0.19	0.10	0.53	0.18	0.00	0.00	0.32	-0.07	-0.32	0.32	-0.09	A-	A-	A+	A-	0.83	0.06	6.40	1.16	6.67	1.28

**Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
133	720165	9	4525	0.52	0.25	0.11	0.12	0.52	0.00	0.00	0.42	-0.13	-0.24	-0.23	0.42	A-	A-	A-	A+	0.85	0.06	1.77	1.04	2.97	1.12
134	720166	9	4525	0.80	0.09	0.80	0.03	0.08	0.00	0.00	0.44	-0.30	0.44	-0.18	-0.20	A-	B-	B-	A+	-0.78	0.07	-0.44	0.98	-1.04	0.92
135	720167	9	4525	0.63	0.13	0.19	0.05	0.63	0.00	0.00	0.48	-0.31	-0.15	-0.32	0.48	A-	A+	A+	A-	0.27	0.06	-2.57	0.93	-2.25	0.90
136	720780	9	4525	0.84	0.05	0.04	0.07	0.84	0.00	0.00	0.54	-0.24	-0.29	-0.33	0.54	A-	A-	A-	A+	-1.14	0.08	-5.53	0.76	-5.01	0.58
137	720783	9	4525	0.81	0.04	0.08	0.06	0.81	0.00	0.00	0.50	-0.27	-0.31	-0.22	0.50	A-	A+	A-	A+	-0.88	0.07	-3.77	0.85	-2.70	0.78
138	720784	9	4525	0.60	0.60	0.16	0.16	0.08	0.00	0.00	0.35	0.35	-0.11	-0.14	-0.29	A+	A+	A+	A+	0.42	0.06	4.04	1.11	2.80	1.13
139	720785	9	4525	0.89	0.04	0.89	0.04	0.03	0.00	0.00	0.54	-0.32	0.54	-0.29	-0.29	A+	A+	A+	A-	-1.57	0.09	-5.20	0.73	-5.71	0.44
140	720787	9	4525	0.87	0.87	0.08	0.03	0.02	0.00	0.00	0.50	0.50	-0.34	-0.26	-0.22	A-	A-	A-	A+	-1.41	0.08	-5.03	0.75	-5.61	0.48
141	720788	9	4525	0.74	0.08	0.04	0.14	0.74	0.00	0.00	0.42	-0.13	-0.29	-0.27	0.42	A-	A+	A+	A+	-0.37	0.07	-1.02	0.96	-0.10	0.99
142	722197	9	4525	0.48	0.48	0.08	0.05	0.39	0.00	0.00	0.39	0.39	-0.27	-0.27	-0.12	A+	A+	A-	A+	1.07	0.06	1.12	1.03	1.12	1.04
143	720156	10	4495	0.44	0.11	0.09	0.44	0.36	0.00	0.00	0.27	-0.11	-0.27	0.27	-0.04	A-	A+	A-	A+	1.30	0.06	7.34	1.18	8.10	1.35
144	720157	10	4495	0.50	0.18	0.18	0.50	0.13	0.00	0.00	0.37	-0.16	-0.18	0.37	-0.15	A-	A-	A-	A-	0.96	0.06	3.31	1.08	3.19	1.12
145	720159	10	4495	0.60	0.60	0.13	0.08	0.19	0.00	0.00	0.30	0.30	-0.24	-0.12	-0.09	A+	A-	A-	A+	0.48	0.06	6.26	1.17	4.69	1.21
146	720160	10	4495	0.83	0.08	0.05	0.04	0.83	0.00	0.00	0.52	-0.32	-0.25	-0.25	0.52	A-	A-	B-	A-	-0.98	0.08	-3.13	0.87	-3.15	0.73
147	720161	10	4495	0.58	0.22	0.58	0.10	0.09	0.00	0.00	0.45	-0.23	0.45	-0.15	-0.27	A+	A+	A-	A-	0.56	0.06	-0.48	0.99	-0.98	0.96
148	720164	10	4495	0.74	0.74	0.08	0.08	0.09	0.00	0.00	0.55	0.55	-0.26	-0.32	-0.27	A-	A-	A-	A-	-0.35	0.07	-5.45	0.82	-5.46	0.68
149	720781	10	4495	0.63	0.16	0.09	0.63	0.12	0.00	0.00	0.49	-0.24	-0.28	0.49	-0.20	A+	B-	B-	A+	0.28	0.06	-2.59	0.93	-2.22	0.90
150	720782	10	4495	0.60	0.22	0.60	0.15	0.03	0.00	0.00	0.43	-0.23	0.43	-0.18	-0.27	A+	A-	A-	A+	0.46	0.06	1.37	1.04	2.40	1.10
151	720786	10	4495	0.69	0.09	0.12	0.69	0.10	0.00	0.00	0.49	-0.24	-0.33	0.49	-0.16	A+	A+	A-	A+	-0.07	0.06	-2.63	0.92	-2.22	0.88
152	720789	10	4495	0.50	0.25	0.17	0.50	0.08	0.00	0.00	0.26	-0.06	-0.23	0.26	-0.07	A-	A-	A-	A+	1.00	0.06	8.03	1.20	8.29	1.34
153	720790	10	4495	0.63	0.09	0.63	0.08	0.19	0.00	0.00	0.34	-0.20	0.34	-0.30	-0.05	A+	A+	A+	A+	0.29	0.06	4.19	1.12	6.15	1.31
154	720792	10	4495	0.73	0.07	0.73	0.07	0.13	0.00	0.00	0.52	-0.27	0.52	-0.24	-0.28	A-	B-	C-	A+	-0.25	0.07	-3.54	0.89	-3.83	0.78
155	721697	11	4505	0.61	0.07	0.28	0.61	0.05	0.00	0.00	0.51	-0.19	-0.43	0.51	-0.04	A-	A-	A+	A+	0.43	0.06	-3.21	0.92	-3.05	0.88
156	721699	11	4505	0.57	0.09	0.18	0.15	0.57	0.00	0.00	0.48	-0.25	-0.25	-0.18	0.48	A-	A-	A+	A+	0.61	0.06	-1.96	0.95	-1.79	0.93
157	721701	11	4505	0.61	0.09	0.12	0.61	0.19	0.00	0.00	0.37	-0.28	-0.26	0.37	-0.05	A-	A-	A-	A+	0.42	0.06	2.51	1.07	2.16	1.09
158	721702	11	4505	0.79	0.08	0.09	0.79	0.04	0.00	0.00	0.48	-0.26	-0.28	0.48	-0.22	A-	A-	A-	A+	-0.67	0.07	-3.10	0.88	-4.04	0.72
159	721703	11	4505	0.53	0.11	0.53	0.20	0.16	0.00	0.00	0.41	-0.19	0.41	-0.25	-0.11	A-	A-	A-	A+	0.84	0.06	0.49	1.01	0.04	1.00
160	721706	11	4505	0.41	0.41	0.35	0.08	0.16	0.00	0.00	0.20	0.20	0.08	-0.22	-0.20	A-	A-	A-	A-	1.44	0.06	9.14	1.23	8.85	1.39
161	722793	11	4505	0.68	0.25	0.03	0.04	0.68	0.00	0.00	0.53	-0.39	-0.19	-0.22	0.53	A-	C-	A-	A+	0.02	0.06	-4.20	0.88	-4.16	0.80
162	722794	11	4505	0.61	0.61	0.08	0.27	0.04	0.00	0.00	0.39	0.39	-0.27	-0.17	-0.20	A+	A-	A-	A+	0.39	0.06	1.27	1.03	0.82	1.03
163	722795	11	4505	0.72	0.11	0.03	0.14	0.72	0.00	0.00	0.36	-0.21	-0.25	-0.15	0.36	A+	A+	A+	A-	-0.22	0.06	1.39	1.04	0.69	1.04
164	722796	11	4505	0.61	0.61	0.17	0.19	0.03	0.00	0.00	0.30	0.30	-0.17	-0.13	-0.21	A+	A-	A-	A+	0.41	0.06	5.92	1.16	5.51	1.25
165	722800	11	4505	0.88	0.88	0.05	0.04	0.02	0.00	0.00	0.50	0.50	-0.29	-0.31	-0.24	A-	A-	A-	A+	-1.50	0.09	-2.84	0.85	-4.69	0.54

**Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
166	722801	11	4505	0.62	0.62	0.07	0.13	0.18	0.00	0.00	0.41	0.41	-0.22	-0.21	-0.19	A-	A-	A-	A+	0.37	0.06	2.09	1.06	1.74	1.08
167	721696	12	4512	0.46	0.35	0.10	0.08	0.46	0.00	0.00	0.23	0.03	-0.20	-0.24	0.23	A-	A-	A-	A+	1.11	0.06	8.79	1.21	8.62	1.33
168	721698	12	4512	0.55	0.05	0.55	0.24	0.15	0.00	0.00	0.35	-0.11	0.35	-0.20	-0.18	A+	A+	A+	A-	0.65	0.06	3.64	1.09	3.69	1.13
169	721704	12	4512	0.68	0.09	0.06	0.17	0.68	0.00	0.00	0.50	-0.25	-0.28	-0.25	0.50	A-	A-	A-	A+	-0.01	0.06	-4.45	0.88	-4.40	0.81
170	721705	12	4512	0.57	0.57	0.12	0.11	0.20	0.00	0.00	0.43	0.43	-0.26	-0.21	-0.15	A-	A-	A-	A-	0.59	0.06	-0.81	0.98	0.39	1.01
171	721707	12	4512	0.63	0.14	0.63	0.10	0.13	0.00	0.00	0.42	-0.25	0.42	-0.22	-0.13	A-	A-	A-	A+	0.24	0.06	0.34	1.01	0.14	1.01
172	721708	12	4512	0.56	0.56	0.15	0.23	0.06	0.00	0.00	0.37	0.37	-0.16	-0.15	-0.25	A-	A+	A+	A-	0.65	0.06	3.69	1.09	3.48	1.13
173	722790	12	4512	0.88	0.88	0.07	0.03	0.02	0.00	0.00	0.48	0.48	-0.30	-0.26	-0.25	A+	A-	A+	A+	-1.56	0.09	-3.28	0.82	-3.19	0.68
174	722792	12	4512	0.72	0.06	0.12	0.72	0.10	0.00	0.00	0.38	-0.20	-0.10	0.38	-0.28	A-	A-	A-	A+	-0.25	0.06	-0.45	0.99	-0.92	0.95
175	722797	12	4512	0.89	0.02	0.06	0.03	0.89	0.00	0.00	0.51	-0.24	-0.31	-0.30	0.51	A+	A-	A-	A-	-1.58	0.09	-4.17	0.77	-4.62	0.55
176	722798	12	4512	0.73	0.11	0.73	0.04	0.12	0.00	0.00	0.41	-0.24	0.41	-0.29	-0.16	A-	A-	A-	A-	-0.30	0.06	-0.13	1.00	1.64	1.09
177	722799	12	4512	0.75	0.14	0.06	0.75	0.05	0.00	0.00	0.35	-0.18	-0.13	0.35	-0.24	A+	A+	A+	A-	-0.44	0.07	1.52	1.05	1.63	1.10
178	723231	12	4512	0.95	0.02	0.95	0.02	0.01	0.00	0.00	0.27	-0.25	0.27	-0.07	-0.13	A+	A-	A-	A+	-2.59	0.12	-1.12	0.89	-0.04	0.98
179	718547	13	4505	0.51	0.51	0.05	0.20	0.24	0.00	0.00	0.44	0.44	-0.25	-0.35	-0.05	A+	A+	A+	A+	0.93	0.06	-1.30	0.97	-1.34	0.95
180	718548	13	4505	0.30	0.24	0.43	0.02	0.30	0.00	0.00	0.30	-0.12	-0.11	-0.21	0.30	C-	B-	A-	A+	2.04	0.06	-0.14	1.00	1.54	1.08
181	718550	13	4505	0.72	0.04	0.72	0.09	0.14	0.00	0.00	0.44	-0.28	0.44	-0.23	-0.21	B-	B-	A-	A+	-0.20	0.06	-1.87	0.94	-2.09	0.89
182	718552	13	4505	0.37	0.23	0.15	0.25	0.37	0.00	0.00	0.24	-0.09	-0.13	-0.06	0.24	A+	A-	A-	A+	1.67	0.06	6.39	1.16	9.12	1.44
183	718553	13	4505	0.65	0.14	0.10	0.65	0.11	0.00	0.00	0.37	-0.21	-0.14	0.37	-0.19	A-	A+	A+	A+	0.20	0.06	2.89	1.08	2.87	1.13
184	718556	13	4505	0.46	0.11	0.46	0.21	0.22	0.00	0.00	0.09	-0.24	0.09	-0.05	0.12	A+	A-	A+	A+	1.18	0.06	9.90	1.43	9.90	1.64
185	721821	13	4505	0.70	0.04	0.05	0.70	0.21	0.00	0.00	0.11	-0.13	-0.21	0.11	0.05	A+	A-	A+	A-	-0.10	0.06	9.90	1.34	9.90	1.69
186	721822	13	4505	0.50	0.34	0.50	0.04	0.12	0.00	0.00	0.42	-0.19	0.42	-0.27	-0.19	A-	A-	A-	A+	0.99	0.06	-1.13	0.97	0.88	1.03
187	721824	13	4505	0.60	0.14	0.06	0.19	0.60	0.00	0.00	0.43	-0.15	-0.23	-0.26	0.43	A-	A-	A-	B+	0.46	0.06	-0.04	1.00	-0.16	0.99
188	721828	13	4505	0.62	0.07	0.20	0.11	0.62	0.00	0.00	0.53	-0.29	-0.25	-0.27	0.53	A-	A-	A-	A+	0.38	0.06	-5.81	0.86	-5.91	0.78
189	721830	13	4505	0.41	0.03	0.47	0.09	0.41	0.00	0.00	0.40	-0.24	-0.21	-0.16	0.40	A-	A-	C-	A+	1.46	0.06	-0.85	0.98	0.70	1.03
190	721832	13	4505	0.76	0.10	0.76	0.08	0.06	0.00	0.00	0.55	-0.31	0.55	-0.28	-0.26	A-	A-	A-	A+	-0.45	0.07	-6.12	0.80	-5.73	0.67
191	718545	14	4499	0.52	0.30	0.12	0.52	0.06	0.00	0.00	0.38	-0.22	-0.19	0.38	-0.11	A-	A-	A-	A+	0.90	0.06	0.14	1.00	1.61	1.06
192	718546	14	4499	0.73	0.09	0.73	0.15	0.02	0.00	0.00	0.39	-0.22	0.39	-0.22	-0.19	B-	B-	A-	A+	-0.27	0.07	-0.77	0.97	-0.14	0.99
193	718549	14	4499	0.59	0.13	0.21	0.59	0.07	0.00	0.00	0.27	-0.13	-0.06	0.27	-0.24	A+	A+	A+	A-	0.50	0.06	7.29	1.19	5.72	1.25
194	718551	14	4499	0.29	0.29	0.32	0.31	0.08	0.00	0.00	0.24	0.24	0.06	-0.14	-0.24	A-	A-	A-	A-	2.10	0.06	2.44	1.07	7.08	1.44
195	718554	14	4499	0.77	0.77	0.15	0.06	0.01	0.00	0.00	0.40	0.40	-0.29	-0.17	-0.18	C-	A-	A-	A+	-0.57	0.07	-0.91	0.97	-0.41	0.97
196	718555	14	4499	0.66	0.04	0.22	0.08	0.66	0.00	0.00	0.18	-0.16	0.04	-0.23	0.18	A+	A+	A+	A-	0.14	0.06	9.90	1.31	9.05	1.50
197	721820	14	4499	0.54	0.32	0.06	0.08	0.54	0.00	0.00	0.35	-0.16	-0.24	-0.15	0.35	A-	A-	A+	A-	0.78	0.06	3.10	1.07	3.00	1.11
198	721825	14	4499	0.76	0.06	0.06	0.76	0.12	0.00	0.00	0.54	-0.26	-0.27	0.54	-0.32	A-	A+	A-	A-	-0.43	0.07	-4.11	0.86	-4.62	0.72

**Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
199	721826	14	4499	0.91	0.02	0.91	0.04	0.03	0.00	0.00	0.42	-0.22	0.42	-0.28	-0.19	A-	A-	A-	B+	-1.81	0.10	-1.62	0.89	-2.96	0.65
200	721827	14	4499	0.72	0.72	0.13	0.07	0.07	0.00	0.00	0.40	0.40	-0.21	-0.23	-0.18	A+	A+	A+	A+	-0.23	0.06	-0.23	0.99	-0.09	0.99
201	721829	14	4499	0.66	0.11	0.66	0.09	0.14	0.00	0.00	0.45	-0.25	0.45	-0.23	-0.19	A-	A-	A-	A+	0.15	0.06	-1.51	0.96	-1.30	0.94
202	721831	14	4499	0.57	0.11	0.15	0.57	0.17	0.00	0.00	0.43	-0.17	-0.26	0.43	-0.18	A-	A-	A-	A+	0.61	0.06	-2.00	0.95	-1.94	0.93
203	721683	15	4438	0.52	0.21	0.09	0.18	0.52	0.00	0.00	0.34	-0.09	-0.27	-0.14	0.34	A-	A+	A+	B-	0.91	0.06	3.91	1.09	4.49	1.16
204	721684	15	4438	0.60	0.13	0.13	0.60	0.14	0.00	0.00	0.31	-0.14	-0.28	0.31	-0.02	A+	A-	A-	A+	0.51	0.06	4.44	1.12	4.56	1.18
205	721685	15	4438	0.83	0.07	0.83	0.03	0.07	0.00	0.00	0.45	-0.23	0.45	-0.29	-0.23	A-	A-	B-	A+	-0.94	0.07	-2.03	0.92	-3.12	0.76
206	721687	15	4438	0.71	0.08	0.11	0.09	0.71	0.00	0.00	0.46	-0.24	-0.21	-0.26	0.46	B-	A-	A-	A+	-0.12	0.06	-1.91	0.94	-2.29	0.88
207	721689	15	4438	0.57	0.20	0.57	0.08	0.15	0.00	0.00	0.47	-0.24	0.47	-0.29	-0.17	A-	A+	A-	A-	0.63	0.06	-2.08	0.95	-0.99	0.96
208	721695	15	4438	0.62	0.62	0.16	0.10	0.12	0.00	0.00	0.37	0.37	-0.09	-0.31	-0.16	A-	A+	A-	A+	0.40	0.06	2.93	1.08	3.41	1.14
209	722018	15	4438	0.58	0.16	0.58	0.19	0.08	0.00	0.00	0.28	-0.23	0.28	-0.06	-0.11	A+	A-	A-	A-	0.60	0.06	8.19	1.21	8.23	1.33
210	722021	15	4438	0.61	0.04	0.61	0.23	0.11	0.00	0.00	0.21	-0.17	0.21	-0.02	-0.18	A-	A-	A+	A+	0.44	0.06	9.69	1.27	8.00	1.35
211	722022	15	4438	0.77	0.77	0.05	0.04	0.14	0.00	0.00	0.33	0.33	-0.26	-0.25	-0.10	A+	A-	A-	A+	-0.52	0.07	0.98	1.03	2.57	1.18
212	722024	15	4438	0.64	0.15	0.09	0.64	0.12	0.00	0.00	0.38	-0.22	-0.19	0.38	-0.15	A-	A-	A-	A+	0.25	0.06	1.58	1.04	0.22	1.01
213	722027	15	4438	0.77	0.08	0.06	0.08	0.77	0.00	0.00	0.50	-0.27	-0.27	-0.25	0.50	A+	A-	A-	A-	-0.50	0.07	-2.99	0.90	-2.47	0.84
214	723039	15	4438	0.58	0.58	0.16	0.08	0.18	0.00	0.00	0.27	0.27	-0.06	-0.23	-0.12	A-	A-	A-	A+	0.60	0.06	7.31	1.19	7.79	1.31
215	721686	16	4547	0.86	0.86	0.03	0.04	0.07	0.00	0.00	0.44	0.44	-0.26	-0.29	-0.19	A-	A-	B-	A+	-1.24	0.08	-2.32	0.88	-2.49	0.76
216	721690	16	4547	0.50	0.14	0.04	0.31	0.50	0.00	0.00	0.22	-0.21	-0.23	0.03	0.22	A-	A+	A-	A+	0.99	0.06	9.90	1.25	9.57	1.37
217	721691	16	4547	0.74	0.74	0.06	0.05	0.15	0.00	0.00	0.51	0.51	-0.28	-0.28	-0.27	C-	B-	B-	A+	-0.31	0.07	-3.82	0.88	-3.70	0.79
218	721692	16	4547	0.80	0.08	0.05	0.80	0.08	0.00	0.00	0.51	-0.29	-0.31	0.51	-0.22	A-	A-	A-	A-	-0.71	0.07	-5.34	0.80	-4.48	0.69
219	721693	16	4547	0.68	0.16	0.68	0.07	0.09	0.00	0.00	0.41	-0.16	0.41	-0.25	-0.24	A-	A-	A-	A-	0.04	0.06	1.55	1.05	1.99	1.10
220	721694	16	4547	0.67	0.67	0.10	0.14	0.09	0.00	0.00	0.38	0.38	-0.26	-0.20	-0.10	A-	A+	A+	A+	0.11	0.06	-1.80	0.95	-0.44	0.98
221	722016	16	4547	0.41	0.41	0.43	0.12	0.04	0.00	0.00	0.12	0.12	0.13	-0.22	-0.26	A-	A-	A+	A-	1.42	0.06	9.90	1.34	9.90	1.51
222	722017	16	4547	0.91	0.03	0.91	0.03	0.04	0.00	0.00	0.38	-0.23	0.38	-0.20	-0.22	A-	A-	A-	A+	-1.83	0.10	-1.86	0.87	-2.42	0.69
223	722019	16	4547	0.60	0.17	0.06	0.16	0.60	0.00	0.00	0.32	-0.08	-0.32	-0.12	0.32	A+	A-	A-	A+	0.47	0.06	3.20	1.08	3.27	1.14
224	722020	16	4547	0.86	0.04	0.06	0.86	0.04	0.00	0.00	0.46	-0.26	-0.28	0.46	-0.22	A-	A+	A+	A+	-1.22	0.08	-1.12	0.94	-1.99	0.80
225	722025	16	4547	0.31	0.35	0.31	0.13	0.21	0.00	0.00	0.04	0.00	0.04	-0.11	0.04	A-	A-	A+	A+	1.95	0.06	9.90	1.32	9.90	1.70
226	722026	16	4547	0.73	0.73	0.13	0.08	0.06	0.00	0.00	0.44	0.44	-0.20	-0.22	-0.28	A+	A+	A+	A+	-0.27	0.07	-0.70	0.98	-0.96	0.94
227	721886	17	4497	0.73	0.11	0.12	0.73	0.04	0.00	0.00	0.39	-0.25	-0.21	0.39	-0.10	A+	A-	A-	A+	-0.21	0.07	1.23	1.04	0.26	1.01
228	721888	17	4497	0.37	0.37	0.15	0.18	0.29	0.01	0.00	0.16	0.16	-0.20	-0.02	0.02	A+	A+	A-	A-	1.68	0.06	9.90	1.30	9.90	1.62
229	721889	17	4497	0.48	0.09	0.48	0.12	0.30	0.00	0.00	0.16	-0.16	0.16	-0.13	0.03	A-	A+	A+	A+	1.10	0.06	9.90	1.27	9.90	1.39
230	721892	17	4497	0.29	0.38	0.19	0.14	0.29	0.00	0.00	0.04	0.12	-0.07	-0.14	0.04	A-	A-	A-	A-	2.13	0.06	9.90	1.32	9.90	1.70
231	721896	17	4497	0.59	0.59	0.10	0.16	0.15	0.00	0.00	0.30	0.30	-0.20	-0.13	-0.11	A-	A-	A-	A+	0.57	0.06	5.68	1.15	4.88	1.20

**Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
232	722840	17	4497	0.63	0.15	0.08	0.13	0.63	0.00	0.00	0.41	-0.21	-0.16	-0.23	0.41	A-	A-	A-	A+	0.36	0.06	-0.32	0.99	-0.75	0.97
233	723409	17	4497	0.83	0.08	0.83	0.05	0.04	0.00	0.00	0.46	-0.27	0.46	-0.25	-0.23	B-	A+	A-	A+	-0.98	0.08	-2.55	0.89	-3.23	0.73
234	723416	17	4497	0.67	0.10	0.14	0.08	0.67	0.00	0.00	0.43	-0.25	-0.10	-0.31	0.43	A+	A-	A-	A-	0.11	0.06	0.58	1.02	1.85	1.09
235	723417	17	4497	0.60	0.21	0.60	0.06	0.13	0.01	0.00	0.50	-0.22	0.50	-0.26	-0.26	A+	A-	A+	A+	0.52	0.06	-3.89	0.90	-3.73	0.86
236	723418	17	4497	0.81	0.81	0.03	0.04	0.12	0.00	0.00	0.42	0.42	-0.29	-0.30	-0.17	B-	A-	A-	A+	-0.76	0.07	-1.22	0.95	-0.99	0.92
237	723420	17	4497	0.63	0.14	0.09	0.14	0.63	0.00	0.00	0.55	-0.26	-0.32	-0.24	0.55	B-	A-	A-	A+	0.36	0.06	-7.83	0.80	-7.43	0.72
238	723581	17	4497	0.87	0.87	0.05	0.04	0.04	0.00	0.00	0.46	0.46	-0.20	-0.30	-0.25	C-	A+	A-	B+	-1.33	0.08	-3.71	0.81	-3.35	0.67
239	721884	18	4520	0.57	0.04	0.23	0.57	0.16	0.00	0.00	0.31	-0.23	-0.12	0.31	-0.15	A-	A-	A-	A+	0.60	0.06	3.56	1.09	2.31	1.08
240	721885	18	4520	0.73	0.07	0.07	0.12	0.73	0.00	0.00	0.40	-0.16	-0.20	-0.24	0.40	A+	A-	A+	A+	-0.31	0.06	1.28	1.04	1.12	1.06
241	721890	18	4520	0.71	0.07	0.12	0.09	0.71	0.00	0.00	0.25	-0.12	-0.08	-0.18	0.25	A+	A-	A-	A-	-0.19	0.06	3.93	1.12	5.88	1.34
242	721893	18	4520	0.41	0.32	0.05	0.41	0.21	0.00	0.00	0.14	0.08	-0.25	0.14	-0.10	A-	A+	A+	A+	1.42	0.06	9.90	1.30	9.90	1.51
243	721895	18	4520	0.55	0.11	0.55	0.22	0.12	0.00	0.00	0.22	-0.16	0.22	-0.02	-0.15	A+	A+	A+	A+	0.71	0.06	9.19	1.23	9.90	1.38
244	722839	18	4520	0.64	0.10	0.64	0.17	0.09	0.00	0.00	0.23	-0.22	0.23	-0.08	-0.05	A-	A-	A-	A+	0.21	0.06	8.13	1.23	7.30	1.33
245	723408	18	4520	0.22	0.61	0.12	0.22	0.05	0.00	0.00	0.01	0.25	-0.26	0.01	-0.18	B-	A-	B-	A-	2.48	0.07	8.28	1.32	9.90	2.24
246	723410	18	4520	0.57	0.57	0.11	0.27	0.05	0.00	0.00	0.34	0.34	-0.17	-0.11	-0.26	A-	A-	A-	A+	0.61	0.06	4.16	1.10	4.81	1.17
247	723411	18	4520	0.51	0.12	0.27	0.09	0.51	0.00	0.00	0.43	-0.21	-0.17	-0.24	0.43	A-	A+	A+	A-	0.87	0.06	-3.17	0.93	-2.24	0.93
248	723412	18	4520	0.58	0.11	0.23	0.58	0.07	0.00	0.00	0.28	-0.08	-0.22	0.28	-0.07	A-	A+	A-	A+	0.55	0.06	5.17	1.13	4.89	1.18
249	723413	18	4520	0.69	0.21	0.69	0.08	0.03	0.00	0.00	0.39	-0.27	0.39	-0.12	-0.23	B-	B-	A-	A+	-0.05	0.06	-0.24	0.99	-0.54	0.97
250	723414	18	4520	0.65	0.65	0.15	0.16	0.05	0.00	0.00	0.53	0.53	-0.18	-0.35	-0.28	A-	A+	A+	A+	0.18	0.06	-6.48	0.84	-6.32	0.76
251	719014	19	4473	0.25	0.39	0.21	0.25	0.15	0.00	0.00	0.20	-0.07	-0.09	0.20	-0.03	A-	A+	A-	A+	2.28	0.07	1.89	1.06	6.50	1.46
252	719018	19	4473	0.53	0.09	0.07	0.53	0.31	0.00	0.00	0.23	-0.21	-0.22	0.23	0.01	A+	A+	A+	A-	0.78	0.06	8.16	1.20	7.94	1.30
253	719022	19	4473	0.62	0.22	0.62	0.08	0.07	0.00	0.00	0.48	-0.18	0.48	-0.36	-0.21	A+	A-	A-	A+	0.33	0.06	-2.14	0.95	-1.79	0.93
254	719024	19	4473	0.43	0.24	0.12	0.43	0.20	0.00	0.00	0.42	-0.19	-0.33	0.42	-0.04	A-	A-	A-	A+	1.29	0.06	-2.03	0.95	0.53	1.02
255	719025	19	4473	0.57	0.57	0.13	0.15	0.14	0.00	0.00	0.45	0.45	-0.20	-0.21	-0.22	A+	A-	A+	A-	0.59	0.06	-2.91	0.93	-3.05	0.89
256	721773	19	4473	0.57	0.57	0.24	0.05	0.14	0.00	0.00	0.32	0.32	-0.15	-0.23	-0.12	A-	A-	A-	A+	0.59	0.06	3.93	1.10	4.01	1.15
257	722099	19	4473	0.56	0.06	0.13	0.56	0.25	0.00	0.00	0.39	-0.19	-0.22	0.39	-0.17	A-	A+	A-	A-	0.64	0.06	2.02	1.05	2.17	1.08
258	722102	19	4473	0.27	0.13	0.47	0.27	0.12	0.00	0.00	0.10	-0.16	0.26	0.10	-0.37	A+	A+	A+	A+	2.17	0.06	9.90	1.33	9.90	1.70
259	722103	19	4473	0.64	0.08	0.64	0.15	0.12	0.00	0.00	0.30	-0.20	0.30	-0.11	-0.13	A-	A+	A+	A+	0.24	0.06	4.09	1.11	5.27	1.24
260	722104	19	4473	0.64	0.64	0.15	0.08	0.14	0.00	0.00	0.34	0.34	-0.25	-0.10	-0.12	A-	A-	A-	A+	0.25	0.06	3.82	1.10	2.46	1.10
261	722106	19	4473	0.53	0.14	0.22	0.53	0.11	0.00	0.00	0.30	-0.18	-0.08	0.30	-0.17	A-	A-	A-	A-	0.79	0.06	4.46	1.11	4.42	1.16
262	722107	19	4473	0.18	0.51	0.14	0.16	0.18	0.00	0.00	0.11	0.19	-0.15	-0.22	0.11	A-	A-	A+	A-	2.77	0.07	2.65	1.11	7.64	1.77
263	719015	20	4504	0.63	0.03	0.63	0.16	0.18	0.00	0.00	0.35	-0.25	0.35	-0.21	-0.11	A-	A-	A-	A+	0.28	0.06	3.66	1.10	4.72	1.23
264	719016	20	4504	0.49	0.49	0.12	0.08	0.31	0.00	0.00	0.36	0.36	-0.11	-0.24	-0.17	A+	A+	A+	A+	1.01	0.06	2.18	1.05	3.08	1.12

**Table J–7 (continued). Literature Multiple-Choice Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
265	719017	20	4504	0.50	0.30	0.11	0.08	0.50	0.00	0.00	0.40	-0.11	-0.23	-0.27	0.40	A+	A+	A+	A+	0.96	0.06	2.66	1.07	3.80	1.15
266	719019	20	4504	0.66	0.10	0.66	0.20	0.05	0.00	0.00	0.41	-0.22	0.41	-0.19	-0.24	B-	C-	B-	B+	0.13	0.06	0.90	1.03	1.72	1.09
267	719020	20	4504	0.53	0.53	0.11	0.04	0.32	0.00	0.00	0.35	0.35	-0.11	-0.29	-0.18	A-	A-	A-	A+	0.79	0.06	4.45	1.11	3.24	1.13
268	719021	20	4504	0.78	0.10	0.09	0.78	0.04	0.00	0.00	0.49	-0.28	-0.26	0.49	-0.24	A-	A-	A-	A+	-0.61	0.07	-5.17	0.82	-3.92	0.73
269	722097	20	4504	0.78	0.78	0.14	0.05	0.03	0.00	0.00	0.30	0.30	-0.12	-0.20	-0.23	A-	A+	A+	A-	-0.65	0.07	2.44	1.09	3.37	1.28
270	722100	20	4504	0.76	0.09	0.04	0.11	0.76	0.00	0.00	0.53	-0.33	-0.25	-0.26	0.53	A-	A-	A-	A+	-0.51	0.07	-4.85	0.84	-3.92	0.75
271	722101	20	4504	0.71	0.07	0.12	0.10	0.71	0.00	0.00	0.52	-0.26	-0.22	-0.32	0.52	A-	A+	A+	A+	-0.18	0.06	-3.65	0.89	-3.21	0.82
272	722105	20	4504	0.60	0.10	0.16	0.14	0.60	0.00	0.00	0.35	-0.16	-0.12	-0.22	0.35	A+	A+	A+	A+	0.45	0.06	4.80	1.13	4.44	1.20
273	722109	20	4504	0.72	0.03	0.72	0.23	0.03	0.00	0.00	0.32	-0.26	0.32	-0.16	-0.20	A-	A-	A-	A+	-0.25	0.06	3.54	1.12	4.88	1.33
274	722904	20	4504	0.85	0.03	0.85	0.04	0.08	0.00	0.00	0.39	-0.23	0.39	-0.24	-0.20	A-	A-	A-	A+	-1.22	0.08	-1.65	0.92	-1.46	0.85
275	634495	21	4584	0.89	0.03	0.89	0.04	0.04	0.00	0.00	0.42	-0.22	0.42	-0.22	-0.24	B-	B-	C-	A+	-1.64	0.09	-2.46	0.86	-1.78	0.79
276	634496	21	4584	0.60	0.15	0.09	0.60	0.15	0.00	0.00	0.41	-0.28	-0.28	0.41	-0.04	A-	A+	A-	A-	0.42	0.06	1.44	1.04	1.75	1.07
277	634497	21	4584	0.63	0.04	0.63	0.24	0.08	0.00	0.00	0.37	-0.22	0.37	-0.22	-0.13	A-	A-	A-	A+	0.28	0.06	2.00	1.05	1.00	1.04
278	634500	21	4584	0.62	0.06	0.21	0.11	0.62	0.00	0.00	0.42	-0.21	-0.15	-0.28	0.42	A-	A-	A-	A+	0.32	0.06	1.42	1.04	1.39	1.06
279	634504	21	4584	0.81	0.81	0.04	0.12	0.04	0.00	0.00	0.50	0.50	-0.27	-0.27	-0.29	A-	A-	A-	A+	-0.85	0.07	-3.06	0.88	-4.07	0.70
280	634505	21	4584	0.59	0.25	0.09	0.59	0.07	0.00	0.00	0.43	-0.18	-0.25	0.43	-0.24	A-	A-	A+	A+	0.50	0.06	-0.50	0.99	-1.32	0.95
281	635610	21	4584	0.70	0.12	0.04	0.70	0.14	0.00	0.00	0.44	-0.29	-0.26	0.44	-0.16	A-	A+	A-	A-	-0.12	0.06	-1.10	0.97	-2.43	0.88
282	635611	21	4584	0.57	0.12	0.19	0.12	0.57	0.00	0.00	0.33	-0.17	-0.14	-0.15	0.33	A+	A+	A+	A-	0.57	0.06	4.17	1.11	4.21	1.17
283	635614	21	4584	0.78	0.78	0.05	0.10	0.07	0.00	0.00	0.43	0.43	-0.19	-0.29	-0.19	A-	A-	A-	A+	-0.68	0.07	-2.23	0.92	-2.82	0.81
284	635615	21	4584	0.87	0.07	0.87	0.05	0.02	0.00	0.00	0.39	-0.23	0.39	-0.25	-0.16	A-	A+	A-	A-	-1.37	0.08	-1.49	0.92	0.63	1.06
285	635619	21	4584	0.76	0.10	0.76	0.05	0.09	0.00	0.00	0.44	-0.20	0.44	-0.23	-0.26	A+	A+	A-	A-	-0.50	0.07	-2.50	0.92	-2.42	0.85
286	637034	21	4584	0.56	0.06	0.24	0.56	0.14	0.00	0.00	0.34	-0.08	-0.21	0.34	-0.17	A-	A-	A-	A-	0.64	0.06	4.38	1.11	4.75	1.19
287	634494	22	4484	0.51	0.19	0.23	0.08	0.51	0.00	0.00	0.40	-0.17	-0.18	-0.20	0.40	A-	A-	A-	A-	1.02	0.06	1.98	1.05	2.28	1.08
288	634498	22	4484	0.56	0.56	0.29	0.05	0.10	0.00	0.00	0.32	0.32	-0.03	-0.30	-0.26	A+	A+	A+	A+	0.73	0.06	5.27	1.13	4.87	1.19
289	634499	22	4484	0.75	0.11	0.08	0.75	0.05	0.00	0.00	0.42	-0.18	-0.29	0.42	-0.19	B-	A+	A-	A-	-0.34	0.07	-0.72	0.98	1.10	1.07
290	634502	22	4484	0.52	0.13	0.52	0.18	0.17	0.00	0.00	0.34	-0.10	0.34	-0.19	-0.15	A-	A-	A-	B+	0.96	0.06	4.02	1.10	4.88	1.18
291	634503	22	4484	0.54	0.11	0.11	0.23	0.54	0.00	0.00	0.42	-0.19	-0.17	-0.22	0.42	A-	A-	A-	A-	0.83	0.06	1.53	1.04	2.70	1.10
292	635607	22	4484	0.72	0.12	0.11	0.05	0.72	0.00	0.00	0.44	-0.29	-0.17	-0.22	0.44	A-	A-	A-	A+	-0.16	0.06	-1.36	0.96	-0.40	0.98
293	635608	22	4484	0.71	0.11	0.12	0.71	0.06	0.00	0.00	0.45	-0.21	-0.27	0.45	-0.19	A+	A+	A+	A+	-0.06	0.06	-1.55	0.95	-2.17	0.89
294	635609	22	4484	0.64	0.19	0.06	0.64	0.11	0.00	0.00	0.43	-0.18	-0.26	0.43	-0.22	A+	A-	A-	A-	0.31	0.06	0.44	1.01	0.45	1.02
295	635612	22	4484	0.91	0.91	0.04	0.02	0.03	0.00	0.00	0.43	0.43	-0.25	-0.22	-0.25	A-	A-	A-	A+	-1.76	0.09	-3.34	0.80	-3.18	0.62
296	635616	22	4484	0.72	0.06	0.06	0.72	0.16	0.00	0.00	0.26	-0.23	-0.14	0.26	-0.08	A+	A+	A+	A+	-0.15	0.06	5.64	1.19	5.37	1.34
297	635617	22	4484	0.48	0.24	0.48	0.11	0.17	0.00	0.00	0.26	-0.07	0.26	-0.17	-0.12	A+	A+	A-	A-	1.14	0.06	8.69	1.21	9.36	1.38

Table J-7 (continued). Literature Multiple-Choice Item Statistics: Spring

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
298	636181	22	4484	0.94	0.02	0.94	0.02	0.01	0.00	0.00	0.37	-0.21	0.37	-0.23	-0.16	A-	A-	A-	A+	-2.35	0.12	-0.55	0.95	-1.09	0.81
299	635903	23	4493	0.37	0.26	0.12	0.24	0.37	0.00	0.00	0.12	-0.06	-0.12	0.02	0.12	A-	A-	A-	A+	1.65	0.06	9.90	1.31	9.90	1.68
300	635904	23	4493	0.76	0.76	0.04	0.02	0.17	0.00	0.00	0.37	0.37	-0.24	-0.25	-0.18	A+	A-	A-	A+	-0.42	0.07	0.90	1.03	3.06	1.21
301	635908	23	4493	0.83	0.05	0.05	0.06	0.83	0.00	0.00	0.52	-0.30	-0.29	-0.24	0.52	A-	A+	A-	A-	-1.00	0.08	-2.71	0.88	-2.85	0.76
302	635909	23	4493	0.36	0.54	0.36	0.04	0.05	0.00	0.00	-0.05	0.27	-0.05	-0.28	-0.24	A-	A+	A+	A-	1.69	0.06	9.90	1.53	9.90	1.91
303	635912	23	4493	0.81	0.07	0.07	0.81	0.05	0.00	0.00	0.47	-0.19	-0.29	0.47	-0.28	A-	B-	B-	B+	-0.80	0.07	-1.88	0.92	-3.21	0.76
304	636536	23	4493	0.47	0.07	0.08	0.38	0.47	0.00	0.00	0.17	-0.25	-0.28	0.11	0.17	A-	A-	A-	A+	1.13	0.06	9.90	1.29	9.90	1.44
305	636537	23	4493	0.47	0.19	0.24	0.47	0.10	0.00	0.00	0.28	-0.01	-0.12	0.28	-0.27	A+	A-	B-	A-	1.14	0.06	7.65	1.18	8.05	1.31
306	636538	23	4493	0.72	0.12	0.72	0.09	0.06	0.00	0.00	0.44	-0.15	0.44	-0.29	-0.26	A-	A-	A-	A+	-0.21	0.07	-0.76	0.97	-0.97	0.94
307	636539	23	4493	0.63	0.09	0.63	0.07	0.21	0.00	0.00	0.45	-0.23	0.45	-0.27	-0.19	A-	A-	A+	A+	0.35	0.06	-1.93	0.95	-2.33	0.91
308	636543	23	4493	0.61	0.61	0.08	0.17	0.14	0.00	0.00	0.45	0.45	-0.33	-0.15	-0.20	B-	A-	A-	A+	0.45	0.06	-2.44	0.94	-2.21	0.91
309	637157	23	4493	0.75	0.11	0.75	0.08	0.06	0.00	0.00	0.41	-0.20	0.41	-0.19	-0.26	A-	B-	B-	A+	-0.37	0.07	-0.06	1.00	-0.57	0.96
310	637176	23	4493	0.59	0.59	0.11	0.12	0.18	0.00	0.00	0.23	0.23	-0.09	-0.14	-0.10	A-	A+	A-	A-	0.54	0.06	8.83	1.24	7.72	1.33
311	635901	24	4560	0.83	0.09	0.04	0.83	0.03	0.00	0.00	0.34	-0.15	-0.25	0.34	-0.18	A+	A-	A-	A+	-0.97	0.07	-0.03	1.00	1.05	1.09
312	635902	24	4560	0.53	0.11	0.22	0.15	0.53	0.00	0.00	0.41	-0.21	-0.21	-0.15	0.41	A-	A-	A-	A+	0.85	0.06	-0.07	1.00	0.65	1.02
313	635905	24	4560	0.45	0.45	0.11	0.37	0.07	0.00	0.00	0.19	0.19	0.02	-0.14	-0.14	A-	A-	A-	A-	1.24	0.06	9.90	1.28	9.90	1.42
314	635906	24	4560	0.75	0.05	0.75	0.09	0.11	0.00	0.00	0.44	-0.19	0.44	-0.24	-0.25	A-	C-	B-	A+	-0.38	0.07	-2.41	0.92	-3.22	0.81
315	635910	24	4560	0.43	0.30	0.43	0.11	0.15	0.00	0.00	0.38	-0.15	0.38	-0.16	-0.18	A+	A-	A+	A+	1.32	0.06	-0.58	0.99	1.58	1.06
316	635911	24	4560	0.93	0.02	0.93	0.02	0.03	0.00	0.00	0.37	-0.24	0.37	-0.20	-0.19	A-	A-	B-	A+	-2.10	0.11	-1.92	0.86	-1.52	0.77
317	636532	24	4560	0.66	0.12	0.14	0.08	0.66	0.00	0.00	0.50	-0.28	-0.18	-0.29	0.50	A-	A-	A-	A-	0.17	0.06	-2.56	0.93	-2.63	0.88
318	636533	24	4560	0.58	0.26	0.09	0.58	0.06	0.00	0.00	0.45	-0.19	-0.29	0.45	-0.22	A-	A-	A-	A+	0.57	0.06	-1.04	0.97	-0.37	0.99
319	636534	24	4560	0.49	0.09	0.49	0.27	0.15	0.00	0.00	0.28	-0.23	0.28	-0.09	-0.07	A-	A-	A+	A+	1.05	0.06	6.59	1.16	6.32	1.24
320	636535	24	4560	0.73	0.06	0.12	0.09	0.73	0.00	0.00	0.51	-0.24	-0.26	-0.28	0.51	A-	B-	A-	A+	-0.30	0.07	-3.79	0.88	-4.37	0.76
321	636541	24	4560	0.88	0.02	0.06	0.88	0.04	0.00	0.00	0.36	-0.22	-0.18	0.36	-0.21	A-	A+	A-	A+	-1.44	0.08	-1.26	0.93	-0.10	0.98
322	636544	24	4560	0.58	0.28	0.58	0.11	0.02	0.00	0.00	0.30	-0.08	0.30	-0.26	-0.19	A-	A-	A-	A+	0.55	0.06	5.43	1.14	5.46	1.22

**Table J-8. Algebra I Multiple-Choice Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	633061	0	274	0.35	0.13	0.20	0.35	0.31	0.00	0.00	0.19	0.04	-0.05	0.19	-0.19	0.94	0.14	1.75	1.10	2.32	1.18
2	651336	0	274	0.64	0.64	0.11	0.12	0.13	0.00	0.00	0.19	0.19	-0.12	-0.18	0.01	-0.23	0.13	0.95	1.04	0.61	1.04
3	661904	0	274	0.76	0.09	0.04	0.11	0.76	0.00	0.00	0.33	-0.14	-0.11	-0.26	0.33	-0.38	0.13	-4.32	0.81	-3.01	0.78
4	661910	0	274	0.73	0.73	0.04	0.10	0.13	0.00	0.00	0.43	0.43	-0.22	-0.28	-0.19	-1.10	0.15	-0.53	0.96	-1.15	0.86
5	663260	0	274	0.23	0.22	0.33	0.22	0.23	0.00	0.00	0.38	-0.13	-0.09	-0.16	0.38	1.23	0.14	-2.72	0.83	-2.09	0.82
6	663710	0	274	0.30	0.22	0.30	0.30	0.18	0.01	0.00	0.33	-0.03	0.33	-0.17	-0.16	1.04	0.14	-1.16	0.93	-0.43	0.96
7	664102	0	274	0.71	0.10	0.12	0.07	0.71	0.00	0.00	0.41	-0.14	-0.22	-0.27	0.41	-0.68	0.14	-2.81	0.85	-2.33	0.79
8	667537	0	274	0.65	0.12	0.11	0.12	0.65	0.00	0.00	0.34	-0.07	-0.21	-0.22	0.34	-0.48	0.13	-0.72	0.96	-0.68	0.94
9	417143	0	274	0.64	0.07	0.24	0.64	0.05	0.00	0.00	0.24	-0.08	-0.20	0.24	-0.05	-0.21	0.13	0.04	1.00	-0.13	0.99
10	417691	0	274	0.59	0.13	0.22	0.59	0.06	0.00	0.00	0.40	-0.18	-0.21	0.40	-0.20	-0.39	0.13	-0.72	0.97	-1.16	0.91
11	417697	0	274	0.34	0.32	0.34	0.29	0.05	0.00	0.00	0.17	-0.01	0.17	-0.12	-0.09	0.94	0.14	1.71	1.10	1.52	1.12
12	418899	0	274	0.51	0.17	0.23	0.51	0.08	0.00	0.00	0.32	-0.13	-0.17	0.32	-0.14	-0.18	0.13	1.34	1.06	0.58	1.04
13	419190	0	274	0.41	0.31	0.41	0.11	0.16	0.00	0.00	0.22	0.02	0.22	-0.19	-0.16	0.50	0.13	1.35	1.06	1.18	1.07
14	419198	0	274	0.45	0.18	0.15	0.23	0.45	0.00	0.00	0.37	-0.19	-0.13	-0.16	0.37	0.31	0.13	-1.40	0.94	-1.19	0.93
15	471080	0	274	0.39	0.17	0.25	0.39	0.20	0.00	0.00	0.08	-0.13	0.01	0.08	0.02	0.87	0.13	3.91	1.22	3.91	1.30
16	495965	0	274	0.43	0.12	0.43	0.14	0.31	0.00	0.00	0.34	-0.16	0.34	-0.29	-0.03	0.47	0.13	-0.80	0.97	-0.35	0.98
17	496387	0	274	0.73	0.12	0.73	0.09	0.06	0.00	0.00	0.35	-0.19	0.35	-0.16	-0.18	-0.65	0.14	-2.30	0.88	-1.98	0.82
18	496410	0	274	0.45	0.45	0.29	0.16	0.10	0.00	0.00	0.32	0.32	-0.12	-0.09	-0.24	-0.02	0.13	0.94	1.04	0.66	1.04
19	496413	0	274	0.27	0.14	0.27	0.37	0.23	0.00	0.00	0.28	-0.02	0.28	-0.11	-0.16	1.17	0.14	-0.88	0.94	-0.27	0.97
20	505125	0	274	0.41	0.18	0.20	0.41	0.22	0.00	0.00	0.34	-0.10	-0.21	0.34	-0.10	0.37	0.13	-0.79	0.97	-0.40	0.98
21	505130	0	274	0.55	0.13	0.26	0.06	0.55	0.00	0.00	0.31	-0.21	-0.09	-0.17	0.31	0.10	0.13	-0.42	0.98	-0.20	0.99
22	505134	0	274	0.67	0.10	0.67	0.20	0.02	0.01	0.00	0.25	-0.08	0.25	-0.18	-0.08	-0.56	0.14	0.57	1.03	0.04	1.00
23	569279	0	274	0.50	0.50	0.23	0.15	0.12	0.00	0.00	0.27	0.27	-0.09	-0.08	-0.21	0.20	0.13	0.79	1.03	0.58	1.03
24	569309	0	274	0.31	0.31	0.35	0.20	0.13	0.00	0.00	0.33	0.33	-0.12	-0.09	-0.18	1.23	0.14	0.37	1.02	0.36	1.03
25	569900	0	274	0.30	0.30	0.32	0.14	0.23	0.00	0.00	0.27	0.27	-0.07	-0.10	-0.13	0.65	0.13	-1.01	0.95	-0.21	0.99
26	574494	0	274	0.42	0.42	0.42	0.08	0.08	0.00	0.00	0.22	0.22	-0.02	-0.22	-0.14	0.66	0.13	1.70	1.08	1.63	1.11
27	574498	0	274	0.50	0.23	0.50	0.13	0.14	0.00	0.00	0.18	-0.18	0.18	-0.08	0.03	-0.10	0.13	3.17	1.14	2.89	1.20
28	577987	0	274	0.50	0.50	0.22	0.19	0.08	0.00	0.00	0.42	0.42	-0.21	-0.14	-0.24	-0.08	0.13	-1.33	0.95	-1.38	0.91
29	579831	0	274	0.65	0.11	0.16	0.65	0.08	0.01	0.00	0.36	-0.16	-0.17	0.36	-0.19	-0.14	0.13	-2.33	0.91	-2.15	0.86
30	580170	0	274	0.79	0.79	0.08	0.14		0.00	0.00	0.34	0.34	-0.27	-0.20		-0.85	0.14	-3.52	0.79	-3.03	0.70
31	584016	0	274	0.36	0.20	0.34	0.10	0.36	0.00	0.00	0.28	-0.12	-0.13	-0.09	0.28	1.30	0.14	2.75	1.20	2.19	1.22
32	584134	0	274	0.54	0.21	0.54	0.11	0.14	0.00	0.00	0.40	-0.29	0.40	-0.16	-0.09	0.23	0.13	-1.97	0.92	-1.52	0.91
33	584278	0	274	0.28	0.23	0.28	0.32	0.16	0.00	0.00	0.19	-0.06	0.19	0.05	-0.22	1.54	0.15	2.21	1.18	2.40	1.29



**Table J-8 (continued). Algebra I Multiple-Choice Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	585778	0	274	0.46	0.04	0.11	0.39	0.46	0.00	0.00	0.17	-0.15	-0.13	-0.02	0.17	0.33	0.13	2.60	1.11	2.14	1.13
35	593287	0	274	0.32	0.19	0.28	0.20	0.32	0.01	0.00	0.29	-0.03	-0.08	-0.22	0.29	0.81	0.13	-0.74	0.96	0.07	1.00
36	593803	0	274	0.70	0.11	0.06	0.70	0.12	0.00	0.00	0.34	-0.19	-0.22	0.34	-0.13	-0.16	0.13	-3.04	0.88	-2.54	0.83

**Table J–9. Biology Multiple-Choice Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	625825	0	210	0.62	0.16	0.11	0.62	0.11	0.00	0.00	0.42	-0.15	-0.27	0.42	-0.21	-0.52	0.15	-0.85	0.95	-1.21	0.86
2	634709	0	210	0.56	0.08	0.31	0.04	0.56	0.00	0.00	0.24	-0.16	-0.13	-0.08	0.24	-0.13	0.15	1.95	1.09	0.81	1.08
3	635330	0	210	0.46	0.14	0.20	0.21	0.46	0.00	0.00	0.41	-0.17	-0.32	-0.05	0.41	0.52	0.15	-0.67	0.96	-0.16	0.99
4	653084	0	210	0.55	0.10	0.17	0.55	0.19	0.00	0.00	0.45	-0.24	-0.26	0.45	-0.14	0.20	0.15	-2.08	0.91	-0.64	0.95
5	654024	0	210	0.34	0.15	0.34	0.40	0.10	0.00	0.00	0.30	-0.07	0.30	-0.12	-0.21	1.25	0.17	1.79	1.16	2.17	1.25
6	654026	0	210	0.67	0.04	0.23	0.67	0.06	0.00	0.00	0.34	-0.19	-0.21	0.34	-0.14	-0.70	0.16	0.10	1.00	-0.63	0.91
7	654027	0	210	0.50	0.15	0.50	0.20	0.16	0.00	0.00	0.33	-0.06	0.33	-0.17	-0.20	0.22	0.15	0.38	1.02	0.21	1.02
8	654028	0	210	0.42	0.10	0.22	0.42	0.25	0.00	0.00	0.44	-0.14	-0.18	0.44	-0.24	0.14	0.15	-1.32	0.94	-1.39	0.88
9	654811	0	210	0.53	0.53	0.08	0.34	0.05	0.00	0.00	0.30	0.30	-0.25	-0.11	-0.12	0.04	0.15	0.82	1.04	0.29	1.02
10	655211	0	210	0.35	0.35	0.26	0.21	0.18	0.00	0.00	0.43	0.43	-0.16	-0.12	-0.22	0.80	0.16	-1.23	0.92	-1.18	0.90
11	655213	0	210	0.51	0.51	0.22	0.16	0.10	0.00	0.00	0.22	0.22	0.00	-0.17	-0.17	0.16	0.15	2.14	1.10	3.71	1.35
12	656237	0	210	0.42	0.18	0.09	0.42	0.31	0.00	0.00	0.07	0.16	-0.21	0.07	-0.07	0.26	0.15	4.84	1.24	5.03	1.47
13	658183	0	210	0.51	0.23	0.14	0.51	0.12	0.00	0.00	0.26	-0.04	-0.19	0.26	-0.15	0.23	0.15	1.49	1.07	2.19	1.19
14	659660	0	210	0.53	0.53	0.15	0.16	0.17	0.00	0.00	0.38	0.38	-0.18	-0.20	-0.14	0.29	0.15	-0.42	0.98	-1.07	0.91
15	659666	0	210	0.53	0.53	0.13	0.21	0.12	0.01	0.00	0.27	0.27	-0.17	-0.03	-0.18	0.09	0.15	1.20	1.05	1.89	1.17
16	659675	0	210	0.47	0.17	0.29	0.47	0.08	0.00	0.00	0.23	-0.17	-0.02	0.23	-0.15	0.34	0.15	2.18	1.11	1.06	1.09
17	659682	0	210	0.39	0.25	0.39	0.22	0.14	0.00	0.00	0.34	-0.05	0.34	-0.30	-0.06	0.43	0.15	-0.02	1.00	0.04	1.00
18	659689	0	210	0.41	0.26	0.41	0.22	0.11	0.00	0.00	0.36	-0.01	0.36	-0.33	-0.11	0.61	0.15	-0.02	1.00	0.18	1.01
19	659737	0	210	0.60	0.07	0.28	0.05	0.60	0.00	0.00	0.45	-0.17	-0.34	-0.12	0.45	-0.25	0.15	-2.29	0.90	-1.88	0.81
20	660371	0	210	0.43	0.10	0.14	0.43	0.32	0.00	0.00	0.29	-0.21	-0.23	0.29	0.00	0.47	0.15	1.08	1.06	1.25	1.10
21	662913	0	210	0.69	0.17	0.69	0.12	0.03	0.00	0.00	0.36	-0.27	0.36	-0.12	-0.15	-0.88	0.16	0.09	1.00	0.99	1.15
22	662916	0	210	0.53	0.53	0.20	0.16	0.11	0.00	0.00	0.45	0.45	-0.26	-0.18	-0.15	-0.10	0.15	-1.78	0.92	-1.58	0.85
23	665927	0	210	0.41	0.24	0.18	0.16	0.41	0.00	0.00	0.48	-0.11	-0.25	-0.26	0.48	0.28	0.15	-2.56	0.88	-2.19	0.83
24	665929	0	210	0.64	0.64	0.09	0.22	0.05	0.00	0.00	0.34	0.34	-0.16	-0.19	-0.18	-0.61	0.16	0.20	1.01	0.07	1.00
25	665977	0	210	0.49	0.10	0.49	0.28	0.13	0.00	0.00	0.50	-0.16	0.50	-0.24	-0.26	0.28	0.15	-2.84	0.87	-2.45	0.81
26	314619	0	210	0.42	0.42	0.20	0.23	0.14	0.00	0.00	0.33	0.33	-0.14	-0.22	-0.04	0.56	0.15	0.49	1.03	-0.10	0.99
27	314933	0	210	0.35	0.22	0.23	0.20	0.35	0.00	0.00	0.35	-0.11	-0.11	-0.19	0.35	0.65	0.15	-0.47	0.97	-0.29	0.97
28	314936	0	210	0.48	0.11	0.24	0.16	0.48	0.01	0.00	0.35	-0.15	-0.07	-0.25	0.35	0.75	0.15	1.86	1.12	1.15	1.10
29	339649	0	210	0.29	0.15	0.18	0.39	0.29	0.00	0.00	0.49	-0.18	-0.20	-0.17	0.49	1.05	0.16	-2.60	0.81	-2.43	0.78
30	341686	0	210	0.41	0.15	0.23	0.41	0.21	0.00	0.00	0.24	-0.12	-0.11	0.24	-0.08	0.76	0.16	2.12	1.14	2.51	1.23
31	409046	0	210	0.61	0.24	0.61	0.07	0.08	0.00	0.00	0.36	-0.30	0.36	-0.15	-0.04	-0.07	0.15	-1.16	0.95	-1.08	0.90
32	410011	0	210	0.53	0.08	0.33	0.06	0.53	0.00	0.00	0.40	-0.11	-0.28	-0.17	0.40	0.31	0.15	-0.71	0.97	-0.81	0.93
33	413689	0	210	0.60	0.60	0.20	0.17	0.04	0.00	0.00	0.31	0.31	-0.14	-0.15	-0.21	-0.09	0.15	-0.21	0.99	1.37	1.13

**Table J–9 (continued). Biology Multiple-Choice Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	413690	0	210	0.47	0.26	0.12	0.15	0.47	0.00	0.00	0.40	-0.11	-0.19	-0.24	0.40	0.54	0.15	-0.22	0.99	-0.56	0.95
35	421156	0	210	0.46	0.23	0.46	0.20	0.09	0.01	0.00	0.18	0.11	0.18	-0.19	-0.22	0.46	0.15	3.13	1.17	2.55	1.22
36	468957	0	210	0.62	0.10	0.20	0.62	0.08	0.00	0.00	0.25	-0.02	-0.13	0.25	-0.23	0.19	0.15	1.70	1.08	1.26	1.11
37	469811	0	210	0.42	0.19	0.42	0.22	0.17	0.00	0.00	0.47	-0.24	0.47	-0.21	-0.13	0.37	0.15	-2.22	0.89	-2.01	0.84
38	469812	0	210	0.31	0.36	0.08	0.24	0.31	0.00	0.00	0.44	-0.18	-0.13	-0.20	0.44	0.96	0.16	-1.88	0.87	-0.83	0.92
39	477342	0	210	0.52	0.19	0.52	0.09	0.20	0.00	0.00	0.23	-0.13	0.23	-0.12	-0.07	0.30	0.15	2.32	1.11	1.60	1.13
40	477344	0	210	0.39	0.10	0.33	0.18	0.39	0.00	0.00	0.42	-0.09	-0.28	-0.11	0.42	0.57	0.15	-1.40	0.92	-0.59	0.95
41	478935	0	210	0.57	0.12	0.09	0.22	0.57	0.00	0.00	0.38	-0.19	-0.17	-0.19	0.38	-0.13	0.15	-0.97	0.96	-0.96	0.91
42	566406	0	210	0.53	0.53	0.11	0.16	0.19	0.00	0.00	0.32	0.32	-0.13	-0.06	-0.23	0.29	0.15	0.73	1.03	-0.13	0.99
43	566508	0	210	0.40	0.29	0.40	0.18	0.14	0.00	0.00	0.50	-0.29	0.50	-0.09	-0.22	0.42	0.15	-2.95	0.86	-2.61	0.80
44	574704	0	210	0.48	0.48	0.17	0.13	0.22	0.00	0.00	0.18	0.18	-0.04	-0.12	-0.09	0.04	0.15	3.48	1.16	2.04	1.19
45	574715	0	210	0.43	0.14	0.16	0.43	0.27	0.00	0.00	0.37	-0.24	-0.15	0.37	-0.10	0.64	0.15	0.23	1.01	-0.28	0.97
46	574718	0	210	0.33	0.33	0.17	0.41	0.09	0.00	0.00	0.35	0.35	-0.27	-0.13	0.01	0.90	0.16	-0.16	0.99	-0.17	0.98
47	575122	0	210	0.62	0.62	0.06	0.14	0.18	0.00	0.00	0.45	0.45	-0.23	-0.21	-0.24	-0.07	0.15	-3.01	0.87	-2.23	0.80
48	578320	0	210	0.49	0.23	0.19	0.49	0.09	0.00	0.00	0.50	-0.23	-0.28	0.50	-0.15	0.38	0.15	-2.52	0.88	-2.31	0.82

**Table J–10. Literature Multiple-Choice Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	621288	0	111	0.50	0.05	0.28	0.50	0.15	0.01	0.00	0.30	-0.05	-0.06	0.30	-0.32	-0.18	0.21	0.78	1.06	0.46	1.04
2	621289	0	111	0.40	0.26	0.13	0.40	0.21	0.01	0.00	0.27	-0.10	-0.13	0.27	-0.13	0.79	0.21	0.86	1.07	0.59	1.06
3	621290	0	111	0.64	0.64	0.11	0.11	0.14	0.01	0.00	0.27	0.27	-0.02	-0.21	-0.20	0.14	0.20	0.68	1.05	0.61	1.05
4	621292	0	111	0.38	0.16	0.38	0.25	0.20	0.01	0.00	0.28	-0.17	0.28	-0.24	0.06	1.10	0.22	1.46	1.14	1.76	1.27
5	621295	0	111	0.49	0.13	0.08	0.30	0.49	0.01	0.00	0.34	-0.15	-0.24	-0.12	0.34	-0.03	0.20	0.01	1.00	-0.24	0.98
6	621299	0	111	0.75	0.05	0.75	0.14	0.05	0.01	0.00	0.20	-0.27	0.20	-0.08	-0.02	-0.62	0.21	-0.40	0.96	-0.44	0.94
7	621300	0	111	0.35	0.35	0.32	0.24	0.07	0.01	0.00	0.09	0.09	0.14	-0.16	-0.18	0.75	0.21	1.81	1.14	1.69	1.20
8	621407	0	111	0.50	0.29	0.14	0.05	0.50	0.01	0.00	0.31	-0.04	-0.29	-0.18	0.31	0.12	0.20	-0.02	1.00	0.30	1.02
9	663965	0	111	0.53	0.14	0.15	0.53	0.18	0.00	0.00	0.40	-0.32	-0.12	0.40	-0.12	-0.66	0.22	1.53	1.14	1.12	1.14
10	663966	0	111	0.66	0.66	0.13	0.12	0.10	0.00	0.00	0.57	0.57	-0.33	-0.21	-0.30	-0.98	0.23	-0.64	0.93	-1.13	0.83
11	663968	0	111	0.39	0.39	0.15	0.15	0.30	0.01	0.00	0.19	0.19	-0.16	-0.16	0.06	0.57	0.21	1.06	1.08	0.98	1.10
12	663969	0	111	0.59	0.15	0.12	0.59	0.14	0.00	0.00	0.29	-0.05	-0.39	0.29	-0.01	-0.13	0.21	0.03	1.00	-0.12	0.99
13	663970	0	111	0.53	0.22	0.53	0.09	0.16	0.00	0.00	0.30	-0.13	0.30	-0.37	0.02	-0.30	0.21	0.81	1.06	0.91	1.09
14	663971	0	111	0.29	0.26	0.29	0.16	0.29	0.00	0.00	0.03	-0.07	-0.05	0.10	0.03	1.34	0.23	2.25	1.26	2.99	1.59
15	663972	0	111	0.49	0.13	0.49	0.24	0.14	0.00	0.00	0.38	-0.03	0.38	-0.39	-0.04	0.19	0.20	-0.86	0.94	-0.79	0.93
16	663975	0	111	0.41	0.41	0.11	0.27	0.20	0.01	0.00	0.13	0.13	-0.21	-0.03	0.04	0.94	0.22	2.87	1.26	2.68	1.38
17	665262	0	111	0.37	0.30	0.16	0.37	0.17	0.00	0.00	0.37	-0.05	-0.35	0.37	-0.07	0.18	0.20	-0.68	0.95	-0.71	0.94
18	303307	0	111	0.41	0.11	0.41	0.37	0.10	0.01	0.00	0.23	-0.23	0.23	-0.05	-0.06	0.17	0.20	0.93	1.06	1.17	1.10
19	303309	0	111	0.42	0.22	0.13	0.23	0.42	0.01	0.00	0.30	-0.12	-0.08	-0.17	0.30	0.67	0.21	0.43	1.03	0.77	1.08
20	303310	0	111	0.27	0.27	0.23	0.36	0.14	0.01	0.00	0.06	0.06	-0.19	0.16	-0.06	0.95	0.22	1.01	1.09	0.94	1.12
21	303311	0	111	0.54	0.15	0.54	0.26	0.05	0.00	0.00	0.18	0.01	0.18	-0.18	-0.08	0.23	0.20	1.73	1.12	2.18	1.20
22	303312	0	111	0.61	0.13	0.19	0.61	0.06	0.01	0.00	0.41	-0.14	-0.24	0.41	-0.24	-0.15	0.21	-1.42	0.90	-1.39	0.88
23	303313	0	111	0.59	0.05	0.15	0.20	0.59	0.01	0.00	0.35	-0.06	-0.23	-0.19	0.35	0.01	0.20	-0.56	0.96	-0.89	0.92
24	303318	0	111	0.64	0.07	0.08	0.20	0.64	0.01	0.00	0.54	-0.35	-0.23	-0.26	0.54	-1.00	0.23	0.09	1.01	-0.56	0.91
25	303319	0	111	0.61	0.61	0.10	0.22	0.07	0.00	0.00	0.47	0.47	-0.31	-0.26	-0.11	-0.42	0.21	-1.71	0.87	-1.66	0.83
26	340120	0	111	0.56	0.56	0.09	0.25	0.10	0.00	0.00	0.17	0.17	-0.05	-0.13	-0.05	0.26	0.20	1.97	1.14	1.85	1.17
27	340121	0	111	0.63	0.02	0.12	0.63	0.23	0.00	0.00	0.41	-0.08	-0.32	0.41	-0.20	-0.17	0.21	-1.66	0.88	-1.29	0.88
28	340122	0	111	0.20	0.42	0.27	0.11	0.20	0.00	0.00	0.10	0.12	-0.13	-0.15	0.10	1.84	0.26	1.21	1.19	1.29	1.32
29	340124	0	111	0.30	0.15	0.05	0.50	0.30	0.00	0.00	0.08	-0.20	-0.14	0.13	0.08	1.33	0.23	1.99	1.23	2.25	1.42
30	340125	0	111	0.32	0.41	0.32	0.14	0.14	0.00	0.00	0.35	-0.21	0.35	-0.07	-0.10	1.00	0.22	-0.61	0.94	-0.70	0.90
31	340126	0	111	0.60	0.60	0.15	0.12	0.13	0.00	0.00	0.15	0.15	-0.15	-0.10	0.04	-0.41	0.21	1.80	1.15	1.74	1.19
32	340128	0	111	0.76	0.09	0.76	0.04	0.12	0.00	0.00	0.41	-0.18	0.41	-0.19	-0.28	-0.74	0.22	-2.36	0.79	-2.20	0.73
33	340129	0	111	0.55	0.14	0.19	0.12	0.55	0.00	0.00	0.13	-0.01	-0.07	-0.10	0.13	-0.10	0.21	2.09	1.15	1.92	1.18

**Table J–10 (continued). Literature Multiple-Choice Item Statistics: Summer**

Ref	PubID	Form	<i>N</i>	Pval	P(A)	P(B)	P(C)	P(D)	P(-)	P(*)	ITCorr	Corr(A)	Corr(B)	Corr(C)	Corr(D)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	340130	0	111	0.44	0.44	0.11	0.32	0.14	0.00	0.00	0.21	0.21	-0.09	-0.13	-0.04	0.26	0.20	1.11	1.08	1.55	1.14

## CONSTRUCTED-RESPONSE ITEMS

Table J–11. Algebra I Constructed-Response Item Statistics: Winter

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	569522	0	54824	0.13	0.54	0.28	0.08	0.02	0.00	0.06	0.66	-0.46	0.30	0.39	0.28	0.14	2.17	0.01	-9.90	0.68	-9.90	0.60
2	497691	0	54824	0.16	0.52	0.27	0.10	0.04	0.01	0.05	0.59	-0.42	0.16	0.30	0.30	0.25	1.80	0.01	-1.43	0.99	-9.90	0.89
3	395477	0	54824	0.32	0.25	0.27	0.19	0.14	0.05	0.08	0.66	-0.40	-0.10	0.19	0.37	0.36	0.82	0.01	-9.90	0.94	-9.90	0.89
4	420160	0	54824	0.21	0.42	0.27	0.13	0.06	0.02	0.07	0.70	-0.45	0.10	0.31	0.37	0.34	1.10	0.01	-9.90	0.70	-9.90	0.64
5	314746	0	54824	0.13	0.59	0.27	0.07	0.02	0.01	0.03	0.66	-0.49	0.25	0.34	0.32	0.26	1.87	0.01	-9.90	0.78	-9.90	0.71
6	339446	0	54824	0.23	0.28	0.51	0.14	0.03	0.01	0.02	0.60	-0.36	0.00	0.32	0.31	0.25	1.73	0.01	-9.90	0.86	-9.90	0.85

Table J–12. Biology Constructed-Response Item Statistics: Winter

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	337696	0	40077	0.12	0.65	0.15	0.06	0.03	0.07	0.62	-0.31	0.24	0.37	0.38	1.22	0.01	-9.90	0.77	-9.90	0.63
2	412911	0	40077	0.17	0.54	0.25	0.08	0.04	0.07	0.67	-0.39	0.19	0.42	0.40	1.20	0.01	-9.90	0.69	-9.90	0.63
3	303514	0	40077	0.32	0.33	0.29	0.20	0.09	0.07	0.63	-0.35	-0.02	0.29	0.47	0.55	0.01	-9.90	0.89	-9.90	0.86
4	469074	0	40077	0.23	0.45	0.31	0.12	0.04	0.05	0.61	-0.35	0.09	0.34	0.41	1.12	0.01	-9.90	0.87	-9.90	0.84
5	566503	0	40077	0.20	0.49	0.29	0.09	0.04	0.06	0.65	-0.40	0.15	0.38	0.42	0.86	0.01	-9.90	0.79	-9.90	0.75
6	469018	0	40077	0.47	0.15	0.16	0.53	0.06	0.06	0.44	-0.22	-0.02	0.13	0.40	0.63	0.01	9.90	1.51	9.90	1.54

Table J–13. Literature Constructed-Response Item Statistics: Winter

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	516559	0	35222	0.29	0.30	0.33	0.20	0.04	0.10	0.73	-0.40	0.04	0.51	0.35	0.92	0.01	-9.90	0.90	-9.90	0.86
2	516557	0	35222	0.35	0.21	0.40	0.23	0.07	0.06	0.75	-0.42	-0.07	0.46	0.42	0.99	0.01	-9.90	0.69	-9.90	0.68
3	512208	0	35222	0.37	0.14	0.44	0.27	0.05	0.08	0.70	-0.36	-0.11	0.48	0.34	0.89	0.02	-9.90	0.82	-9.90	0.81
4	512098	0	35222	0.38	0.15	0.41	0.28	0.06	0.07	0.73	-0.37	-0.14	0.47	0.38	0.82	0.01	-9.90	0.69	-9.90	0.68
5	512079	0	35222	0.44	0.10	0.41	0.33	0.08	0.05	0.71	-0.34	-0.27	0.42	0.42	1.08	0.01	-9.90	0.77	-9.90	0.76
6	528771	0	35222	0.46	0.09	0.39	0.38	0.07	0.04	0.68	-0.33	-0.29	0.42	0.38	0.88	0.02	-9.90	0.85	-9.90	0.84

**Table J-14. Algebra I Constructed-Response Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	324670	0	118915	0.19	0.49	0.25	0.12	0.06	0.02	0.04	0.77	-0.60	0.18	0.40	0.41	0.28					1.95	0.01	-9.90	0.79	-9.90	0.69
2	418194	0	118915	0.26	0.41	0.25	0.17	0.10	0.04	0.03	0.80	-0.61	0.01	0.36	0.42	0.35					1.53	0.01	-9.90	0.88	-9.90	0.81
3	420997	0	118915	0.31	0.44	0.14	0.12	0.16	0.09	0.04	0.73	-0.56	0.01	0.17	0.41	0.41					1.34	0.01	9.90	1.30	9.90	1.22
4	421535	0	118915	0.44	0.12	0.31	0.29	0.20	0.07	0.01	0.81	-0.39	-0.44	0.07	0.49	0.44					0.87	0.02	-9.90	0.73	-9.90	0.73
5	569437	0	118915	0.33	0.33	0.25	0.18	0.14	0.08	0.02	0.81	-0.58	-0.11	0.22	0.42	0.46					1.31	0.01	-9.90	0.88	-9.90	0.82
6	580100	0	118915	0.29	0.16	0.52	0.16	0.08	0.02	0.05	0.71	-0.35	-0.18	0.35	0.42	0.26					1.74	0.01	-9.90	0.85	-9.90	0.83
7	630782	1	1493	0.19	0.54	0.20	0.12	0.07	0.03	0.03	0.72	-0.58	0.14	0.35	0.42	0.28	A+	NA	NA	A-	1.95	0.07	-3.38	0.86	-5.01	0.72
8	719077	1	1493	0.21	0.60	0.13	0.09	0.07	0.08	0.02	0.73	-0.62	0.19	0.22	0.30	0.49	A-	NA	NA	A+	1.56	0.07	-1.13	0.95	-3.29	0.75
9	667463	2	1496	0.23	0.53	0.19	0.09	0.11	0.06	0.02	0.73	-0.60	0.08	0.27	0.42	0.36	A+	NA	NA	A-	1.80	0.07	-1.07	0.95	-3.18	0.81
10	719955	2	1496	0.31	0.43	0.19	0.15	0.13	0.09	0.01	0.80	-0.65	0.00	0.19	0.43	0.46	A-	NA	NA	A-	1.37	0.07	-5.84	0.78	-6.75	0.69
11	664310	3	1493	0.38	0.18	0.33	0.29	0.12	0.06	0.02	0.70	-0.43	-0.20	0.14	0.39	0.39	A-	NA	NA	A-	1.09	0.07	-3.50	0.88	-3.92	0.87
12	580099	3	1493	0.42	0.20	0.14	0.35	0.27	0.01	0.02	0.73	-0.51	-0.22	0.08	0.56	0.18	A-	NA	NA	A-	1.48	0.07	-9.11	0.72	-7.96	0.72
13	719078	4	1495	0.34	0.31	0.21	0.29	0.13	0.05	0.01	0.76	-0.56	-0.15	0.27	0.38	0.40	A+	A-	NA	A-	1.31	0.07	-7.13	0.77	-7.73	0.74
14	721018	4	1495	0.18	0.46	0.31	0.16	0.02	0.01	0.02	0.72	-0.58	0.19	0.49	0.24	0.19	A+	A-	NA	A-	2.49	0.06	-5.15	0.81	-6.69	0.73
15	623098	5	1494	0.12	0.55	0.35	0.06	0.00	0.00	0.03	0.67	-0.59	0.50	0.34	0.10	0.07	A-	A+	NA	A+	3.80	0.06	-5.54	0.80	-6.34	0.74
16	708674	5	1494	0.50	0.14	0.18	0.26	0.25	0.14	0.02	0.67	-0.37	-0.23	-0.08	0.21	0.54	A+	A-	NA	A+	0.44	0.08	0.32	1.01	0.90	1.03
17	615354	6	1493	0.28	0.36	0.28	0.17	0.10	0.04	0.03	0.75	-0.58	0.02	0.33	0.39	0.31	A-	NA	NA	A+	1.65	0.06	-3.69	0.86	-5.09	0.80
18	708672	6	1493	0.31	0.31	0.29	0.17	0.13	0.05	0.03	0.80	-0.56	-0.12	0.30	0.46	0.38	A+	NA	NA	A-	1.46	0.07	-9.58	0.68	-9.90	0.63
19	666624	7	1495	0.35	0.26	0.27	0.29	0.10	0.06	0.01	0.75	-0.48	-0.23	0.25	0.39	0.40	A+	NA	NA	A+	1.20	0.07	-5.35	0.82	-5.43	0.82
20	719073	7	1495	0.14	0.58	0.30	0.07	0.02	0.01	0.02	0.60	-0.48	0.25	0.34	0.24	0.21	A-	NA	NA	A-	2.80	0.06	-1.36	0.94	-0.62	0.97
21	708673	8	1494	0.16	0.57	0.23	0.10	0.05	0.02	0.02	0.67	-0.56	0.26	0.26	0.36	0.30	A+	NA	NA	A-	2.27	0.06	-2.44	0.89	-3.55	0.80
22	719995	8	1494	0.21	0.52	0.21	0.11	0.08	0.04	0.02	0.74	-0.57	0.12	0.30	0.37	0.40	A+	NA	NA	A+	1.84	0.07	-4.09	0.83	-4.99	0.73
23	720107	9	1491	0.31	0.32	0.30	0.20	0.11	0.05	0.02	0.68	-0.42	-0.12	0.17	0.41	0.38	A+	NA	NA	A-	1.43	0.06	-0.07	1.00	-0.44	0.98
24	720776	9	1491	0.31	0.35	0.25	0.21	0.13	0.04	0.02	0.66	-0.46	-0.11	0.21	0.40	0.33	A-	NA	NA	A+	1.54	0.07	1.46	1.05	0.82	1.03
25	720106	10	1492	0.22	0.37	0.35	0.16	0.06	0.00	0.04	0.73	-0.55	0.10	0.44	0.36	0.12	A+	NA	NA	A+	2.65	0.06	-6.04	0.79	-7.01	0.74
26	721021	10	1492	0.68	0.05	0.14	0.17	0.23	0.39	0.02	0.63	-0.24	-0.33	-0.21	-0.01	0.56	A+	NA	NA	A+	-0.44	0.11	-0.70	0.97	2.80	1.15
27	630779	11	1494	0.26	0.41	0.19	0.30	0.08	0.00	0.02	0.72	-0.60	0.04	0.45	0.36	0.10	A+	A-	NA	A-	2.28	0.06	-5.37	0.82	-6.34	0.75
28	651614	11	1494	0.16	0.53	0.30	0.07	0.04	0.01	0.03	0.66	-0.52	0.28	0.32	0.36	0.17	A+	A+	NA	A-	2.43	0.06	-3.66	0.84	-5.19	0.76
29	719953	12	1499	0.30	0.34	0.30	0.19	0.07	0.08	0.02	0.70	-0.55	-0.03	0.34	0.25	0.39	A+	NA	A+	A+	1.35	0.06	-0.23	0.99	-2.15	0.91
30	720104	12	1499	0.30	0.32	0.26	0.24	0.14	0.01	0.02	0.65	-0.47	-0.09	0.30	0.43	0.15	A+	NA	A-	A-	2.06	0.07	0.35	1.01	-0.40	0.98
31	618262	13	1496	0.09	0.73	0.16	0.05	0.03	0.00	0.02	0.58	-0.47	0.30	0.33	0.29	0.12	A+	NA	NA	A+	3.06	0.07	-3.22	0.82	-3.44	0.72
32	719076	13	1496	0.27	0.40	0.22	0.24	0.05	0.05	0.02	0.68	-0.53	0.02	0.35	0.28	0.33	A+	NA	NA	A+	1.60	0.06	0.81	1.03	-0.58	0.97

**Table J–14 (continued). Algebra I Constructed-Response Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
33	720108	14	1494	0.28	0.44	0.20	0.11	0.11	0.09	0.03	0.75	-0.56	0.04	0.24	0.34	0.47	A-	NA	A-	A-	1.39	0.07	-3.02	0.88	-4.14	0.79
34	721020	14	1494	0.34	0.36	0.22	0.19	0.15	0.08	0.01	0.75	-0.59	-0.10	0.24	0.38	0.41	A-	NA	A-	A-	1.17	0.07	-3.86	0.86	-4.87	0.80
35	719952	15	1494	0.13	0.63	0.17	0.10	0.04	0.01	0.03	0.69	-0.56	0.33	0.38	0.31	0.23	A-	NA	NA	A-	2.65	0.07	-5.38	0.76	-7.30	0.53
36	719954	15	1494	0.42	0.22	0.25	0.17	0.22	0.11	0.02	0.82	-0.51	-0.28	0.08	0.47	0.48	A-	NA	NA	A-	0.89	0.07	-9.90	0.65	-9.90	0.64
37	719506	16	1496	0.28	0.35	0.36	0.13	0.05	0.08	0.02	0.77	-0.56	-0.01	0.32	0.30	0.47	A+	NA	NA	A+	1.49	0.06	-7.11	0.73	-8.12	0.69
38	719950	16	1496	0.21	0.37	0.40	0.10	0.08	0.00	0.03	0.72	-0.55	0.19	0.34	0.43	0.07	A+	NA	NA	A-	2.88	0.06	-4.90	0.82	-6.47	0.76
39	708676	17	1495	0.19	0.37	0.45	0.12	0.01	0.01	0.02	0.66	-0.52	0.24	0.41	0.14	0.20	A-	A-	NA	A-	2.51	0.06	-3.96	0.85	-4.55	0.84
40	719991	17	1495	0.19	0.52	0.25	0.10	0.06	0.02	0.02	0.70	-0.60	0.24	0.35	0.34	0.26	A+	A-	NA	A-	2.17	0.06	-2.02	0.91	-4.17	0.79
41	719074	18	1497	0.39	0.31	0.20	0.14	0.14	0.16	0.03	0.76	-0.55	-0.08	0.10	0.29	0.53	A+	NA	NA	A-	0.86	0.07	-1.76	0.93	-2.76	0.87
42	720109	18	1497	0.20	0.47	0.26	0.16	0.05	0.02	0.03	0.74	-0.59	0.19	0.39	0.37	0.23	A+	NA	NA	A+	2.27	0.06	-5.55	0.79	-6.89	0.69
43	708675	19	1494	0.39	0.11	0.42	0.27	0.15	0.04	0.01	0.76	-0.37	-0.44	0.27	0.42	0.38	A+	NA	NA	A-	1.04	0.09	-8.19	0.73	-8.31	0.73
44	575216	19	1494	0.40	0.35	0.14	0.12	0.20	0.15	0.02	0.80	-0.63	-0.07	0.02	0.38	0.54	A+	NA	NA	A-	0.85	0.07	-6.35	0.77	-5.89	0.72
45	615073	20	1497	0.44	0.18	0.27	0.21	0.17	0.14	0.02	0.78	-0.45	-0.30	0.08	0.34	0.53	A+	NA	NA	A-	0.65	0.07	-6.72	0.78	-5.64	0.80
46	721022	20	1497	0.23	0.49	0.22	0.10	0.08	0.06	0.03	0.76	-0.60	0.11	0.30	0.39	0.40	A-	NA	NA	A+	1.70	0.07	-5.66	0.77	-6.61	0.66
47	618977	21	1493	0.52	0.16	0.19	0.22	0.21	0.20	0.01	0.68	-0.44	-0.25	-0.02	0.22	0.49	A-	NA	NA	A-	0.32	0.08	0.60	1.02	1.59	1.06
48	624024	21	1493	0.17	0.42	0.43	0.10	0.01	0.00	0.02	0.69	-0.55	0.32	0.46	0.12	0.10	A-	NA	NA	A-	3.06	0.06	-6.77	0.77	-8.12	0.73
49	615355	22	1493	0.43	0.25	0.14	0.26	0.22	0.10	0.02	0.75	-0.52	-0.15	0.05	0.36	0.47	A+	NA	NA	A-	0.83	0.07	-5.22	0.83	-4.85	0.82
50	721023	22	1493	0.26	0.38	0.28	0.21	0.08	0.02	0.02	0.71	-0.50	-0.05	0.35	0.42	0.26	A+	NA	NA	A-	1.90	0.06	-3.25	0.88	-4.52	0.82
51	720774	23	1492	0.17	0.60	0.15	0.12	0.06	0.02	0.03	0.72	-0.60	0.19	0.35	0.39	0.28	A+	A-	NA	A-	2.19	0.07	-4.31	0.81	-5.38	0.65
52	720779	23	1492	0.30	0.37	0.27	0.16	0.11	0.07	0.02	0.74	-0.55	-0.03	0.21	0.37	0.45	A+	A+	NA	A-	1.37	0.06	-3.09	0.88	-3.43	0.86
53	632299	24	1495	0.39	0.29	0.21	0.18	0.19	0.11	0.02	0.70	-0.50	-0.14	0.13	0.36	0.40	A+	NA	NA	A-	1.02	0.07	2.25	1.09	0.99	1.04
54	719996	24	1495	0.23	0.51	0.18	0.15	0.09	0.04	0.02	0.75	-0.61	0.06	0.31	0.41	0.36	A+	NA	NA	A-	1.87	0.07	-5.06	0.80	-5.82	0.69



**Table J–15. Biology Constructed-Response Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	635235	0	112360	0.35	0.33	0.36	0.19	0.11	0.01	0.64	-0.47	-0.03	0.31	0.43					0.79	0.03	0.26	1.00	0.31	1.00
2	656079	0	112360	0.40	0.26	0.29	0.26	0.13	0.03	0.72	-0.48	-0.09	0.36	0.47					0.78	0.03	-9.90	0.78	-9.90	0.81
3	658169	0	112360	0.41	0.27	0.31	0.24	0.15	0.02	0.65	-0.44	-0.09	0.27	0.45					0.59	0.03	1.08	1.02	2.99	1.05
4	342566	0	112360	0.37	0.23	0.40	0.24	0.08	0.04	0.59	-0.20	-0.26	0.36	0.44					0.73	0.03	-0.82	0.99	-2.31	0.97
5	578642	0	112360	0.35	0.41	0.23	0.16	0.16	0.03	0.72	-0.51	-0.01	0.27	0.55					0.85	0.03	-6.48	0.91	-4.90	0.91
6	583809	0	112360	0.42	0.29	0.28	0.24	0.17	0.01	0.74	-0.55	-0.10	0.30	0.52					0.67	0.03	-9.90	0.84	-9.90	0.83
7	632268	1	1496	0.14	0.67	0.15	0.07	0.04	0.04	0.61	-0.43	0.32	0.37	0.31	A+	NA	NA	A+	2.14	0.07	-2.63	0.87	-4.80	0.63
8	634551	1	1496	0.31	0.43	0.18	0.23	0.10	0.03	0.59	-0.42	0.01	0.38	0.34	A+	NA	NA	A+	1.25	0.06	4.12	1.16	2.03	1.11
9	617705	2	1496	0.34	0.31	0.37	0.19	0.09	0.02	0.69	-0.51	0.02	0.38	0.40	A-	C-	NA	A-	1.20	0.06	-4.43	0.85	-5.24	0.82
10	630602	2	1496	0.52	0.10	0.29	0.44	0.13	0.02	0.65	-0.30	-0.38	0.33	0.41	A+	B-	NA	A-	0.38	0.08	-4.85	0.85	-4.37	0.86
11	632269	3	1498	0.36	0.35	0.22	0.26	0.11	0.03	0.72	-0.51	0.01	0.35	0.48	A-	A-	NA	A-	1.12	0.06	-5.18	0.83	-5.34	0.79
12	635246	3	1498	0.26	0.52	0.13	0.24	0.05	0.04	0.69	-0.54	0.12	0.48	0.36	A+	A-	NA	A-	1.70	0.07	-3.42	0.87	-4.94	0.72
13	625836	4	1494	0.71	0.04	0.17	0.33	0.43	0.02	0.57	-0.22	-0.30	-0.21	0.56	A+	NA	NA	A+	-0.65	0.11	-0.43	0.98	0.00	1.00
14	630622	4	1494	0.34	0.26	0.41	0.23	0.05	0.04	0.37	-0.26	0.07	0.21	0.16	A-	NA	NA	A-	1.38	0.06	9.90	1.41	9.90	1.41
15	623006	5	1495	0.30	0.41	0.29	0.13	0.11	0.03	0.67	-0.49	0.12	0.30	0.45	A+	NA	NA	A-	1.31	0.06	-2.44	0.91	-3.43	0.85
16	635454	5	1495	0.58	0.10	0.25	0.32	0.28	0.03	0.48	-0.26	-0.18	0.09	0.36	A+	NA	NA	A+	0.01	0.08	6.84	1.25	8.25	1.34
17	618335	6	1493	0.59	0.14	0.22	0.29	0.32	0.02	0.58	-0.39	-0.21	0.13	0.43	A+	NA	NA	A+	0.03	0.08	1.67	1.06	4.70	1.20
18	635329	6	1493	0.48	0.19	0.28	0.32	0.18	0.03	0.63	-0.44	-0.14	0.23	0.44	A+	NA	NA	A+	0.54	0.07	-0.72	0.98	-0.55	0.98

**Table J–16. Literature Constructed-Response Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	666342	0	108475	0.50	0.12	0.31	0.37	0.15	0.03	0.75	-0.44	-0.24	0.37	0.43					0.98	0.02	-9.90	0.79	-9.90	0.78
2	666767	0	108475	0.49	0.09	0.34	0.35	0.14	0.05	0.73	-0.38	-0.26	0.39	0.41					1.21	0.02	-9.90	0.81	-9.90	0.81
3	667361	0	108475	0.50	0.12	0.28	0.40	0.14	0.04	0.77	-0.50	-0.21	0.39	0.41					0.74	0.02	-9.90	0.82	-9.90	0.81
4	667363	0	108475	0.55	0.08	0.29	0.41	0.18	0.02	0.78	-0.45	-0.34	0.32	0.47					0.69	0.02	-9.90	0.73	-9.90	0.73
5	584216	0	108475	0.51	0.07	0.37	0.40	0.12	0.03	0.72	-0.35	-0.35	0.41	0.38					1.12	0.02	-6.91	0.95	-7.47	0.95
6	594417	0	108475	0.60	0.05	0.26	0.43	0.23	0.02	0.75	-0.37	-0.40	0.23	0.49					0.44	0.02	-9.90	0.89	-9.90	0.89
7	711011	1	1497	0.56	0.07	0.27	0.50	0.14	0.01	0.72	-0.41	-0.38	0.36	0.39	B+	NA	A-	A-	0.66	0.10	-8.45	0.73	-8.30	0.73
8	711098	1	1497	0.58	0.06	0.25	0.50	0.16	0.01	0.75	-0.41	-0.38	0.33	0.41	A+	NA	A-	A-	0.60	0.10	-9.78	0.69	-9.75	0.69
9	711006	2	1492	0.52	0.08	0.37	0.40	0.13	0.01	0.72	-0.40	-0.37	0.38	0.41	A+	NA	NA	A-	0.82	0.10	-8.99	0.72	-9.08	0.72
10	711100	2	1492	0.52	0.10	0.34	0.37	0.16	0.02	0.74	-0.43	-0.32	0.32	0.46	A+	NA	NA	A+	0.82	0.09	-9.13	0.72	-9.38	0.71
11	721730	3	1497	0.47	0.11	0.38	0.38	0.10	0.02	0.73	-0.43	-0.24	0.45	0.34	A+	A-	NA	A+	1.17	0.09	-9.52	0.71	-9.54	0.71
12	723243	3	1497	0.49	0.12	0.36	0.36	0.13	0.01	0.74	-0.50	-0.21	0.40	0.39	A+	A+	NA	A-	1.05	0.08	-9.82	0.70	-9.79	0.70
13	721735	4	1499	0.51	0.13	0.27	0.46	0.11	0.02	0.77	-0.52	-0.27	0.46	0.36	A+	NA	NA	A-	1.10	0.08	-9.90	0.67	-9.90	0.67
14	723245	4	1499	0.50	0.10	0.36	0.43	0.09	0.01	0.72	-0.44	-0.35	0.48	0.33	A+	NA	NA	A-	1.09	0.09	-8.79	0.72	-8.83	0.72
15	721746	5	1499	0.55	0.09	0.30	0.39	0.19	0.01	0.72	-0.45	-0.28	0.28	0.44	A+	NA	A-	A+	0.71	0.09	-8.53	0.74	-8.42	0.74
16	723282	5	1499	0.59	0.03	0.32	0.43	0.20	0.01	0.70	-0.32	-0.45	0.27	0.44	A+	NA	A+	A+	0.34	0.12	-8.55	0.74	-8.56	0.74
17	721747	6	1494	0.52	0.12	0.31	0.35	0.18	0.02	0.75	-0.44	-0.28	0.33	0.46	A+	A-	NA	A+	0.90	0.08	-9.90	0.69	-9.90	0.68
18	723271	6	1494	0.57	0.07	0.29	0.43	0.19	0.01	0.75	-0.46	-0.35	0.28	0.46	B+	A-	NA	A+	0.57	0.10	-9.90	0.67	-9.90	0.67
19	721771	7	1499	0.42	0.19	0.37	0.30	0.10	0.03	0.74	-0.53	-0.08	0.44	0.36	A+	A+	A+	A-	1.47	0.07	-9.90	0.70	-9.90	0.68
20	722267	7	1499	0.50	0.12	0.30	0.38	0.14	0.04	0.76	-0.46	-0.23	0.37	0.44	A+	A-	A-	A+	1.09	0.08	-9.90	0.68	-9.90	0.67
21	721772	8	1497	0.47	0.14	0.34	0.34	0.13	0.02	0.76	-0.49	-0.20	0.40	0.42	B+	NA	NA	A+	1.15	0.08	-9.90	0.66	-9.90	0.66
22	722266	8	1497	0.51	0.14	0.27	0.39	0.16	0.02	0.75	-0.46	-0.25	0.35	0.45	A+	NA	NA	A+	0.96	0.08	-9.83	0.70	-9.89	0.69
23	722198	9	1495	0.51	0.14	0.29	0.38	0.16	0.01	0.78	-0.52	-0.22	0.34	0.47	A+	NA	NA	A-	0.95	0.08	-9.90	0.66	-9.90	0.65
24	722298	9	1495	0.51	0.10	0.30	0.43	0.13	0.02	0.76	-0.44	-0.26	0.38	0.42	B+	NA	NA	A-	0.97	0.09	-9.65	0.70	-9.48	0.70
25	722200	10	1498	0.57	0.07	0.28	0.44	0.18	0.01	0.75	-0.36	-0.42	0.31	0.47	B+	NA	NA	A-	0.63	0.10	-9.42	0.71	-9.51	0.71
26	722296	10	1498	0.54	0.07	0.30	0.46	0.14	0.02	0.74	-0.40	-0.35	0.37	0.41	B+	NA	NA	A-	0.80	0.10	-8.87	0.72	-8.86	0.72
27	722739	11	1498	0.45	0.15	0.35	0.35	0.10	0.02	0.74	-0.52	-0.15	0.47	0.33	B+	NA	NA	A+	1.32	0.08	-8.95	0.73	-9.18	0.72
28	723237	11	1498	0.53	0.04	0.40	0.39	0.14	0.01	0.71	-0.33	-0.40	0.37	0.40	A+	NA	NA	A-	0.71	0.11	-8.34	0.74	-8.37	0.74
29	722740	12	1497	0.44	0.13	0.40	0.36	0.07	0.02	0.74	-0.44	-0.23	0.50	0.32	A+	NA	A-	A-	1.33	0.08	-9.90	0.67	-9.90	0.67
30	723236	12	1497	0.51	0.08	0.37	0.40	0.11	0.02	0.72	-0.40	-0.33	0.41	0.38	A+	NA	A-	A-	0.88	0.09	-9.75	0.70	-9.76	0.70
31	720732	13	1498	0.44	0.17	0.35	0.31	0.12	0.02	0.71	-0.46	-0.15	0.42	0.37	A+	NA	NA	A+	1.35	0.07	-7.94	0.75	-8.14	0.74
32	723234	13	1498	0.55	0.10	0.29	0.35	0.21	0.02	0.77	-0.45	-0.28	0.28	0.50	A+	NA	NA	A+	0.77	0.09	-9.90	0.68	-9.90	0.67
33	720734	14	1496	0.42	0.21	0.33	0.30	0.11	0.03	0.73	-0.53	-0.08	0.39	0.41	A+	NA	NA	A+	1.45	0.07	-9.90	0.70	-9.90	0.67

**Table J–16 (continued). Literature Constructed-Response Item Statistics: Spring**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	M/F	W/B	W/H	O/P	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
34	723235	14	1496	0.57	0.09	0.27	0.39	0.21	0.02	0.73	-0.42	-0.29	0.24	0.48	B+	NA	NA	A-	0.66	0.09	-7.56	0.76	-7.96	0.75
35	722329	15	1496	0.50	0.11	0.35	0.37	0.14	0.02	0.76	-0.48	-0.30	0.38	0.44	A+	NA	NA	A-	0.94	0.08	-9.90	0.68	-9.90	0.68
36	723041	15	1496	0.60	0.07	0.22	0.47	0.22	0.01	0.78	-0.46	-0.41	0.25	0.50	A+	NA	NA	A+	0.44	0.10	-9.90	0.63	-9.90	0.63
37	722331	16	1499	0.51	0.11	0.31	0.37	0.16	0.03	0.77	-0.44	-0.27	0.37	0.45	A+	NA	NA	A+	1.01	0.08	-9.90	0.65	-9.90	0.64
38	723069	16	1499	0.50	0.12	0.30	0.39	0.15	0.03	0.76	-0.47	-0.25	0.36	0.45	A+	NA	NA	A-	1.07	0.08	-9.90	0.66	-9.90	0.66
39	722843	17	1497	0.43	0.18	0.36	0.34	0.09	0.02	0.71	-0.49	-0.12	0.43	0.34	A+	NA	NA	A-	1.54	0.07	-8.56	0.74	-8.77	0.73
40	723583	17	1497	0.56	0.09	0.27	0.43	0.18	0.01	0.73	-0.40	-0.35	0.33	0.43	A+	NA	NA	A+	0.80	0.09	-8.61	0.73	-8.42	0.74
41	722841	18	1499	0.50	0.12	0.32	0.40	0.12	0.02	0.73	-0.48	-0.22	0.36	0.41	A+	NA	NA	A-	1.01	0.08	-9.39	0.71	-9.40	0.71
42	723584	18	1499	0.55	0.07	0.27	0.47	0.14	0.02	0.73	-0.39	-0.35	0.35	0.42	A+	NA	NA	A+	0.73	0.09	-9.80	0.70	-9.72	0.70
43	721775	19	1498	0.48	0.13	0.36	0.34	0.13	0.03	0.75	-0.46	-0.23	0.42	0.41	A+	A-	NA	A-	1.08	0.08	-9.90	0.68	-9.90	0.67
44	722905	19	1498	0.53	0.11	0.26	0.41	0.16	0.03	0.76	-0.45	-0.26	0.38	0.42	B+	A-	NA	A-	0.87	0.08	-9.90	0.67	-9.90	0.67
45	721774	20	1497	0.45	0.15	0.37	0.35	0.09	0.02	0.76	-0.48	-0.21	0.47	0.38	A+	NA	A-	A+	1.33	0.08	-9.90	0.67	-9.90	0.67
46	722908	20	1497	0.51	0.10	0.31	0.42	0.13	0.01	0.75	-0.46	-0.28	0.38	0.42	B+	NA	A+	A-	0.94	0.09	-9.34	0.71	-8.95	0.72
47	634506	21	1499	0.51	0.11	0.32	0.35	0.17	0.03	0.75	-0.42	-0.32	0.34	0.48	A+	NA	A-	A+	0.87	0.08	-9.90	0.69	-9.90	0.68
48	637103	21	1499	0.43	0.19	0.36	0.30	0.11	0.02	0.74	-0.52	-0.11	0.42	0.40	A+	NA	A-	A-	1.37	0.07	-9.90	0.69	-9.90	0.68
49	636185	22	1497	0.48	0.12	0.36	0.35	0.13	0.02	0.76	-0.48	-0.21	0.39	0.42	A+	NA	A-	A+	1.14	0.08	-9.90	0.67	-9.90	0.66
50	637036	22	1497	0.56	0.06	0.32	0.40	0.19	0.02	0.75	-0.36	-0.41	0.31	0.46	A+	NA	A-	A+	0.61	0.10	-9.85	0.70	-9.76	0.70
51	636540	23	1497	0.51	0.09	0.33	0.39	0.14	0.02	0.76	-0.44	-0.30	0.42	0.40	A+	NA	NA	A+	1.03	0.09	-9.90	0.65	-9.90	0.65
52	637186	23	1497	0.53	0.10	0.31	0.41	0.15	0.01	0.73	-0.41	-0.34	0.35	0.43	A+	NA	NA	A+	0.91	0.09	-9.12	0.72	-9.31	0.71
53	637158	24	1496	0.48	0.13	0.31	0.39	0.12	0.03	0.76	-0.47	-0.26	0.43	0.41	A+	A-	NA	A+	1.15	0.08	-9.90	0.65	-9.90	0.65
54	637185	24	1496	0.51	0.08	0.37	0.39	0.12	0.02	0.74	-0.42	-0.35	0.41	0.41	B+	A-	NA	A+	0.93	0.09	-9.90	0.69	-9.90	0.69

**Table J–17. Algebra I Constructed-Response Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(4)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Corr(4)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	505393	0	274	0.27	0.28	0.43	0.12	0.09	0.04	0.02	0.66	-0.39	-0.03	0.17	0.35	0.40	1.19	0.14	-3.15	0.74	-3.08	0.73
2	664847	0	274	0.43	0.08	0.30	0.45	0.13	0.03	0.00	0.53	-0.25	-0.28	0.17	0.19	0.35	0.90	0.19	-2.40	0.82	-2.45	0.81
3	623096	0	274	0.28	0.23	0.50	0.19	0.05	0.02	0.01	0.61	-0.34	-0.16	0.29	0.36	0.29	1.42	0.15	-3.13	0.75	-3.47	0.73
4	420057	0	274	0.14	0.58	0.30	0.08	0.02	0.01	0.01	0.52	-0.39	0.20	0.18	0.31	0.25	2.26	0.14	-1.93	0.81	-1.64	0.81
5	663749	0	274	0.26	0.28	0.47	0.14	0.07	0.03	0.01	0.62	-0.39	-0.03	0.22	0.43	0.22	1.25	0.14	-3.00	0.75	-3.03	0.74
6	664307	0	274	0.30	0.39	0.23	0.14	0.15	0.07	0.02	0.46	-0.34	0.03	0.06	0.18	0.35	1.08	0.14	2.51	1.24	2.42	1.32

**Table J–18. Biology Constructed-Response Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	659539	0	210	0.57	0.13	0.23	0.36	0.26	0.02	0.52	-0.27	-0.17	-0.03	0.48	0.03	0.18	-2.41	0.82	-2.10	0.78
2	621826	0	210	0.60	0.11	0.22	0.42	0.24	0.00	0.50	-0.28	-0.24	0.04	0.42	0.25	0.19	-0.28	0.97	0.26	1.02
3	656090	0	210	0.32	0.39	0.29	0.19	0.10	0.03	0.60	-0.33	-0.10	0.18	0.56	1.08	0.16	-0.27	0.97	-0.25	0.97
4	579565	0	210	0.25	0.54	0.20	0.10	0.11	0.04	0.67	-0.42	0.07	0.14	0.61	1.10	0.16	-1.22	0.87	-1.31	0.83
5	578194	0	210	0.31	0.36	0.33	0.17	0.09	0.03	0.55	-0.23	-0.08	0.18	0.51	1.01	0.16	0.16	1.01	-0.13	0.98
6	566401	0	210	0.39	0.29	0.28	0.26	0.12	0.02	0.64	-0.41	-0.06	0.25	0.47	0.60	0.16	-2.70	0.78	-1.44	0.86

**Table J–19. Literature Constructed-Response Item Statistics: Summer**

Ref	PubID	Form	N	Pval	P(0)	P(1)	P(2)	P(3)	P(B)	ITCorr	Corr(0)	Corr(1)	Corr(2)	Corr(3)	Meas	SEM	z-Infit	MS-Infit	z-Outfit	MS-Outfit
1	621896	0	111	0.35	0.11	0.52	0.25	0.01	0.06	0.61	-0.42	0.17	0.34	0.20	0.82	0.25	-4.13	0.58	-3.98	0.59
2	342070	0	111	0.44	0.08	0.41	0.45	0.01	0.04	0.60	-0.30	-0.25	0.55	0.02	0.91	0.24	-3.07	0.67	-3.20	0.66
3	304751	0	111	0.36	0.13	0.45	0.26	0.04	0.09	0.66	-0.36	-0.03	0.43	0.31	1.16	0.23	-2.41	0.73	-2.54	0.71
4	621888	0	111	0.40	0.09	0.50	0.34	0.01	0.04	0.62	-0.35	-0.14	0.52	0.02	0.71	0.27	-4.32	0.56	-4.23	0.56
5	665267	0	111	0.35	0.14	0.47	0.25	0.03	0.06	0.63	-0.28	-0.03	0.43	0.27	0.75	0.24	-3.76	0.62	-3.50	0.62
6	665265	0	111	0.32	0.19	0.38	0.27	0.02	0.09	0.57	-0.31	0.12	0.38	0.19	0.80	0.24	-1.72	0.80	-1.72	0.80

## APPENDIX K: RAW-TO-SCALE SCORE CONVERSION TABLES

---

Appendix K provides the raw-to-scaled score conversion tables for each administration and content area. The scaled score conversions are administration specific and therefore cannot be used for highest total scaled score to date.

**Table K-1. Raw-to-Scaled Score Conversion Tables**

Column Heading	Definition
Raw	Raw score
SS	Scaled score
CSEM	Conditional standard error of measurement
LCI	Lower confidence interval
UCI	Upper confidence interval

## WINTER

Table K-2. Algebra I Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1209	92	1200	1301
1	1271	51	1220	1322
2	1308	37	1271	1345
3	1330	31	1299	1361
4	1347	27	1320	1374
5	1360	25	1335	1385
6	1371	23	1348	1394
7	1381	21	1360	1402
8	1390	20	1370	1410
9	1398	19	1379	1417
10	1405	19	1386	1424
11	1412	18	1394	1430
12	1418	18	1400	1436
13	1424	17	1407	1441
14	1430	17	1413	1447
15	1435	16	1419	1451
16	1440	16	1424	1456
17	1445	16	1429	1461
18	1450	15	1435	1465
19	1455	15	1440	1470
20	1460	15	1445	1475
21	1464	15	1449	1479
22	1468	15	1453	1483
23	1473	15	1458	1488
24	1477	14	1463	1491
25	1481	14	1467	1495
26	1485	14	1471	1499
27	1489	14	1475	1503
28	1493	14	1479	1507
29	1497	14	1483	1511
30	1501	14	1487	1515
31	1505	14	1491	1519
32	1509	14	1495	1523
33	1514	14	1500	1528
34	1518	14	1504	1532
35	1522	14	1508	1536
36	1526	14	1512	1540
37	1530	15	1515	1545
38	1534	15	1519	1549
39	1539	15	1524	1554
40	1543	15	1528	1558
41	1548	15	1533	1563

**Table K-2 (continued). Algebra I Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
42	1553	15	1538	1568
43	1557	16	1541	1573
44	1562	16	1546	1578
45	1568	16	1552	1584
46	1573	17	1556	1590
47	1579	17	1562	1596
48	1585	18	1567	1603
49	1591	18	1573	1609
50	1598	19	1579	1617
51	1605	20	1585	1625
52	1613	21	1592	1634
53	1622	22	1600	1644
54	1632	23	1609	1655
55	1644	25	1619	1669
56	1658	28	1630	1686
57	1675	31	1644	1706
58	1698	38	1660	1736
59	1736	52	1684	1788
60	1799	92	1707	1800

**Table K-3. Biology Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
0	1221	92	1200	1313
1	1281	51	1230	1332
2	1317	36	1281	1353
3	1338	30	1308	1368
4	1354	26	1328	1380
5	1366	24	1342	1390
6	1376	22	1354	1398
7	1385	20	1365	1405
8	1393	19	1374	1412
9	1400	18	1382	1418
10	1406	18	1388	1424
11	1412	17	1395	1429
12	1418	16	1402	1434
13	1423	16	1407	1439
14	1428	15	1413	1443
15	1432	15	1417	1447
16	1437	15	1422	1452
17	1441	14	1427	1455
18	1445	14	1431	1459
19	1449	14	1435	1463
20	1453	14	1439	1467
21	1457	14	1443	1471
22	1460	13	1447	1473
23	1464	13	1451	1477
24	1468	13	1455	1481
25	1471	13	1458	1484
26	1474	13	1461	1487
27	1478	13	1465	1491
28	1481	13	1468	1494
29	1484	13	1471	1497
30	1488	13	1475	1501
31	1491	13	1478	1504
32	1494	13	1481	1507
33	1497	13	1484	1510
34	1500	13	1487	1513
35	1504	13	1491	1517
36	1507	13	1494	1520
37	1510	13	1497	1523
38	1513	13	1500	1526
39	1517	13	1504	1530
40	1520	13	1507	1533
41	1523	13	1510	1536
42	1527	13	1514	1540



**Table K-3 (continued). Biology Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
43	1530	13	1517	1543
44	1534	13	1521	1547
45	1537	14	1523	1551
46	1541	14	1527	1555
47	1545	14	1531	1559
48	1549	14	1535	1563
49	1553	14	1539	1567
50	1557	15	1542	1572
51	1561	15	1546	1576
52	1566	15	1551	1581
53	1571	16	1555	1587
54	1576	16	1560	1592
55	1581	17	1564	1598
56	1587	17	1570	1604
57	1593	18	1575	1611
58	1600	19	1581	1619
59	1608	20	1588	1628
60	1617	22	1595	1639
61	1627	24	1603	1651
62	1639	26	1613	1665
63	1655	30	1625	1685
64	1676	36	1640	1712
65	1712	51	1661	1763
66	1773	92	1681	1800

**Table K-4. Literature Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
0	1206	92	1200	1298
1	1267	51	1216	1318
2	1304	37	1267	1341
3	1326	30	1296	1356
4	1342	27	1315	1369
5	1355	24	1331	1379
6	1366	23	1343	1389
7	1375	21	1354	1396
8	1384	20	1364	1404
9	1392	19	1373	1411
10	1399	19	1380	1418
11	1405	18	1387	1423
12	1412	18	1394	1430
13	1418	17	1401	1435
14	1423	17	1406	1440
15	1429	16	1413	1445
16	1434	16	1418	1450
17	1439	16	1423	1455
18	1444	16	1428	1460
19	1449	16	1433	1465
20	1454	16	1438	1470
21	1459	15	1444	1474
22	1464	15	1449	1479
23	1468	15	1453	1483
24	1473	15	1458	1488
25	1478	15	1463	1493
26	1482	15	1467	1497
27	1487	15	1472	1502
28	1492	15	1477	1507
29	1497	15	1482	1512
30	1501	16	1485	1517
31	1506	16	1490	1522
32	1511	16	1495	1527
33	1517	16	1501	1533
34	1522	16	1506	1538
35	1527	17	1510	1544
36	1533	17	1516	1550
37	1539	17	1522	1556
38	1545	18	1527	1563
39	1551	18	1533	1569
40	1558	19	1539	1577
41	1565	19	1546	1584
42	1572	20	1552	1592

**Table K-4 (continued). Literature Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
43	1580	21	1559	1601
44	1589	21	1568	1610
45	1599	23	1576	1622
46	1610	24	1586	1634
47	1622	26	1596	1648
48	1636	28	1608	1664
49	1654	32	1622	1686
50	1678	38	1640	1716
51	1716	52	1664	1768
52	1779	92	1687	1800

## SPRING

Table K-5. Algebra I Raw-to-Scaled Score Conversion Table

Raw	SS	CSEM	LCI	UCI
0	1217	92	1200	1309
1	1278	51	1227	1329
2	1315	37	1278	1352
3	1337	30	1307	1367
4	1353	27	1326	1380
5	1366	24	1342	1390
6	1377	22	1355	1399
7	1386	21	1365	1407
8	1395	20	1375	1415
9	1402	19	1383	1421
10	1409	18	1391	1427
11	1416	18	1398	1434
12	1422	17	1405	1439
13	1428	17	1411	1445
14	1433	16	1417	1449
15	1439	16	1423	1455
16	1444	16	1428	1460
17	1449	15	1434	1464
18	1453	15	1438	1468
19	1458	15	1443	1473
20	1462	15	1447	1477
21	1467	15	1452	1482
22	1471	15	1456	1486
23	1475	14	1461	1489
24	1479	14	1465	1493
25	1483	14	1469	1497
26	1487	14	1473	1501
27	1491	14	1477	1505
28	1495	14	1481	1509
29	1499	14	1485	1513
30	1503	14	1489	1517
31	1507	14	1493	1521
32	1511	14	1497	1525
33	1515	14	1501	1529
34	1519	14	1505	1533
35	1523	14	1509	1537
36	1527	14	1513	1541
37	1531	14	1517	1545
38	1535	14	1521	1549
39	1539	15	1524	1554
40	1543	15	1528	1558
41	1548	15	1533	1563

**Table K-5 (continued). Algebra I Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
42	1552	15	1537	1567
43	1557	15	1542	1572
44	1562	16	1546	1578
45	1567	16	1551	1583
46	1572	16	1556	1588
47	1578	17	1561	1595
48	1583	17	1566	1600
49	1589	18	1571	1607
50	1596	18	1578	1614
51	1603	19	1584	1622
52	1611	20	1591	1631
53	1619	21	1598	1640
54	1629	23	1606	1652
55	1640	24	1616	1664
56	1653	27	1626	1680
57	1669	30	1639	1699
58	1691	37	1654	1728
59	1728	51	1677	1779
60	1789	92	1697	1800

**Table K-6. Biology Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
0	1220	92	1200	1312
1	1281	51	1230	1332
2	1317	36	1281	1353
3	1338	30	1308	1368
4	1354	26	1328	1380
5	1366	24	1342	1390
6	1376	22	1354	1398
7	1385	20	1365	1405
8	1393	19	1374	1412
9	1400	18	1382	1418
10	1406	18	1388	1424
11	1412	17	1395	1429
12	1417	16	1401	1433
13	1422	16	1406	1438
14	1427	15	1412	1442
15	1432	15	1417	1447
16	1436	15	1421	1451
17	1441	14	1427	1455
18	1445	14	1431	1459
19	1449	14	1435	1463
20	1452	14	1438	1466
21	1456	14	1442	1470
22	1460	13	1447	1473
23	1463	13	1450	1476
24	1467	13	1454	1480
25	1470	13	1457	1483
26	1473	13	1460	1486
27	1477	13	1464	1490
28	1480	13	1467	1493
29	1483	13	1470	1496
30	1486	13	1473	1499
31	1489	12	1477	1501
32	1492	12	1480	1504
33	1496	12	1484	1508
34	1499	12	1487	1511
35	1502	12	1490	1514
36	1505	12	1493	1517
37	1508	12	1496	1520
38	1511	12	1499	1523
39	1514	12	1502	1526
40	1517	13	1504	1530
41	1520	13	1507	1533
42	1523	13	1510	1536

**Table K-6 (continued). Biology Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
43	1527	13	1514	1540
44	1530	13	1517	1543
45	1533	13	1520	1546
46	1537	13	1524	1550
47	1540	13	1527	1553
48	1544	14	1530	1558
49	1548	14	1534	1562
50	1552	14	1538	1566
51	1556	14	1542	1570
52	1560	15	1545	1575
53	1565	15	1550	1580
54	1570	16	1554	1586
55	1575	16	1559	1591
56	1580	17	1563	1597
57	1586	18	1568	1604
58	1593	19	1574	1612
59	1600	20	1580	1620
60	1608	21	1587	1629
61	1618	23	1595	1641
62	1630	26	1604	1656
63	1645	29	1616	1674
64	1666	36	1630	1702
65	1701	50	1651	1751
66	1762	91	1671	1800

**Table K-7. Literature Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
0	1202	92	1200	1294
1	1263	51	1212	1314
2	1299	37	1262	1336
3	1321	30	1291	1351
4	1338	27	1311	1365
5	1351	24	1327	1375
6	1361	23	1338	1384
7	1371	21	1350	1392
8	1380	20	1360	1400
9	1387	19	1368	1406
10	1395	19	1376	1414
11	1401	18	1383	1419
12	1408	18	1390	1426
13	1414	17	1397	1431
14	1419	17	1402	1436
15	1425	17	1408	1442
16	1430	16	1414	1446
17	1435	16	1419	1451
18	1441	16	1425	1457
19	1446	16	1430	1462
20	1450	16	1434	1466
21	1455	16	1439	1471
22	1460	15	1445	1475
23	1465	15	1450	1480
24	1470	15	1455	1485
25	1474	15	1459	1489
26	1479	15	1464	1494
27	1484	15	1469	1499
28	1489	16	1473	1505
29	1494	16	1478	1510
30	1498	16	1482	1514
31	1503	16	1487	1519
32	1508	16	1492	1524
33	1514	16	1498	1530
34	1519	16	1503	1535
35	1524	17	1507	1541
36	1530	17	1513	1547
37	1536	17	1519	1553
38	1542	18	1524	1560
39	1548	18	1530	1566
40	1555	18	1537	1573
41	1562	19	1543	1581
42	1569	20	1549	1589



**Table K-7 (continued). Literature Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
43	1577	20	1557	1597
44	1586	21	1565	1607
45	1595	22	1573	1617
46	1606	24	1582	1630
47	1618	25	1593	1643
48	1632	28	1604	1660
49	1650	32	1618	1682
50	1673	38	1635	1711
51	1711	52	1659	1763
52	1774	92	1682	1800

## SUMMER

**Table K-8. Algebra Raw-to-Scaled Score Conversion Table**

Raw	SS	CSEM	LCI	UCI
0	1217	92	1200	1309
1	1278	51	1227	1329
2	1314	36	1278	1350
3	1336	30	1306	1366
4	1352	26	1326	1378
5	1365	24	1341	1389
6	1375	22	1353	1397
7	1385	21	1364	1406
8	1393	20	1373	1413
9	1401	19	1382	1420
10	1407	18	1389	1425
11	1414	18	1396	1432
12	1420	17	1403	1437
13	1426	17	1409	1443
14	1431	16	1415	1447
15	1436	16	1420	1452
16	1441	16	1425	1457
17	1446	15	1431	1461
18	1451	15	1436	1466
19	1455	15	1440	1470
20	1460	15	1445	1475
21	1464	15	1449	1479
22	1468	14	1454	1482
23	1473	14	1459	1487
24	1477	14	1463	1491
25	1481	14	1467	1495
26	1485	14	1471	1499
27	1488	14	1474	1502
28	1492	14	1478	1506
29	1496	14	1482	1510
30	1500	14	1486	1514
31	1504	14	1490	1518
32	1507	14	1493	1521
33	1511	14	1497	1525
34	1515	14	1501	1529
35	1519	14	1505	1533
36	1522	14	1508	1536
37	1526	14	1512	1540
38	1530	14	1516	1544
39	1534	14	1520	1548
40	1538	14	1524	1552
41	1542	15	1527	1557

**Table K-8 (continued). Algebra Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
42	1547	15	1532	1562
43	1551	15	1536	1566
44	1556	15	1541	1571
45	1560	16	1544	1576
46	1565	16	1549	1581
47	1571	16	1555	1587
48	1576	17	1559	1593
49	1582	17	1565	1599
50	1588	18	1570	1606
51	1595	19	1576	1614
52	1602	20	1582	1622
53	1611	21	1590	1632
54	1620	22	1598	1642
55	1630	24	1606	1654
56	1643	27	1616	1670
57	1659	30	1629	1689
58	1681	37	1644	1718
59	1718	51	1667	1769
60	1779	92	1687	1800

**Table K-9. Biology Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
0	1228	92	1200	1320
1	1289	50	1239	1339
2	1325	36	1289	1361
3	1346	30	1316	1376
4	1361	26	1335	1387
5	1373	23	1350	1396
6	1383	22	1361	1405
7	1392	20	1372	1412
8	1399	19	1380	1418
9	1406	18	1388	1424
10	1412	17	1395	1429
11	1418	17	1401	1435
12	1423	16	1407	1439
13	1428	16	1412	1444
14	1433	15	1418	1448
15	1438	15	1423	1453
16	1442	14	1428	1456
17	1446	14	1432	1460
18	1450	14	1436	1464
19	1454	14	1440	1468
20	1457	13	1444	1470
21	1461	13	1448	1474
22	1464	13	1451	1477
23	1468	13	1455	1481
24	1471	13	1458	1484
25	1474	13	1461	1487
26	1477	13	1464	1490
27	1481	13	1468	1494
28	1484	12	1472	1496
29	1487	12	1475	1499
30	1490	12	1478	1502
31	1493	12	1481	1505
32	1496	12	1484	1508
33	1499	12	1487	1511
34	1502	12	1490	1514
35	1505	12	1493	1517
36	1508	12	1496	1520
37	1511	12	1499	1523
38	1514	12	1502	1526
39	1517	12	1505	1529
40	1521	13	1508	1534
41	1524	13	1511	1537
42	1527	13	1514	1540

**Table K-9 (continued). Biology Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
43	1530	13	1517	1543
44	1534	13	1521	1547
45	1537	13	1524	1550
46	1540	13	1527	1553
47	1544	14	1530	1558
48	1548	14	1534	1562
49	1552	14	1538	1566
50	1556	14	1542	1570
51	1560	15	1545	1575
52	1564	15	1549	1579
53	1569	15	1554	1584
54	1574	16	1558	1590
55	1579	16	1563	1595
56	1584	17	1567	1601
57	1591	18	1573	1609
58	1597	19	1578	1616
59	1605	20	1585	1625
60	1613	21	1592	1634
61	1623	23	1600	1646
62	1635	26	1609	1661
63	1650	30	1620	1680
64	1671	36	1635	1707
65	1707	50	1657	1757
66	1767	92	1675	1800

**Table K-10. Literature Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
0	1209	92	1200	1301
1	1271	51	1220	1322
2	1307	36	1271	1343
3	1328	30	1298	1358
4	1344	26	1318	1370
5	1357	24	1333	1381
6	1368	22	1346	1390
7	1377	21	1356	1398
8	1385	20	1365	1405
9	1393	19	1374	1412
10	1399	18	1381	1417
11	1406	18	1388	1424
12	1412	17	1395	1429
13	1418	17	1401	1435
14	1423	16	1407	1439
15	1428	16	1412	1444
16	1433	16	1417	1449
17	1438	16	1422	1454
18	1443	15	1428	1458
19	1448	15	1433	1463
20	1453	15	1438	1468
21	1457	15	1442	1472
22	1462	15	1447	1477
23	1466	15	1451	1481
24	1471	15	1456	1486
25	1475	15	1460	1490
26	1480	15	1465	1495
27	1484	15	1469	1499
28	1489	15	1474	1504
29	1493	15	1478	1508
30	1498	15	1483	1513
31	1503	15	1488	1518
32	1508	16	1492	1524
33	1513	16	1497	1529
34	1518	16	1502	1534
35	1523	16	1507	1539
36	1528	17	1511	1545
37	1534	17	1517	1551
38	1540	17	1523	1557
39	1546	18	1528	1564
40	1552	18	1534	1570
41	1559	19	1540	1578
42	1567	19	1548	1586

**Table K-10 (continued). Literature Raw-to-Scaled Score Conversion Table**

<b>Raw</b>	<b>SS</b>	<b>CSEM</b>	<b>LCI</b>	<b>UCI</b>
43	1574	20	1554	1594
44	1583	21	1562	1604
45	1592	22	1570	1614
46	1603	23	1580	1626
47	1614	25	1589	1639
48	1628	28	1600	1656
49	1645	31	1614	1676
50	1669	37	1632	1706
51	1706	51	1655	1757
52	1768	92	1676	1800

# APPENDIX L: PRE-EQUATING VERIFICATION RESULTS

Appendix L contains additional information and results based on the data used for the pre-equating verification. Results are presented for the fully-anchored pre-equating solution (hereinafter “pre-equating”) and the partially anchored pre-equating solution when misfitting items were freely calibrated (hereinafter “post-equating”). The results presented in this appendix provide support for utilizing the pre-equated solution for all student reporting. A complete description of the pre-equating verification process is discussed in Chapter Fifteen.

Appendix L includes person fit summary box plots, normalized scale score difference plots, and raw to scale score table comparisons generated during the pre-equating verification process. Figures L–1, L–3, and L–5 present the person infit boxplots for the pre-equated (left) and post-equated (right) solutions by content area for winter, spring, and summer, respectively. Results are presented for each of the following subgroups: race/ethnicity, gender, IEP, and EL status. Figures L–2, L–4, and L–6 present the normalized scale score difference plots that show this difference between the pre-equated and post-equated solutions by content area for winter, spring, and summer, respectively. The horizontal reference line represents no differences in normalized scaled scores between the pre-equated and post-equated solutions. The vertical dashed lines depict the raw cut-scores based on the pre-equated solution.

Tables L–1 through L–9 show both the pre-equated and post-equated raw-to-scaled score conversion tables. Tables provide details for both pre-equated (Pre) and post-equated (Post) solutions with respect to the scaled score (SS), standard error of measurement (SEM), performance level (PL), overall proportion of students earning each raw score, and whether there were differences in the performance level classification for pre- and post-equated solutions (Same PL). All student reporting is based on the tables in Appendix K, which use the pre-equated solutions.

Results are presented for each administration separately.

## WINTER

Figure L–1. Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

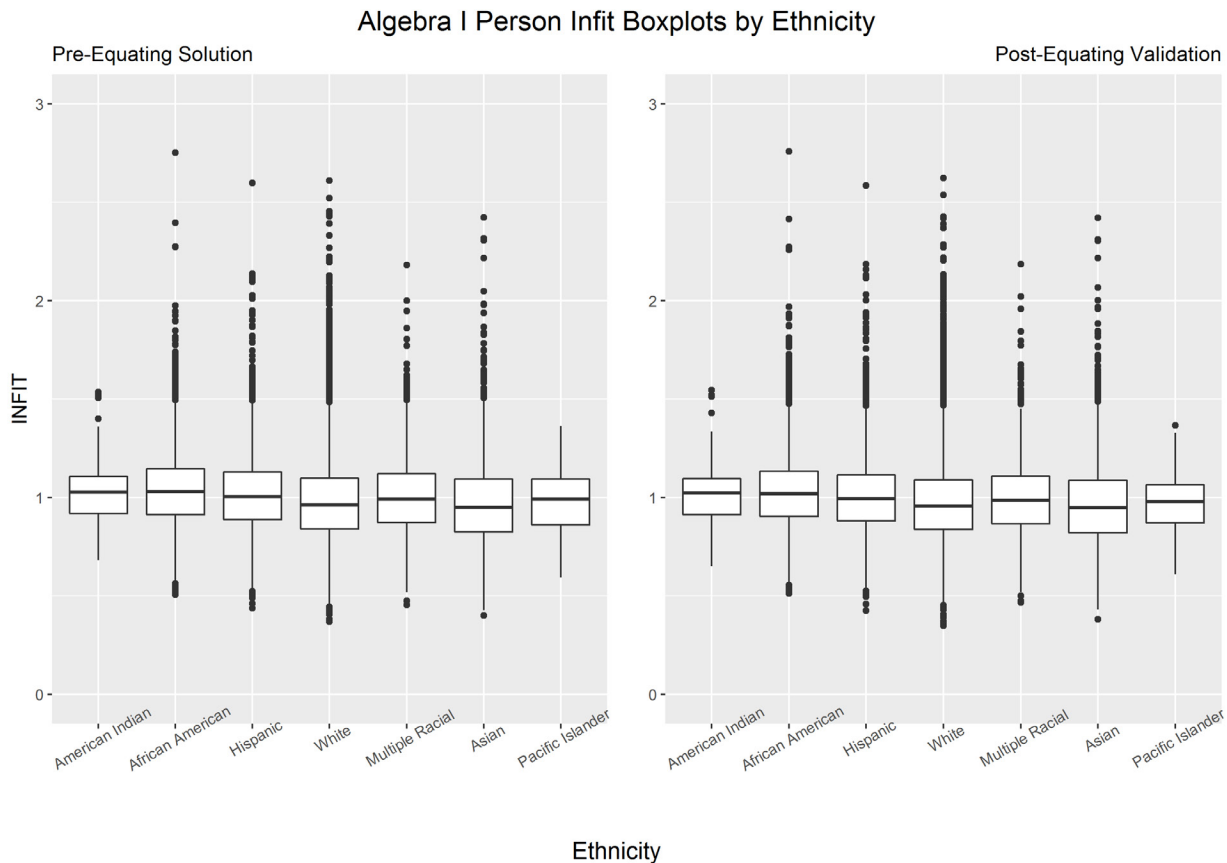




Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

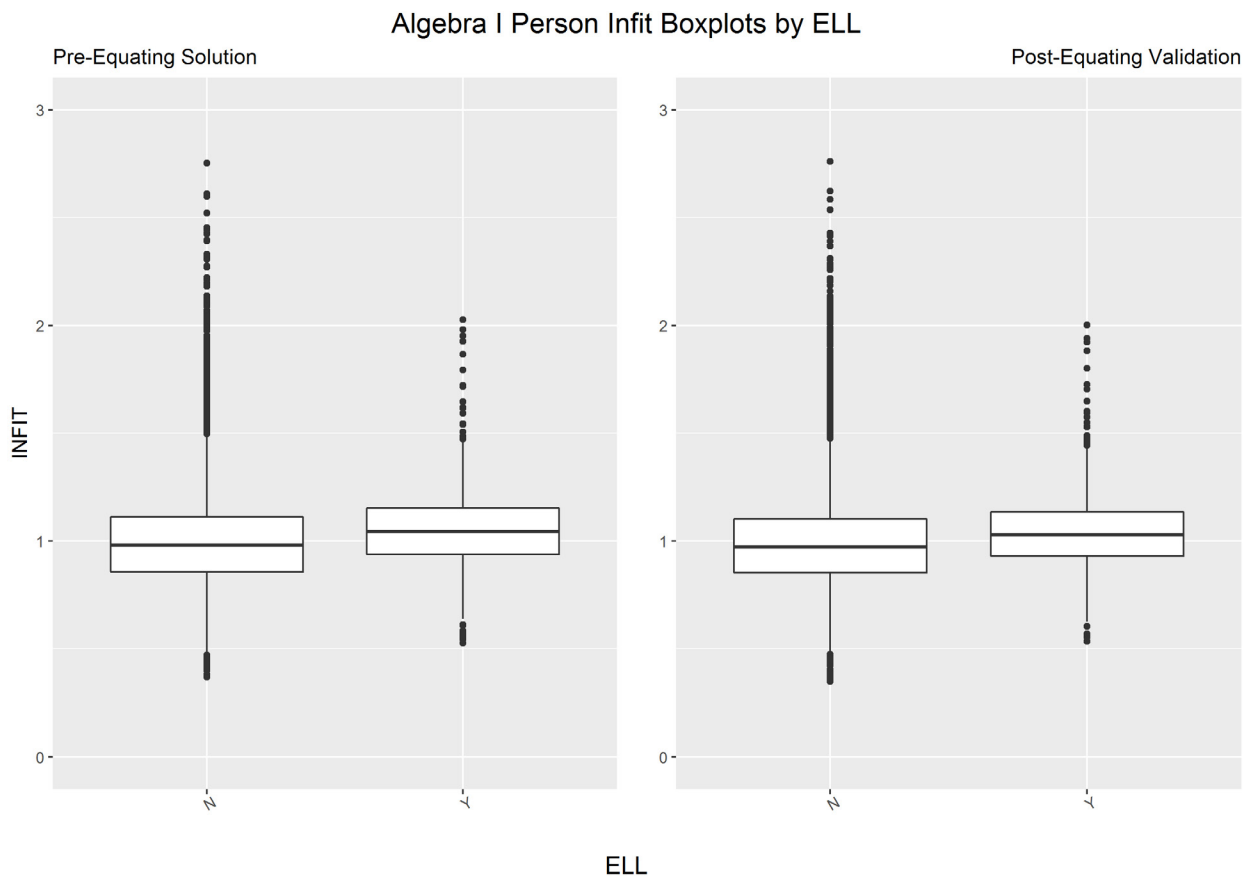
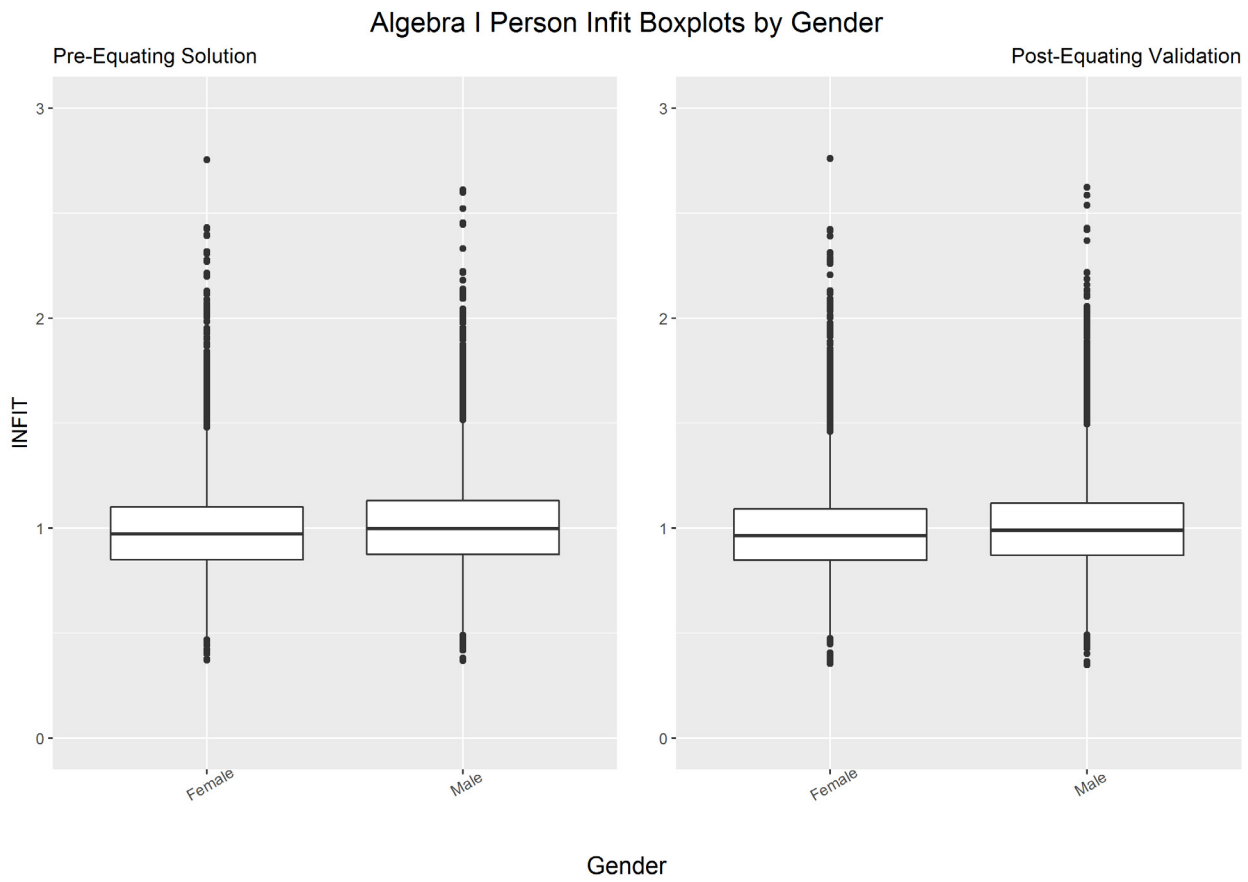
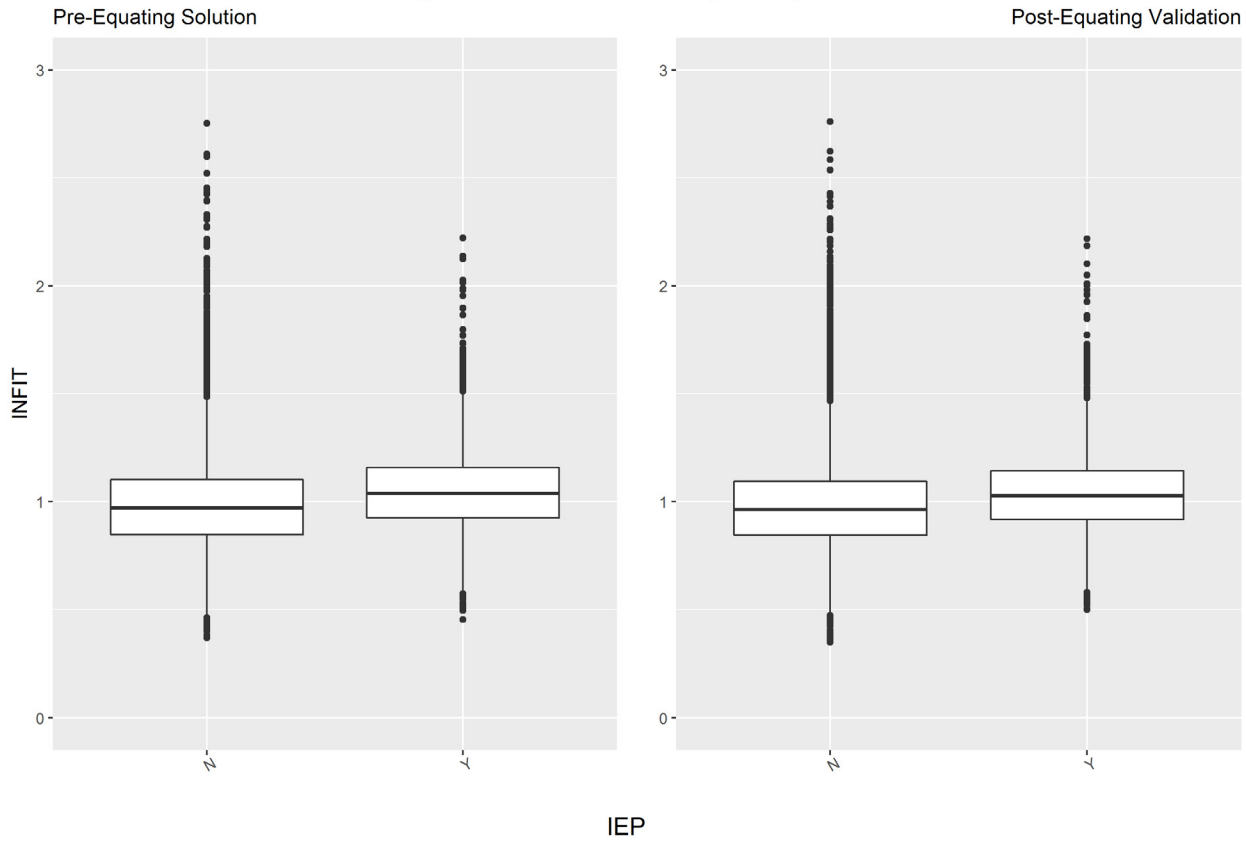
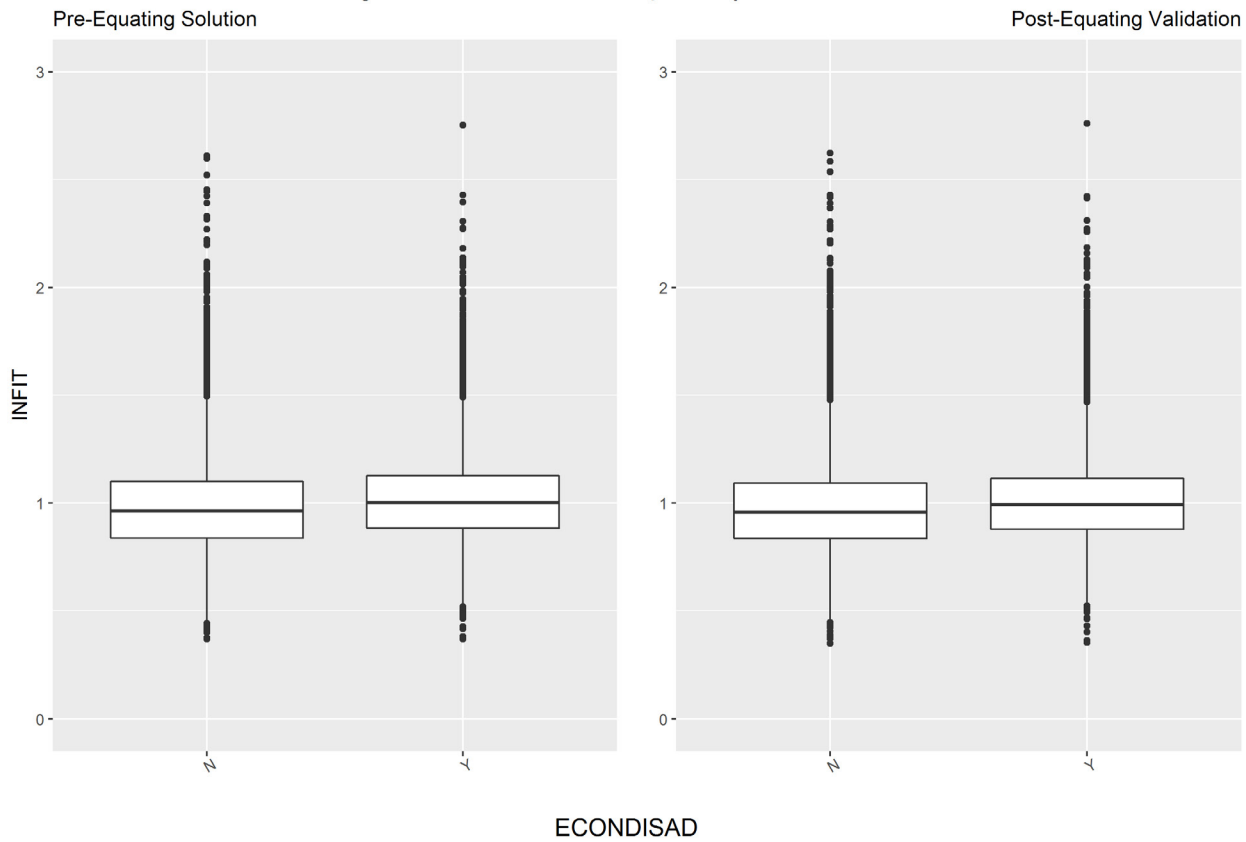


Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

Algebra I Person Infit Boxplots by IEP

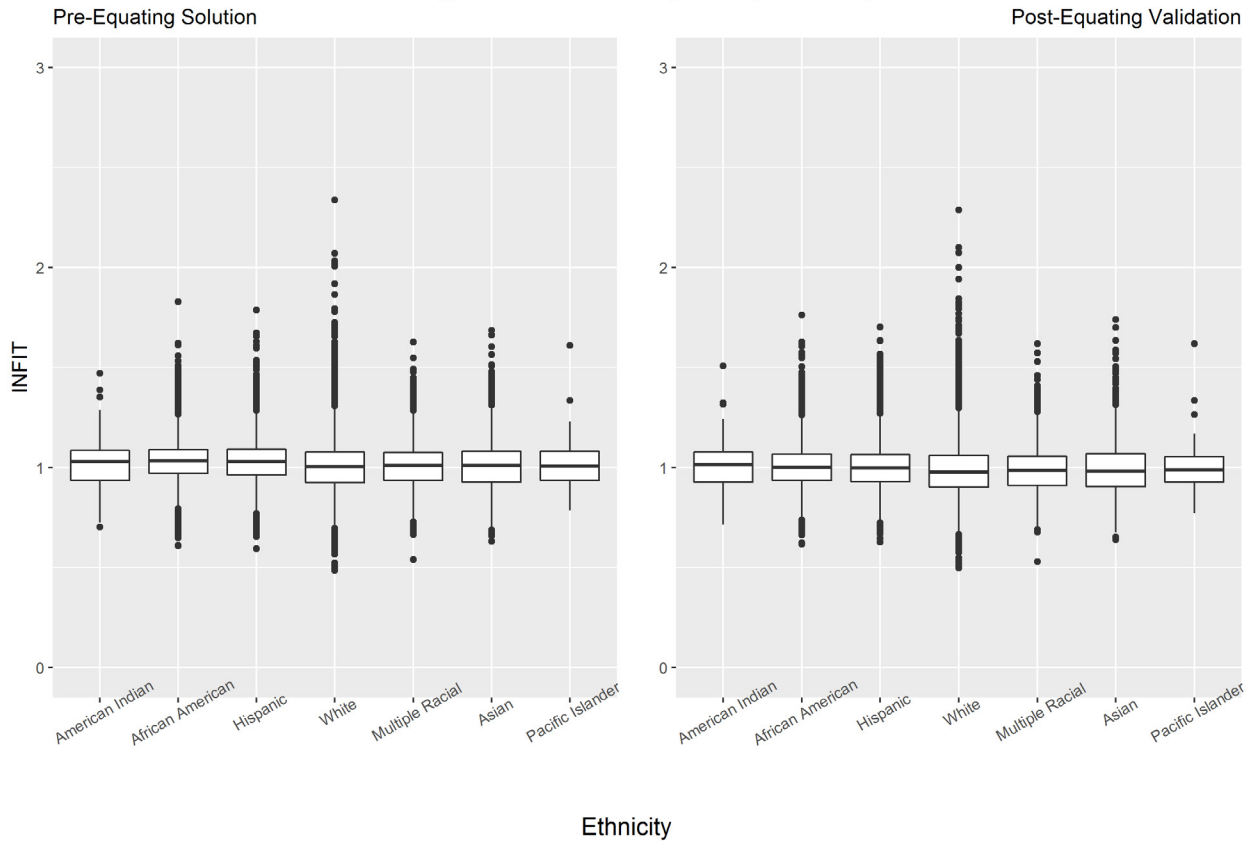


Algebra I Person Infit Boxplots by ECONDISAD



**Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter**

**Biology Person Infit Boxplots by Ethnicity**



**Biology Person Infit Boxplots by Gender**

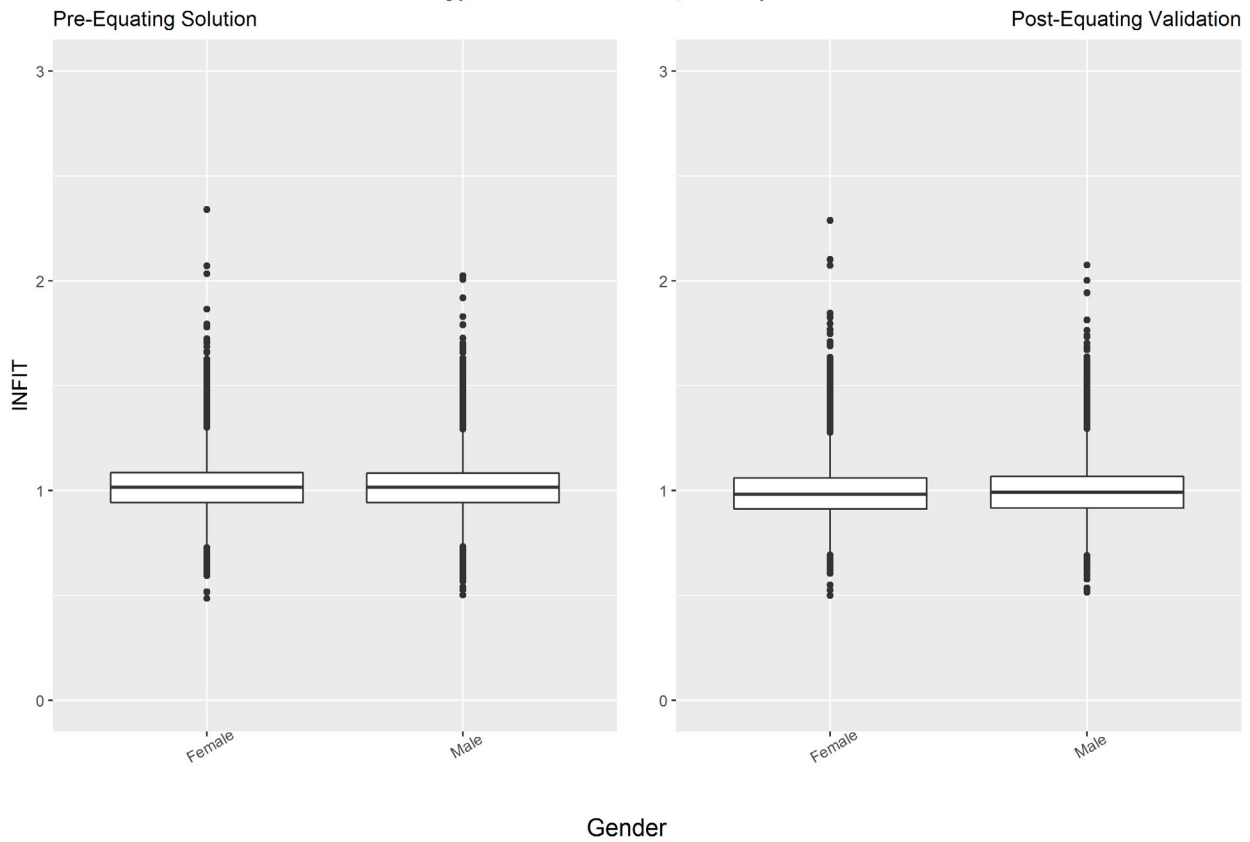
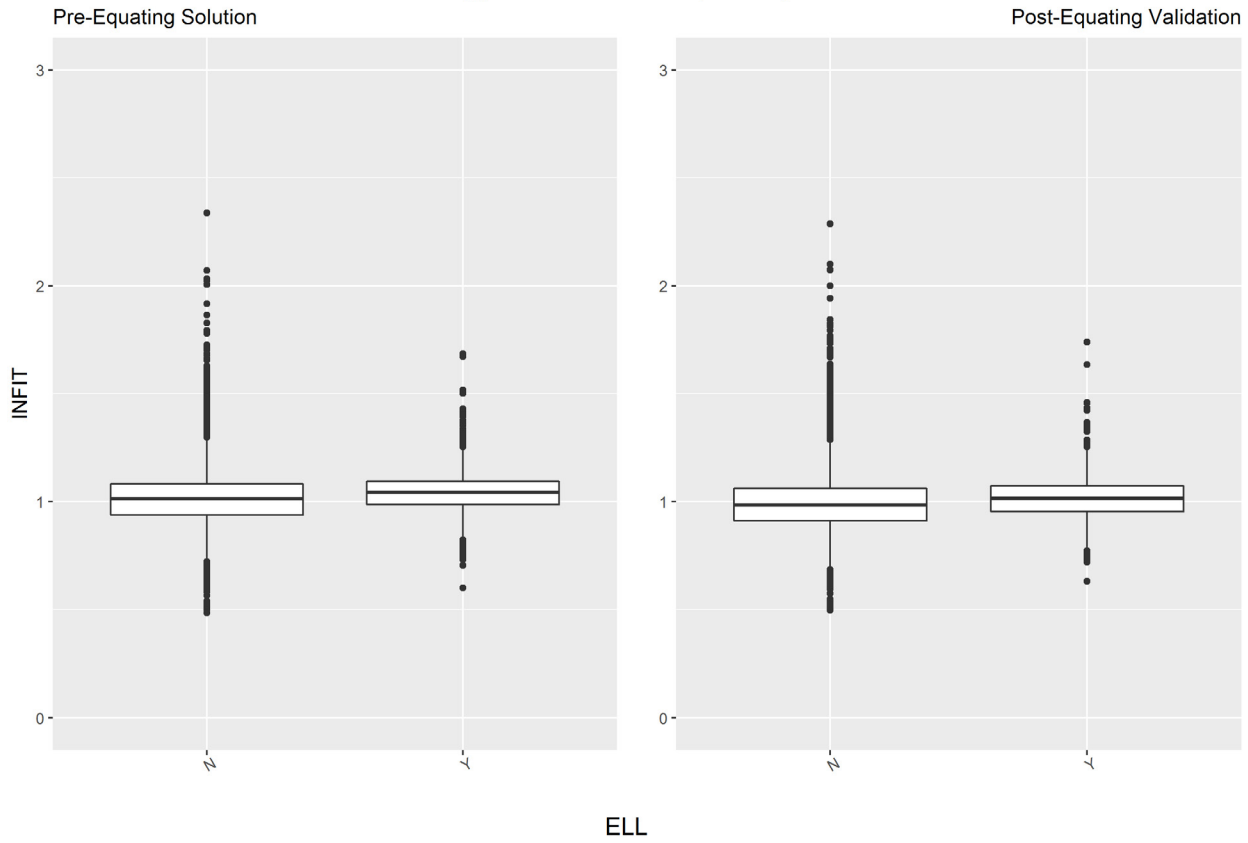


Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

Biology Person Infit Boxplots by ELL



Biology Person Infit Boxplots by IEP

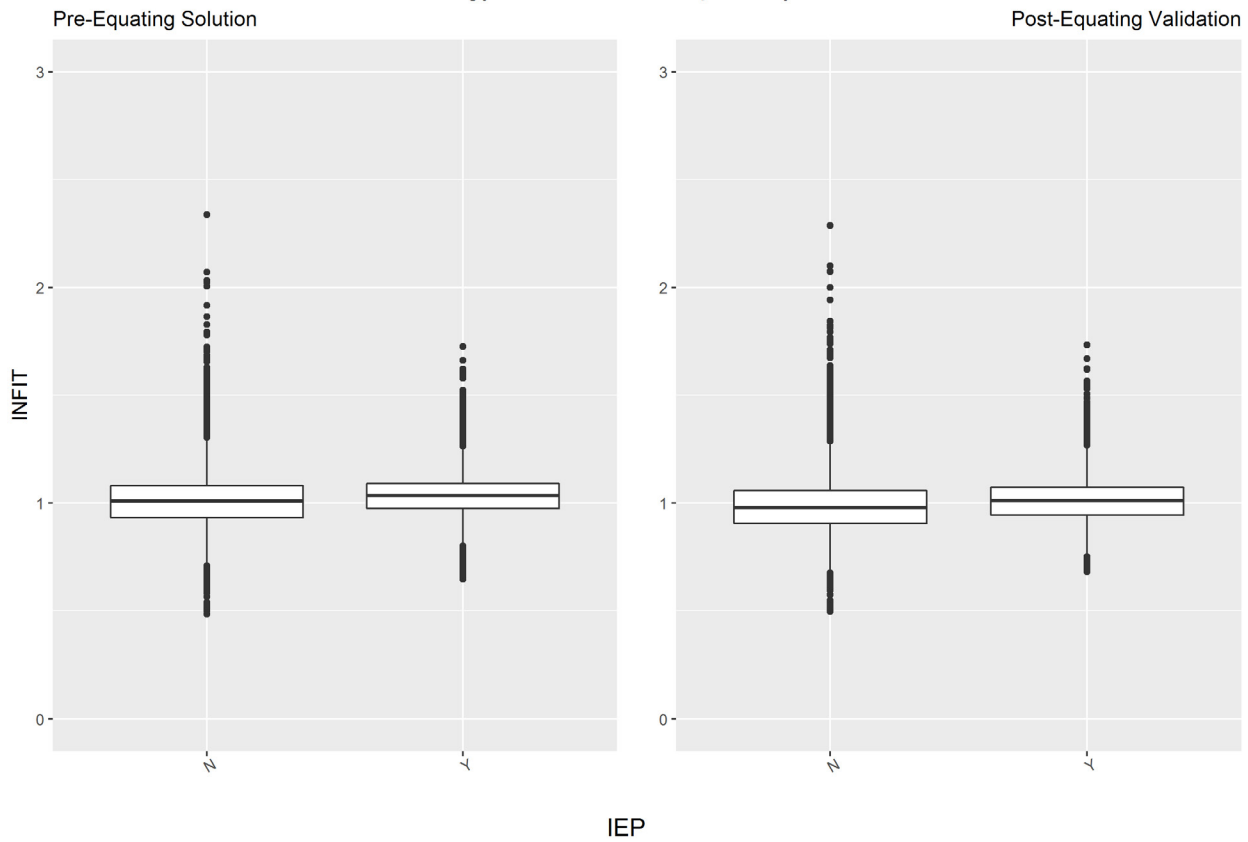


Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

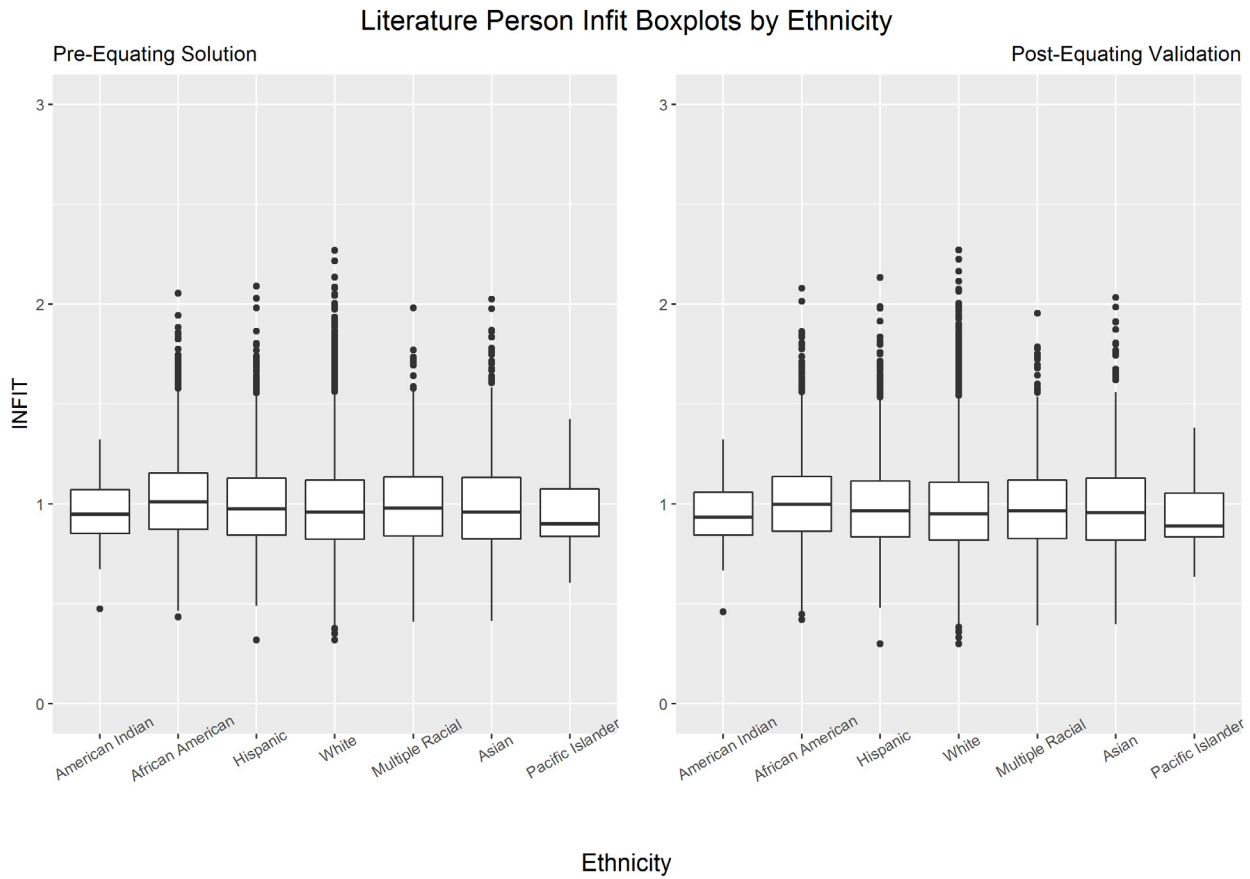
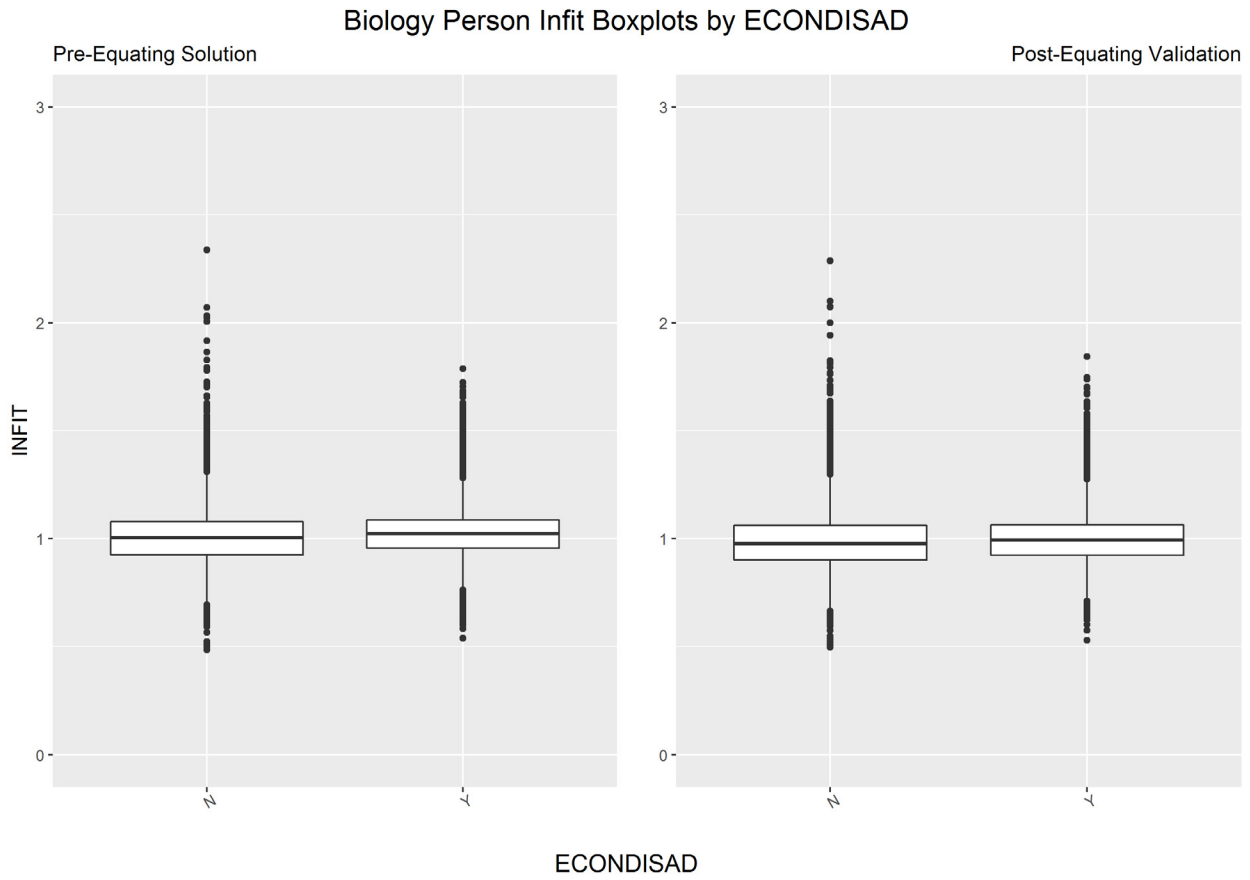


Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

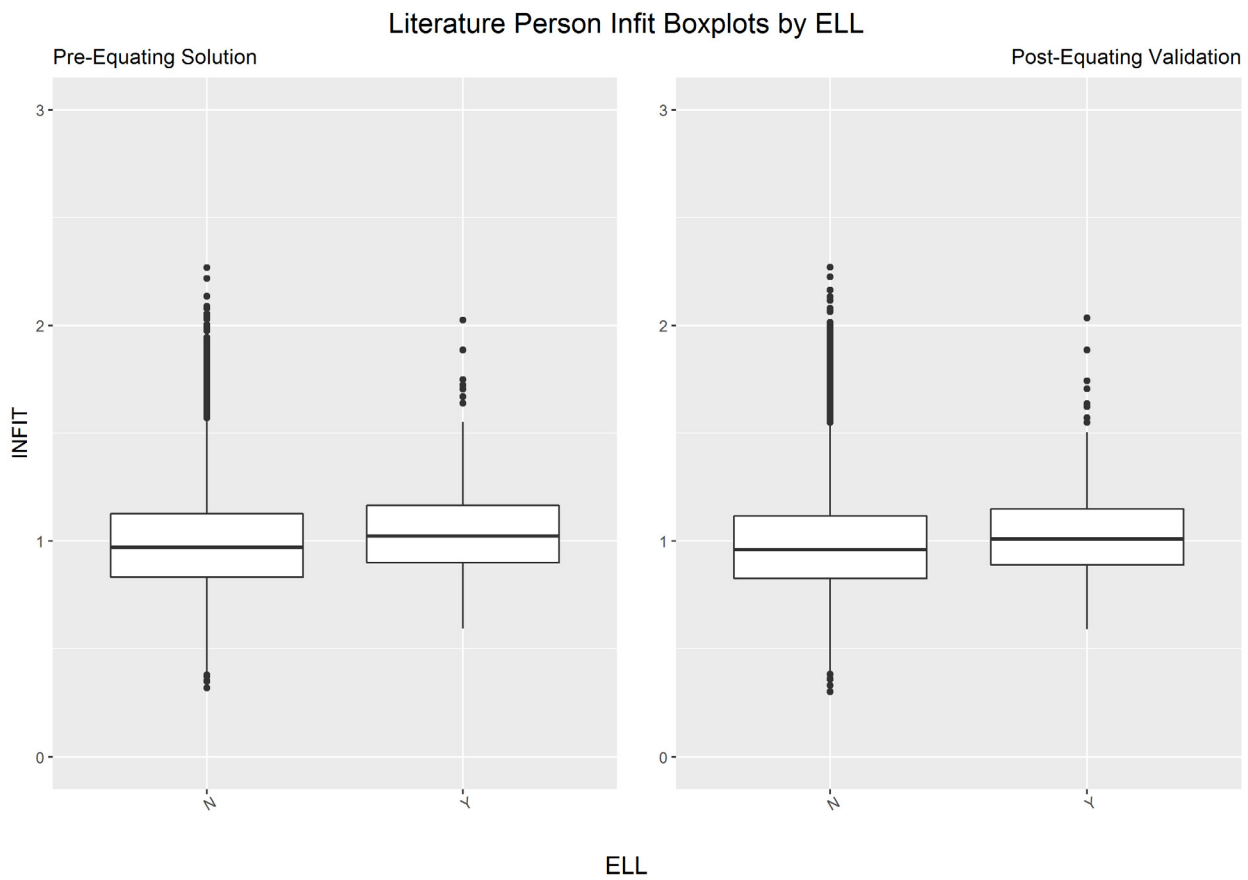
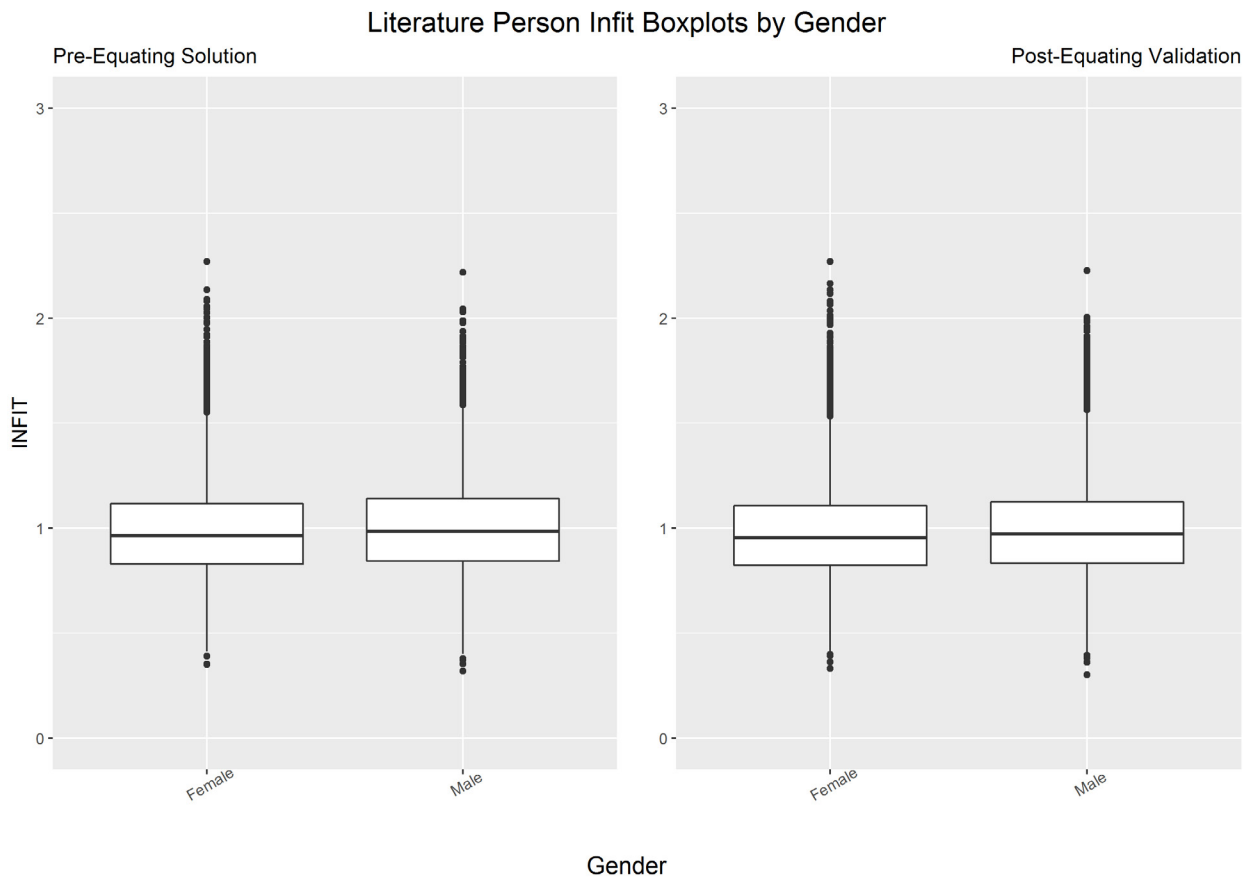


Figure L-1 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Winter

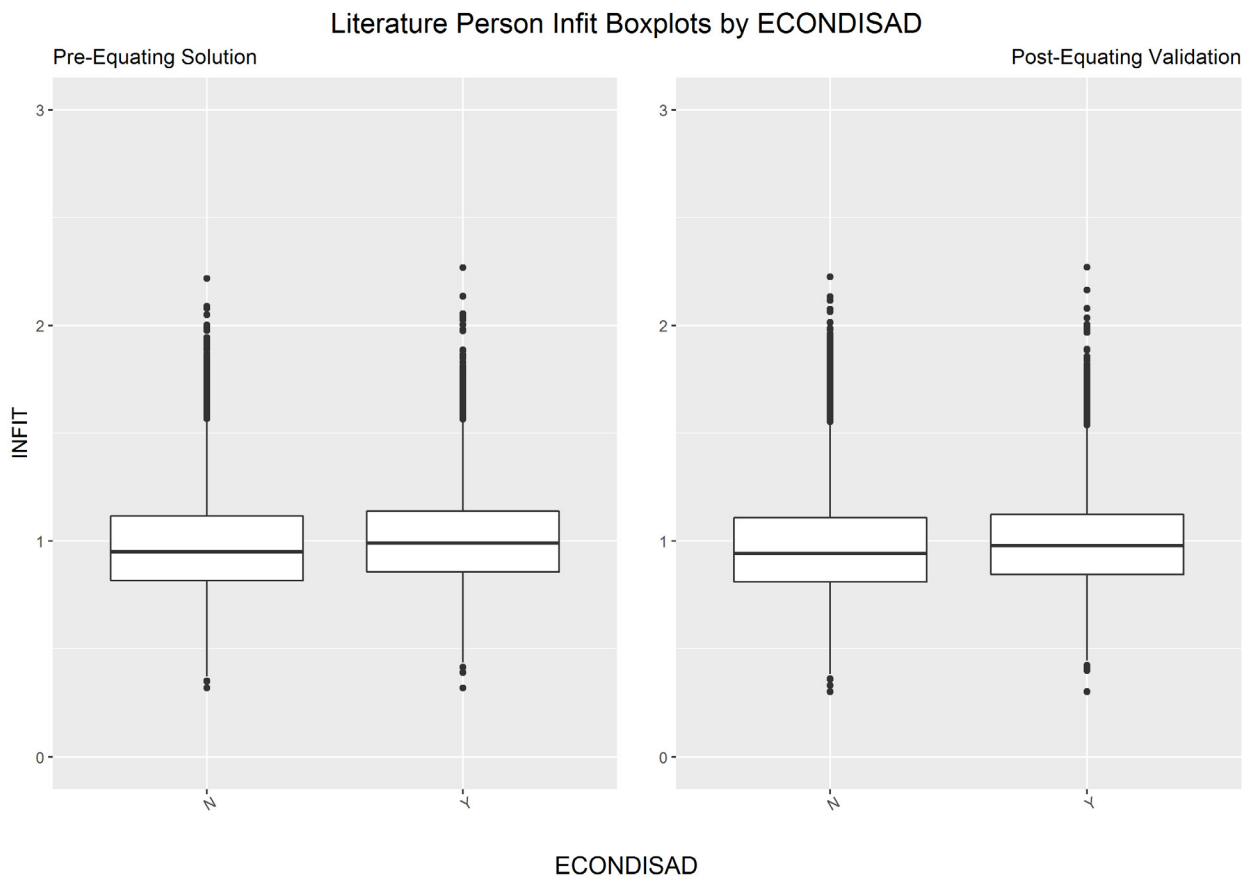
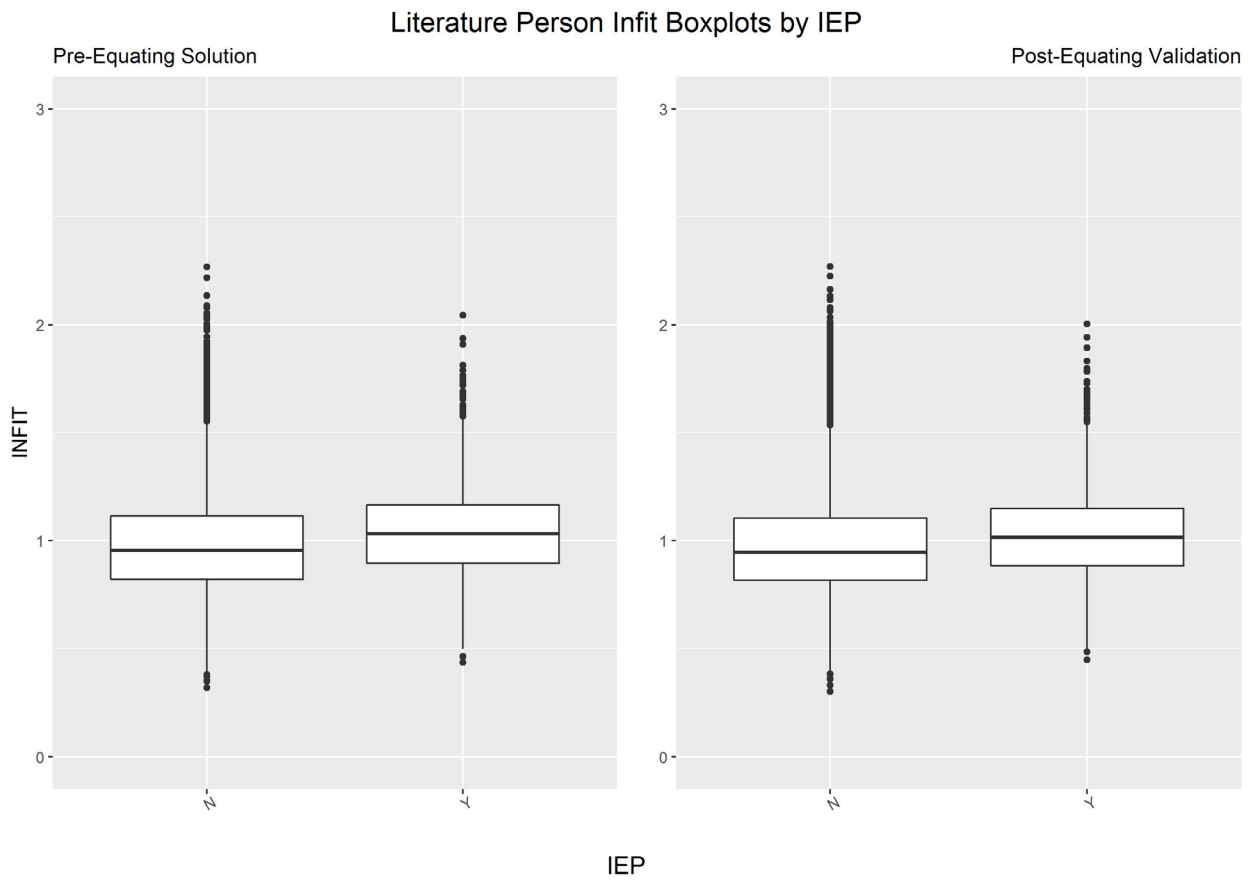
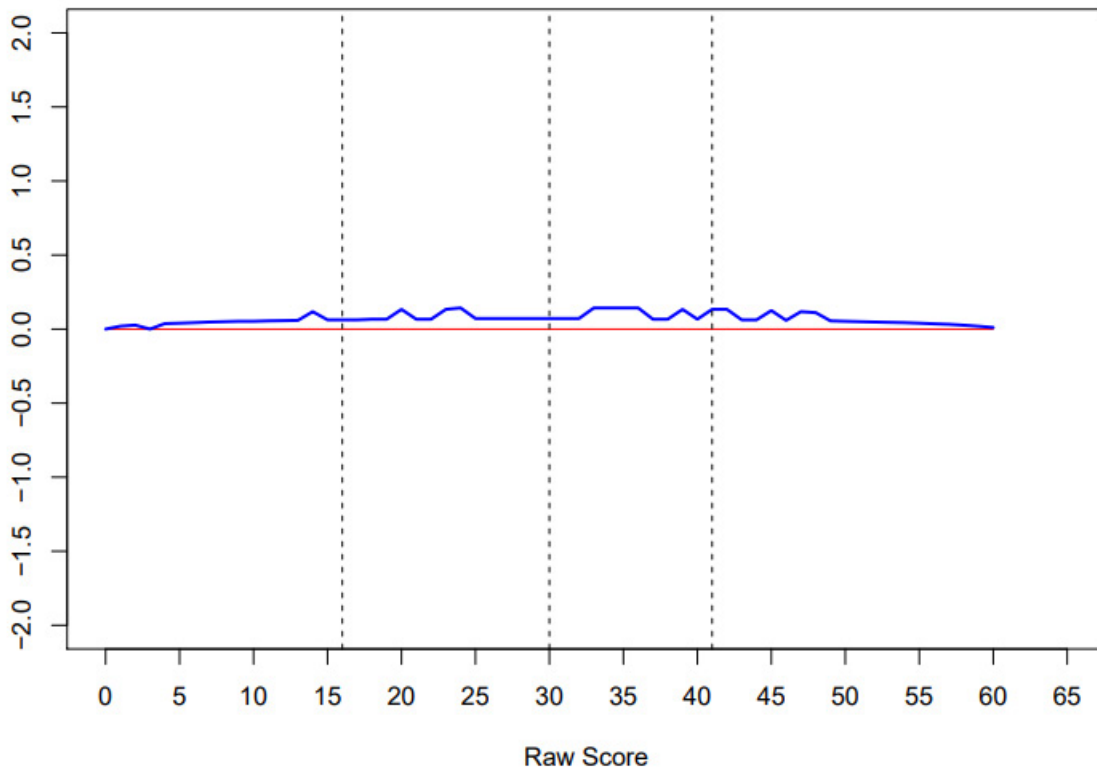


Figure L-2. Normalized Scale Score Differences by Content Area: Winter  
Algebra I



Biology

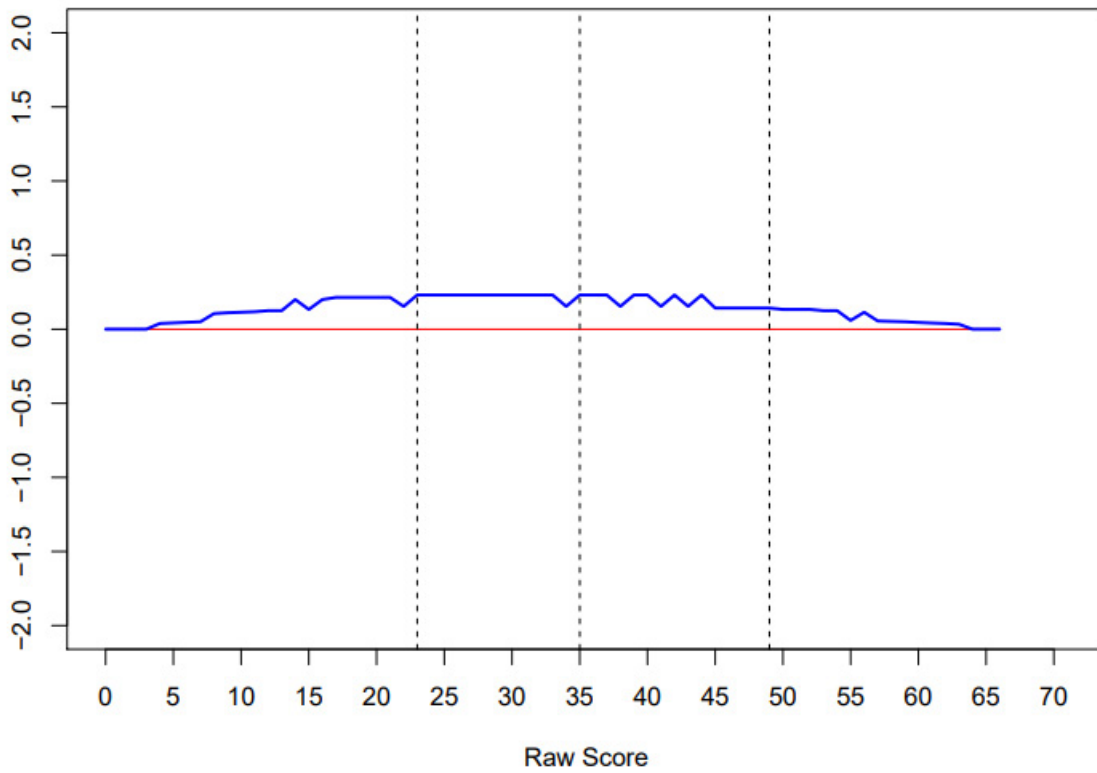
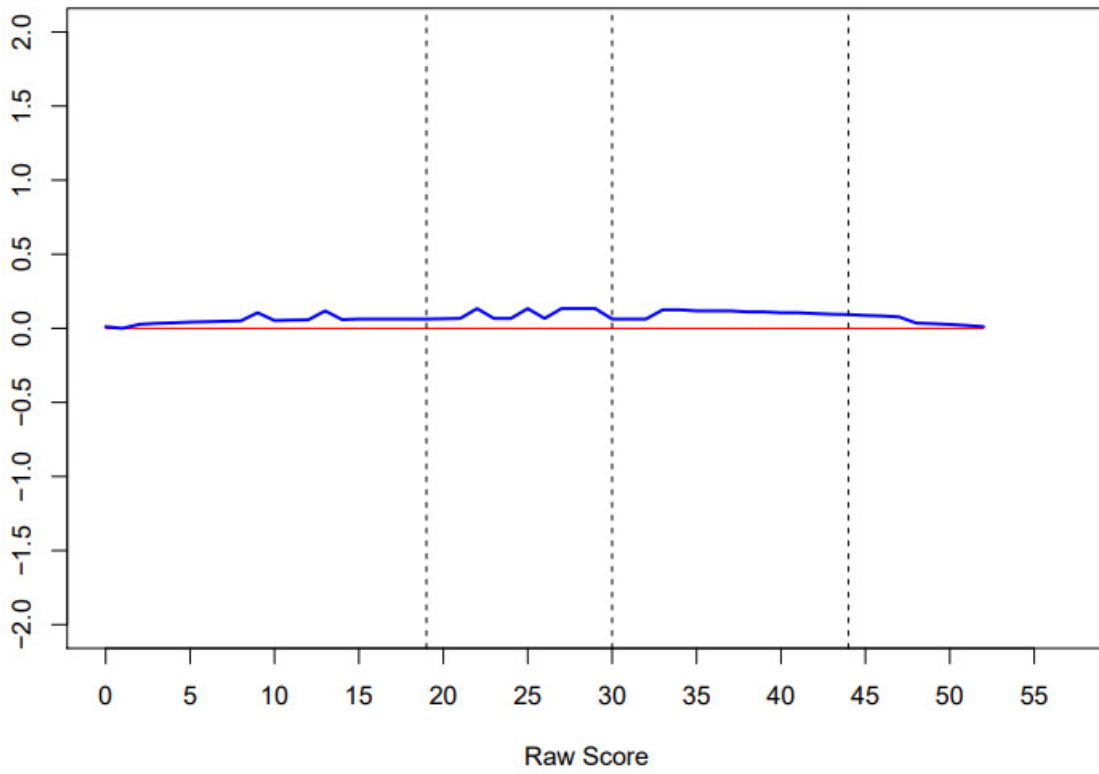




Figure L-2 (continued). Normalized Scale Score Differences by Content Area: Winter

Literature



**Table L-1. Winter Algebra I Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1209	1209	92	92	BB	BB	0.0	TRUE
1	1271	1270	51	51	BB	BB	0.0	TRUE
2	1308	1307	37	37	BB	BB	0.0	TRUE
3	1330	1330	31	31	BB	BB	0.1	TRUE
4	1347	1346	27	27	BB	BB	0.3	TRUE
5	1360	1359	25	25	BB	BB	0.7	TRUE
6	1371	1370	23	23	BB	BB	1.1	TRUE
7	1381	1380	21	21	BB	BB	1.9	TRUE
8	1390	1389	20	20	BB	BB	2.7	TRUE
9	1398	1397	19	19	BB	BB	3.5	TRUE
10	1405	1404	19	19	BB	BB	3.9	TRUE
11	1412	1411	18	18	BB	BB	4.5	TRUE
12	1418	1417	18	17	BB	BB	4.9	TRUE
13	1424	1423	17	17	BB	BB	4.8	TRUE
14	1430	1428	17	17	BB	BB	4.4	TRUE
15	1435	1434	16	16	BB	BB	4.5	TRUE
16	1440	1439	16	16	B	B	4.3	TRUE
17	1445	1444	16	16	B	B	4.2	TRUE
18	1450	1449	15	15	B	B	4.1	TRUE
19	1455	1454	15	15	B	B	3.7	TRUE
20	1460	1458	15	15	B	B	3.6	TRUE
21	1464	1463	15	15	B	B	3.4	TRUE
22	1468	1467	15	15	B	B	3.4	TRUE
23	1473	1471	15	15	B	B	3.2	TRUE
24	1477	1475	14	14	B	B	3.0	TRUE
25	1481	1480	14	14	B	B	2.9	TRUE
26	1485	1484	14	14	B	B	2.6	TRUE
27	1489	1488	14	14	B	B	2.4	TRUE
28	1493	1492	14	14	B	B	2.4	TRUE
29	1497	1496	14	14	B	B	2.1	TRUE
30	1501	1500	14	14	P	P	1.9	TRUE
31	1505	1504	14	14	P	P	1.8	TRUE
32	1509	1508	14	14	P	P	1.5	TRUE
33	1514	1512	14	14	P	P	1.4	TRUE
34	1518	1516	14	14	P	P	1.2	TRUE
35	1522	1520	14	14	P	P	1.1	TRUE
36	1526	1524	14	14	P	P	1.0	TRUE
37	1530	1529	15	15	P	P	0.8	TRUE
38	1534	1533	15	15	P	P	0.8	TRUE
39	1539	1537	15	15	P	P	0.7	TRUE
40	1543	1542	15	15	P	P	0.6	TRUE
41	1548	1546	15	15	A	A	0.5	TRUE

**Table L-1 (continued). Winter Algebra I Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
42	1553	1551	15	15	A	A	0.5	TRUE
43	1557	1556	16	16	A	A	0.4	TRUE
44	1562	1561	16	16	A	A	0.4	TRUE
45	1568	1566	16	16	A	A	0.3	TRUE
46	1573	1572	17	17	A	A	0.3	TRUE
47	1579	1577	17	17	A	A	0.3	TRUE
48	1585	1583	18	18	A	A	0.2	TRUE
49	1591	1590	18	18	A	A	0.2	TRUE
50	1598	1597	19	19	A	A	0.2	TRUE
51	1605	1604	20	20	A	A	0.2	TRUE
52	1613	1612	21	21	A	A	0.2	TRUE
53	1622	1621	22	22	A	A	0.1	TRUE
54	1632	1631	23	23	A	A	0.1	TRUE
55	1644	1643	25	25	A	A	0.1	TRUE
56	1658	1657	28	28	A	A	0.1	TRUE
57	1675	1674	31	31	A	A	0.1	TRUE
58	1698	1697	38	38	A	A	0.0	TRUE
59	1736	1735	52	52	A	A	0.0	TRUE
60	1799	1798	92	92	A	A	0.0	TRUE

**Table L-2. Winter Biology Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1221	1221	92	92	BB	BB	0.0	TRUE
1	1281	1281	51	50	BB	BB	0.0	TRUE
2	1317	1317	36	36	BB	BB	0.0	TRUE
3	1338	1338	30	30	BB	BB	0.0	TRUE
4	1354	1353	26	26	BB	BB	0.0	TRUE
5	1366	1365	24	23	BB	BB	0.1	TRUE
6	1376	1375	22	21	BB	BB	0.2	TRUE
7	1385	1384	20	20	BB	BB	0.4	TRUE
8	1393	1391	19	19	BB	BB	0.8	TRUE
9	1400	1398	18	18	BB	BB	1.3	TRUE
10	1406	1404	18	17	BB	BB	1.7	TRUE
11	1412	1410	17	17	BB	BB	2.5	TRUE
12	1418	1416	16	16	BB	BB	3.2	TRUE
13	1423	1421	16	16	BB	BB	3.6	TRUE
14	1428	1425	15	15	BB	BB	4.0	TRUE
15	1432	1430	15	15	BB	BB	4.3	TRUE
16	1437	1434	15	15	BB	BB	4.4	TRUE
17	1441	1438	14	14	BB	BB	4.6	TRUE
18	1445	1442	14	14	BB	BB	4.5	TRUE
19	1449	1446	14	14	BB	BB	4.2	TRUE
20	1453	1450	14	14	BB	BB	4.2	TRUE
21	1457	1454	14	14	BB	BB	3.9	TRUE
22	1460	1458	13	13	B	BB	3.8	FALSE
23	1464	1461	13	13	B	B	3.6	TRUE
24	1468	1465	13	13	B	B	3.3	TRUE
25	1471	1468	13	13	B	B	3.2	TRUE
26	1474	1471	13	13	B	B	2.9	TRUE
27	1478	1475	13	13	B	B	2.8	TRUE
28	1481	1478	13	13	B	B	2.6	TRUE
29	1484	1481	13	13	B	B	2.4	TRUE
30	1488	1485	13	13	B	B	2.2	TRUE
31	1491	1488	13	13	B	B	2.0	TRUE
32	1494	1491	13	13	B	B	1.8	TRUE
33	1497	1494	13	13	B	B	1.5	TRUE
34	1500	1498	13	13	P	B	1.5	FALSE
35	1503	1502	13	12	P	P	2.0	TRUE
35	1504	1501	13	13	P	P	1.3	TRUE
36	1507	1504	13	13	P	P	1.2	TRUE
37	1510	1507	13	13	P	P	1.0	TRUE
38	1513	1511	13	13	P	P	1.0	TRUE
39	1517	1514	13	13	P	P	0.9	TRUE
40	1520	1517	13	13	P	P	0.8	TRUE

**Table L-2 (continued). Winter Biology Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
41	1523	1521	13	13	P	P	0.8	TRUE
42	1527	1524	13	13	P	P	0.8	TRUE
43	1530	1528	13	13	P	P	0.7	TRUE
44	1534	1531	13	13	P	P	0.7	TRUE
45	1537	1535	14	14	P	P	0.6	TRUE
46	1541	1539	14	14	P	P	0.6	TRUE
47	1545	1543	14	14	P	P	0.6	TRUE
48	1549	1547	14	14	A	P	0.6	FALSE
49	1553	1551	14	14	A	A	0.6	TRUE
50	1557	1555	15	15	A	A	0.6	TRUE
51	1561	1559	15	15	A	A	0.6	TRUE
52	1566	1564	15	15	A	A	0.6	TRUE
53	1571	1569	16	16	A	A	0.5	TRUE
54	1576	1574	16	16	A	A	0.5	TRUE
55	1581	1580	17	17	A	A	0.5	TRUE
56	1587	1585	17	18	A	A	0.5	TRUE
57	1593	1592	18	18	A	A	0.5	TRUE
58	1600	1599	19	19	A	A	0.4	TRUE
59	1608	1607	20	20	A	A	0.3	TRUE
60	1617	1616	22	22	A	A	0.2	TRUE
61	1627	1626	24	24	A	A	0.2	TRUE
62	1639	1638	26	26	A	A	0.2	TRUE
63	1655	1654	30	30	A	A	0.1	TRUE
64	1676	1676	36	36	A	A	0.1	TRUE
65	1712	1712	51	51	A	A	0.0	TRUE
66	1773	1773	92	92	A	A	0.0	TRUE

**Table L-3. Winter Literature Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1206	1205	92	92	BB	BB	0.0	TRUE
1	1267	1267	51	51	BB	BB	0.0	TRUE
2	1304	1303	37	37	BB	BB	0.0	TRUE
3	1326	1325	30	30	BB	BB	0.1	TRUE
4	1342	1341	27	27	BB	BB	0.2	TRUE
5	1355	1354	24	24	BB	BB	0.4	TRUE
6	1366	1365	23	22	BB	BB	0.8	TRUE
7	1375	1374	21	21	BB	BB	1.2	TRUE
8	1384	1383	20	20	BB	BB	1.6	TRUE
9	1392	1390	19	19	BB	BB	2.0	TRUE
10	1399	1398	19	19	BB	BB	2.2	TRUE
11	1405	1404	18	18	BB	BB	2.6	TRUE
12	1412	1411	18	17	BB	BB	2.9	TRUE
13	1418	1416	17	17	BB	BB	3.0	TRUE
14	1423	1422	17	17	BB	BB	3.0	TRUE
15	1429	1428	16	16	BB	BB	3.0	TRUE
16	1434	1433	16	16	BB	BB	2.9	TRUE
17	1439	1438	16	16	BB	BB	2.8	TRUE
18	1444	1443	16	16	B	BB	3.1	FALSE
19	1449	1448	16	16	B	B	3.2	TRUE
20	1454	1453	16	15	B	B	3.1	TRUE
21	1459	1458	15	15	B	B	3.1	TRUE
22	1464	1462	15	15	B	B	3.2	TRUE
23	1468	1467	15	15	B	B	3.4	TRUE
24	1473	1472	15	15	B	B	3.2	TRUE
25	1478	1476	15	15	B	B	3.4	TRUE
26	1482	1481	15	15	B	B	3.0	TRUE
27	1487	1485	15	15	B	B	3.1	TRUE
28	1492	1490	15	15	B	B	3.0	TRUE
29	1497	1495	15	15	B	B	2.8	TRUE
30	1501	1500	16	16	P	P	2.8	TRUE
31	1506	1505	16	16	P	P	2.8	TRUE
32	1511	1510	16	16	P	P	2.7	TRUE
33	1517	1515	16	16	P	P	2.4	TRUE
34	1522	1520	16	16	P	P	2.3	TRUE
35	1527	1525	17	17	P	P	2.1	TRUE
36	1533	1531	17	17	P	P	2.1	TRUE
37	1539	1537	17	17	P	P	2.2	TRUE
38	1545	1543	18	18	P	P	1.8	TRUE
39	1551	1549	18	18	P	P	1.8	TRUE
40	1558	1556	19	19	P	P	1.7	TRUE
41	1565	1563	19	19	P	P	1.6	TRUE

**Table L-3 (continued). Winter Literature Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
42	1572	1570	20	20	P	P	1.5	TRUE
43	1580	1578	21	21	P	P	1.3	TRUE
44	1589	1587	21	22	A	A	1.2	TRUE
45	1599	1597	23	23	A	A	1.0	TRUE
46	1610	1608	24	24	A	A	0.8	TRUE
47	1622	1620	26	26	A	A	0.6	TRUE
48	1636	1635	28	28	A	A	0.5	TRUE
49	1654	1653	32	32	A	A	0.3	TRUE
50	1678	1677	38	38	A	A	0.2	TRUE
51	1716	1715	52	52	A	A	0.1	TRUE
52	1779	1778	92	92	A	A	0.0	TRUE

# SPRING

Figure L-3. Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring

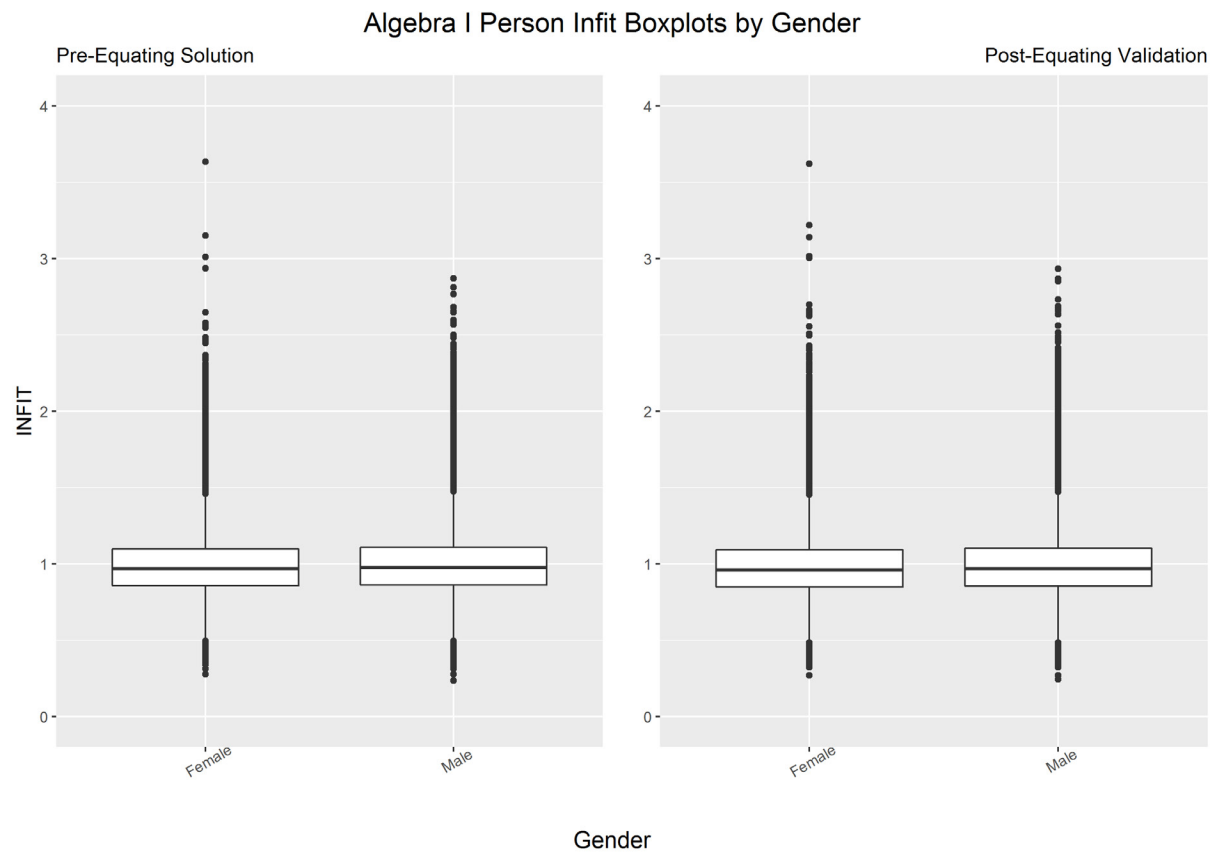
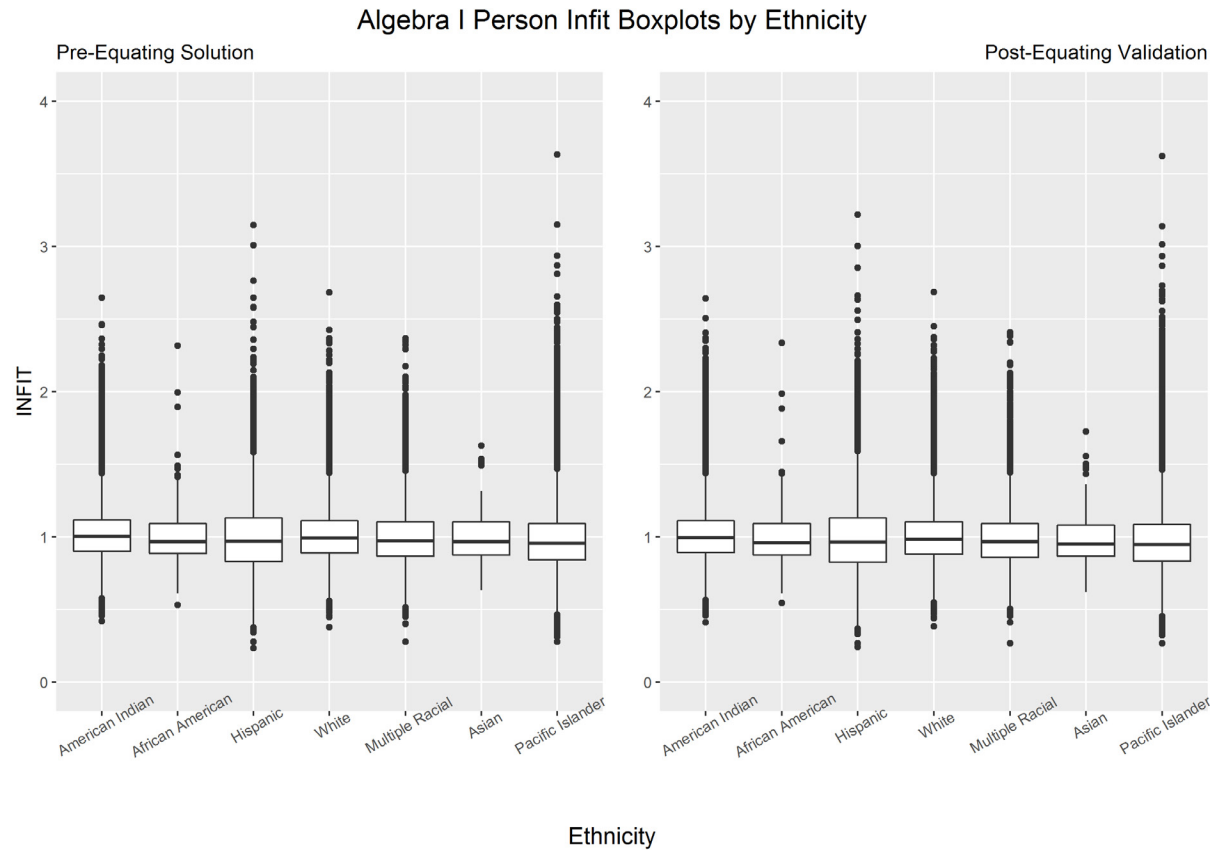
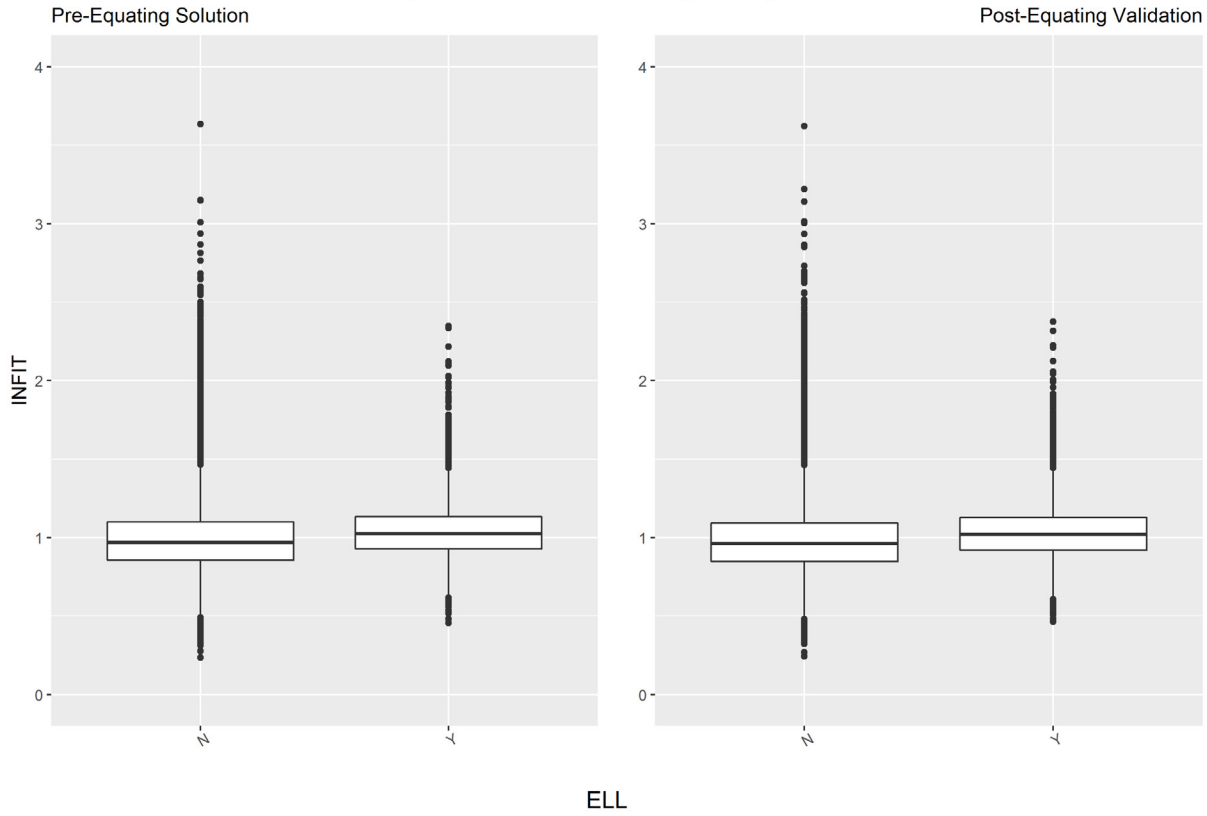




Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring

Algebra I Person Infit Boxplots by ELL



Algebra I Person Infit Boxplots by IEP

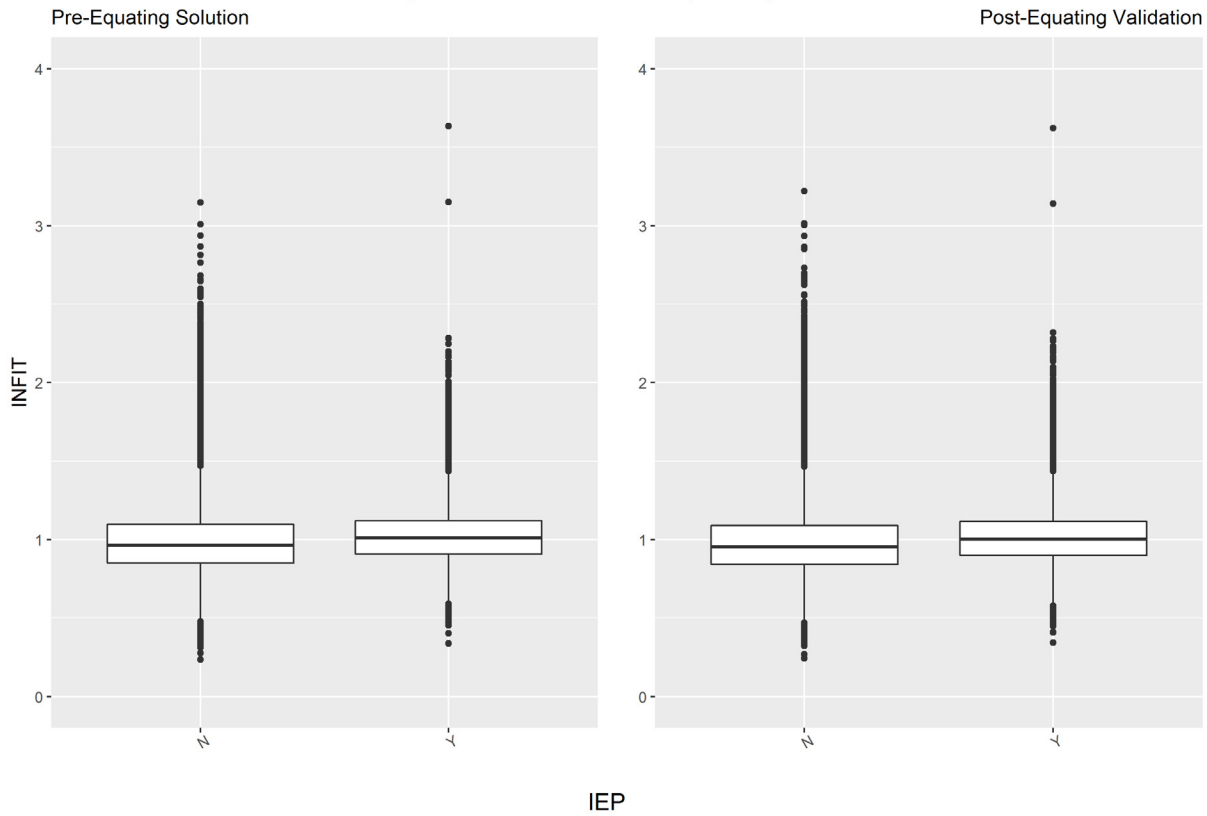
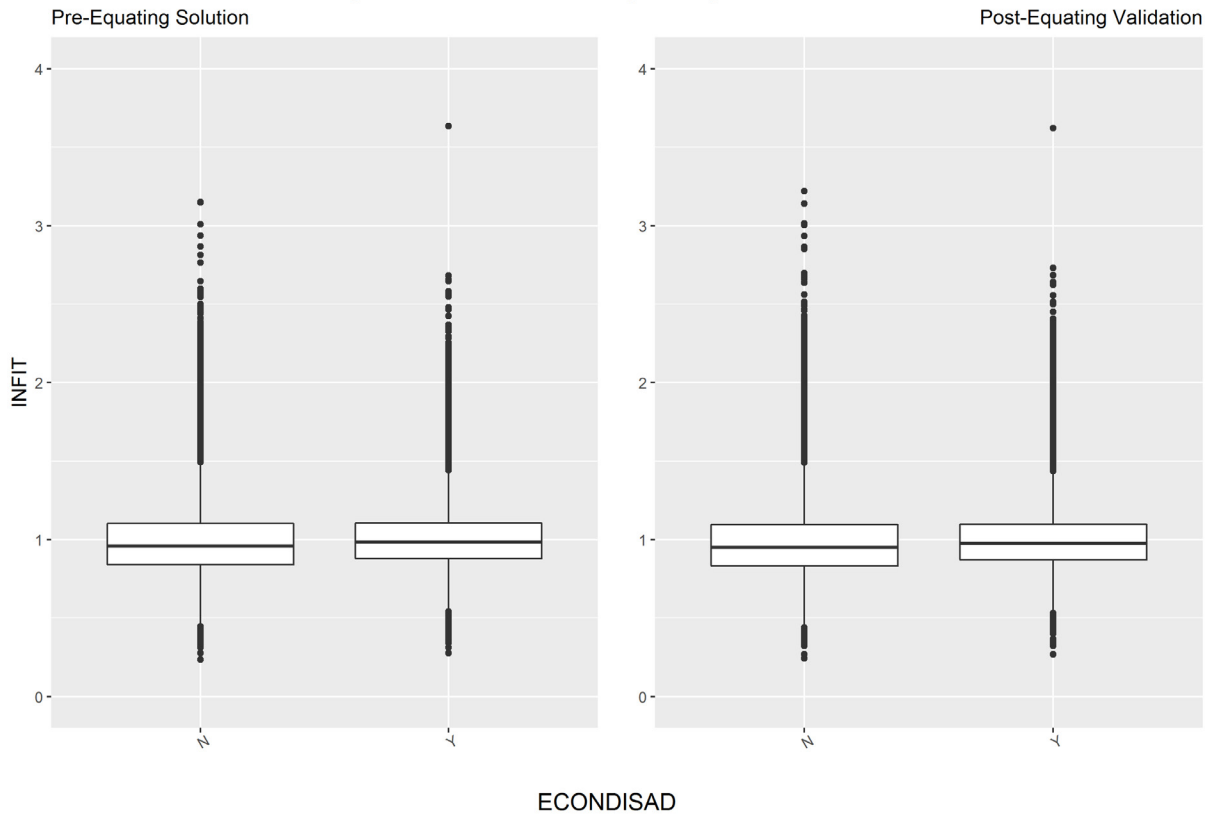


Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring

Algebra I Person Infit Boxplots by ECONDISAD



Biology Person Infit Boxplots by Ethnicity

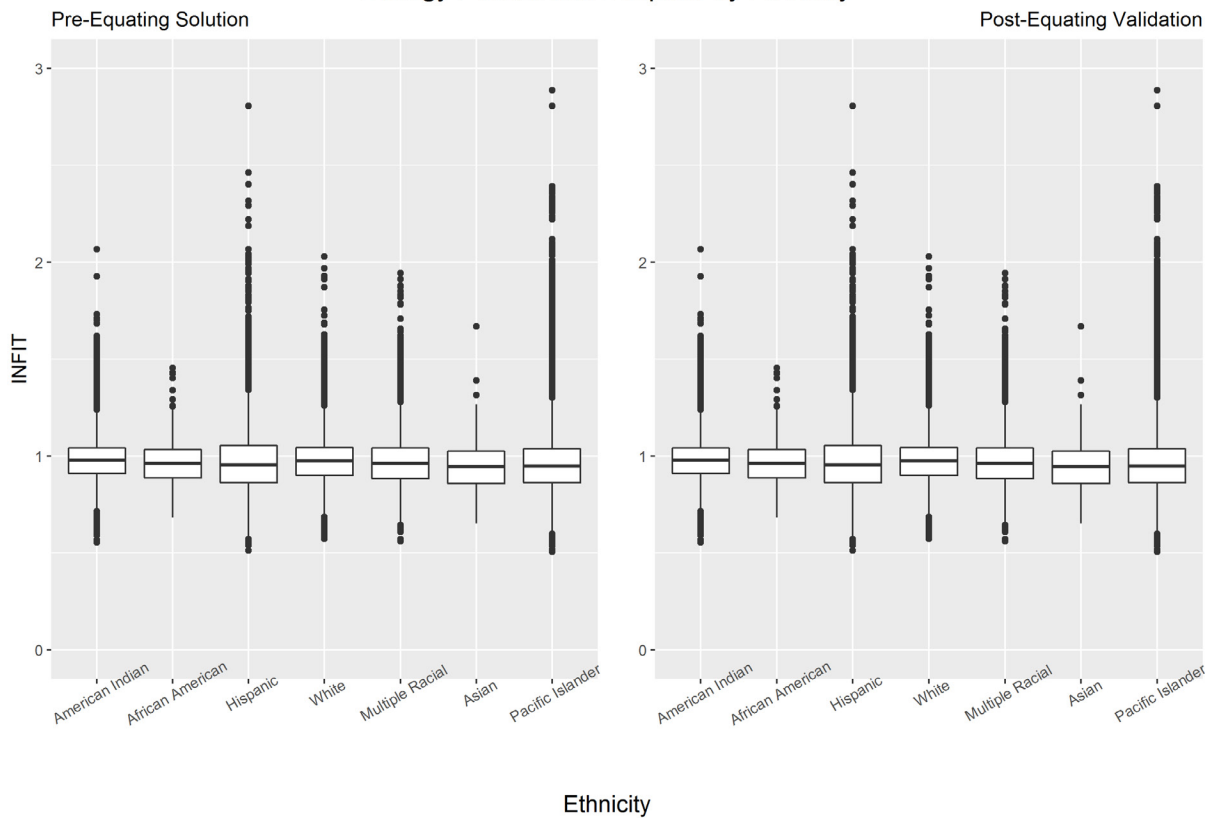


Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring

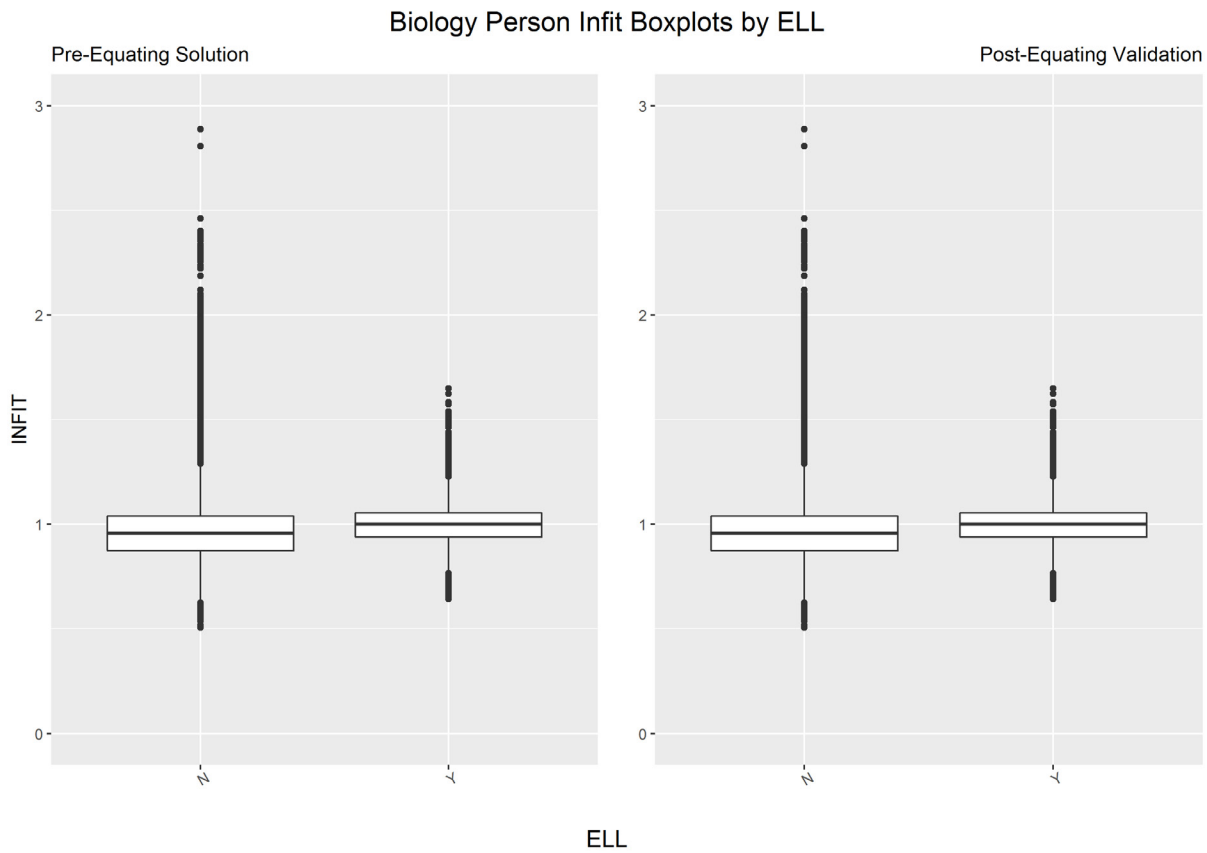
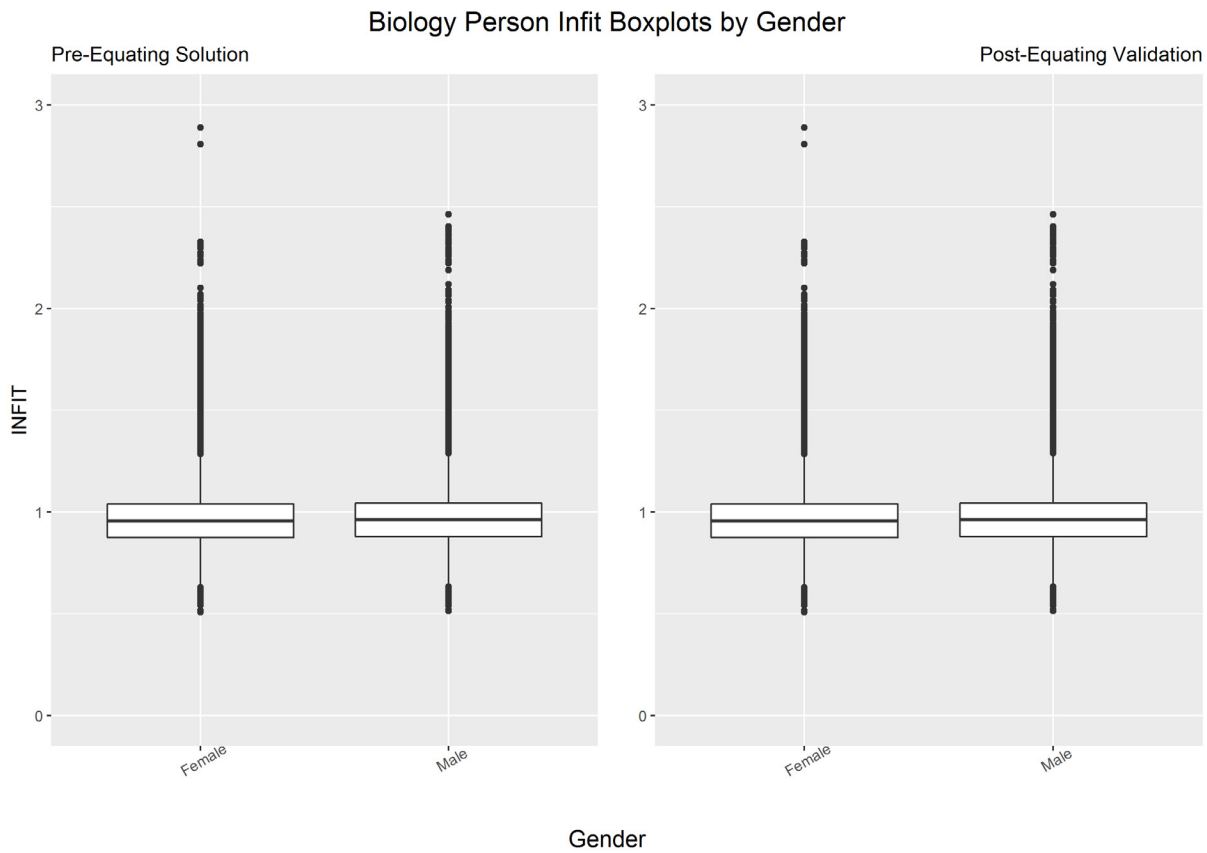
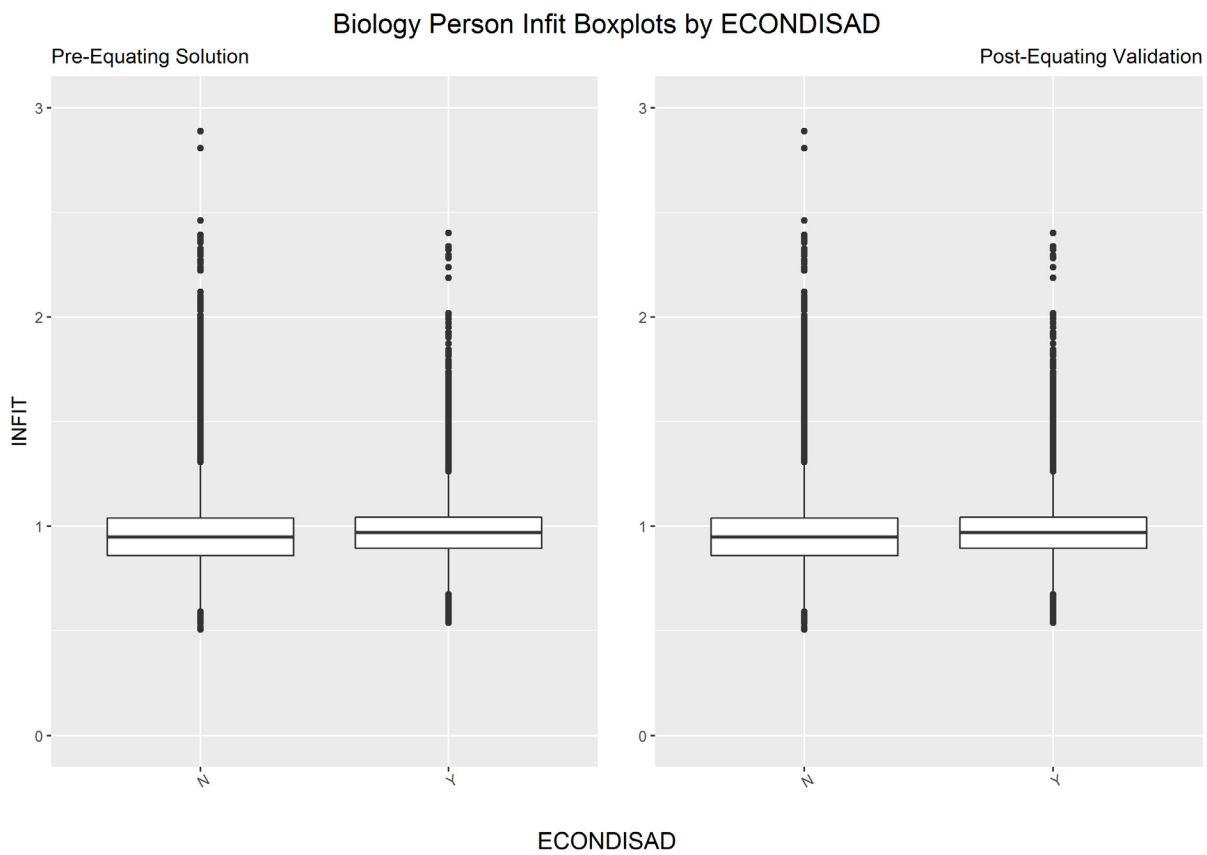
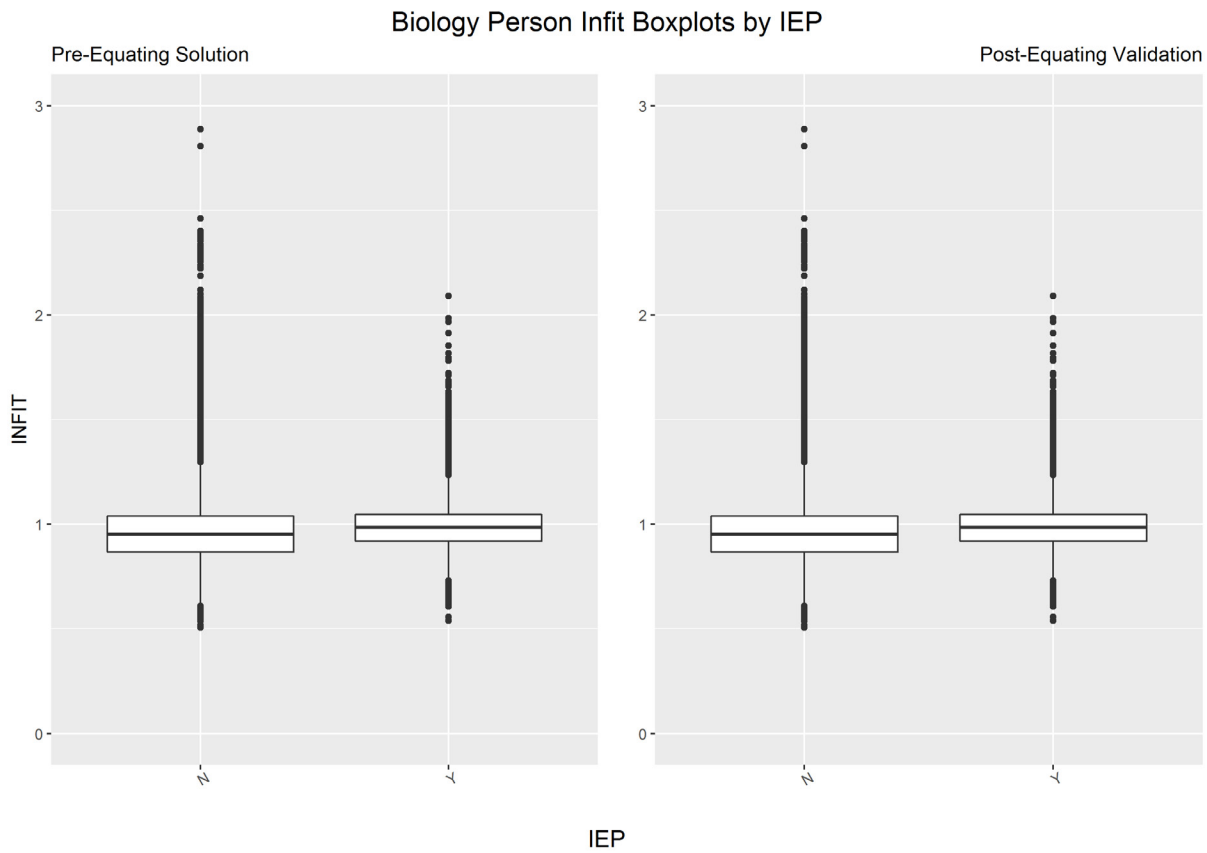


Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring



**Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring**

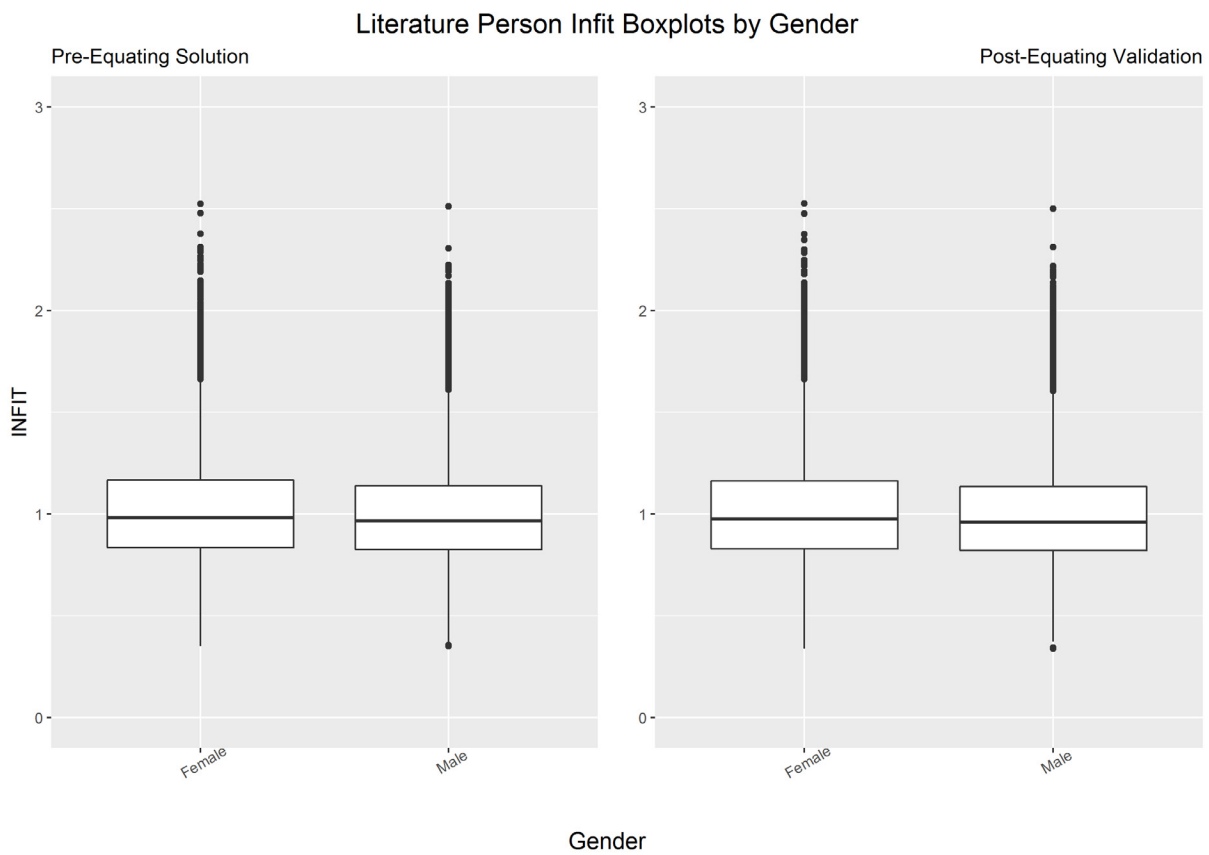
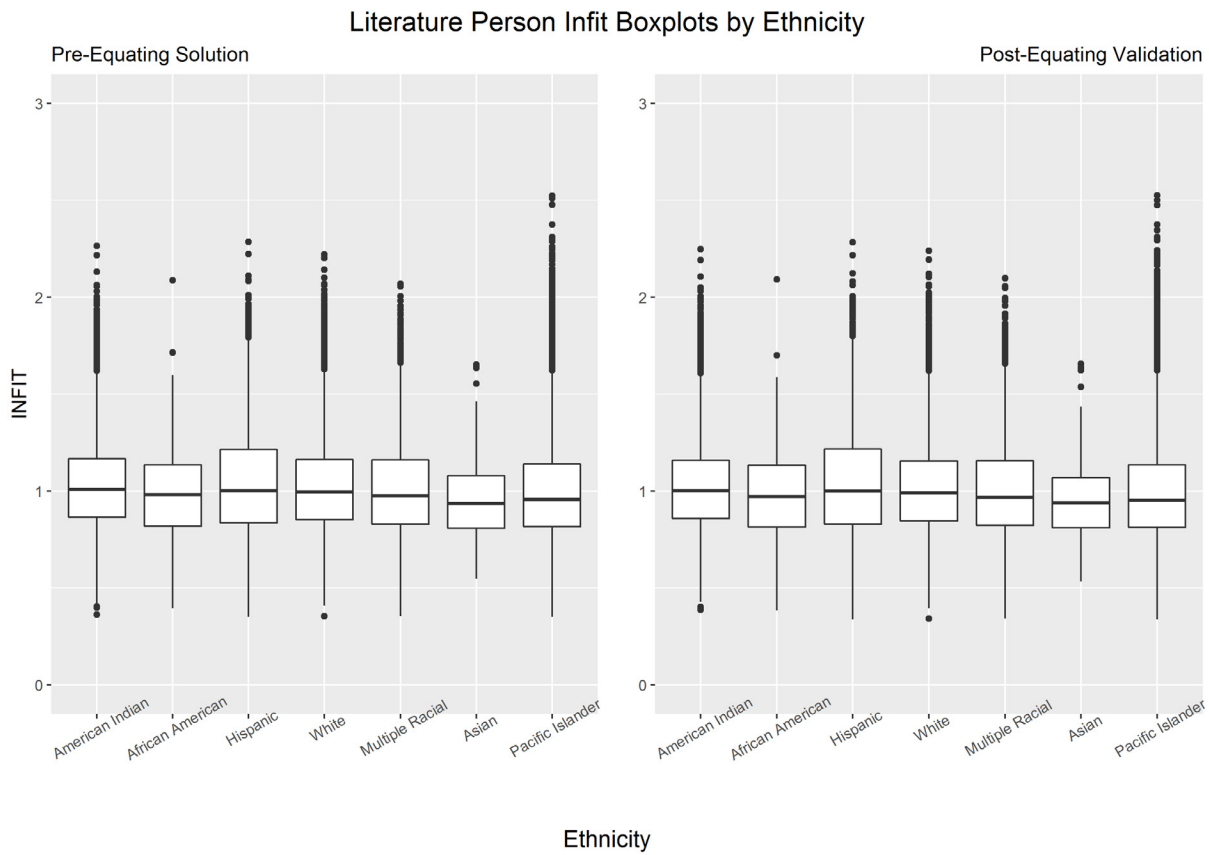


Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring

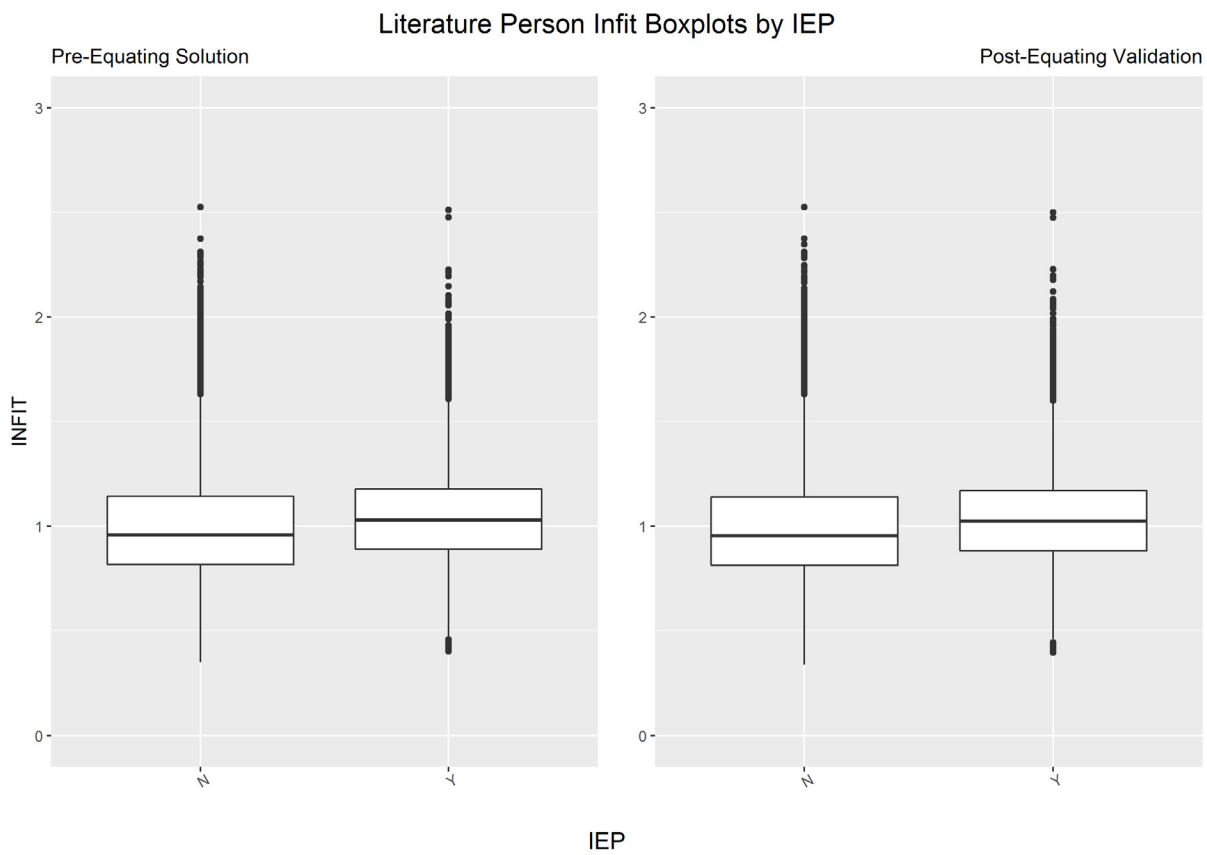
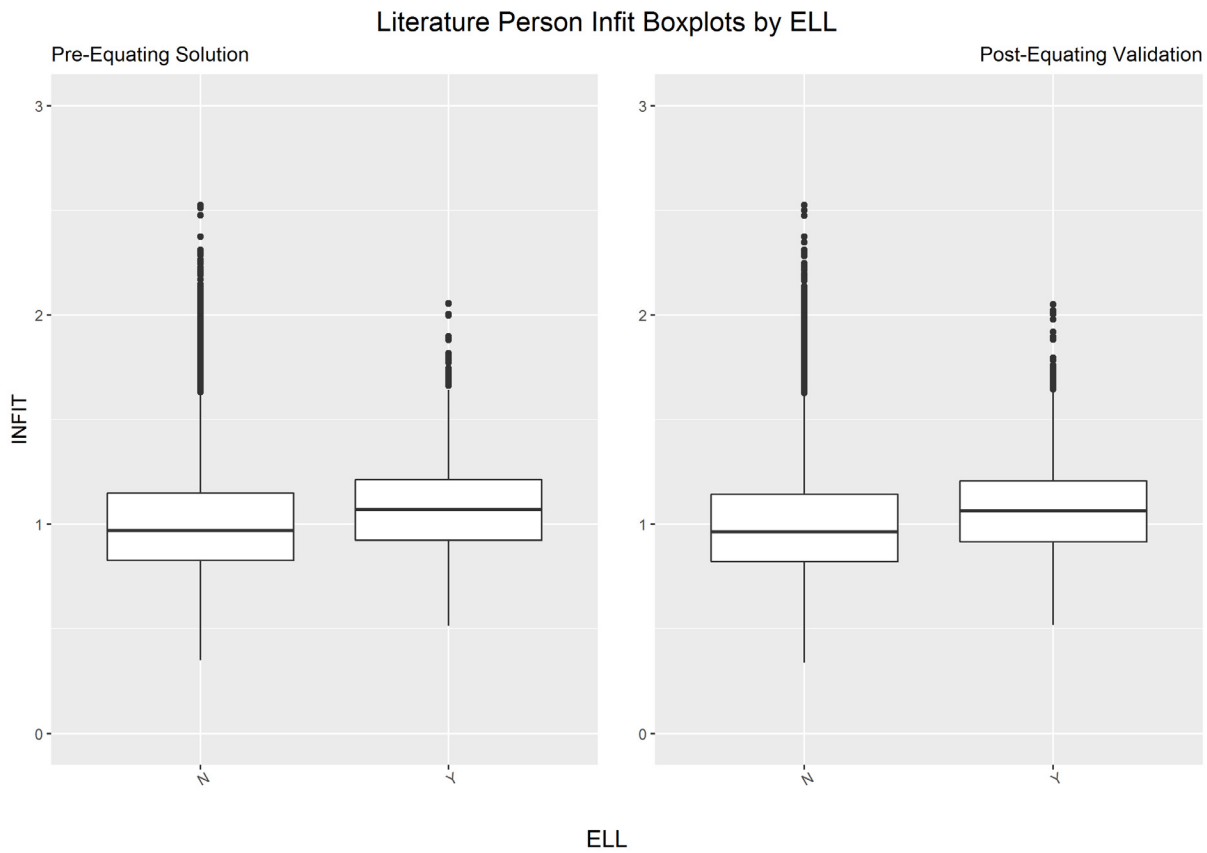


Figure L-3 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Spring

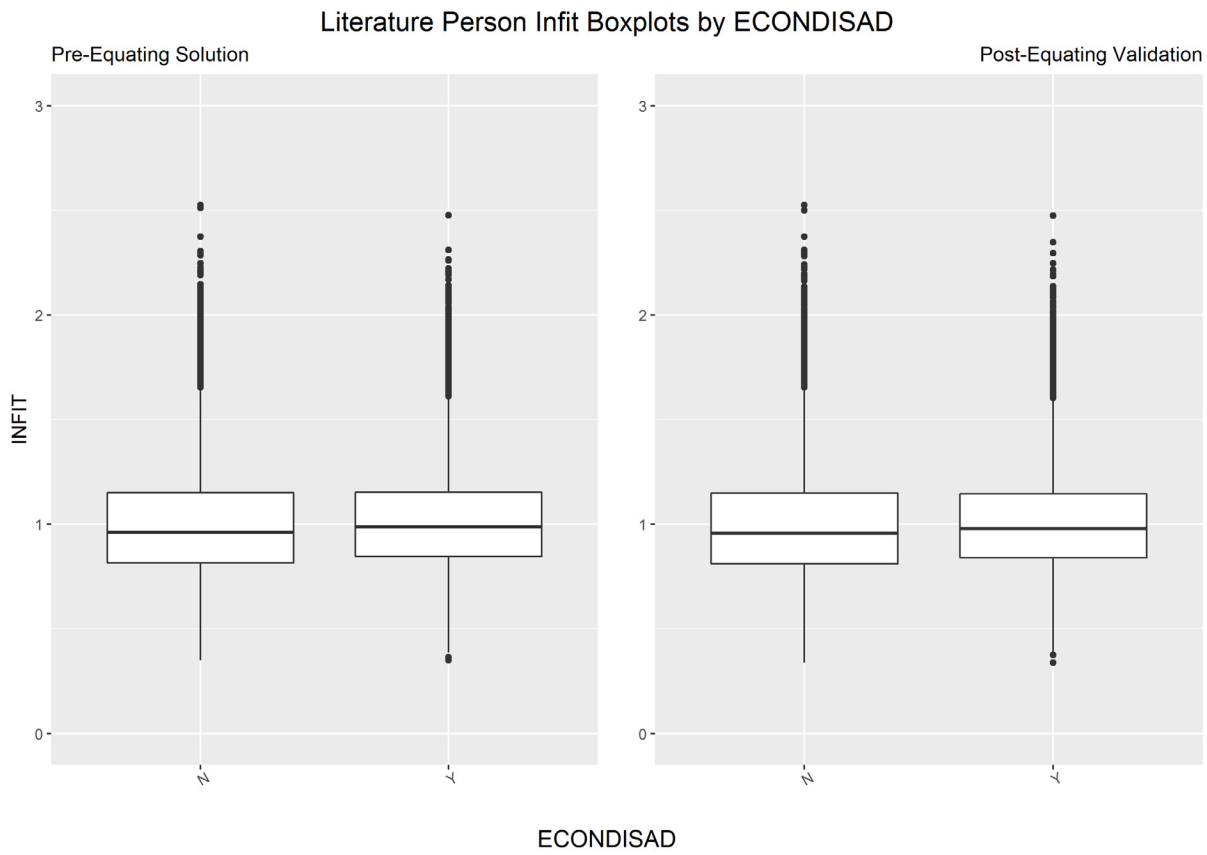
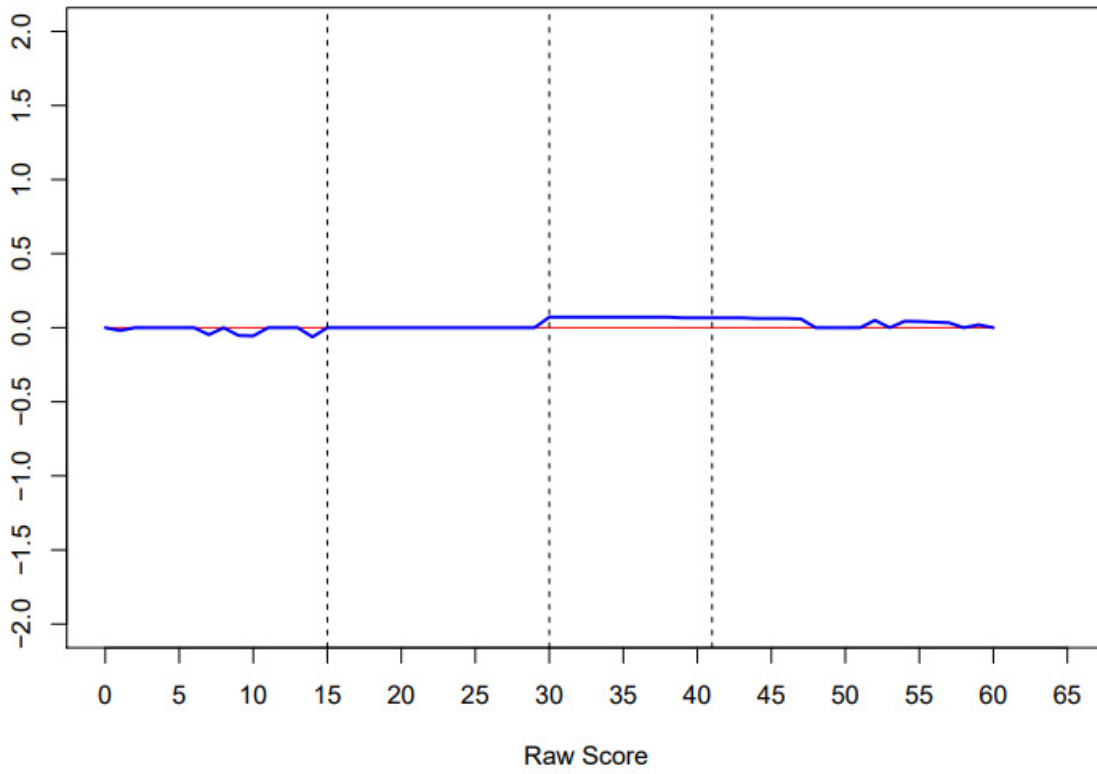


Figure L-4. Normalized Scale Score Distributions by Content Area: Spring

### Algebra I



### Biology

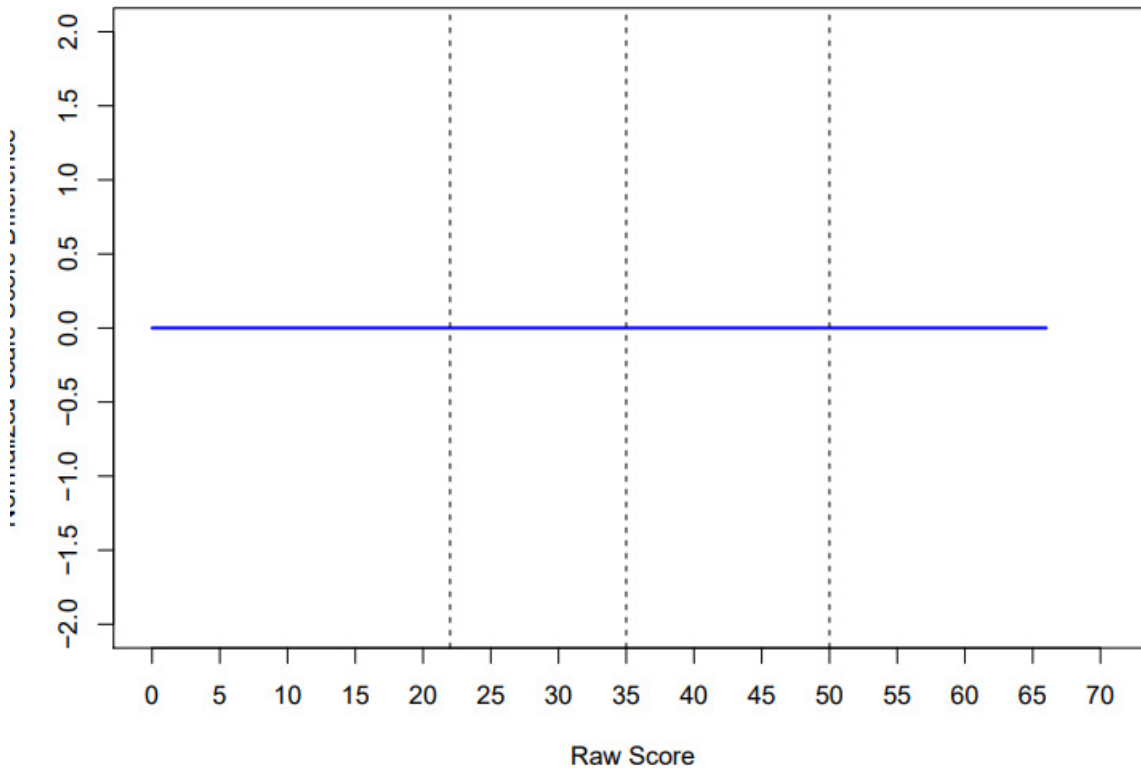
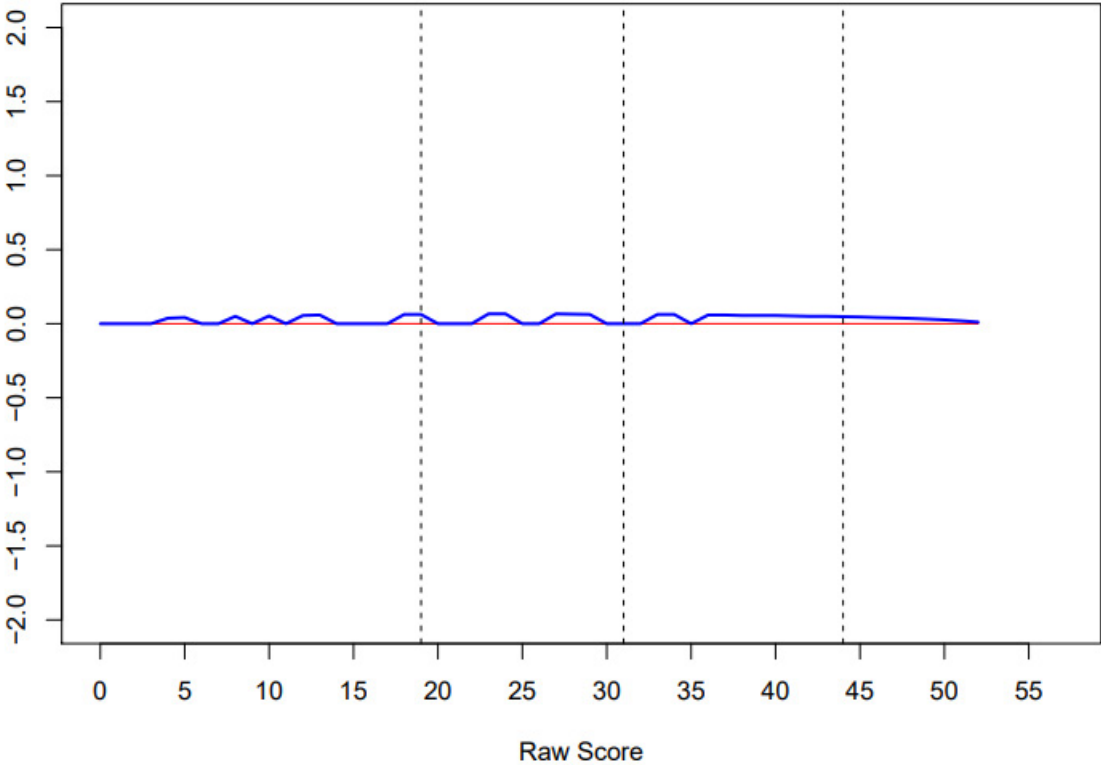




Figure L-4 (continued). Normalized Scale Score Distributions by Content Area: Spring

Literature



**Table L-4. Spring Algebra I Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1217	1217	92	92	BB	BB	0.0	TRUE
1	1278	1279	51	51	BB	BB	0.0	TRUE
2	1315	1315	37	37	BB	BB	0.0	TRUE
3	1337	1337	30	30	BB	BB	0.1	TRUE
4	1353	1353	27	27	BB	BB	0.2	TRUE
5	1366	1366	24	24	BB	BB	0.4	TRUE
6	1377	1377	22	22	BB	BB	0.8	TRUE
7	1386	1387	21	21	BB	BB	1.3	TRUE
8	1395	1395	20	20	BB	BB	2.0	TRUE
9	1402	1403	19	19	BB	BB	2.7	TRUE
10	1409	1410	18	18	BB	BB	3.5	TRUE
11	1416	1416	18	18	BB	BB	4.0	TRUE
12	1422	1422	17	17	BB	BB	4.2	TRUE
13	1428	1428	17	17	BB	BB	4.3	TRUE
14	1433	1434	16	16	BB	BB	4.2	TRUE
15	1439	1439	16	16	B	B	4.1	TRUE
16	1444	1444	16	16	B	B	3.8	TRUE
17	1449	1449	15	15	B	B	3.6	TRUE
18	1453	1453	15	15	B	B	3.5	TRUE
19	1458	1458	15	15	B	B	3.3	TRUE
20	1462	1462	15	15	B	B	3.0	TRUE
27	1479	1478	1481	13	B	B	2.4	TRUE
28	1481	1482	13	13	B	B	2.3	TRUE
21	1467	1467	15	15	B	B	2.9	TRUE
22	1471	1471	15	14	B	B	2.7	TRUE
23	1475	1475	14	14	B	B	2.5	TRUE
24	1479	1479	14	14	B	B	2.4	TRUE
25	1483	1483	14	14	B	B	2.4	TRUE
26	1487	1487	14	14	B	B	2.3	TRUE
27	1491	1491	14	14	B	B	2.1	TRUE
28	1495	1495	14	14	B	B	2.0	TRUE
29	1499	1499	14	14	B	B	2.0	TRUE
30	1503	1502	14	14	P	P	1.9	TRUE
31	1507	1506	14	14	P	P	1.8	TRUE
32	1511	1510	14	14	P	P	1.7	TRUE
33	1515	1514	14	14	P	P	1.7	TRUE
34	1519	1518	14	14	P	P	1.6	TRUE
35	1523	1522	14	14	P	P	1.6	TRUE
36	1527	1526	14	14	P	P	1.5	TRUE
37	1531	1530	14	14	P	P	1.4	TRUE
38	1535	1534	14	14	P	P	1.3	TRUE
39	1539	1538	15	15	P	P	1.3	TRUE

**Table L-4 (continued). Spring Algebra I Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
40	1543	1542	15	15	P	P	1.2	TRUE
41	1548	1547	15	15	A	A	1.2	TRUE
42	1552	1551	15	15	A	A	1.2	TRUE
43	1557	1556	15	15	A	A	1.1	TRUE
44	1562	1561	16	16	A	A	1.1	TRUE
45	1567	1566	16	16	A	A	1.1	TRUE
46	1572	1571	16	16	A	A	0.9	TRUE
47	1578	1577	17	17	A	A	0.9	TRUE
48	1583	1583	17	17	A	A	0.8	TRUE
49	1589	1589	18	18	A	A	0.8	TRUE
50	1596	1596	18	18	A	A	0.7	TRUE
51	1603	1603	19	19	A	A	0.6	TRUE
52	1611	1610	20	20	A	A	0.5	TRUE
53	1619	1619	21	21	A	A	0.5	TRUE
54	1629	1628	23	23	A	A	0.4	TRUE
55	1640	1639	24	24	A	A	0.3	TRUE
56	1653	1652	27	27	A	A	0.2	TRUE
57	1669	1668	30	30	A	A	0.2	TRUE
58	1691	1691	37	37	A	A	0.1	TRUE
59	1728	1727	51	51	A	A	0.0	TRUE
60	1789	1789	92	92	A	A	0.0	TRUE

**Table L-5. Spring Biology Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1220	1220	92	92	BB	BB	0.0	TRUE
1	1281	1281	51	51	BB	BB	0.0	TRUE
2	1317	1317	36	36	BB	BB	0.0	TRUE
3	1338	1338	30	30	BB	BB	0.0	TRUE
4	1354	1354	26	26	BB	BB	0.0	TRUE
5	1366	1366	24	24	BB	BB	0.1	TRUE
6	1376	1376	22	22	BB	BB	0.1	TRUE
7	1385	1385	20	20	BB	BB	0.3	TRUE
8	1393	1393	19	19	BB	BB	0.5	TRUE
9	1400	1400	18	18	BB	BB	0.8	TRUE
10	1406	1406	18	18	BB	BB	1.2	TRUE
11	1412	1412	17	17	BB	BB	1.7	TRUE
12	1417	1417	16	16	BB	BB	2.2	TRUE
13	1422	1422	16	16	BB	BB	2.6	TRUE
14	1427	1427	15	15	BB	BB	3.0	TRUE
15	1432	1432	15	15	BB	BB	3.2	TRUE
16	1436	1436	15	15	BB	BB	3.2	TRUE
17	1441	1441	14	14	BB	BB	3.4	TRUE
18	1445	1445	14	14	BB	BB	3.3	TRUE
19	1449	1449	14	14	BB	BB	3.1	TRUE
20	1452	1452	14	14	BB	BB	2.9	TRUE
21	1456	1456	14	14	BB	BB	2.9	TRUE
21	14	1464	15	15	B	B	2.9	TRUE
22	1460	1460	13	13	B	B	2.6	TRUE
23	1463	1463	13	13	B	B	2.5	TRUE
24	1467	1467	13	13	B	B	2.4	TRUE
25	1470	1470	13	13	B	B	2.2	TRUE
26	1473	1473	13	13	B	B	2.2	TRUE
27	1477	1477	13	13	B	B	2.1	TRUE
28	1480	1480	13	13	B	B	2.0	TRUE
29	1483	1483	13	13	B	B	1.9	TRUE
30	1486	1486	13	13	B	B	1.8	TRUE
31	1489	1489	12	12	B	B	1.8	TRUE
32	1492	1492	12	12	B	B	1.8	TRUE
33	1496	1496	12	12	B	B	1.8	TRUE
34	1499	1499	12	12	B	B	1.7	TRUE
35	1502	1502	12	12	P	P	1.7	TRUE
36	1505	1505	12	12	P	P	1.6	TRUE
37	1508	1508	12	12	P	P	1.5	TRUE
38	1511	1511	12	12	P	P	1.6	TRUE
39	1514	1514	12	12	P	P	1.5	TRUE
40	1517	1517	13	13	P	P	1.6	TRUE

**Table L-5 (continued). Spring Biology Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
41	1520	1520	13	13	P	P	1.5	TRUE
42	1523	1523	13	13	P	P	1.6	TRUE
43	1527	1527	13	13	P	P	1.5	TRUE
44	1530	1530	13	13	P	P	1.4	TRUE
45	1533	1533	13	13	P	P	1.5	TRUE
46	1537	1537	13	13	P	P	1.5	TRUE
47	1540	1540	13	13	P	P	1.4	TRUE
48	1544	1544	14	14	P	P	1.4	TRUE
49	1548	1548	14	14	P	P	1.4	TRUE
50	1552	1552	14	14	A	A	1.3	TRUE
51	1556	1556	14	14	A	A	1.3	TRUE
52	1560	1560	15	15	A	A	1.3	TRUE
53	1565	1565	15	15	A	A	1.3	TRUE
54	1570	1570	16	16	A	A	1.2	TRUE
55	1575	1575	16	16	A	A	1.3	TRUE
56	1580	1580	17	17	A	A	1.2	TRUE
57	1586	1586	18	18	A	A	1.2	TRUE
58	1593	1593	19	19	A	A	1.1	TRUE
59	1600	1600	20	20	A	A	1.0	TRUE
60	1608	1608	21	21	A	A	0.9	TRUE
61	1618	1618	23	23	A	A	0.8	TRUE
62	1630	1630	26	26	A	A	0.7	TRUE
63	1645	1645	29	29	A	A	0.6	TRUE
64	1666	1666	36	36	A	A	0.4	TRUE
65	1701	1701	50	50	A	A	0.2	TRUE
66	1762	1762	91	91	A	A	0.1	TRUE
55	1647	1647	26	26	A	A	0.9	TRUE
56	16	1652	29	29	A	A	0.2	TRUE
50	1681	1681	1402	32	A	A	0.1	TRUE
58	1705	17	39	9	A	A	0.8	TRUE
59	15	1745	1653	53	A	A	0.0	TRUE
60	1800	1800	93	93	A	A	4.0	TRUE

**Table L-6. Spring Literature Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1202	1202	92	92	BB	BB	0.0	TRUE
1	1263	1263	51	51	BB	BB	0.0	TRUE
2	1299	1299	37	37	BB	BB	0.0	TRUE
3	1321	1321	30	30	BB	BB	0.0	TRUE
4	1338	1337	27	27	BB	BB	0.1	TRUE
5	1351	1350	24	24	BB	BB	0.2	TRUE
6	1361	1361	23	23	BB	BB	0.4	TRUE
7	1371	1371	21	21	BB	BB	0.6	TRUE
8	1380	1379	20	20	BB	BB	0.8	TRUE
9	1387	1387	19	19	BB	BB	1.0	TRUE
10	1395	1394	19	19	BB	BB	1.1	TRUE
11	1401	1401	18	18	BB	BB	1.3	TRUE
12	1408	1407	18	18	BB	BB	1.4	TRUE
13	1414	1413	17	17	BB	BB	1.5	TRUE
14	1419	1419	17	17	BB	BB	1.6	TRUE
15	1425	1425	17	16	BB	BB	1.5	TRUE
16	1430	1430	16	16	BB	BB	1.7	TRUE
17	1435	1435	16	16	BB	BB	1.8	TRUE
18	1441	1440	16	16	BB	BB	1.9	TRUE
19	1446	1445	16	16	B	B	1.9	TRUE
20	1450	1450	16	16	B	B	2.0	TRUE
21	1455	1455	16	15	B	B	2.1	TRUE
22	1460	1460	15	15	B	B	2.3	TRUE
23	1465	1464	15	15	B	B	2.3	TRUE
24	1470	1469	15	15	B	B	2.5	TRUE
25	1474	1474	15	15	B	B	2.6	TRUE
26	1479	1479	15	15	B	B	2.7	TRUE
27	1484	1483	15	15	B	B	2.7	TRUE
28	1489	1488	16	15	B	B	2.9	TRUE
29	1494	1493	16	16	B	B	3.0	TRUE
30	1498	1498	16	16	B	B	3.1	TRUE
31	1503	1503	16	16	P	P	3.2	TRUE
32	1508	1508	16	16	P	P	3.3	TRUE
33	1514	1513	16	16	P	P	3.5	TRUE
34	1519	1518	16	16	P	P	3.5	TRUE
35	1524	1524	17	17	P	P	3.6	TRUE
36	1530	1529	17	17	P	P	3.8	TRUE
37	1536	1535	17	17	P	P	3.7	TRUE
38	1542	1541	18	18	P	P	3.6	TRUE
39	1548	1547	18	18	P	P	3.6	TRUE
40	1555	1554	18	18	P	P	3.4	TRUE
41	1562	1561	19	19	P	P	3.2	TRUE

**Table L-6 (continued). Spring Literature Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
42	1569	1568	20	20	P	P	2.9	TRUE
43	1577	1576	20	20	P	P	2.7	TRUE
44	1586	1585	21	21	A	A	2.3	TRUE
45	1595	1594	22	22	A	A	2.0	TRUE
46	1606	1605	24	24	A	A	1.5	TRUE
47	1618	1617	25	26	A	A	1.2	TRUE
48	1632	1631	28	28	A	A	0.8	TRUE
49	1650	1649	32	32	A	A	0.5	TRUE
50	1673	1672	38	38	A	A	0.3	TRUE
51	1711	1710	52	52	A	A	0.1	TRUE
52	1774	1773	92	92	A	A	0.0	TRUE

# SUMMER

Figure L-5. Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer

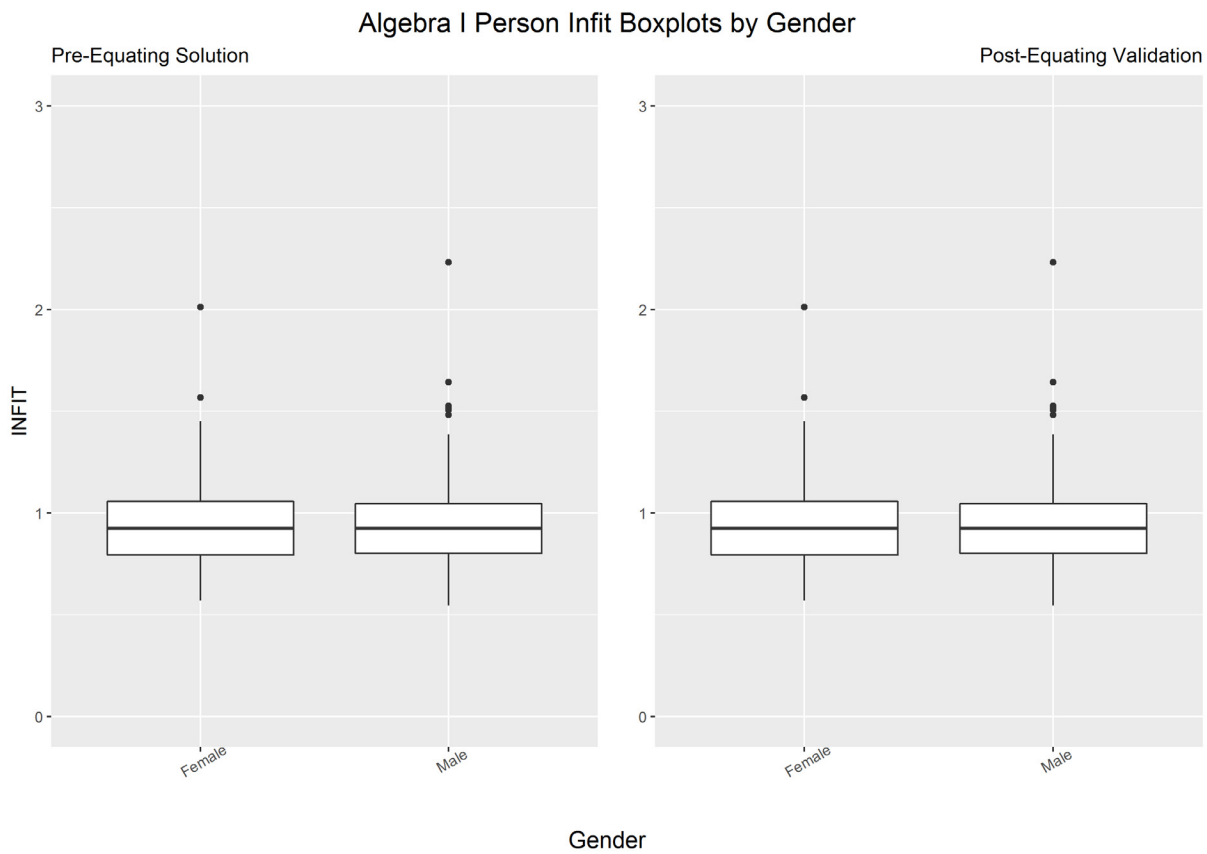
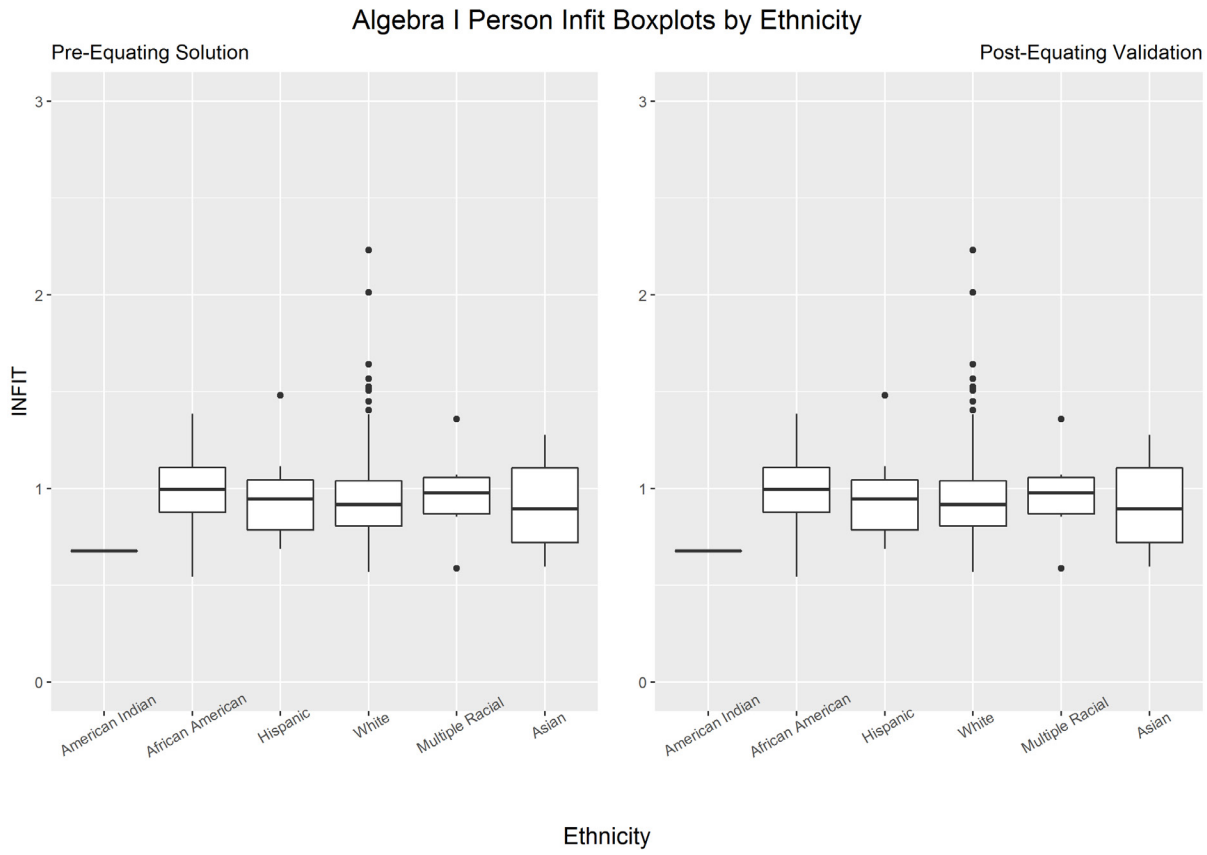
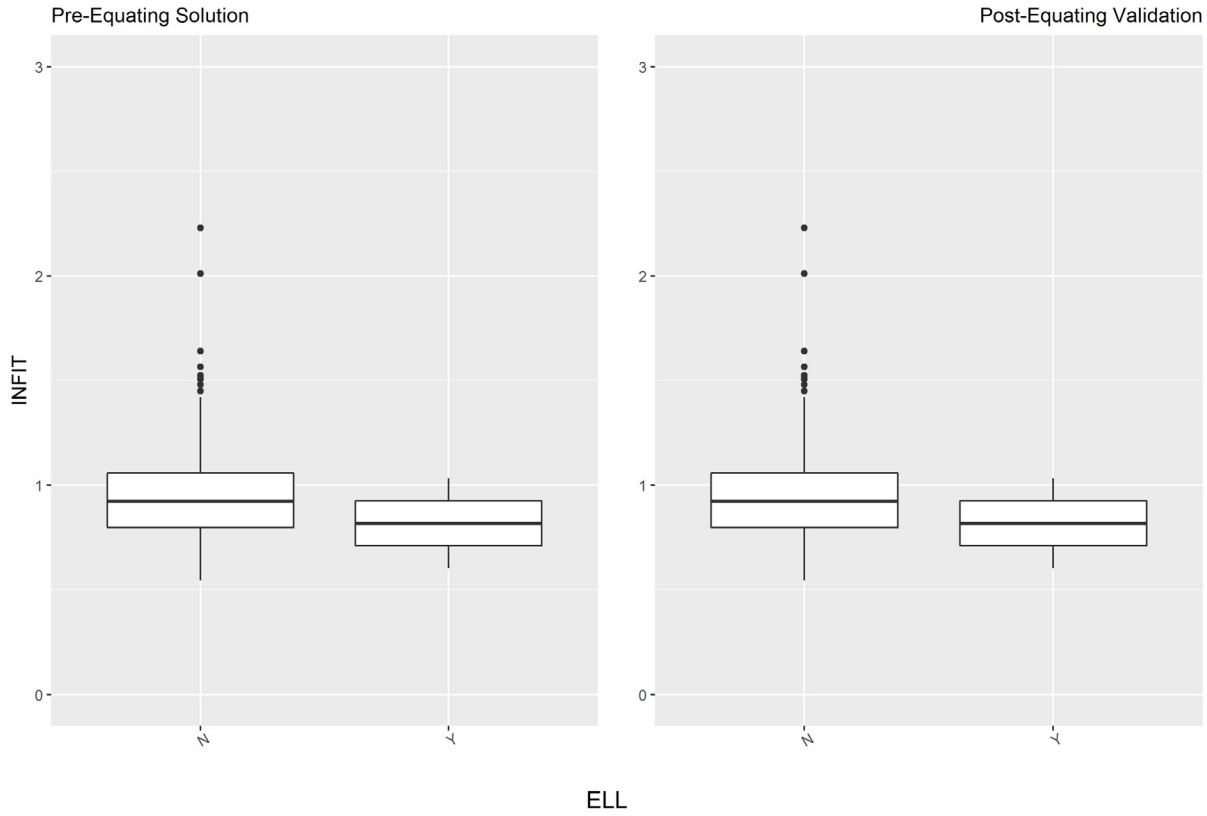


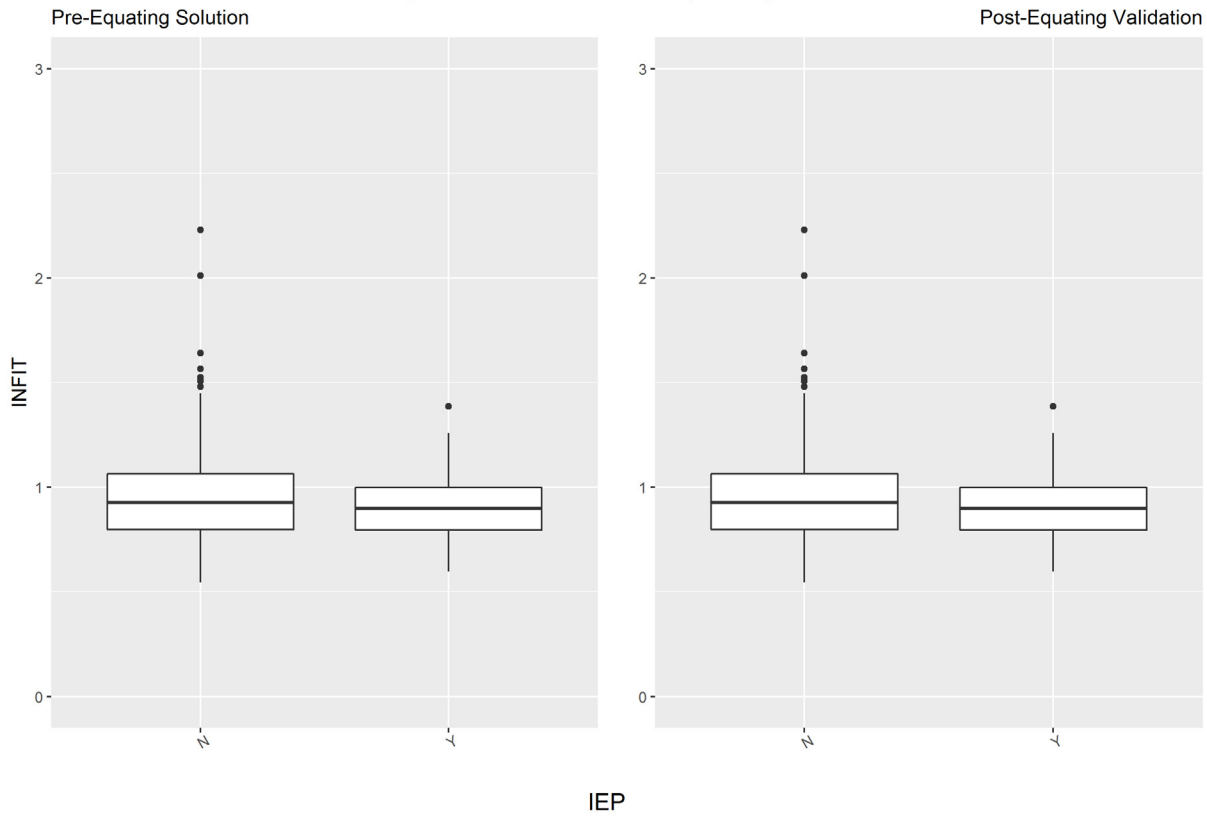


Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer

Algebra I Person Infit Boxplots by ELL

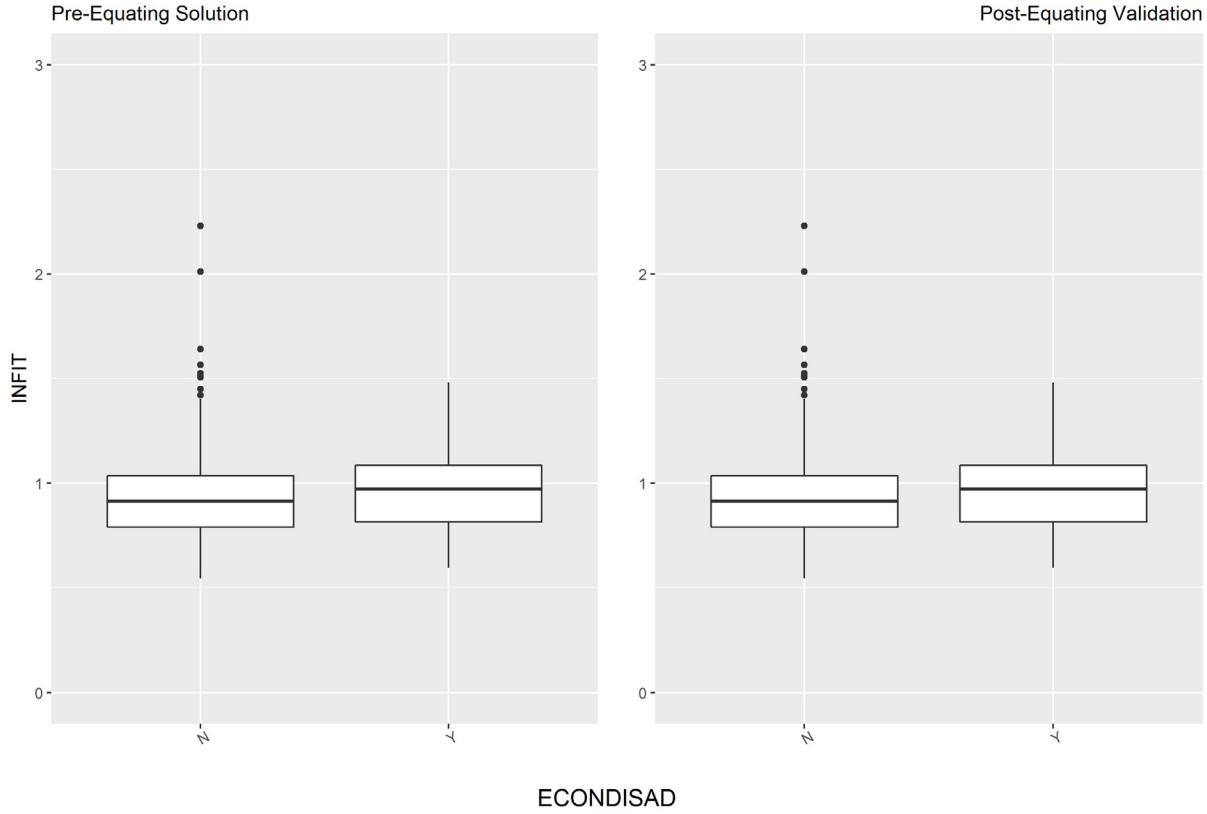


Algebra I Person Infit Boxplots by IEP

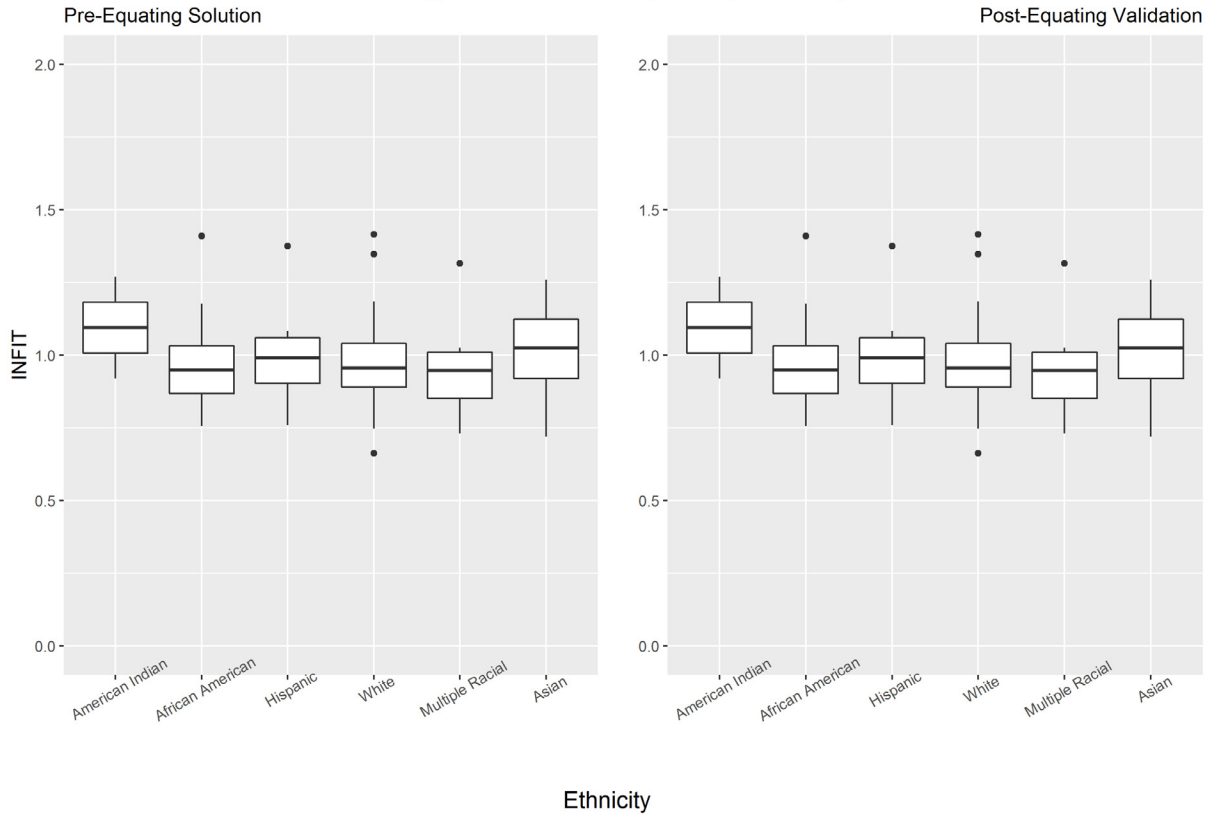


**Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer**

**Algebra I Person Infit Boxplots by ECONDISAD**



**Biology Person Infit Boxplots by Ethnicity**



**Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer**

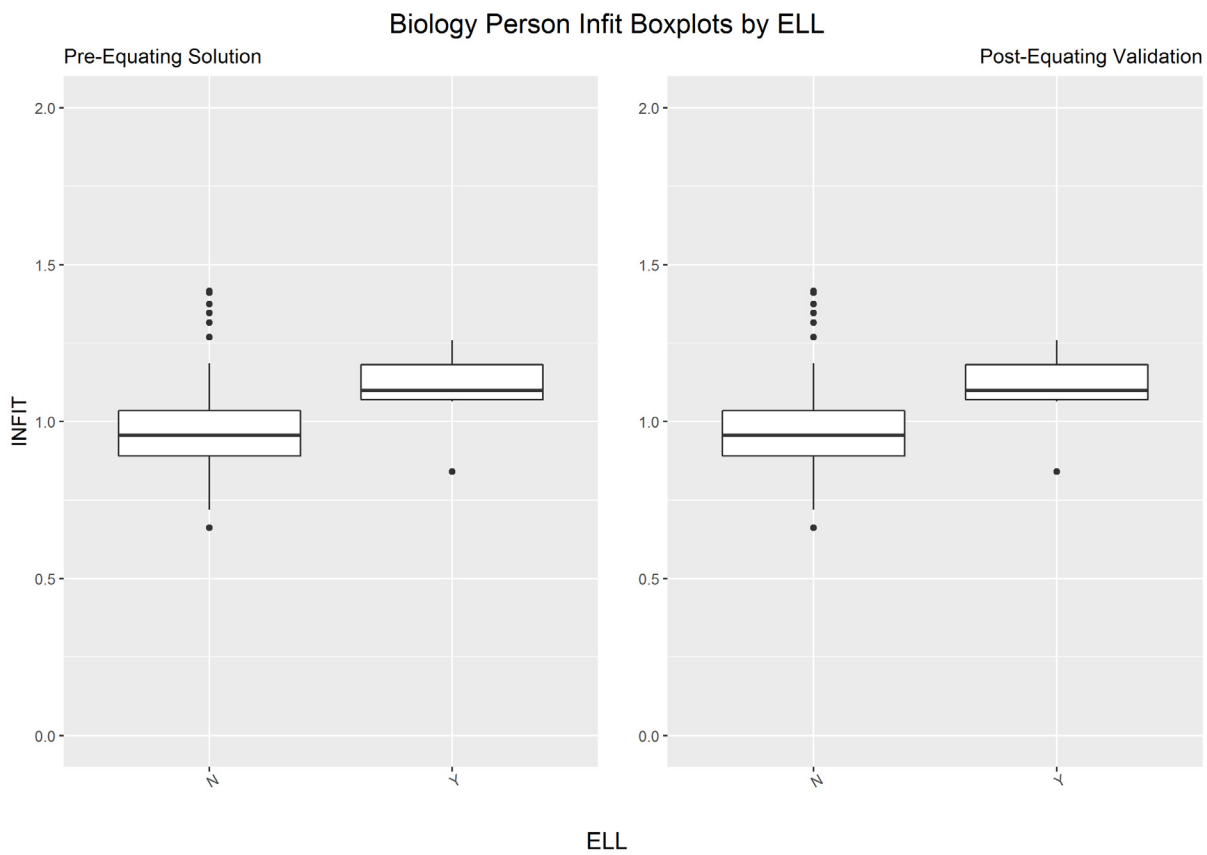
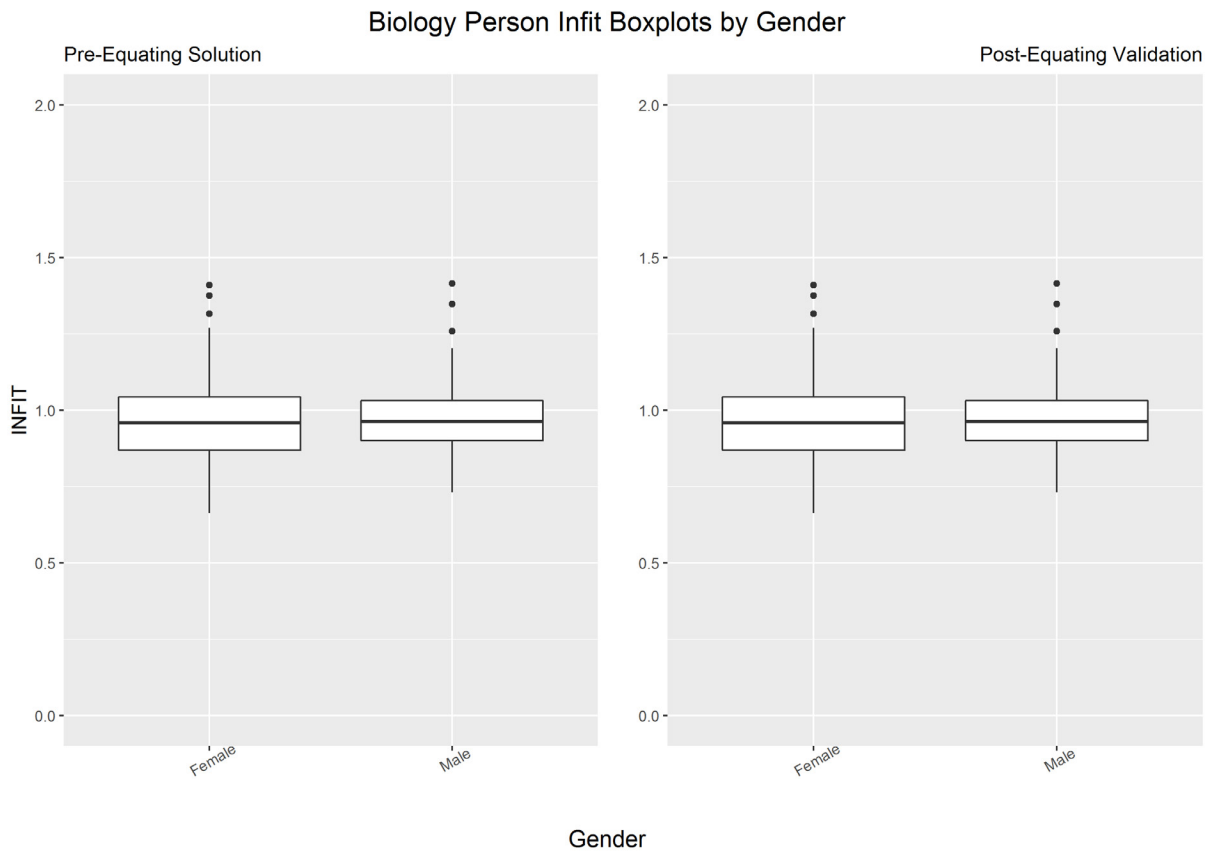
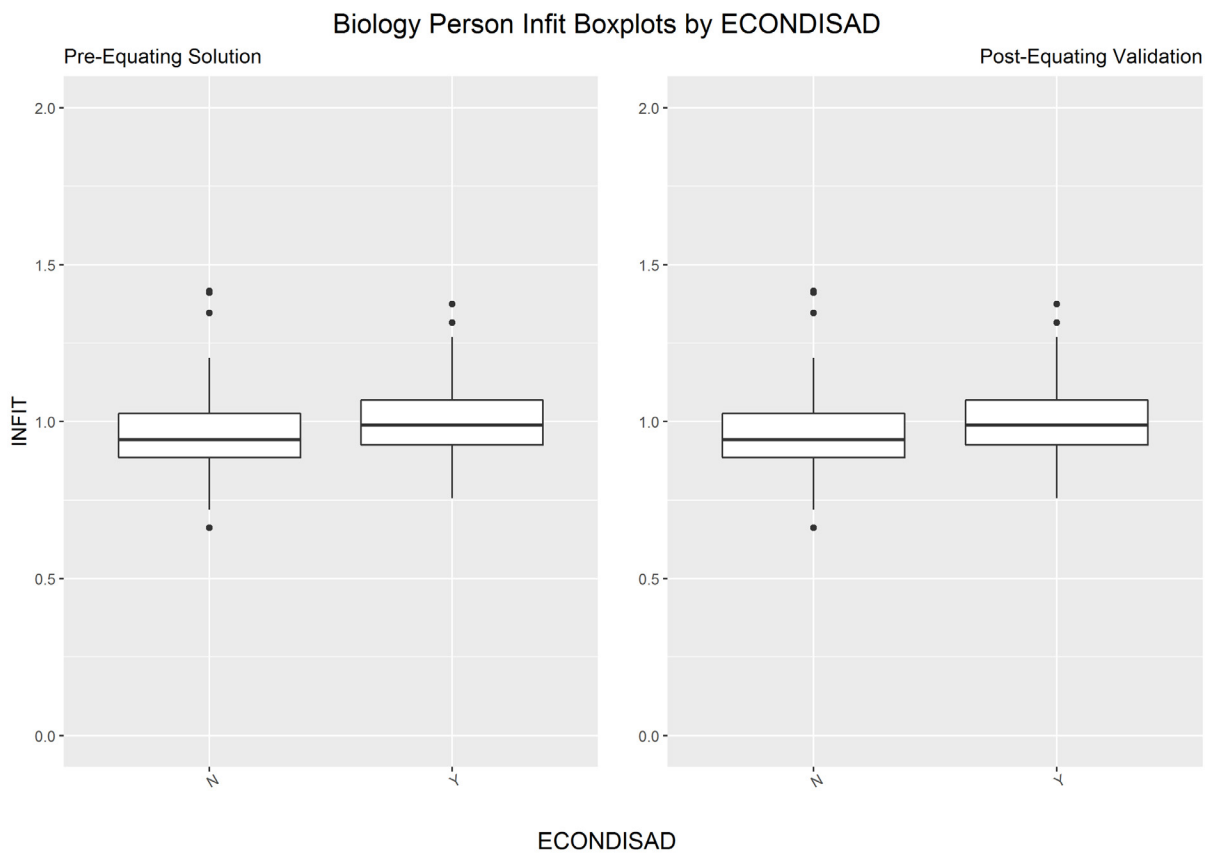
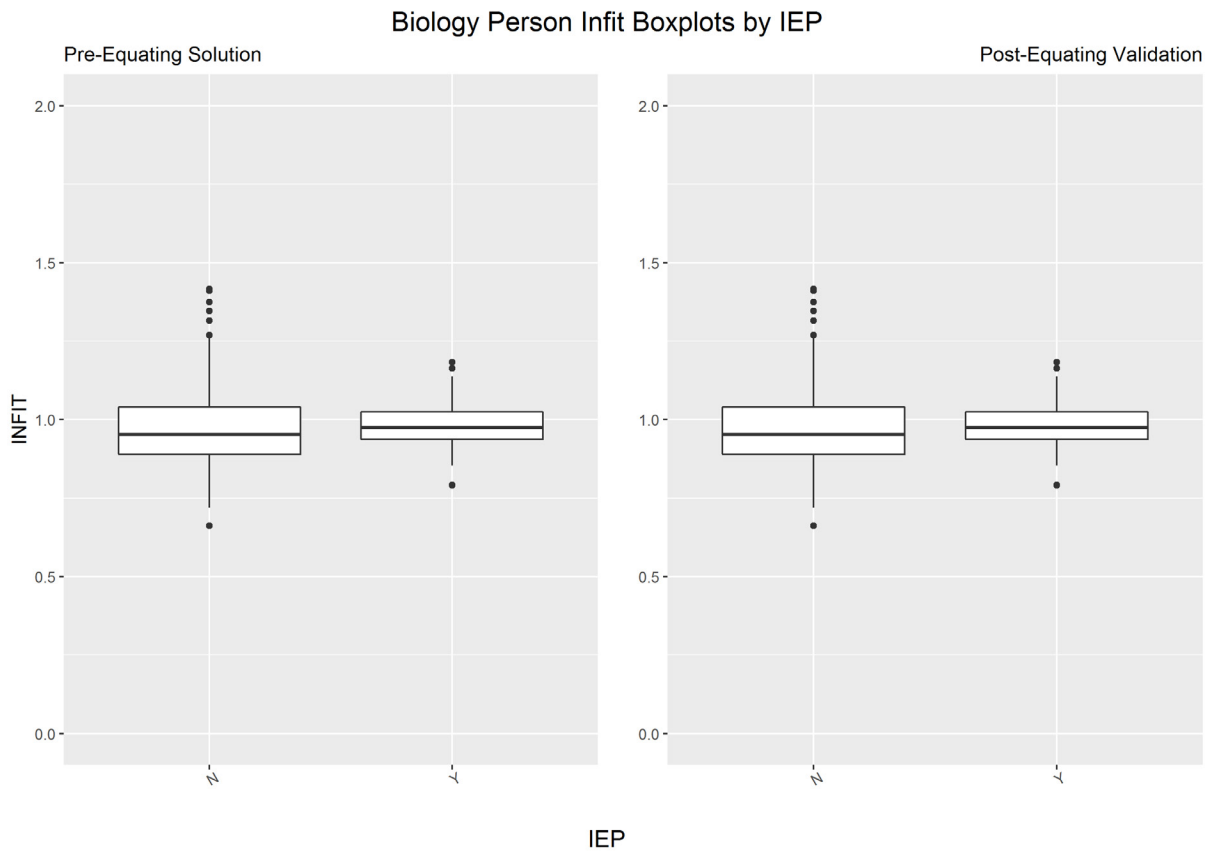


Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer



**Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer**

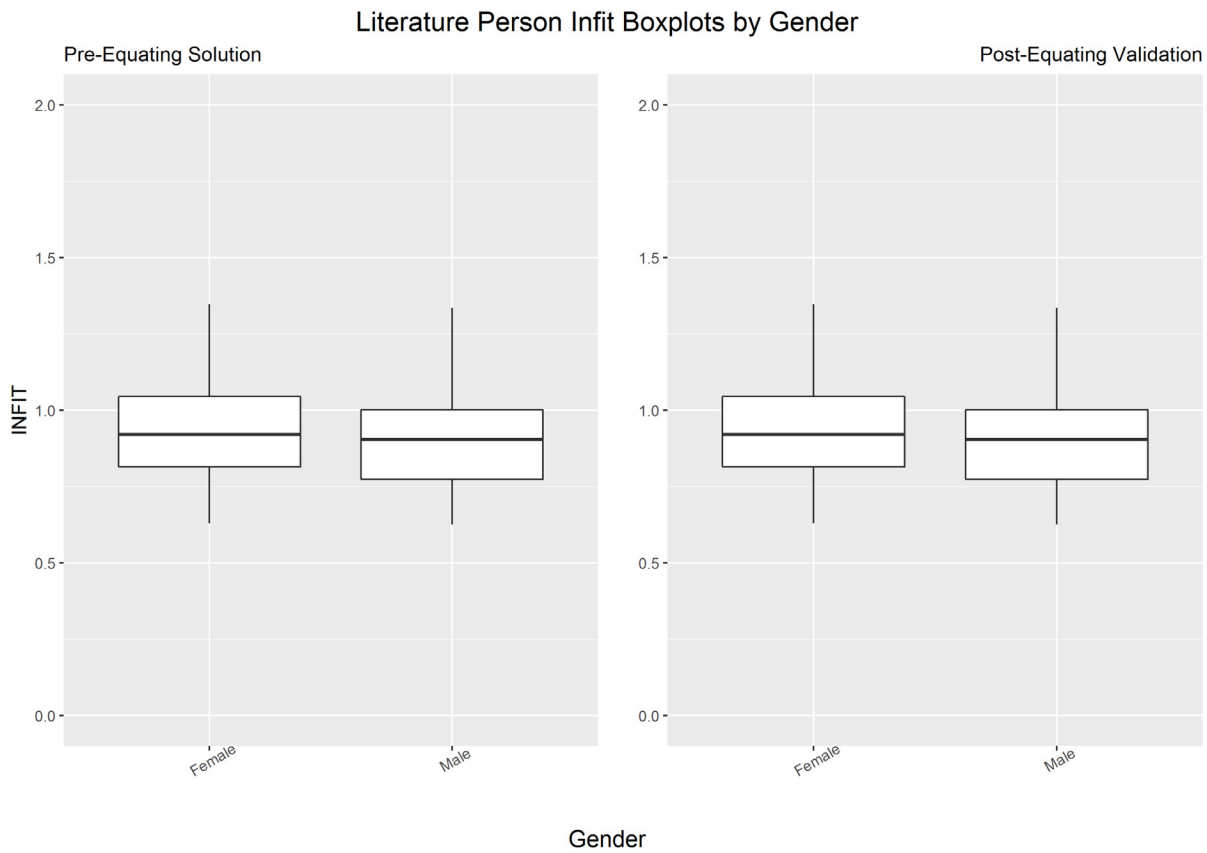
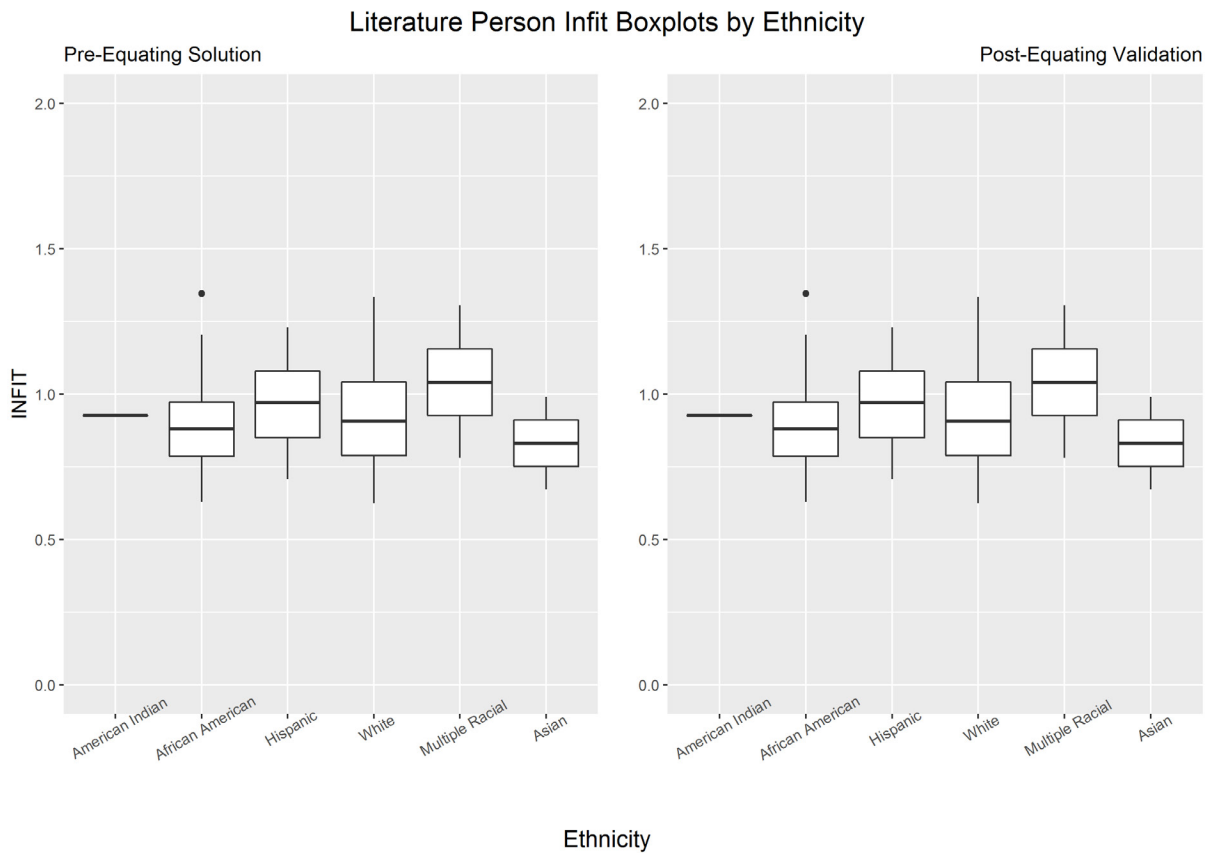


Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer

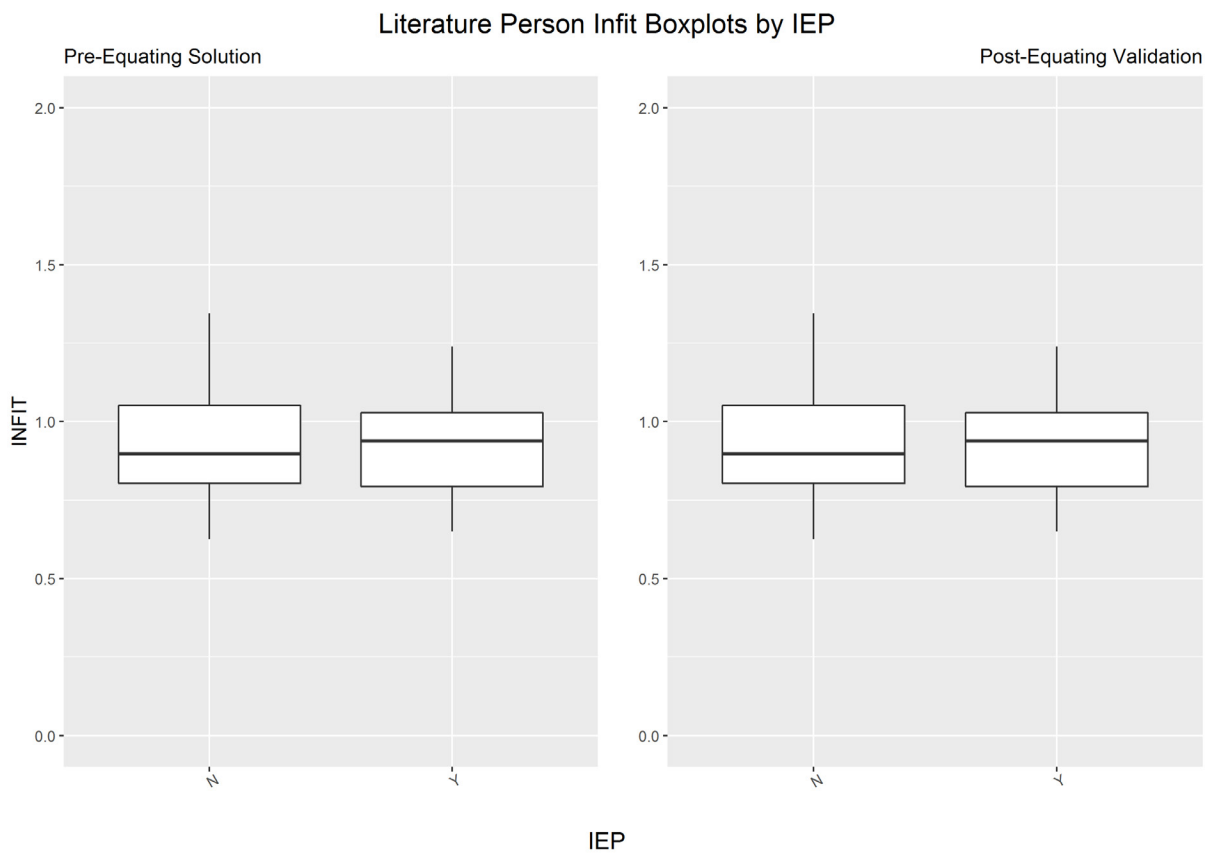
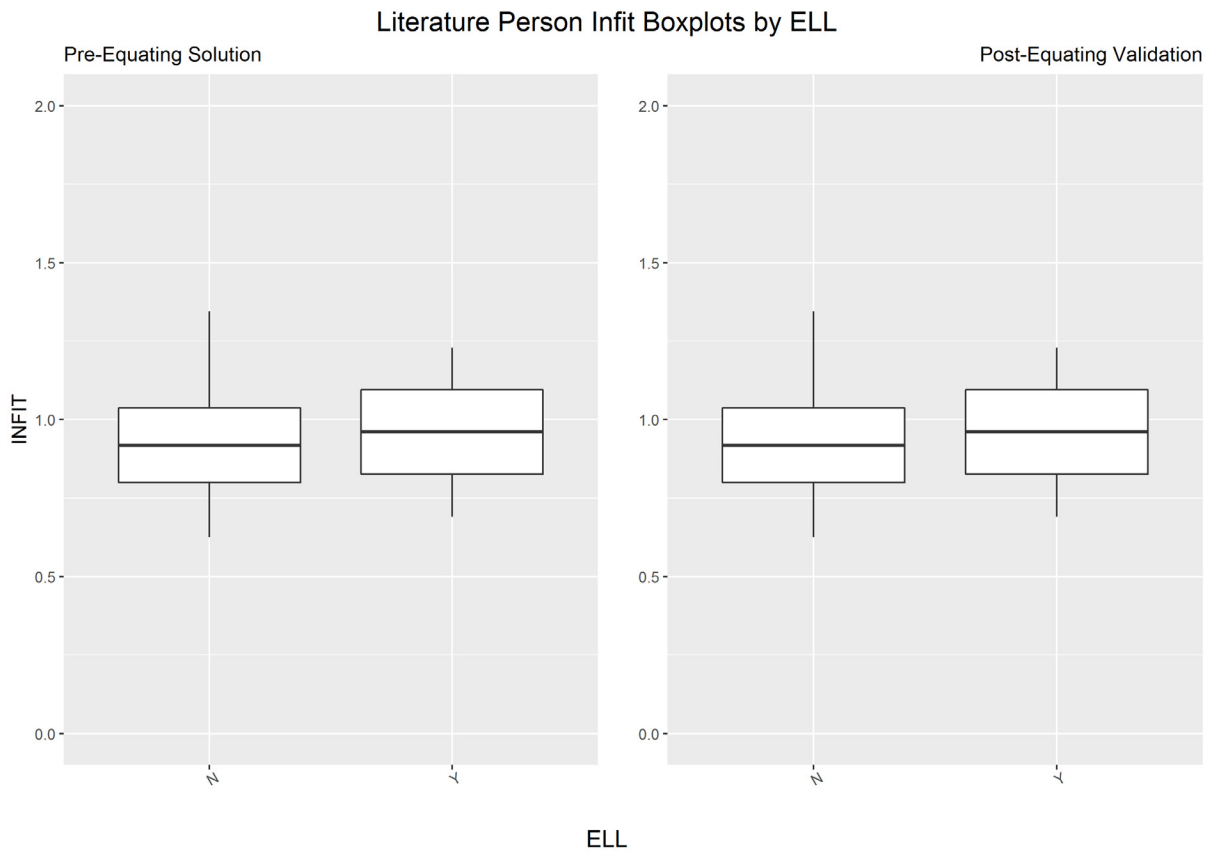


Figure L-5 (continued). Person Infit Boxplots by Content Area for Pre- and Post-Equated Solutions: Summer

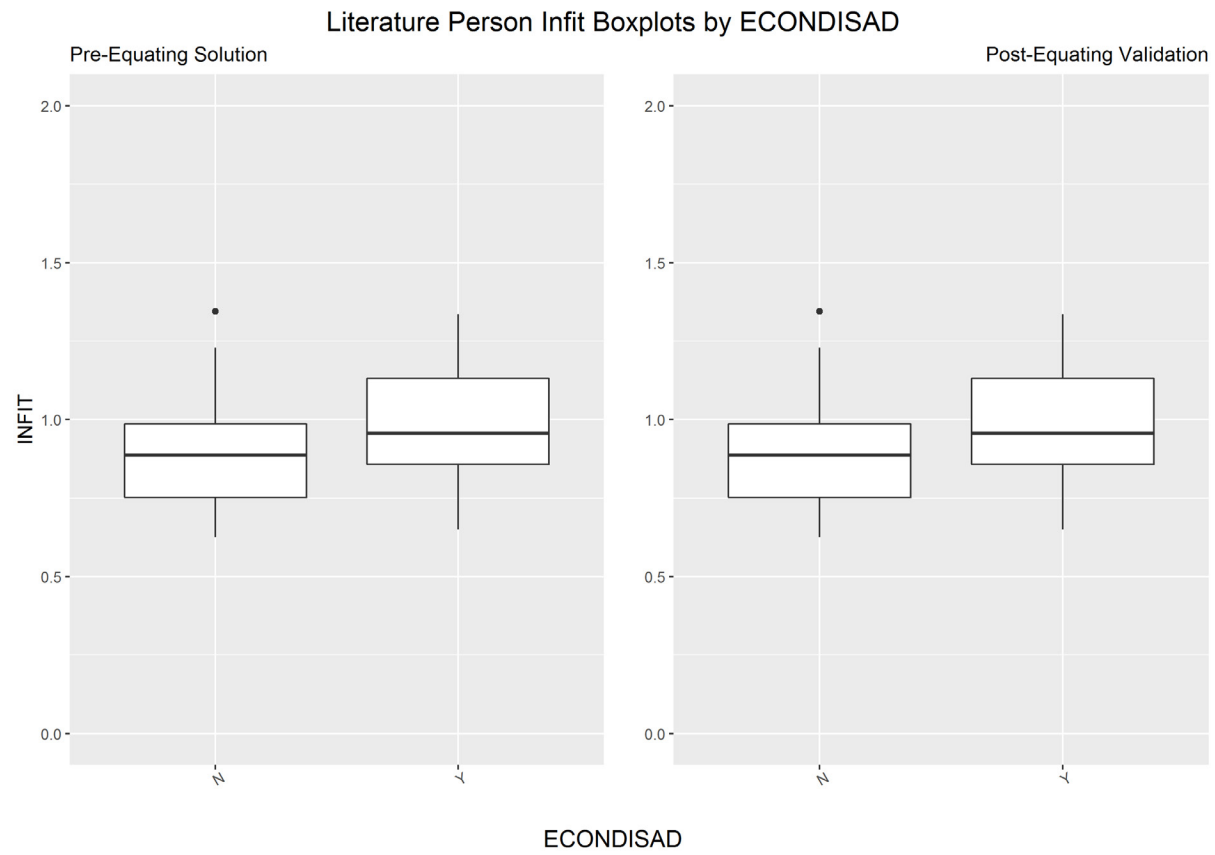
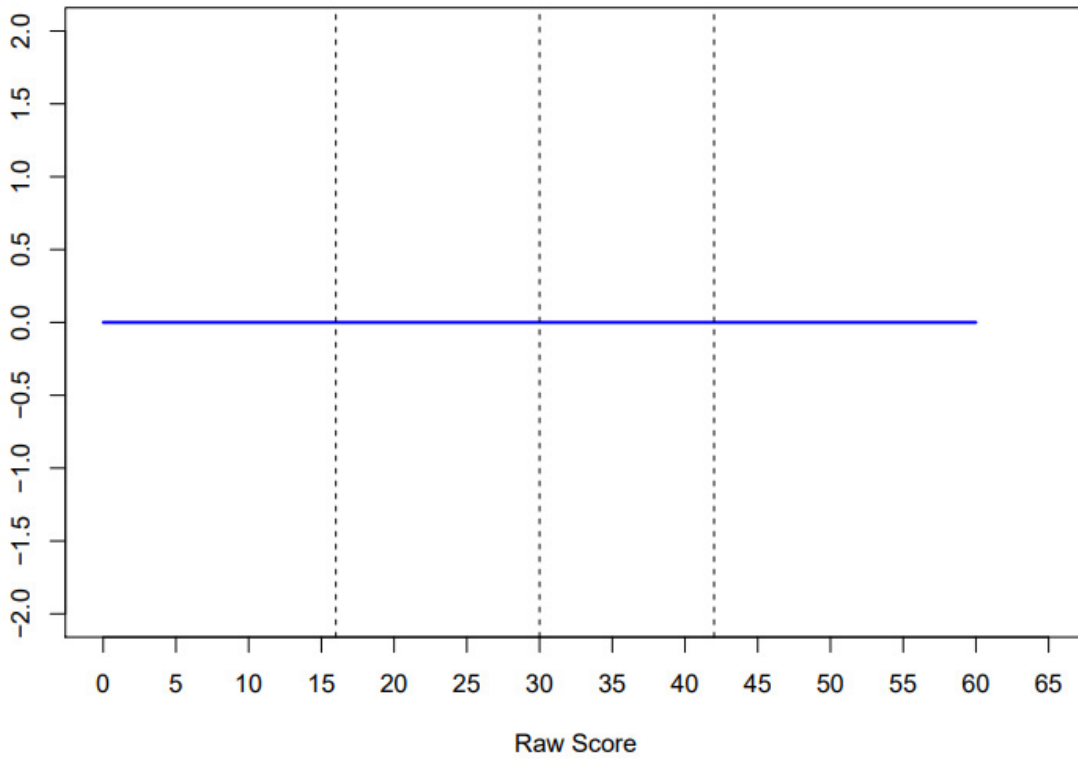


Figure L-6. Normalized Scale Score Differences by Content Area: Summer

### Algebra I



### Biology

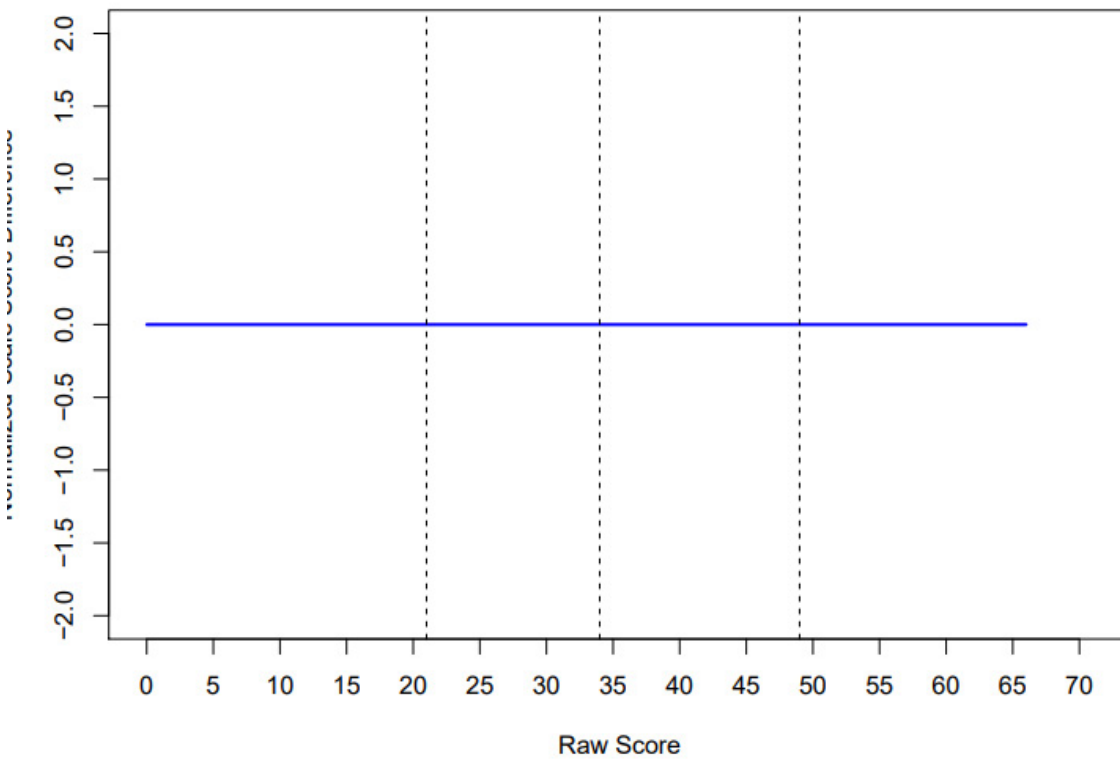
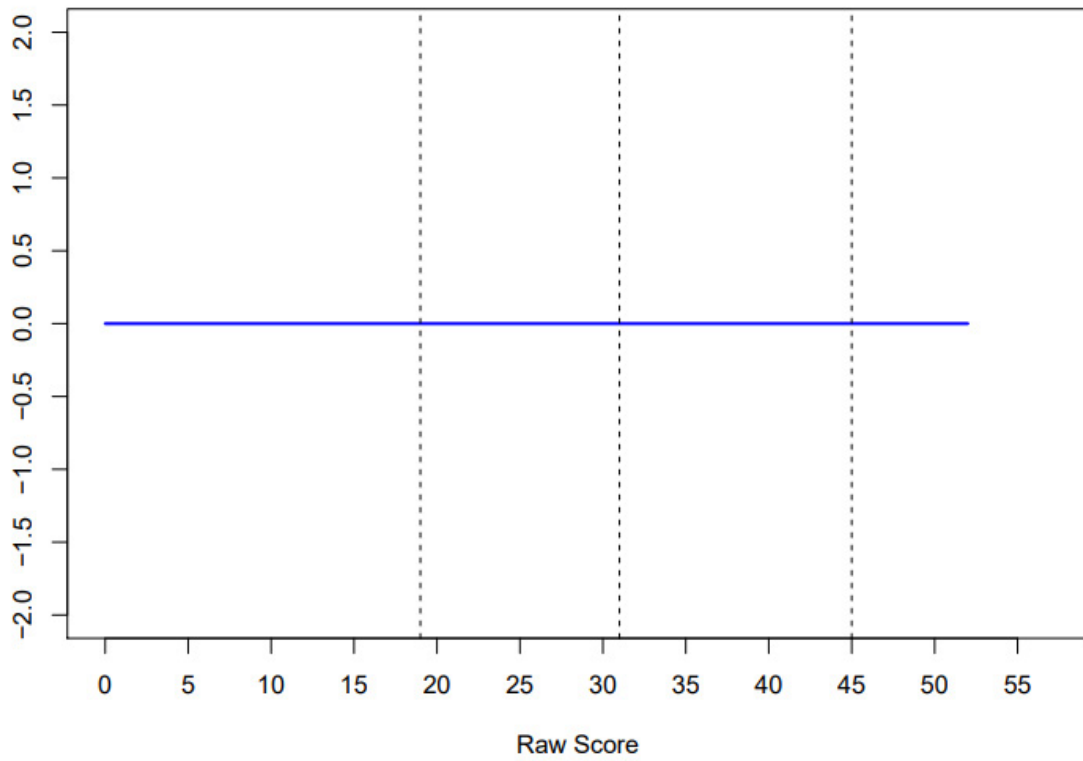




Figure L-6 (continued). Normalized Scale Score Differences by Content Area: Summer  
**Literature**



**Table L-7. Summer Algebra I Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1217	1217	92	92	BB	BB	0.0	TRUE
1	1278	1278	51	51	BB	BB	0.0	TRUE
2	1314	1314	36	36	BB	BB	0.0	TRUE
3	1336	1336	30	30	BB	BB	0.0	TRUE
4	1352	1352	26	26	BB	BB	0.0	TRUE
5	1365	1365	24	24	BB	BB	0.7	TRUE
6	1375	1375	22	22	BB	BB	0.4	TRUE
7	1385	1385	21	21	BB	BB	0.4	TRUE
8	1393	1393	20	20	BB	BB	0.7	TRUE
9	1401	1401	19	19	BB	BB	0.0	TRUE
10	1407	1407	18	18	BB	BB	0.7	TRUE
11	1414	1414	18	18	BB	BB	2.2	TRUE
12	1420	1420	17	17	BB	BB	2.2	TRUE
13	1426	1426	17	17	BB	BB	2.6	TRUE
14	1431	1431	16	16	BB	BB	3.3	TRUE
15	1436	1436	16	16	BB	BB	2.6	TRUE
16	1441	1441	16	16	B	B	4.0	TRUE
17	1446	1446	15	15	B	B	3.3	TRUE
18	1451	1451	15	15	B	B	4.4	TRUE
19	1455	1455	15	15	B	B	6.2	TRUE
20	1460	1460	15	15	B	B	5.5	TRUE
21	1464	1464	15	15	B	B	5.1	TRUE
22	1468	1468	14	14	B	B	2.9	TRUE
23	1473	1473	14	14	B	B	5.1	TRUE
24	1477	1477	14	14	B	B	4.7	TRUE
25	1481	1481	14	14	B	B	5.1	TRUE
26	1485	1485	14	14	B	B	1.8	TRUE
27	1488	1488	14	14	B	B	2.9	TRUE
28	1492	1492	14	14	B	B	2.2	TRUE
29	1496	1496	14	14	B	B	3.3	TRUE
30	1500	1500	14	14	P	P	2.2	TRUE
31	1504	1504	14	14	P	P	3.6	TRUE
32	1507	1507	14	14	P	P	1.8	TRUE
33	1511	1511	14	14	P	P	2.2	TRUE
34	1515	1515	14	14	P	P	3.6	TRUE
35	1519	1519	14	14	P	P	2.2	TRUE
36	1522	1522	14	14	P	P	1.1	TRUE
37	1526	1526	14	14	P	P	1.1	TRUE
38	1530	1530	14	14	P	P	0.4	TRUE
39	1534	1534	14	14	P	P	1.5	TRUE
40	1538	1538	14	14	P	P	2.2	TRUE
41	1542	1542	15	15	P	P	0.7	TRUE

**Table L-7 (continued). Summer Algebra I Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
42	1547	1547	15	15	A	A	0.0	TRUE
43	1551	1551	15	15	A	A	0.4	TRUE
44	1556	1556	15	15	A	A	1.1	TRUE
45	1560	1560	16	16	A	A	0.4	TRUE
46	1565	1565	16	16	A	A	0.4	TRUE
47	1571	1571	16	16	A	A	0.7	TRUE
48	1576	1576	17	17	A	A	0.0	TRUE
49	1582	1582	17	17	A	A	0.4	TRUE
50	1588	1588	18	18	A	A	0.7	TRUE
51	1595	1595	19	19	A	A	0.4	TRUE
52	1602	1602	20	20	A	A	0.0	TRUE
53	1611	1611	21	21	A	A	0.0	TRUE
54	1620	1620	22	22	A	A	0.0	TRUE
55	1630	1630	24	24	A	A	0.4	TRUE
56	1643	1643	27	27	A	A	0.0	TRUE
57	1659	1659	30	30	A	A	0.4	TRUE
58	1681	1681	37	37	A	A	0.0	TRUE
59	1718	1718	51	51	A	A	0.0	TRUE
60	1779	1779	92	92	A	A	0.0	TRUE

**Table L-8. Summer Biology Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1228	1228	92	92	BB	BB	0.0	TRUE
1	1289	1289	50	50	BB	BB	0.0	TRUE
2	1325	1325	36	36	BB	BB	0.0	TRUE
3	1346	1346	30	30	BB	BB	0.0	TRUE
4	1361	1361	26	26	BB	BB	0.0	TRUE
5	1373	1373	23	23	BB	BB	0.0	TRUE
6	1383	1383	22	22	BB	BB	0.0	TRUE
7	1392	1392	20	20	BB	BB	0.0	TRUE
8	1399	1399	19	19	BB	BB	0.5	TRUE
9	1406	1406	18	18	BB	BB	0.5	TRUE
10	1412	1412	17	17	BB	BB	0.5	TRUE
11	1418	1418	17	17	BB	BB	1.0	TRUE
12	1423	1423	16	16	BB	BB	0.5	TRUE
13	1428	1428	16	16	BB	BB	1.0	TRUE
14	1433	1433	15	15	BB	BB	2.4	TRUE
15	1438	1438	15	15	BB	BB	2.4	TRUE
16	1442	1442	14	14	BB	BB	2.4	TRUE
17	1446	1446	14	14	BB	BB	1.9	TRUE
18	1450	1450	14	14	BB	BB	1.4	TRUE
19	1454	1454	14	14	BB	BB	2.9	TRUE
20	1457	1457	13	13	BB	BB	5.2	TRUE
21	1461	1461	13	13	B	B	3.3	TRUE
22	1464	1464	13	13	B	B	2.4	TRUE
23	1468	1468	13	13	B	B	3.8	TRUE
24	1471	1471	13	13	B	B	4.8	TRUE
25	1474	1474	13	13	B	B	3.3	TRUE
26	1477	1477	13	13	B	B	1.4	TRUE
27	1481	1481	13	13	B	B	2.9	TRUE
28	1484	1484	12	12	B	B	4.3	TRUE
29	1487	1487	12	12	B	B	5.2	TRUE
30	1490	1490	12	12	B	B	4.3	TRUE
31	1493	1493	12	12	B	B	5.2	TRUE
32	1496	1496	12	12	B	B	3.3	TRUE
33	1499	1499	12	12	B	B	1.9	TRUE
34	1502	1502	12	12	P	P	4.8	TRUE
35	1505	1505	12	12	P	P	2.4	TRUE
36	1508	1508	12	12	P	P	2.4	TRUE
37	1511	1511	12	12	P	P	1.4	TRUE
38	1514	1514	12	12	P	P	1.4	TRUE
39	1517	1517	12	12	P	P	0.5	TRUE
40	1521	1521	13	13	P	P	0.5	TRUE
41	1524	1524	13	13	P	P	1.0	TRUE

**Table L-8 (continued). Summer Biology Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
42	1527	1527	13	13	P	P	0.5	TRUE
43	1530	1530	13	13	P	P	1.0	TRUE
44	1534	1534	13	13	P	P	1.0	TRUE
45	1537	1537	13	13	P	P	0.5	TRUE
46	1540	1540	13	13	P	P	0.0	TRUE
47	1544	1544	14	14	P	P	1.0	TRUE
48	1548	1548	14	14	P	P	1.0	TRUE
49	1552	1552	14	14	A	A	0.5	TRUE
50	1556	1556	14	14	A	A	1.0	TRUE
51	1560	1560	15	15	A	A	1.0	TRUE
52	1564	1564	15	15	A	A	0.5	TRUE
53	1569	1569	15	15	A	A	0.0	TRUE
54	1574	1574	16	16	A	A	0.5	TRUE
55	1579	1579	16	16	A	A	0.5	TRUE
56	1584	1584	17	17	A	A	0.5	TRUE
57	1591	1591	18	18	A	A	0.5	TRUE
58	1597	1597	19	19	A	A	1.4	TRUE
59	1605	1605	20	20	A	A	1.4	TRUE
60	1613	1613	21	21	A	A	1.4	TRUE
61	1623	1623	23	23	A	A	1.4	TRUE
62	1635	1635	26	26	A	A	1.0	TRUE
63	1650	1650	30	30	A	A	0.5	TRUE
64	1671	1671	36	36	A	A	0.5	TRUE
65	1707	1707	50	50	A	A	0.0	TRUE
66	1767	1767	92	92	A	A	0.0	TRUE

**Table L-9. Summer Literature Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
0	1209	1209	92	92	BB	BB	0.0	TRUE
1	1271	1271	51	51	BB	BB	0.0	TRUE
2	1307	1307	36	36	BB	BB	0.0	TRUE
3	1328	1328	30	30	BB	BB	0.0	TRUE
4	1344	1344	26	26	BB	BB	0.0	TRUE
5	1357	1357	24	24	BB	BB	0.9	TRUE
6	1368	1368	22	22	BB	BB	0.0	TRUE
7	1377	1377	21	21	BB	BB	0.9	TRUE
8	1385	1385	20	20	BB	BB	0.9	TRUE
9	1393	1393	19	19	BB	BB	0.9	TRUE
10	1399	1399	18	18	BB	BB	3.6	TRUE
11	1406	1406	18	18	BB	BB	3.6	TRUE
12	1412	1412	17	17	BB	BB	1.8	TRUE
13	1418	1418	17	17	BB	BB	4.5	TRUE
14	1423	1423	16	16	BB	BB	3.6	TRUE
15	1428	1428	16	16	BB	BB	1.8	TRUE
16	1433	1433	16	16	BB	BB	0.9	TRUE
17	1438	1438	16	16	BB	BB	1.8	TRUE
18	1443	1443	15	15	BB	BB	1.8	TRUE
19	1448	1448	15	15	B	B	2.7	TRUE
20	1453	1453	15	15	B	B	6.3	TRUE
21	1457	1457	15	15	B	B	4.5	TRUE
22	1462	1462	15	15	B	B	3.6	TRUE
23	1466	1466	15	15	B	B	3.6	TRUE
24	1471	1471	15	15	B	B	4.5	TRUE
25	1475	1475	15	15	B	B	3.6	TRUE
26	1480	1480	15	15	B	B	4.5	TRUE
27	1484	1484	15	15	B	B	2.7	TRUE
28	1489	1489	15	15	B	B	7.2	TRUE
29	1493	1493	15	15	B	B	3.6	TRUE
30	1498	1498	15	15	B	B	6.3	TRUE
31	1503	1503	15	15	P	P	2.7	TRUE
32	1508	1508	16	16	P	P	5.4	TRUE
33	1513	1513	16	16	P	P	1.8	TRUE
34	1518	1518	16	16	P	P	1.8	TRUE
35	1523	1523	16	16	P	P	0.9	TRUE
36	1528	1528	17	17	P	P	0.9	TRUE
37	1534	1534	17	17	P	P	0.9	TRUE
38	1540	1540	17	17	P	P	2.7	TRUE
39	1546	1546	18	18	P	P	0.0	TRUE
40	1552	1552	18	18	P	P	1.8	TRUE
41	1559	1559	19	19	P	P	0.0	TRUE

**Table L-9 (continued). Summer Literature Raw-to-Scaled Score Comparison for Pre-Equated and Post-Equated Solutions**

Raw Score	Pre-SS	Post-SS	Pre-SEM	Post-SEM	Pre-PL	Post-PL	Proportion (%)	Same PL
42	1567	1567	19	19	P	P	0.9	TRUE
43	1574	1574	20	20	P	P	0.0	TRUE
44	1583	1583	21	21	P	P	0.0	TRUE
45	1592	1592	22	22	A	A	0.0	TRUE
46	1603	1603	23	23	A	A	0.0	TRUE
47	1614	1614	25	25	A	A	0.0	TRUE
48	1628	1628	28	28	A	A	0.0	TRUE
49	1645	1645	31	31	A	A	0.0	TRUE
50	1669	1669	37	37	A	A	0.0	TRUE
51	1706	1706	51	51	A	A	0.0	TRUE
52	1768	1768	92	92	A	A	0.0	TRUE

## APPENDIX M: RELIABILITIES

---

Table M-1. Reliabilities

Column Heading	Definition
Level	Total test or module level
Group	Student group: all students or subgroup
Pts.	Max points possible
Len.	Test length
<i>N</i>	Number of students
Mean	Mean of raw score
SD	Standard deviation of raw score
<i>r</i>	Reliability coefficient: Cronbach's alpha
<i>SEM</i>	Standard error of measurement

Note. "DNR" in the tables below represents "Do Not Report," which occurs when the *N* count was too small to report.



**Table M–2. Winter: Algebra I Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	60	42	54824	20.31	9.81	0.88	3.35
	Module 1	All	30	21	54824	10.30	5.09	0.78	2.40
	Module 2	All	30	21	54824	10.02	5.29	0.81	2.33
Gender	Total	Female	60	42	27137	20.51	9.56	0.88	3.33
		Male	60	42	27676	20.12	10.05	0.89	3.36
	Module 1	Female	30	21	27137	10.51	5.03	0.77	2.40
		Male	30	21	27676	10.09	5.14	0.78	2.40
	Module 2	Female	30	21	27137	10.00	5.08	0.79	2.31
		Male	30	21	27676	10.03	5.48	0.82	2.35
Ethnicity	Total	African American	60	42	10408	15.73	7.67	0.83	3.13
		American Indian	60	42	105	18.93	8.98	0.86	3.33
		Asian	60	42	1703	26.12	12.56	0.92	3.54
		Hispanic	60	42	8962	17.72	8.73	0.86	3.23
		Multi-racial	60	42	2400	20.01	9.77	0.88	3.35
		Native Hawaiian/ Pacific Islander	60	42	49	21.08	9.75	0.89	3.31
		White	60	42	31187	22.30	9.83	0.88	3.41
	Module 1	African American	30	21	10408	8.05	4.24	0.71	2.26
		American Indian	30	21	105	9.70	4.73	0.73	2.44
		Asian	30	21	1703	13.23	6.33	0.84	2.52
		Hispanic	30	21	8962	9.07	4.67	0.75	2.33
		Multi-racial	30	21	2400	10.14	5.08	0.78	2.41
		Native Hawaiian/ Pacific Islander	30	21	49	11.10	5.49	0.81	2.39
		White	30	21	31187	11.25	5.05	0.77	2.44
	Module 2	African American	30	21	10408	7.68	4.06	0.72	2.16
		American Indian	30	21	105	9.24	4.79	0.78	2.26
		Asian	30	21	1703	12.88	6.68	0.86	2.49
		Hispanic	30	21	8962	8.65	4.65	0.77	2.24
		Multi-racial	30	21	2400	9.87	5.25	0.80	2.33
		Native Hawaiian/ Pacific Islander	30	21	49	9.98	4.94	0.79	2.27
		White	30	21	31187	11.05	5.37	0.80	2.38
EL	Total	All	60	42	2790	14.29	7.19	0.82	3.01
	Module 1	All	30	21	2790	7.42	4.06	0.72	2.17
	Module 2	All	30	21	2790	6.87	3.77	0.69	2.09
IEP	Total	All	60	42	9912	14.62	7.18	0.82	3.06
	Module 1	All	30	21	9912	7.40	3.92	0.69	2.19
	Module 2	All	30	21	9912	7.22	3.94	0.71	2.14
ED	Total	All	60	42	28057	18.03	8.67	0.86	3.25
	Module 1	All	30	21	28057	9.19	4.62	0.74	2.34
	Module 2	All	30	21	28057	8.84	4.66	0.77	2.25

**Table M–3. Winter: Biology Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	66	54	40077	24.85	11.90	0.91	3.67
	Module 1	All	33	27	40077	11.99	6.58	0.85	2.57
	Module 2	All	33	27	40077	12.86	5.96	0.81	2.60
Gender	Total	Female	66	54	19297	25.39	11.74	0.90	3.68
		Male	66	54	20774	24.34	12.03	0.91	3.65
	Module 1	Female	33	27	19297	12.24	6.52	0.84	2.59
		Male	33	27	20774	11.75	6.62	0.85	2.56
	Module 2	Female	33	27	19297	13.15	5.85	0.80	2.61
		Male	33	27	20774	12.59	6.05	0.82	2.59
Ethnicity	Total	African American	66	54	7854	19.85	8.69	0.83	3.54
		American Indian	66	54	75	24.43	13.21	0.92	3.67
		Asian	66	54	1266	30.63	14.19	0.93	3.74
		Hispanic	66	54	6170	21.53	9.79	0.86	3.60
		Multi-racial	66	54	1829	24.54	11.72	0.90	3.65
		Native Hawaiian/ Pacific Islander	66	54	54	26.37	12.28	0.91	3.74
		White	66	54	22823	27.16	12.45	0.91	3.70
	Module 1	African American	33	27	7854	9.42	4.91	0.75	2.46
		American Indian	33	27	75	11.83	7.44	0.89	2.52
		Asian	33	27	1266	15.07	7.88	0.89	2.64
		Hispanic	33	27	6170	10.18	5.47	0.79	2.52
		Multi-racial	33	27	1829	11.86	6.48	0.84	2.56
		Native Hawaiian/ Pacific Islander	33	27	54	12.57	6.38	0.83	2.66
		White	33	27	22823	13.20	6.88	0.86	2.61
	Module 2	African American	33	27	7854	10.43	4.58	0.69	2.54
		American Indian	33	27	75	12.60	6.31	0.82	2.66
		Asian	33	27	1266	15.56	6.88	0.85	2.64
		Hispanic	33	27	6170	11.35	5.05	0.74	2.57
		Multi-racial	33	27	1829	12.68	5.88	0.81	2.59
		Native Hawaiian/ Pacific Islander	33	27	54	13.80	6.34	0.83	2.63
		White	33	27	22823	13.97	6.19	0.82	2.62
EL	Total	All	66	54	2021	17.95	7.39	0.78	3.48
	Module 1	All	33	27	2021	8.42	4.19	0.67	2.41
	Module 2	All	33	27	2021	9.53	4.08	0.62	2.50
IEP	Total	All	66	54	8464	19.00	8.33	0.82	3.52
	Module 1	All	33	27	8464	9.00	4.72	0.73	2.44
	Module 2	All	33	27	8464	9.99	4.44	0.68	2.53
ED	Total	All	66	54	20750	21.96	9.87	0.87	3.61
	Module 1	All	33	27	20750	10.51	5.54	0.79	2.53
	Module 2	All	33	27	20750	11.46	5.08	0.74	2.57

**Table M–4. Winter: Literature Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	52	40	35222	24.80	10.61	0.91	3.13
	Module 1	All	26	20	35222	12.50	5.48	0.84	2.23
	Module 2	All	26	20	35222	12.30	5.63	0.85	2.20
Gender	Total	Female	52	40	16145	26.57	10.45	0.91	3.13
		Male	52	40	19071	23.30	10.52	0.91	3.11
	Module 1	Female	26	20	16145	13.44	5.41	0.83	2.23
		Male	26	20	19071	11.71	5.42	0.83	2.21
	Module 2	Female	26	20	16145	13.13	5.54	0.84	2.19
		Male	26	20	19071	11.59	5.61	0.85	2.18
Ethnicity	Total	African American	52	40	6732	21.08	9.03	0.88	3.12
		American Indian	52	40	68	22.71	10.40	0.91	3.05
		Asian	52	40	1125	29.43	11.11	0.92	3.08
		Hispanic	52	40	5651	22.34	9.72	0.90	3.11
		Multi-racial	52	40	1595	25.00	10.66	0.91	3.13
		Native Hawaiian/ Pacific Islander	52	40	35	24.86	10.85	0.92	3.09
		White	52	40	20011	26.47	10.83	0.92	3.13
	Module 1	African American	26	20	6732	10.72	4.78	0.79	2.20
		American Indian	26	20	68	11.38	5.53	0.85	2.14
		Asian	26	20	1125	14.69	5.69	0.85	2.22
		Hispanic	26	20	5651	11.38	5.14	0.82	2.20
		Multi-racial	26	20	1595	12.59	5.50	0.84	2.22
		Native Hawaiian/ Pacific Islander	26	20	35	12.17	5.69	0.85	2.17
		White	26	20	20011	13.29	5.57	0.84	2.22
	Module 2	African American	26	20	6732	10.36	4.87	0.80	2.20
		American Indian	26	20	68	11.32	5.46	0.84	2.15
		Asian	26	20	1125	14.74	5.84	0.87	2.14
		Hispanic	26	20	5651	10.96	5.14	0.82	2.19
		Multi-racial	26	20	1595	12.41	5.69	0.85	2.20
		Native Hawaiian/ Pacific Islander	26	20	35	12.69	5.57	0.84	2.19
		White	26	20	20011	13.18	5.73	0.85	2.19
EL	Total	All	52	40	1759	17.08	7.21	0.82	3.07
	Module 1	All	26	20	1759	8.68	3.97	0.70	2.16
	Module 2	All	26	20	1759	8.41	3.92	0.69	2.17
IEP	Total	All	52	40	7965	17.96	8.13	0.86	3.05
	Module 1	All	26	20	7965	9.19	4.38	0.76	2.15
	Module 2	All	26	20	7965	8.77	4.41	0.76	2.15
ED	Total	All	52	40	18403	22.26	9.65	0.90	3.13
	Module 1	All	26	20	18403	11.30	5.07	0.81	2.21
	Module 2	All	26	20	18403	10.96	5.15	0.82	2.20

**Table M–5. Spring: Algebra Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	60	42	162971	23.59	12.26	0.92	3.55
	Module 1	All	30	21	162971	11.83	5.94	0.84	2.37
	Module 2	All	30	21	162971	11.76	6.83	0.85	2.63
Gender	Total	Female	60	42	79631	23.66	12.05	0.91	3.55
		Male	60	42	83257	23.53	12.47	0.92	3.53
	Module 1	Female	30	21	79631	12.02	5.79	0.83	2.36
		Male	30	21	83257	11.64	6.07	0.85	2.38
	Module 2	Female	30	21	79631	11.63	6.76	0.85	2.65
		Male	30	21	83257	11.88	6.89	0.86	2.61
Ethnicity	Total	African American	60	42	26455	16.41	8.39	0.85	3.22
		American Indian	60	42	283	20.80	11.32	0.91	3.42
		Asian	60	42	7134	33.30	14.19	0.93	3.70
		Hispanic	60	42	24646	18.17	9.68	0.88	3.31
		Multi-racial	60	42	7342	22.63	12.03	0.91	3.52
		Native Hawaiian/ Pacific Islander	60	42	163	22.52	11.67	0.91	3.54
		White	60	42	96856	26.30	12.15	0.91	3.59
	Module 1	African American	30	21	26455	8.53	4.26	0.73	2.20
		American Indian	30	21	283	10.46	5.57	0.83	2.31
		Asian	30	21	7134	16.51	6.96	0.87	2.49
		Hispanic	30	21	24646	9.33	4.78	0.78	2.25
		Multi-racial	30	21	7342	11.42	5.82	0.83	2.37
		Native Hawaiian/ Pacific Islander	30	21	163	11.17	5.60	0.83	2.27
		White	30	21	96856	13.06	5.88	0.83	2.39
	Module 2	African American	30	21	26455	7.88	4.77	0.76	2.34
		American Indian	30	21	283	10.33	6.24	0.84	2.53
		Asian	30	21	7134	16.79	7.67	0.87	2.74
		Hispanic	30	21	24646	8.84	5.49	0.80	2.42
		Multi-racial	30	21	7342	11.21	6.72	0.85	2.61
		Native Hawaiian/ Pacific Islander	30	21	163	11.35	6.77	0.84	2.69
		White	30	21	96856	13.24	6.80	0.84	2.68
EL	Total	All	60	42	9197	15.03	8.01	0.85	3.12
	Module 1	All	30	21	9197	7.87	4.16	0.73	2.18
	Module 2	All	30	21	9197	7.16	4.49	0.75	2.23
IEP	Total	All	60	42	28393	15.59	8.24	0.85	3.15
	Module 1	All	30	21	28393	8.06	4.17	0.73	2.18
	Module 2	All	30	21	28393	7.53	4.70	0.77	2.26
ED	Total	All	60	42	78500	19.39	10.11	0.89	3.37
	Module 1	All	30	21	78500	9.89	4.96	0.79	2.28
	Module 2	All	30	21	78500	9.50	5.73	0.81	2.48

**Table M–6. Spring: Biology Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	66	54	137797	31.22	15.03	0.94	3.70
	Module 1	All	33	27	137797	15.42	7.56	0.88	2.58
	Module 2	All	33	27	137797	15.80	7.95	0.89	2.65
Gender	Total	Female	66	54	67443	31.41	14.69	0.94	3.71
		Male	66	54	70329	31.03	15.34	0.94	3.68
	Module 1	Female	33	27	67443	15.40	7.35	0.88	2.59
		Male	33	27	70329	15.44	7.76	0.89	2.56
	Module 2	Female	33	27	67443	16.01	7.83	0.88	2.66
		Male	33	27	70329	15.59	8.05	0.89	2.63
Ethnicity	Total	African American	66	54	244	28.14	13.33	0.92	3.71
		American Indian	66	54	20892	22.29	10.90	0.89	3.59
		Asian	66	54	6126	41.50	16.29	0.95	3.58
		Hispanic	66	54	19269	24.59	12.32	0.91	3.64
		Multi-racial	66	54	5984	29.50	14.76	0.94	3.69
		Native Hawaiian/ Pacific Islander	66	54	125	31.18	14.96	0.94	3.72
		White	66	54	85129	34.30	14.83	0.94	3.70
	Module 1	African American	33	27	20892	11.08	5.62	0.80	2.52
		American Indian	33	27	244	14.02	6.77	0.85	2.58
		Asian	33	27	6126	20.44	8.26	0.91	2.50
		Hispanic	33	27	19269	12.21	6.33	0.84	2.55
		Multi-racial	33	27	5984	14.57	7.45	0.88	2.57
		Native Hawaiian/ Pacific Islander	33	27	125	15.54	7.33	0.87	2.60
		White	33	27	85129	16.91	7.46	0.88	2.57
	Module 2	African American	33	27	20892	11.21	5.91	0.81	2.55
		American Indian	33	27	244	14.12	7.15	0.86	2.66
		Asian	33	27	6126	21.06	8.43	0.91	2.55
		Hispanic	33	27	19269	12.38	6.57	0.84	2.60
		Multi-racial	33	27	5984	14.93	7.80	0.89	2.64
		Native Hawaiian/ Pacific Islander	33	27	125	15.64	8.07	0.89	2.66
		White	33	27	85129	17.39	7.86	0.89	2.65
EL	Total	All	66	54	6752	19.40	9.15	0.85	3.50
	Module 1	All	33	27	6752	9.66	4.90	0.75	2.46
	Module 2	All	33	27	6752	9.74	4.94	0.74	2.49
IEP	Total	All	66	54	25164	21.18	10.77	0.89	3.55
	Module 1	All	33	27	25164	10.63	5.58	0.80	2.50
	Module 2	All	33	27	25164	10.56	5.82	0.81	2.51
ED	Total	All	66	54	64475	25.87	12.85	0.92	3.67
	Module 1	All	33	27	64475	12.84	6.54	0.85	2.57
	Module 2	All	33	27	64475	13.03	6.87	0.86	2.61

**Table M-7. Spring: Literature Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	52	40	125553	30.08	10.53	0.91	3.16
	Module 1	All	26	20	125553	15.30	5.49	0.84	2.22
	Module 2	All	26	20	125553	14.78	5.55	0.84	2.23
Gender	Total	Female	52	40	60556	31.98	9.89	0.90	3.13
		Male	52	40	64975	28.30	10.79	0.92	3.15
	Module 1	Female	26	20	60556	16.52	5.11	0.82	2.18
		Male	26	20	64975	14.16	5.60	0.84	2.23
	Module 2	Female	26	20	60556	15.46	5.31	0.82	2.23
		Male	26	20	64975	14.14	5.69	0.85	2.21
Ethnicity	Total	African American	52	40	18706	24.58	9.68	0.89	3.22
		American Indian	52	40	225	27.72	10.29	0.90	3.19
		Asian	52	40	5551	35.92	9.56	0.90	2.96
		Hispanic	52	40	17232	25.80	10.22	0.90	3.21
		Multi-racial	52	40	5315	29.17	10.48	0.91	3.18
		Native Hawaiian/ Pacific Islander	52	40	125	31.12	9.82	0.90	3.10
		White	52	40	78370	31.99	10.02	0.90	3.10
	Module 1	African American	26	20	18706	12.67	5.23	0.81	2.27
		American Indian	26	20	225	13.97	5.44	0.83	2.24
		Asian	26	20	5551	17.83	4.97	0.82	2.11
		Hispanic	26	20	17232	13.14	5.49	0.83	2.26
		Multi-racial	26	20	5315	14.91	5.52	0.84	2.24
		Native Hawaiian/ Pacific Islander	26	20	125	15.75	5.48	0.84	2.19
		White	26	20	78370	16.25	5.21	0.83	2.18
	Module 2	African American	26	20	18706	11.91	5.10	0.80	2.26
		American Indian	26	20	225	13.75	5.39	0.82	2.26
		Asian	26	20	5551	18.09	5.03	0.83	2.07
		Hispanic	26	20	17232	12.66	5.32	0.82	2.26
		Multi-racial	26	20	5315	14.26	5.49	0.83	2.25
		Native Hawaiian/ Pacific Islander	26	20	125	15.37	4.81	0.79	2.19
		White	26	20	78370	15.73	5.32	0.83	2.20
EL	Total	All	52	40	5318	19.93	8.52	0.86	3.20
	Module 1	All	26	20	5318	9.81	4.63	0.76	2.27
	Module 2	All	26	20	5318	10.12	4.56	0.76	2.25
IEP	Total	All	52	40	23755	21.05	9.18	0.88	3.16
	Module 1	All	26	20	23755	10.81	5.03	0.80	2.24
	Module 2	All	26	20	23755	10.24	4.80	0.79	2.22
ED	Total	All	52	40	57746	26.53	10.11	0.90	3.20
	Module 1	All	26	20	57746	13.61	5.39	0.82	2.25
	Module 2	All	26	20	57746	12.92	5.31	0.82	2.26

**Table M–8. Summer: Algebra Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	60	42	274	24.53	9.55	0.87	3.50
	Module 1	All	30	21	274	13.01	5.27	0.76	2.58
	Module 2	All	30	21	274	11.51	4.87	0.76	2.37
Gender	Total	Female	60	42	145	24.70	9.83	0.87	3.48
		Male	60	42	128	24.20	9.17	0.85	3.50
	Module 1	Female	30	21	145	13.34	5.46	0.78	2.58
		Male	30	21	128	12.59	5.03	0.74	2.56
	Module 2	Female	30	21	145	11.36	4.93	0.77	2.34
		Male	30	21	128	11.61	4.75	0.75	2.40
Ethnicity	Total	African American	60	42	20	22.20	7.41	0.78	3.46
		American Indian	60	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	60	42	17	33.24	13.07	0.92	3.73
		Hispanic	60	42	22	24.50	9.07	0.85	3.50
		Multi-racial	60	DNR	DNR	DNR	DNR	DNR	DNR
		Native Hawaiian/ Pacific Islander	60	DNR	DNR	DNR	DNR	DNR	DNR
		White	60	42	204	23.99	9.09	0.86	3.46
	Module 1	African American	30	21	20	12.30	4.71	0.70	2.58
		American Indian	30	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	30	21	17	17.35	7.89	0.88	2.71
		Hispanic	30	21	22	13.36	5.24	0.75	2.64
		Multi-racial	30	DNR	DNR	DNR	DNR	DNR	DNR
		Native Hawaiian/ Pacific Islander	30	DNR	DNR	DNR	DNR	DNR	DNR
		White	30	21	204	12.66	4.94	0.74	2.54
	Module 2	African American	30	21	20	9.90	3.67	0.61	2.29
		American Indian	30	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	30	21	17	15.88	6.24	0.84	2.48
		Hispanic	30	21	22	11.14	4.23	0.70	2.32
		Multi-racial	30	DNR	DNR	DNR	DNR	DNR	DNR
		Native Hawaiian/ Pacific Islander	30	DNR	DNR	DNR	DNR	DNR	DNR
		White	30	21	204	11.33	4.72	0.75	2.36
EL	Total	All	60	DNR	DNR	DNR	DNR	DNR	DNR
	Module 1	All	30	DNR	DNR	DNR	DNR	DNR	DNR
	Module 2	All	30	DNR	DNR	DNR	DNR	DNR	DNR
IEP	Total	All	60	42	22	23.09	9.47	0.87	3.44
	Module 1	All	30	21	22	12.23	5.26	0.79	2.44
	Module 2	All	30	21	22	10.86	4.83	0.75	2.42
ED	Total	All	60	42	85	23.36	10.50	0.89	3.54
	Module 1	All	30	21	85	12.52	5.81	0.80	2.62
	Module 2	All	30	21	85	10.85	5.25	0.79	2.39

**Table M–9. Summer: Biology Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	66	54	210	30.59	12.88	0.91	3.85
	Module 1	All	33	27	210	15.05	6.89	0.84	2.72
	Module 2	All	33	27	210	15.54	6.58	0.83	2.73
Gender	Total	Female	66	54	106	30.82	12.45	0.91	3.82
		Male	66	54	104	30.36	13.36	0.92	3.87
	Module 1	Female	33	27	106	15.13	6.65	0.83	2.71
		Male	33	27	104	14.96	7.15	0.86	2.71
	Module 2	Female	33	27	106	15.69	6.29	0.82	2.70
		Male	33	27	104	15.39	6.89	0.84	2.75
Ethnicity	Total	African American	66	54	22	25.00	7.78	0.75	3.86
		American Indian	66	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	66	54	17	46.82	13.09	0.92	3.71
		Hispanic	66	54	14	27.50	9.88	0.85	3.77
		Multi-racial	66	54	14	35.64	16.21	0.94	3.83
		Native Hawaiian/ Pacific Islander	66	DNR	DNR	DNR	DNR	DNR	DNR
		White	66	54	141	29.38	11.96	0.90	3.83
	Module 1	African American	33	27	22	11.86	4.62	0.65	2.73
		American Indian	33	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	33	27	17	22.88	7.39	0.88	2.59
		Hispanic	33	27	14	12.71	5.28	0.75	2.62
		Multi-racial	33	27	14	17.36	8.56	0.90	2.71
		Native Hawaiian/ Pacific Islander	33	DNR	DNR	DNR	DNR	DNR	DNR
		White	33	27	141	14.61	6.44	0.82	2.71
	Module 2	African American	33	27	22	13.14	4.00	0.53	2.74
		American Indian	33	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	33	27	17	23.94	6.62	0.84	2.62
		Hispanic	33	27	14	14.79	5.18	0.72	2.72
		Multi-racial	33	27	14	18.29	7.91	0.88	2.73
		Native Hawaiian/ Pacific Islander	33	DNR	DNR	DNR	DNR	DNR	DNR
		White	33	27	141	14.77	6.14	0.81	2.70
EL	Total	All	66	DNR	DNR	DNR	DNR	DNR	DNR
	Module 1	All	33	DNR	DNR	DNR	DNR	DNR	DNR
	Module 2	All	33	DNR	DNR	DNR	DNR	DNR	DNR
IEP	Total	All	66	54	33	25.61	10.07	0.85	3.86
	Module 1	All	33	27	33	12.79	5.48	0.76	2.71
	Module 2	All	33	27	33	12.82	5.50	0.75	2.73
ED	Total	All	66	54	74	26.26	9.11	0.82	3.90
	Module 1	All	33	27	74	12.74	4.97	0.69	2.76
	Module 2	All	33	27	74	13.51	4.89	0.68	2.75



**Table M–10. Summer: Literature Reliabilities**

	Level	Group	Pts.	Len.	N	Mean	SD	r	SEM
Overall	Total	All	52	40	111	23.42	8.41	0.86	3.14
	Module 1	All	26	20	111	11.95	4.17	0.73	2.18
	Module 2	All	26	20	111	11.47	4.93	0.79	2.24
Gender	Total	Female	52	40	55	23.02	8.42	0.86	3.17
		Male	52	40	56	23.82	8.45	0.87	3.11
	Module 1	Female	26	20	55	11.85	4.17	0.72	2.20
		Male	26	20	56	12.05	4.21	0.74	2.16
	Module 2	Female	26	20	55	11.16	4.93	0.79	2.27
		Male	26	20	56	11.77	4.96	0.80	2.21
Ethnicity	Total	African American	52	40	20	24.45	8.29	0.85	3.18
		American Indian	52	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	52	DNR	DNR	DNR	DNR	DNR	DNR
		Hispanic	52	40	16	20.06	9.92	0.91	3.06
		Multi-racial	52	DNR	DNR	DNR	DNR	DNR	DNR
		Native Hawaiian/ Pacific Islander	52	DNR	DNR	DNR	DNR	DNR	DNR
		White	52	40	68	23.62	8.08	0.85	3.15
	Module 1	African American	26	20	20	13.00	4.10	0.72	2.17
		American Indian	26	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	26	DNR	DNR	DNR	DNR	DNR	DNR
		Hispanic	26	20	16	10.00	4.44	0.76	2.20
		Multi-racial	26	DNR	DNR	DNR	DNR	DNR	DNR
		Native Hawaiian/ Pacific Islander	26	DNR	DNR	DNR	DNR	DNR	DNR
		White	26	20	20	13.00	4.10	0.72	2.17
	Module 2	African American	26	20	20	11.45	4.78	0.76	2.33
		American Indian	26	DNR	DNR	DNR	DNR	DNR	DNR
		Asian	26	DNR	DNR	DNR	DNR	DNR	DNR
		Hispanic	26	20	16	10.06	5.92	0.87	2.11
		Multi-racial	26	DNR	DNR	DNR	DNR	DNR	DNR
		Native Hawaiian/ Pacific Islander	26	DNR	DNR	DNR	DNR	DNR	DNR
		White	26	20	68	11.71	4.66	0.76	2.26
EL	Total	All	52	DNR	DNR	DNR	DNR	DNR	DNR
	Module 1	All	26	DNR	DNR	DNR	DNR	DNR	DNR
	Module 2	All	26	DNR	DNR	DNR	DNR	DNR	DNR
IEP	Total	All	52	40	30	21.27	6.63	0.78	3.10
	Module 1	All	26	20	30	10.87	3.48	0.60	2.19
	Module 2	All	26	20	30	10.40	3.80	0.67	2.20
ED	Total	All	52	40	46	19.57	8.11	0.85	3.16
	Module 1	All	26	20	46	10.17	3.94	0.68	2.23
	Module 2	All	26	20	46	9.39	4.72	0.78	2.24

## REFERENCES

---

- Allman, C. (2004). *Test access: Making tests accessible for students with visual impairments—A guide for test publishers, test developers, and state assessment personnel* (2nd ed.). Louisville, KY: American Printing House for the Blind. Retrieved from [www.aph.org](http://www.aph.org).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, L.W. and Krathwohl, D.R. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. New York: Longman.
- Bloom, B.S. (1956). *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. New York: Longman.
- Brennan, R. L. (1998). Misconceptions at the intersection of measurement theory and practice. *Educational Measurement: Issues and Practice*, 17(1), 5–9.
- Brennan, R. L. (2004). BB-Class (Version 1.0) [Computer Software]. Retrieved from <http://www.education.uiowa.edu/casma>.
- Center for Disease Control (July 24, 2020). Health Equity Considerations and Racial and Ethnic Minority Groups. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html>.
- Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289.
- Connell, B. R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., et al. (1997). *The principles of universal design*. Raleigh: North Carolina University, College of Design.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.
- Cronbach, L. J., & Shavelson, R. L. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391–418.
- Data Recognition Corporation. (2010). *Fairness in testing: Training manual for issues of bias, fairness, and sensitivity*. Maple Grove, MN.
- Dorans, N., Schmitt, A., & Bleistein, C. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309–319.
- Eignor, D. R. (1985). *An investigation of the feasibility and practical outcomes of pre-equating the SAT verbal and mathematical sections* (Research Report 85-10). Princeton, NJ: Educational Testing Service.
- Eignor, D. R., & Stocking, M. L. (1986). *An investigation of the possible causes for the inadequacy of IRT pre-equating* (Research Report 86-14). Princeton, NJ: Educational Testing Service.
- Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 24(3), 21–28.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons.
- Hambleton, R., & Novick, M. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10, 159–170.

- Hambleton, R., Swaminathan, H., & Rogers, J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score theory models. *Journal of Educational Measurement*, 27(4), 345–359.
- Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practices*, 10(2), 33–41.
- Hess, K. (2004). *Applying Webb's depth-of-knowledge levels in reading*. [online] available: [www.nciea.org](http://www.nciea.org).
- Ho D.E., Imai K., King G., Stuart E.A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8), 1–28. <https://www.jstatsoft.org/v42/i08/>.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: methods and practices* (2nd ed.). New York: Springer.
- Kolen, M. J., & Harris, D. J. (1990). Comparison of item preequating and random groups equating using IRT and equipercentile methods. *Journal of Educational Measurement*, 27 (1), 27–39.
- Lane, S. (1999). *Validity evidence for assessments*. Paper presented at the 1999 Edward F. Reidy, Interactive Lecture Series, Providence, RI.
- Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 23–30.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A bookmark approach*. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Linacre, J. M. (2018). *A user's guide to WINSTEPS MINNISTEP Rasch-model computer programs*. Chicago: Winsteps.
- Linacre, J. M. (2019). *Winsteps® Rasch measurement* [computer program]. Beaverton, OR: Winsteps.com.
- Livingston, S., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, 9(3), 200–215.
- McDonald, R. P. (1979). The structural analysis of multivariate data: A sketch of a general theory. *Multivariate Behavioral Research*, 14, 21–38.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 104–142). New York: Academic Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 3–104). Washington, DC: American Council on Education.
- National Center for Education Statistics (2020a). *School District Characteristic 2018–2019* [Data file]. Retrieved from <https://data-nces.opendata.arcgis.com/datasets>
- National Center for Education Statistics (2020b). *The Common Education Data Standards (CEDs)*. Retrieved from <http://ceds.ed.gov/>

- Pennsylvania Department of Education. (2010). *Psychometric analysis report for the fall 2010 Keystone field tests*. Harrisburg, PA: PDE.
- Pennsylvania Department of Education. (2011). *Algebra I Keystone Feb 2011 item and scoring sampler*. Retrieved November 11, 2011, from [www.pdesas.org/Assessment/Keystone](http://www.pdesas.org/Assessment/Keystone).
- Pennsylvania Department of Education. (2011). *Biology Keystone Feb 2011 item and scoring sampler*. Retrieved November 11, 2011, from [www.pdesas.org/Assessment/Keystone](http://www.pdesas.org/Assessment/Keystone).
- Pennsylvania Department of Education. (2011). *Keystone Exams score report focus group findings*. Harrisburg, PA: PDE.
- Pennsylvania Department of Education. (2011). *Keystone standard setting technical report: Algebra I, Biology, and Literature*. Harrisburg, PA: PDE.
- Pennsylvania Department of Education. (2023). *2022 Item and Scoring Sampler – Algebra I*. Retrieved February 16, 2023 from [www.education.pa.gov](http://www.education.pa.gov).
- Pennsylvania Department of Education. (2023). *2022 Item and Scoring Sampler – Biology*. Retrieved February 16, 2023 from [www.education.pa.gov](http://www.education.pa.gov).
- Pennsylvania Department of Education. (2023). *2022 Item and Scoring Sampler – Literature*. Retrieved February 16, 2023 from [www.education.pa.gov](http://www.education.pa.gov).
- Pennsylvania Department of Education. (2023). *2023 Accommodations Guidelines* (PDE, revised 9/2022). Retrieved April 18, 2023 from [www.education.pa.gov](http://www.education.pa.gov).
- Pennsylvania Department of Education. (2023). *2023 Accommodations Guidelines for ELs* (PDE, revised 9/2022). Retrieved April 18, 2023 from [www.education.pa.gov](http://www.education.pa.gov).
- Pennsylvania Department of Education. (2023). *2023 Pennsylvania System of School Assessment: Handbook for Assessment Coordinators*. Retrieved April 18, 2023 from [www.education.pa.gov](http://www.education.pa.gov).
- Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 221–262). Washington, DC: American Council on Education.
- R (Version 3.6) [Computer software]. Vienna, Austria: R Core Team.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 17(1), pp. 41–55.
- SAS (Version 7) [Computer software]. Cary, NC: SAS Institute, Inc.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- Staffa, S. & Zurakowski, D. (2018). Five steps to successfully implement and evaluate propensity score matching in clinical research studies. *Anesthesia & Analgesia*, 127(1).
- Stocking, M. L., & Eignor, D. R. (1986). The impact of different ability distributions on IRT preequating (Research Report No. 86–14). Princeton, NJ: Educational Testing Service.

- Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2004, April 28). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*. Washington, DC: Author.
- Valencia, S.W. and Wixson, K.K. (2000). *Policy-oriented research on literacy standards and assessment*. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, and R. Barr (Eds.), *Handbook of Reading Research: Vol. III*. Mahwah, NJ: Lawrence Erlbaum.
- Webb, N. L. (1997). *Criteria for alignment of expectations and tests in mathematics and science education*. Research Monograph No. 6). Madison: University of Wisconsin–Madison, National Institute for Science Education. Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (1997; 2006). *Research monograph number 6: Criteria for alignment of expectations and assessments on mathematics and science education*. Washington, D.C.: CCSSO.
- Webb, N. L. (1999). *18: Alignment of science and mathematics standards and assessments in four states*. Research Monograph No. Madison, WI: National Institute for Science Education.
- Webb, N. L. (1999). *Research monograph No. 18: Alignment of science and mathematics standards and assessments in four states*. Washington, D.C.: CCSSO.
- Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and tests for four states: State collaborative on test and state standards (SCASS)*. Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research.
- Webb, N. L. (November, 2005). *Depth-of-Knowledge levels for four content areas*. Presentation to the Florida Education Research Association, 50th Annual Meeting, Miami, Florida.
- Webb, N. L. (2006). *Web alignment tool* [Computer Software]. Madison: Wisconsin Center of Educational Research. University of Wisconsin-Madison.
- Webb Alignment Tool (WAT) Training Manual retrieved from <http://www.wcer.wisc.edu/WAT/index.aspx>.
- Wright, B.D., & Linacre, M. (1994). Reasonable Mean-Square Fit Statistics. *Rasch Measurement Transactions*, 8(3), 370. Retrieved from: <https://www.rasch.org/rmt/rmt83b.htm>
- Wright, B., & Masters, G. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187–213.
- Zwick, R., & Erickson, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, 26, 55–66.