

# TECHNICAL REPORT

## Pennsylvania Alternate System of Assessment

### A Statewide Assessment for Students with Significant Cognitive Disabilities

2012 Reading and Mathematics  
Grades 3/4, 5/6, 7/8 and 11

2012 Science  
Grades 4, 8 and 11

Provided by

University of Pittsburgh, PASA Project  
October 2012



## Table of Contents

Chapter 1: Overview of the Pennsylvania Alternate System of Assessment (PASA) Reading, Math and Science .....	8
Purpose of the PASA .....	8
Historical Perspective .....	9
Description of the Alternate Assessment .....	9
Chapter 2: Description of Students taking the PASA .....	11
Eligibility .....	11
Description of Examinees .....	12
Assignment of Students to Levels.....	15
Level Changes.....	15
Chapter 3: Test and Item Development .....	17
Development of the Alternate Academic Standards and Alternate Assessment Anchors and Eligible Content .....	17
Development of PASA Reading and Math, Historical Perspective .....	18
Development of PASA Science, Historical Perspective .....	20
Bias, Fairness and Sensitivity Reviews, Historical Perspective .....	21
Alignment of the <i>PASA Reading, Math and Science</i> to the Alternate Standards and Assessment Anchors and Eligible Content.....	21
New Item Development .....	22
Restructuring of the 2012 PASA .....	22
Chapter 4: Additional Versions of the PASA.....	26
Accommodations .....	26
Modifications.....	28
Operational Test Forms for the 2012 PASA.....	28
Chapter 5: Test Administration Procedures.....	29
Materials Required for PASA Administration .....	29
Test Booklets.....	29
Test Kits.....	29
Preparation of Test Administrators.....	29
Chapter 6: Processing and Scoring the PASA.....	31
Processing Completed Assessments .....	31
Scoring Rubric.....	31
Scoring Conferences .....	32

Recruitment and Selection of Scorers .....	32
Training Scorers.....	33
Reliability to Standard .....	33
Team-to-Team Reliability .....	34
Chapter 7: Score Reporting.....	36
Chapter 8: Statistical Analyses of Item Data .....	37
Descriptive Statistics and Internal Consistency.....	37
Item Correlations.....	37
Item Performance .....	40
Percentages of Items at Each Score Level.....	40
Differential Item Performance .....	41
Chapter 9: Item Performance on Supplemental ‘Test’ Items .....	42
Chapter 10: Setting Standards: Performance Level Descriptions and Cut Scores .....	44
Standard Settings for the PASA Reading, Math and Science .....	44
Brief Overview of the Most Recent Standard Setting Workshops for the PASA Reading, Math, and Science .....	44
Establishing Cut Score for the 2012 Administration of the PASA.....	46
Chapter 11: Validity Argument for the <i>PASA</i> .....	53
Validity Studies for the Pennsylvania Alternate System of Assessment ( <i>PASA</i> ) .....	53
Chapter 12: Future Activities .....	58
On-Line Submission of Student Assessments .....	58
On-Line Training of Scorers.....	58
On-Line Training of Test Administrators.....	58
Proficiency Study: Across Grade Learning Progressions .....	58
Enacted Curriculum Study .....	59
Impact Study Phase 1, PASA Math, Reading and Science.....	59
Impact Study Phase 2, PASA Science .....	59

## List of Tables

Table 1. <i>PASA Reading, Math and Science Participation in 2012: Counts by Grade and as Percent of Students Participating in Statewide Assessment</i> .....	12
Table 2. <i>Demographic Data for Students Assessed by the 2012 PASA Reading and Math</i> .....	13
Table 3. <i>Demographic Data for Students Assessed by the 2012 PASA Science</i> .....	14
Table 4. <i>Counts of Level Change Requests by Content Area and Grade Level for the 2012 PASA</i> .....	16
Table 5. <i>Counts of Examinees taking the PASA Reading, Math and Science Overall and by Grade and Test Level</i> .....	16
Table 6. <i>Distribution of Items on the 2012 PASA Reading by Grade Span, Level and PSSA Reporting Category</i> .....	24
Table 7. <i>Distribution of Items on the 2012 PASA Math by Grade Span, Level and PSSA Reporting Category</i> .....	25
Table 8. <i>Distribution of Items on the 2012 PASA Science by Grade, Level and PSSA Reporting Category</i> .....	25
Table 9. <i>Placement of the Supplemental Test Items on the 2012 PASA Reading, Math and Science</i> .....	25
Table 10. <i>Use of Accommodations for the 2012 PASA Math, Reading, and Science as Reported by the Teacher</i> .....	27
Table 11. <i>Reliability to Standard - Average Percentage of Items with Exact Agreement</i> .....	34
Table 12. <i>Team-to-Team Reliability Counts and Percentages</i> .....	34
Table 13. <i>Grades 3 and 4: Percent Agreement for Item Scores</i> .....	35
Table 14. <i>Grades 5 and 6: Percent Agreement for Item Scores</i> .....	35
Table 15. <i>Grades 7 and 8: Percent Agreement for Item Scores</i> .....	35
Table 16. <i>Grade 11: Percent Agreement for Item Scores</i> .....	35
Table 17. <i>Descriptive Statistics, Cronbach’s Alpha and SEM for the 2012 PASA Reading (20 Items) for All Students Completing the PASA</i> .....	38
Table 18. <i>Descriptive Statistics, Cronbach’s Alpha and SEM for the 2012 PASA Math for All Students Completing the PASA</i> .....	39
Table 19. <i>Descriptive Statistics, Cronbach’s Alpha and SEM for the 2012 PASA Science (20 Items) for All Students Completing the PASA</i> .....	40
Table 20. <i>Score Ranges for the 2012 PASA Reading, Math and Science</i> .....	48
Table 21. <i>Score Ranges for the 2011 PASA Reading, Math and Science</i> .....	49
Table 22. <i>Percent of PASA Students at Each Performance Levels for the 2012 PASA Reading</i> .....	50

Table 23. <i>Percent of PASA Students at Each Performance Levels for the 2012 PASA Math</i> .....	51
Table 24. <i>Percent of PASA Students at Each Performance Levels for the 2012 PASA Science</i> .....	52
Table 25. <i>Sources of Validity Evidence for the PASA Reading, Math, and Science</i> .....	54
Table 26. <i>Additional Sources of Validity Evidence for PASA Reading</i> .....	55
Table 27. <i>Additional Sources of Validity Evidence for PASA Science</i> .....	56
Table 28. <i>Sources of Evidence Related to Technical Quality of the Assessment Measuring Student Performance</i> .....	56

## List of Figures

Figure 1. <i>Questions Guiding Decisions about Participation in PASA</i> .....	11
Figure 2. <i>Theoretical representation of growth trajectories for students assigned to Levels A, B, or C of the PASA</i> .....	18
Figure 3. <i>Rubric for Scoring the 2012 PASA</i> .....	32

## Preface

The purpose of this report is to document the development, administration, scoring and technical quality of the 2011-12 administration of the *Pennsylvania Alternate System of Assessment (PASA) Reading, Math, and Science* assessments. The 2011-12 administration is referred to as the 2012 administration throughout this report. The report is organized into 12 chapters. The chapters describe the background, purpose, design and development, administration, scoring, and score reporting of the *PASA*.

Several of the chapters are updates from previous reports to reflect the current year's data. Additionally, most of the appendices contain similar content to that presented in the appendices of previous *PASA* technical reports. Enrollment forms, content standards, administrator manuals, and summaries of statistical analyses conducted on the item level data from the 2012 *PASA* are included in Appendices to this technical report, as they have been in the past.

There are also several sections of the 2012 technical report that have been changed from previous reports. A significant section that can be found in Chapter 3 is called 'Restructuring the 2012 *PASA*'. The 2012 administration of the *PASA* has required some restructuring due to recommendations of the PA Technical Advisory Committee (TAC) and because a sequence of two large-scale validity studies is being conducted during the 2012 and 2013 *PASA* administrations. The restructuring involves some of the forms of the *PASA* being shortened from 25 to 20 items. At the recommendation of the TAC, new standard setting workshops were not conducted, but revised cut scores were needed. Therefore, a section on establishing revised cut scores has also been added to Chapter 10.

Chapter 11 has been revised to include a brief summary of the validity studies and other technical studies that have been carried out for the *PASA*. In addition, a complete supplemental report summarizing several of the validity studies for the *PASA* that have been conducted since 2007-08 is provided in addition to this technical report. The "Pennsylvania Alternate System of Assessment Supplemental Report on Validity Studies," along with eight detailed appendices, supplements this technical report.

Finally, this report concludes with a discussion of activities that are currently being planned to further the development of *PASA Reading, Math and Science* and enhance the technical characteristics of the assessment.

# Chapter 1: Overview of the Pennsylvania Alternate System of Assessment (PASA) Reading, Math and Science

## Purpose of the PASA

As outlined in Chapter 4 of the State Board of Education regulations, the purposes of the statewide assessment (referring to the *Pennsylvania System of School Assessment (PSSA)*) are as follows:

- Provide students, parents, educators, and citizens with an understanding of student and school performance
- Determine the degree to which school programs enable students to attain proficiency of the Academic Standards
- Provide information to state policy makers, including the State Senate, the General Assembly, and the State Board, on how effective Pennsylvania schools are in promoting and demonstrating student proficiency of the Academic Standards
- Provide information to the general public on school performance
- Provide results to school districts based on aggregate performance of all students, for students with an Individualized Education Plan (IEP) and for those without an IEP.

The *PASA* was designed to fulfill the same purposes, and ensures that Pennsylvania students with the most significant cognitive disabilities are included in the state accountability system. An additional purpose of the *PASA* is to provide information to teachers and schools to guide the improvement of curricula and instruction to enable students with the most significant cognitive disabilities to reach proficiency in the Alternate Academic Standards and Eligible Content.

*PASA Reading, Math and Science* were designed in response to federal laws requiring that all students, including those with the most significant disabilities, participate in the statewide accountability process (see Section 504 of The Rehabilitation Act, Title II of the Americans with Disabilities Act of 1990, Title I of the Elementary and Secondary Education Act, and the Individuals with Disabilities Education Act (IDEA) Amendments of 1997). More recently, *PASA Reading, Math and Science* have been used to meet the requirements of No Child Left Behind (NCLB).

Until 2008, *PASA Reading, Math, and Science* test scores were aggregated to the district of residence and used as part of the determination of Adequate Yearly Progress (AYP) at the district level. In 2008, all *PASA Reading, Math and Science* scores were attributed to the students' home school (i.e., the school the student would have attended in the district of residence had the student not had a disability). Since 2009, *PASA Reading, Math and Science* scores for students served in district-run special education programs were attributed to the students' home school, but *PASA Reading, Math and Science* scores for students in Intermediate Unit-run programs or in Approved Private Schools were attributed to the students' district of residence. Although student proficiency scores are reported to parents, *PASA Reading, Math and Science* were not designed to be used to make individual student-level instructional decisions.

The *PASA Reading, Math and Science* measure skills related to the Pennsylvania Alternate Assessment Anchors, which were derived from the Pennsylvania Chapter 4 Content Standards. Like the *PSSA*, the *PASA* is designed to take a snapshot of students' typical performance on a small

sample of academic skills derived from the PA academic standards.

## Historical Perspective

*PASA Reading and Math* were introduced as the statewide alternate assessment during the 2000-2001 school year. Initially, the *PASA* was designed for use in grades 5, 8, and 11. It began as a set of four extended performance tasks, two related to reading and two related to mathematics. Extended performance tasks were divided into component steps, and performance on each step was scored separately. In the process of pilot testing and validating of the conceptual framework for the assessment, it was recommended by administrators and confirmed by raters evaluating student performance that administration and scoring as well as student performance were hindered by the extended task format. Consequently, *PASA Reading and Math* ceased to consist of an extended set of performance tasks and instead became an assessment comprised of several independent items or small item sets linked to one piece of stimuli (e.g., a paragraph of text). In the 2002-03 school year, grade 3 was added to the set of grades tested. In the 2004-05 school year, grades 4, 6, and 7 were included in a mandatory pilot test; in 2005-06, *PASA Reading and Math* scores were included in the calculation of AYP in grades 3 through 8 and 11. The *PASA Reading and Math* are administered in grade spans of 3/4, 5/6, 7/8 and 11, meaning that students in both grades (i.e., 3 and 4) receive the same *PASA* assessment for a given administration. The *PASA Science* was administered for the first time in 2007-08 to students in grades 4, 8 and 11. The format of the *PASA Science* models the format of the *PASA Reading and Math*.

## Description of the Alternate Assessment

The *PASA* assessments have 3 distinct levels of difficulty for reading and math in each grade span 3/4, 5/6, 7/8, and 11, and in Science for grades 4, 8, and 11. The test levels at each grade or grade span, labeled A, B and C, represent three levels of complexity or “cognitive demand,” with Level A being the simplest and most concrete, Level B more complex, and Level C the most complex.

The content of the *PASA Reading, Math, and Science* items link to the Alternate Assessment Anchors and Eligible Content. The Alternate Assessment Anchors are extensions of the PA Assessment Anchors and Eligible Content, which are linked to the PA Academic Standards.

Items on all levels of the *PASA* for a given content area are written to the same Alternate Assessment Anchors and Eligible Content. However, the skills that are evaluated and the format of the items vary on the A, B and C level tests.

Level A items might include tasks across the content areas in which students match objects, pictures, or sets, or where they select a requested object, picture, or amount. Items on the Level A assessments are multiple choice with 2 distractors. Reading content on the Level A assessments involves reading simple pictures and comprehension refers to listening comprehension. Level A math items can involve distinguishing between items based on mathematical concepts such as size, shape, amount and function, and Level A Science items require students to distinguish between items based on scientific concepts.

The Level B assessments at each grade span or grade fall at an intermediate level of complexity. At Level B, students would not be required to draw inferences in order to reach solutions. The

Level B assessments consist of mainly selected response items with limited production items. Selection items on the Level B assessments have 3 distractors, two of which are related. In Level B reading, tasks might include working with complex pictures, icons, or sight words. In Level B math and science, students might be asked to solve problems and answer questions using manipulatives.

Level C items are the most complex. In Level C reading, items require that the student read and answer questions about the information read. In Level C math, items require solving arithmetic problems, using tools to measure, using money, etc. Level C science items might require students to describe or select a scientific function or process. In some cases, items require that students make inferences in order to determine the solution. The level C assessments contain a majority of production items, and the selected response items have 4 distractors. Further, option choices for the selected response items on the level C assessments are more closely related than are option choices on the other test levels.

A summary of the characterizations of the PASA test levels is:

- Level A: Concrete tasks, related to personal experiences, orienting, matching, sorting, receptive labeling, objects as test material, extensive prompting and assistance
- Level B: Representations using pictures, limited test and content area vocabulary, related to familiar surroundings and experiences, receptive and expressive labeling, sorting, classifying, identifying function, limited assistance
- Level C: Symbolic representation using text and complex pictures, extensive test and content area vocabulary, related to experience beyond personal familiarity, labeling, applying knowledge, producing responses to open ended questions

Each form of the *PASA* has consisted of between 20-25 items over the past several administrations. The *PASA Reading, Math and Science* are administered to students by their teachers or another special educator who knows them well, on a one-to-one basis. The assessments consist of a series of independent or small item sets linked to one piece of stimuli (e.g., a paragraph of text). Teachers are provided with a suggested script to follow with each test item, as well as the text, graphics, and most objects that they will need to administer the test.

Student performance is recorded via video or digital media or through detailed narrative notes and these recordings or notes of student performance are submitted for scoring along with Supporting Documentation. The Supporting Documentation form is completed by the test administrator, and is found in Appendix A. The form provides additional background information about the student that can help scorers in interpreting the student's performance.

The *PASA Reading and Math* were administered during a six-week window from February to March of the school year. The *PASA Science* was administered during a four-week window in May of the school year.

Scoring of the *PASA* takes place at annual scoring conferences. Individuals who complete an on-line training and pass a proficiency test can apply to be a scorer. At the scoring conferences, teams of scorers (comprised of practicing teachers, supervisory/administrative school personnel, and college/University faculty) view the media are or are trained to read the narrative notes and use a specific rubric to score students' performances on each item. Scoring of the 2012 *PASA* took place in the spring and summer of 2012.

## Chapter 2: Description of Students taking the PASA

### Eligibility

The *PASA Reading, Math and Science* permit students with the most significant disabilities, who are unable to participate in the *PSSA* or *PSSA-M* (even with accommodations), to demonstrate mastery of skills and attainment of knowledge on the alternate academic anchor standards. The *PASA* is appropriate for students who have significant cognitive disabilities, and who require intensive instruction and extensive support in order to perform and/or participate meaningfully and productively in the everyday activities of integrated school, home, community and work environments. These students require substantial modifications of the general education curriculum as well as instruction in areas not presently assessed by the *PSSA*.

The *PASA* is administered to students with significant disabilities in the equivalent of grades 3, 4, 5, 6, 7, 8, and 11 for Math and Reading, and grades 4, 8 and 11 for Science who meet the criteria that have been established and disseminated by the Bureau of Special Education, Pennsylvania Department of Education. The decision about participation in the regular or alternate statewide assessment is made by the student’s IEP Team, and is based on these criteria. The questions found in Figure 1 guide their decision-making. If the answer was “Yes” to all of these questions, it would be appropriate for the IEP team to consider assessing the student with the *PASA*. If the answer was “No” to any of the above questions, the *PASA* would not be the appropriate statewide assessment for the student.

1. By September 1 of the school year in which this IEP will be operative, will the student be in grade 3, 4, 5, 6, 7, 8, or 11?	YES	NO
AND		
2. Does the student have significant cognitive disabilities?	YES	NO
AND		
3. Does the student require intensive instruction to learn?	YES	NO
AND		
4. Does the student require extensive adaptation and support in order to perform and/or participate meaningfully and productively in the everyday life activities of integrated school, home, community and work environments?	YES	NO
AND		
5. Does the student require substantial modifications of the general education curriculum?	YES	NO
AND		
6. Does the student’s participation in the general education curriculum differ substantially in form and/or substance from that of most other students (i.e., different objectives, materials, or activities)?	YES	NO

Figure 1. *Questions Guiding Decisions about Participation in PASA*

To document a student’s enrollment in the *PASA*, an enrollment packet with forms listing the criteria for eligibility for taking the *PASA* and for enrolling students into the *PASA* were sent to the assessment coordinator of each district and each service provider in September. For students who were previously administered the *PASA*, pre-printed enrollment forms were provided with space to make changes. For students who were not previously administered the

PASA, a “New Student Enrollment Form” was included. A copy of the enrollment packet is found in Appendix A.

## Description of Examinees

In 2012, a total of 14,610 students in grades 3, 4, 5, 6, 7, 8, and 11 participated in *PASA Reading and Math* and a total of 5,894 students participated in *PASA Science*. Table 1 provides the participation counts by grade level and as a percent of total participation in statewide assessment at that grade level. A total of 439 students who initially enrolled in the *PASA Math, Reading or Science* were not administered the assessment. The reasons are as follows: Religious reasons (n=25), extended absence (n=72), deceased/medical emergency (n=46) and other (n=296).

Table 1. *PASA Reading, Math and Science Participation in 2012: Counts by Grade and as Percent of Students Participating in Statewide Assessment*

Grade	Math		Reading		Science	
	N	% Statewide	N	% Statewide	n	% Statewide
3	2,068	1.61%	2,068	1.61%	-	-
4	2,067	1.66%	2,067	1.67%	1,990	1.56%
5	2,183	1.72%	2,183	1.73%	-	-
6	2,123	1.65%	2,123	1.66%	-	-
7	2,140	1.66%	2,140	1.66%	-	-
8	2,076	1.62%	2,076	1.62%	2,022	1.58%
11	1,953	1.54%	1,953	1.53%	1,822	1.48%

Table 2 summarizes participation counts overall and by the demographic variables of gender, ethnicity, and primary disability for the *PASA Reading and Math*. Table 3 presents the corresponding information the *PASA Science*.

Table 2. *Demographic Data for Students Assessed by the 2012 PASA Reading and Math*

	Overall		3		4		5		6		7		8		11	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
<b>Gender</b>																
Female	4,912	33.6	653	31.6	673	32.6	719	32.9	753	35.5	727	34.0	676	32.6	711	36.4
Male	9,698	66.4	1,415	68.4	1,394	67.4	1,464	67.1	1,370	64.5	1,413	66.0	1,400	67.4	1,242	63.6
<b>Ethnicity</b>																
American Indian or Alaskan Native	18	0.1	4	0.2	4	0.2	0	0.0	2	0.1	0	0.0	4	0.2	4	0.2
Asian or Pacific Islander	290	2.0	57	2.8	42	2.0	50	2.3	37	1.7	32	1.5	44	2.1	28	1.4
Black or African American not Hispanic	3,015	20.6	431	20.8	439	21.2	490	22.4	453	21.3	390	18.2	433	20.9	379	19.4
Hispanic or Latino	1,341	9.2	208	10.1	199	9.6	200	9.2	191	9.0	194	9.1	186	9.0	163	8.3
Multi-racial	388	2.7	58	2.8	63	3.0	53	2.4	57	2.7	67	3.1	56	2.7	34	1.7
White (not Hispanic)	9,558	65.4	1,310	63.3	1,320	63.9	1,390	63.7	1,383	65.1	1,457	68.1	1,353	65.2	1,345	68.9
<b>Primary Disability</b>																
Autism	3,985	27.3	744	36.0	678	32.8	678	31.1	577	27.2	509	23.8	471	22.7	328	16.8
Deaf/Blindness	20	0.1	2	0.1	3	0.1	1	0.0	3	0.1	2	0.1	6	0.3	3	0.2
Deafness/Hearing Impairment	93	0.6	12	0.6	12	0.6	22	1.0	18	0.8	9	0.4	9	0.4	11	0.6
Emotional Disturbance	262	1.8	23	1.1	40	1.9	33	1.5	35	1.6	43	2.0	40	1.9	48	2.5
Intellectual Disability	6,531	44.7	753	36.4	800	38.7	896	41.0	940	44.3	1,030	48.1	1,028	49.5	1,084	55.5
Learning Disability	1,332	9.1	153	7.4	164	7.9	187	8.6	211	9.9	212	9.9	195	9.4	210	10.8
Multiple Disabilities	1,266	8.7	177	8.6	177	8.6	195	8.9	174	8.2	182	8.5	188	9.1	173	8.9
Other Health Impairment	698	4.8	123	5.9	123	6.0	105	4.8	104	4.9	105	4.9	83	4.0	55	2.8
Orthopedic Impairment	132	0.9	22	1.1	22	1.1	23	1.1	19	0.9	10	0.5	23	1.1	13	0.7
Speech/Language Impairment	89	0.6	24	1.2	25	1.2	12	0.5	8	0.4	12	0.6	6	0.3	2	0.1
Traumatic Brain Injury	77	0.5	9	0.4	9	0.4	12	0.5	10	0.5	11	0.5	13	0.6	13	0.7
Visual Impairment	125	0.9	26	1.3	14	0.7	19	0.9	24	1.1	15	0.7	14	0.7	13	0.7

Table 3. *Demographic Data for Students Assessed by the 2012 PASA Science*

	Overall		4		8		11	
	n	%	n	%	n	%	n	%
<b>Gender</b>								
Female	1,998	33.9	642	32.3	667	33.0	689	36.6
Male	3,895	66.1	1,348	67.7	1,354	67.0	1,193	63.4
<b>Ethnicity</b>								
American Indian or Alaskan Native	10	0.2	3	0.2	4	0.2	3	0.2
Asian or Pacific Islander	106	1.8	38	1.9	42	2.1	26	1.4
Black or African American (not Hispanic)	1,172	19.9	417	21.0	410	20.3	345	18.3
Hispanic or Latino	522	8.9	191	9.6	180	8.9	151	8.0
Multi-racial	149	2.5	62	3.1	54	2.7	33	1.8
White (not Hispanic)	3,934	66.8	1279	64.3	1,331	65.9	1,324	70.4
<b>Primary Disability</b>								
Autism	1,428	24.2	651	32.7	460	22.8	317	16.8
Deaf/Blindness	12	0.2	3	0.2	6	0.3	3	0.2
Deafness/Hearing Impairment	31	0.5	12	0.6	8	0.4	11	0.6
Emotional Disturbance	122	2.1	40	2.0	37	1.8	45	2.4
Intellectual Disability	2,814	47.8	773	38.8	1,004	49.7	1,037	55.1
Learning Disability	545	9.2	156	7.8	189	9.4	200	10.6
Multiple Disabilities	523	8.9	171	8.6	181	9.0	171	9.1
Orthopedic Impairment	253	4.3	117	5.9	82	4.1	54	2.9
Other Health Impairments	55	0.9	20	1.0	21	1.0	14	0.7
Speech/Language Impairment	32	0.5	24	1.2	6	0.3	2	0.1
Traumatic Brain Injury	38	0.6	9	0.5	14	0.7	15	0.8
Visual Impairment including Blindness	40	0.7	14	0.7	13	0.6	13	0.7

## Assignment of Students to Levels

As was described, the *PASA Reading, Math and Science* have three levels at each grade. Grade-span appropriate Skills Checklists were created in 2005, and have since been used, with past performance data when it is available, in a prediction model for the assignment of the appropriate level of the *PASA* to administer. That is, using a combination of past performance data and responses to the teacher-administered Skills Checklists, levels of difficulty for the *PASA* tests are pre-assigned. Example Skills Checklists for Grades 3 and 11 Reading and Mathematics are found in Appendix A. The assignment of each student to a particular level for the *PASA Science* was based on the Skills Checklist for Mathematics.

The rationale for using the Skills Checklist for Mathematics to assign the level of the *PASA Science* follows. For the 2007-08 administration of the *PASA Science*, each student's test level was assigned to be the same as the student's 2007-08 *PASA Math* assessment. The same criteria of using the *PASA Math* level to assign the *PASA Science* level was used again for the 2008-09 assessment. Additionally, science checklists were developed and completed by the teachers at each grade level (4, 8, and 11) for each student assessed with *PASA Science* in 2009, to further facilitate the assignment of students to *PASA Science* levels for future administrations. The checklist was developed based on the Performance Level Descriptors (PLDs) that were established as part of the Science Standard Setting Study. Analyses of 2009 performance/ achievement data and checklist data indicated that Science Checklist items were not significantly better at predicting performance on the *PASA Science* than the Math Skills Checklist. As a result, and to save teachers additional tasks, the Math Skills Checklist data were and continue to be used to guide assignments for the *PASA Science*.

## Level Changes

In the Administrator Manual, in the administration training sessions, on the web site, and in the packets of testing materials distributed, test administrators are reminded that, if they believe the pre-assigned level of the *PASA* is not appropriate (too high or too low) there is a mechanism for requesting a level change. A form is available on the *PASA* web site on which a level change request and justification can be submitted to *PASA* headquarters.

For the 2012 administration of the *PASA*, if a student's previous performance level (2011) was Proficient or Advanced, teachers could NOT request a level change *down* for the 2012 administration without substantial justification (for example, of a degenerating disability). In addition, students who received a perfect score on the 2011 administration of the *PASA* were automatically assigned a higher level of the assessment. The intent of this policy was to support high expectations for students. The number of requests for level changes for the 2012 administration is summarized in Table 4. A total of 1,320 level change submissions were received, including 594 for math, 634 for reading, and 92 for science. The total column of Table 4 reflects the number of level changes by content area and grade level.

Table 4. *Counts of Level Change Requests by Content Area and Grade Level for the 2012 PASA*

	Grade														Tot
	3		4		5		6		7		8		11		
Content	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	
Math	8	48	20	20	13	32	21	25	12	18	12	16	8	44	297
Reading	13	51	18	16	12	34	21	20	8	31	7	21	13	52	317
Science	-	-	3	4	-	-	-	-	-	-	0	7	2	30	46

Table 5 presents the counts of examinees at each grade level who took the *PASA Reading, Math and Science* overall, in each grade, and at each test level.

Table 5. *Counts of Examinees taking the PASA Reading, Math and Science Overall and by Grade and Test Level*

	Reading				Math				Science			
	Total	A	B	C	Total	A	B	C	Total	A	B	C
Grade	14,610	4,750	6,051	3,809	14,610	5,074	5,580	3,956	5,893	2,084	2,290	1,519
3	2,068	925	762	381	2,068	930	730	408	-	-	-	-
4	2,067	695	958	414	2,067	768	865	434	1,990	752	830	408
5	2,183	670	942	571	2,183	706	829	648	-	-	-	-
6	2,123	613	932	578	2,123	632	853	638	-	-	-	-
7	2,140	589	919	632	2,140	676	791	673	-	-	-	-
8	2,076	590	882	604	2,076	675	781	620	2,021	656	765	600
11	1,953	668	656	629	1,953	687	731	535	1,882	676	695	511

## Chapter 3: Test and Item Development

### Development of the Alternate Academic Standards and Alternate Assessment Anchors and Eligible Content

Before developing the alternate assessment, the Pennsylvania Academic Standards in reading, mathematics and science (22 PA Code § 4.12.) were reinterpreted in ways that would make them more meaningful and relevant for students with significant disabilities. A summary of the process follows.

For the content areas of reading and mathematics, the essence of each standard was first identified. Then, in conjunction with teachers, more specific skills were identified. This process of identifying the essence of the standards and associated skills is reflected on the charts that appear as Appendix B.

These sets of skills were reviewed and discussed by subject matter experts and experts in low incidence disabilities, and comments were collected regarding the validity of the standards and skills. The Advisory Committee to the *PASA Reading and Math*, consisting of stakeholders such as Pennsylvania Department of Education employees, parents of students with disabilities, and technical consultants from the *PSSA*, also reviewed the re-interpreted standards. After incorporating their suggestions, comments were solicited from any interested parties through a posting on the Pennsylvania Department of Education/*PASA* website. All feedback received was reviewed and incorporated into the standards to the fullest extent possible. The *Pennsylvania Alternate Academic Content Standards for Reading and for Mathematics* were formally adopted by the Pennsylvania Board of Education in June, 2005, and are found in Appendix C.

In 2005, Assessment Anchor Content Standards (Assessment Anchors) were developed to clarify the structure and improve the articulation between instruction and the *PSSA*. During 2007-08, a set of comparable Alternate Assessment Anchors and were developed for the *PASA* to parallel the alignment context of the *PSSA* and to improve the articulation between instruction and the *PASA*. These *PASA* Alternate Assessment Anchors also serve to communicate the eligible content or range of knowledge and skills from which the *PASA* is designed each year. Since the Alternate Assessment Anchors and Eligible Content encompass the grade spans 3/4, 5/6, 7/8, and 11, the document informs the test design for all grade levels and levels of complexity within grade levels. The Alternate Assessment Anchors and Eligible Content for Reading, Mathematics, and Science are provided in Appendix D.

*PASA Science* was developed over a three year period prior to its implementation in 2007-08. The development involved a series of focus groups, pilots and field tests. One of the initial products developed was the delineation of Alternate Anchors and Eligible Content developed directly from the regular Pennsylvania Science *PSSA* anchors and eligible content. Once eligible content was delineated, specific assessment items were developed which are used to assess students.

The *PASA Science* Alternate Anchors and Eligible Content include content drawn from the four major reporting areas of the *PSSA* Science assessment: the nature of science, biological sciences, physical science and earth and space science. The *PASA Science* Alternate Anchors and Eligible Content were approved by the Pennsylvania State Board in November, 2008. The Alternate Assessment Anchors and Eligible Content for Science are provided in Appendix D.

During the test design stage, developers reviewed available research literature pertaining to

alternate assessments of science and other related assessment and curricular materials for assessing and teaching science-related skills to students with significant cognitive disabilities. Subsequently, a higher education science content expert (who also has classroom teaching experience teaching science to students with significant disabilities) was retained by the project and several months were spent reviewing and reinterpreting the *PSSA* anchors and eligible content to draft alternate anchors and eligible content. Groups of special education teachers in the Pittsburgh and central Pennsylvania areas were then recruited to further review, develop and validate the alternate anchors and eligible content and to begin to develop ideas for tasks which might be used to assess the targeted skills.

The developers then secured the participation of two science content experts employed by Allegheny Intermediate Unit 3 (who also are teacher trainers in elementary and secondary science) to further refine the alternate anchors and eligible content and to develop tasks to be incorporated into the assessment.

### ***Development of PASA Reading and Math, Historical Perspective***

Prior to the 2006-07 spring administration of *PASA Reading and Math*, a study was completed to verify the adequacy of the conceptual framework underlying *PASA*, initially derived from the general skill areas identified in Appendix B. Specific skills were generated within the general skill areas by teachers, and were reviewed by subject matter experts, experts in the field of low incidence disabilities, and measurement experts. Specific skills ranged from items as simplistic as orienting toward materials, to reading extended blocks of text or solving word problems using addition, subtraction, multiplication, or division. Based on judgments of the cognitive complexity and age appropriateness of each skill, items were placed at specific grade levels and levels of difficulty (A, B, or C). After skills were placed, they were again reviewed by experts to determine if those within a grade were age-appropriate. Items were also examined to determine if the levels of difficulty within a grade differed appropriately and if skills within one level of difficulty (e.g., Level A at grades 3, 5, 8, and 11) changed across grades in a way that was consistent with the *theoretical* pattern of growth for students with severe cognitive disabilities represented in the model shown in Figure 2.

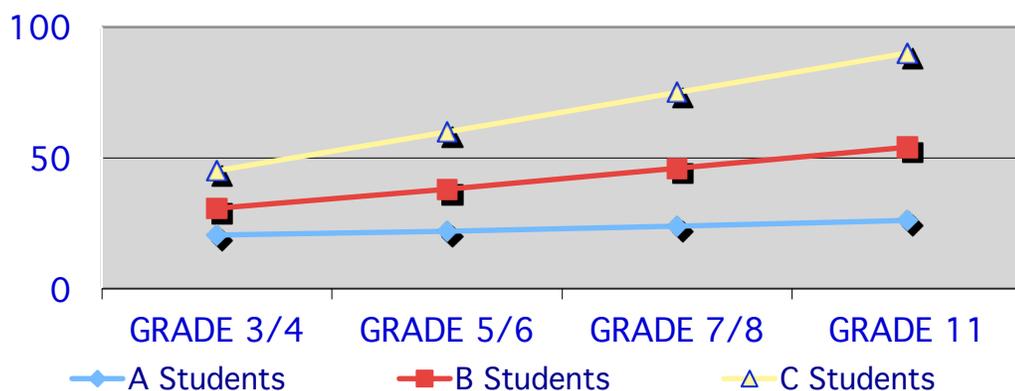


Figure 2. *Theoretical representation of growth trajectories for students assigned to Levels A, B, or C of the PASA*

All items identified by experts were then evaluated in a pilot test conducted in the fall of 2004. In this study, between 31 and 64 students at each grade level from grades 3 through 8 and 11

were selected and given a set of reading and/or math items. Teachers were required to complete a Skills Checklist for each participant. In addition, teachers were asked to read 3 profiles that had been written to describe the typical expected performance of a student who would take levels A, B, or C at a given grade level. Using a combination of patterns of endorsement on the Skills Checklist, the Skills Checklist total score, and the profile selected by a teacher to describe his/her student, students were assigned to receive a Level A, B, or C form in Reading and/or in Math. Teachers were then provided with one of several test booklets. They were instructed to administer the items as they typically would, providing the necessary accommodations and supports to bring out optimal student performance. The teachers were also asked to write down their comments about items. The performance of students was videotaped, and this, along with teacher comments on items, was submitted for scoring. Scorers were asked to rate each item from 0-5 using the usual rubric. They were also asked to select from a list the reason for the score that included a true reflection of the student's ability, a lucky guess, or an error in the administration of the item.

Performance of students was analyzed considering both the score (0-5) and the reason for the score on each item. Items were flagged if they appeared to be too easy or too difficult. In addition, items were examined within a test form to ensure that they were reasonably consistent in terms of difficulty. Items were further examined across forms to make sure that the characteristics identified as increasing the difficulty of items across grades within a level worked effectively.

Across grade levels, existing items and formats that exhibited appropriate characteristics were identified. Extensive reviews were conducted on items that did not perform as desired. Skills that appeared to be fundamentally flawed were eliminated from the conceptual framework. Skills that could potentially be altered and be successful were identified and revisions were made. These items were included as field test items in the subsequent spring administrations of the *PASA Reading and Math*.

Non-field test items were sampled across the domain of possible skills. Consideration was given to general categories (e.g., addition/subtraction, multiplication/division) in order to ensure that the breadth of the content area was covered. When more skills were represented in the Alternate Assessment Anchor Standards than could be assessed in a single 25-item test, skills were tested systematically so that over a two year period all eligible content has been rotated through the assessment.

From the start, individual test items were generated by professionals in special education and in reading and mathematics. Items were reviewed by special education, content, and technical experts in order to identify items that were likely to be biased for or against certain subgroups. Specifically, items were reviewed in an effort to identify those that appeared biased on the basis of gender and also on the basis of setting (rural, urban, or suburban). Items that seemed particularly problematic were revised or eliminated. Considering the nature of the cognitive disabilities for students taking *PASA Reading and Math*, care was taken to present information about subjects with which students were likely to have direct experience (e.g., items that involve activities in the home or in school). In addition, the range of disabilities of students taking *PASA Reading and Math* and the nature of the stimuli and responses were considered and a description of acceptable accommodations was developed. The description appears in the administrator's manual found in Appendix G.

After extensive review by individuals in special education, reading and mathematics, and

measurement, items were tried out by undergraduate and graduate students working toward degrees in special education. This review process helped to identify instructions that were unclear, typographical errors, and materials that were unwieldy.

### ***Development of PASA Science, Historical Perspective***

The *PASA Science* was in development for three years before its implementation in 2007-08. Initially, the developers reviewed available research literature pertaining to alternate assessments of science and other related assessment and curricular materials for assessing and teaching science-related skills to students with significant cognitive disabilities. Subsequently, a higher education science content expert (who also has classroom teaching experience teaching science to students with significant disabilities) was retained by the project and several months were spent reviewing and reinterpreting the *PSSA* anchors and eligible content to draft alternate anchors and eligible content. Groups of special education teachers in the Pittsburgh and central Pennsylvania areas were then recruited to further review, develop and validate the alternate anchors and eligible content and to begin to develop ideas of tasks which might be used to assess the targeted skills.

The developers then secured the participation of two science content experts employed by Allegheny Intermediate Unit 3 (who also are trainers in elementary and secondary science) to further refine the alternate anchors and eligible content and to develop tasks to be incorporated into the assessment. Tasks developed were reviewed and revised by the *PASA* team and materials and kits were prepared for pilot testing. In an effort to ensure that the *PASA-Science* measured science-related content, two separate statewide pilots of the *PASA-Science* were conducted in the spring and fall of 2007.

To collect information about whether the assessment items reflected science-related content, approximately 450 teacher surveys were collected in late spring 2007. Teachers who administered the assessment in the spring of 2007 also provided written feedback on the assessment items on various aspects and characteristic of the items. Additionally, a scoring conference was conducted to score the pilot items and to collect feedback on the assessment items from the participants in the scoring conference. Teacher feedback was collected in the form of Likert type ratings on 17 questions focused on: the value and appropriateness of the items, the structure and quality of the assessment and the potential biases contained within the assessment. Survey questions asked participants to rate their agreement with questions related to validity, tem design, and biases. Of particular note to the question of test validity, test administrators and scorers were asked to rate their agreement with the statement, "This item assesses a science-based skill" for each item of the test level that they administered or scored. Teachers and scorers were also asked to rate their agreement with the statement, "This item assesses a skill that is important for the student to know/be able to do" for each item of the test level that they administered or scored. Survey data were used to make decisions about the design of items, the validity of items, and how to reduce biases within the items prior to the administration of the *PASA-Science* to all the students in grades 4, 8, and 11 in the spring of 2008.

Specifically, over the summer of 2007, the performance data, the scoring reliability data and the scoring conference participant feedback data were reviewed by the developers and the *PASA* staff. As a result of this review, some items were thrown out, some were revised and some were accepted without revision. The developers and the science consultants then developed more items; these items were then reviewed and revised by the *PASA* team and the

(new) items were then pilot tested locally. During the final quarter of 2007 the assessment was assembled together with the materials and the testing kits. Just as in *PASA Reading and Math*, the *PASA Science* skills vary systematically across grades, and within grades across levels A, B, and C.

### ***Bias, Fairness and Sensitivity Reviews, Historical Perspective***

The *PASA* item review team considered each test item and the objects or graphics associated with each item. The team's primary responsibility was to evaluate every test item as to acceptability for students with vision impairments (size and color of objects; clarity of graphics), students who used sign language for communication (signing the instructions did not give away the answer; synonyms were available for vocabulary words to be signed), students whose work table would be no wider than a wheelchair tray (number and size of objects or pictures placed for student consideration), etc. The team also considered issues of familiarity with the content for students educated in separate schools or residential treatment centers, especially items in Level A of the test at any grade span. After item by item discussions, all consensus comments were compiled and actions taken were recorded for future reference.

Expert review by teachers specializing in the instruction of students with hearing impairments was conducted. Certain items that were judged as being flawed, (e.g., irrelevant to students with hearing impairments or flawed for students using ASL) were identified and were eliminated from the computation of these students' final scores.

The final step in item development and fairness review involved selection of objects for test kits as well as graphics, drawings, and photographs for the test booklets. For objects, availability was one consideration, but more importantly, the team considered familiarity, functionality, size, and color. Sample item arrays were reviewed until consensus was reached on the suitability of objects for each test item. The same review process was undertaken for all graphics, pictures, and photographs prepared for the test booklets.

### **Alignment of the *PASA Reading, Math and Science* to the Alternate Standards and Assessment Anchors and Eligible Content**

During 2007-08, the alignment between the *PASA* test items on the 2007 test and the PA Academic Standards and Alternate Academic Standards was examined through an Enhanced Assessment Instruments grant awarded to Pennsylvania, Georgia, Washington State, and Wyoming, with research partners from Measured Progress and the Universities of Oregon, North Carolina at Charlotte, and Western Carolina. The eight criteria used in this alignment study were developed by a collaboration of content experts, special educators, and measurement experts at UNC Charlotte (see Browder, Wakeman, Flowers, Rickleman, Pugalee, & Karvonen, 2006). While some of the alignment criteria are similar to other alignment methods (e.g., Webb, Surveys of Enacted Curriculum, and Achieve), additional criteria (criteria 5-8) were designed especially as value indicators for students with significant cognitive disabilities. All assessment items from each of the 12 Reading and 12 Mathematics grade span assessments for 2006/07 (3&4, 5&6, 7&8, and 11 at each level A, B, and C) were examined (n=267 Reading items; n=283 Math items, n=550 total items). Only items that were included in the grade span assessment for 2006-07 were rated, not the entire bank of items. The Executive Summary of the alignment report was provided as an appendix for the 2009 *PASA* Technical Manual. The

alignment study results suggested that Pennsylvania's extended standards and alternate assessments for students with significant cognitive disabilities are linked to academic content in the majority of instances. The study found some areas for improvement, including decreasing the number of items that were foundational, and providing a test blueprint to more explicitly show the link between *PASA* items and the alternate eligible content. As a result, tables showing the content link between the items and alternate eligible content were added within the technical reports and as appendices. The August 2008 alignment report dealt only with Reading and Math, as *PASA Science* was not administered in 2007.

Two alignment studies, one internal and one external, have been conducted for the *PASA Science*. The results of the internal alignment study were provided as part of a Peer Review packet, and the results of the external alignment study were provided as an appendix to the 2011 Technical Manual. Generally, the alignment studies found that there was alignment associated with the majority of content on the *PASA Science*, across test forms and grade levels. The studies also found areas for improvement. Results from the internal alignment study suggested that an alternate assessment anchor was not linked to the regular education assessment anchors, and items related to this non-linked assessment anchor were phased out of the *PASA Science*. The external alignment study also found that the majority of items linked to the content standards, but did find that some content at certain grades, mainly 11<sup>th</sup> grade Level C, was not linked. As a result of this study, new items following new item writing procedures were added to the *PASA*.

## **New Item Development**

New items had been generated each year by staff (in the cases of reading and math) or commissioned from teachers in the field (in the case of science). The items generated have traditionally reflected the test blueprints, with each item being written to the task/skill specifications, as presented in the blueprints found in Appendix F. Each item was then reviewed by the *PASA* staff for adherence to the grid specifications, for wording and clarity, for graphics or materials, and for absence of bias regarding race, locale, gender, or disability. For new items written for the 11<sup>th</sup> grade Level C *PASA Science* for 2012, external item writers were utilized.

As a result of the external alignment study, items on the 11<sup>th</sup> grade Level C *PASA Science* needed to be revised. General education science content teachers were recruited by the *PASA* team in conjunction with PDE. The teachers were provided with background information about the student population who takes the *PASA*, background information about the assessment including the Alternate Eligible Content and general education Assessment Anchors and Eligible Content, and information about the content and format of assessment items administered previously. The guidelines for new item development were shared, and the item writers were asked to submit items to be considered for the *PASA*. As a result, new items were incorporated into the 11<sup>th</sup> grade Level C *PASA Science*.

## **Restructuring of the 2012 PASA**

A number of changes were made to the assessment forms for the 2012 administration of the *PASA Reading, Math and Science*. These changes resulted from suggestions by the PA Technical Advisory Committee (TAC) as well as suggestions by members from a technical advisory group (TAG) for the *PASA*. The TAG met with representatives from PDE and the *PASA* staff in September 2010. The purpose of the TAG meeting was to review the history and

current status of the *PASA*, and for members of the TAG to make recommendations for areas of exploration or improvement of the assessment program. Both the TAC and the TAG have made suggestions regarding the *PASA* in the last several years, and some of these suggestions, including conducting studies to examine the learning progressions underlying the *PASA*, are being implemented beginning with the 2012 administration. The incorporation of the TAC and TAG suggestions, in combination with the validity studies that are being conducted, has required some restructuring of the *PASA*.

First, TAC members had suggested that the item base that had been developed for the *PASA* was broad enough for the *PASA* staff to consider selecting items from previous years' assessments to administer on a rotating, two year basis. Beginning with the 2011-2012 school year (the 2012 administration), two forms of the *PASA* at each grade level, test level, and content area were developed from the bank of items from past administrations. This pattern of administering alternate forms will continue for subsequent administrations. The first set of forms for all of the grade level and content area combinations was developed and administered in 2012. The second set of forms will be administered in 2013. Items for the alternate, rotating forms were selected based on content characteristics, match to alternate anchors and eligible content, task type, placement in the Performance Level Descriptors, and item level statistics.

In addition to the adoption of two alternate forms of the *PASA* to be administered in alternating years, results from a pilot study conducted in the fall of 2010 intended to examine student performance across levels of the *PASA Reading, Math and Science* also impacted the structure of the *PASA*. The pilot study involved students who had been assigned to and were proficient on the Level A *PASA* (with proficiency being determined by the reported performance level on the previous year's *PASA*) taking a subset of items found on both the Level A and B tests; and students who had been assigned to and were proficient on the Level B *PASA* taking a subset of items found on both the Level B and C tests. The study was carried out for all three content areas. This pilot study involved a special administration of the *PASA* in the fall of 2010 to 304 students. Assessments were scored during a scoring conference in the summer of 2011. Results of the study showed there were some skills on the higher level test that more than 50% of the students who had taken the lower level *PASA* could perform correctly. However, there were several limitations to the pilot study that prevented generalization of these results. First, the sample of students who participated in the pilot study had all been classified as proficient on the previous year's *PASA*, and therefore it was not known how students who were *not* classified as proficient on the Level A *PASA* would perform on the Level B *PASA* items, or how students who were not classified as proficient on the Level B *PASA* would perform on the Level C *PASA* items. Further, the proficiency study involved only a small sample of students in each grade level and content area, including 304 of the approximately 15,000 students served by the *PASA*. Finally, only a subset of the items on the higher level *PASA* were administered, so it was not known if results would generalize to other items measuring different content and skills.

Therefore, based on preliminary results and recommendations from the TAC, a larger scale study was planned to examine the performance of students assigned to take the Level A *PASA* on Level B items, and students assigned to take the Level B *PASA* on Level C items. This study involved the administration of five items from a higher level of the *PASA* to every student taking the 2011-12 *PASA Reading, Math and Science*. For instance, all students taking the Level A *PASA Reading* in each grade span (3/4, 5/6, 7/8, and 11) were administered an assessment that had 20 scored items from the Level A assessment and 5 additional items from the Level B assessment at the respective grade span that were not counted in scoring.

Additionally, all students administered the Level B *PASA Reading* were administered the Level B form of the assessment, which contained 20 scored items, as well as 5 additional items from the Level C assessment that did not count towards scoring. Finally, students taking the Level C *PASA* were administered a 20-item assessment that did not contain additional test items.

Test forms were designed such that the more difficult items were interspersed throughout the forms. At each test level, multiple forms were developed so that all of the alternate eligible content assessed on the higher level *PASA* was represented on at least one test form. Different sets of ‘test’ items from the higher level *PASA* were included on different test forms in order to cover the breadth of content assessed on the higher level *PASA*. That is, all of the alternate eligible content assessed on the Level B *PASA* was represented on at least one Level A form; and all of the alternate eligible content assessed on the Level C *PASA* was represented on at least one Level B form. Results from this study are summarized in the ‘Pennsylvania Alternate System of Assessment Supplemental Report on Validity Studies, with tables of frequency distributions being provided in Appendix B of the Supplemental Report.

Tables 6 through 8 below present summaries of the item distributions for the content areas of Reading, Math, and Science, respectively, by Grade Span, Level, and PSSA Reporting Category. These tables show a total of 20 items being administered on each form of the *PASA*. In addition to these 20 items that were scored, an additional five Level B items were embedded into each Level A form, and an additional five Level C items were embedded into each Level B form, for each content area assessment. Additional items were not embedded into the Level C forms. Table 9 shows the positions of the test items in each form.

Table 6. *Distribution of Items on the 2012 PASA Reading by Grade Span, Level and PSSA Reporting Category*

Grade and Level	Comprehension and Reading Skills		Interpretation and Analysis of Fictional and Nonfictional Text	Total
	Reading	Comprehension		
3/4 A	5	15		20
3/4 B	7	13		20
3/4 C	8	12		20
5/6 A	5	15		20
5/6 B	9	11		20
5/6 C	0	20		20
7/8 A	5	15		20
7/8 B	7	13		20
7/8 C	0	20		20
11A	5	15		20
11B	5	15		20
11C	0	20		20

Table 7. Distribution of Items on the 2012 *PASA Math* by Grade Span, Level and PSSA Reporting Category

Grade and Level	Numbers and Operations	Measurement	Geometry	Algebra	Data Analysis and Probability	Total
3/4 A	5	13	1	1	0	20
3/4 B	7	7	1	1	4	20
3/4 C	10	4	1	2	3	20
5/6 A	5	13	1	1	0	20
5/6 B	8	7	1	1	3	20
5/6 C	9	5	1	2	3	20
7/8 A	5	13	1	1	0	20
7/8 B	9	6	1	1	3	20
7/8 C	9	4	1	3	3	20
11A	5	13	1	1	0	20
11B	10	5	1	1	3	20
11C	9	4	1	3	3	20

Table 8. Distribution of Items on the 2012 *PASA Science* by Grade, Level and PSSA Reporting Category

Grade and Level	Nature of Science	Biological Sciences	Physical Sciences	Earth and Space Sciences	Total
4A	5	5	4	6	20
4B	4	5	6	5	20
4C	4	5	6	5	20
8A	5	5	4	6	20
8B	4	5	5	6	20
8C	4	5	5	6	20
11A	6	5	3	6	20
11B	6	5	3	6	20
11C	4	6	4	6	20

Table 9. Placement of the Supplemental Test Items on the 2012 *PASA Reading, Math and Science*

Subject	Test/Grade	'Test' Item Position	Total Number of 'Test' Items
Reading	3/4 A, 5/6 A, 7/8 A, 11 A	4, 8, 12, 16 & 20	5
	3/4 B, 11 B	8, 12, 13, 14 & 20	5
	5/6 B, 7/8 B	8, 12, 13, 20 & 24	5
Math	3 /4 A, 5/6 A, 7/8 A, 11 A, 3/4 B, 11 B, 5/6 B, 7/8 B	4, 8, 12, 16, 20	5
Science	4A, 8 A, 11A, 4B, 8B, 11B	4, 8, 12, 16, 20	5

## Chapter 4: Additional Versions of the PASA

Although the *PASA* has been designed to be accessible to students with the most severe cognitive disabilities, some students would benefit from additional adaptations to the test. A section in the Administrator Manual and in each test booklet included information about some of the ways the test can be adapted.

### Accommodations

Accommodations are changes to the skills that are designed to reduce or remove the construct-irrelevant variance that arises as a function of disability. Accommodations for an item do not change what is being assessed and do not affect the difficulty level of the math, reading, or science assessed by the items. Given the heterogeneous nature of the population, it was necessary to allow items to be accommodated in a variety of ways. Common types of accommodations included using enlarged pictures or fonts; adaptive or prosthetic equipment; assistive technology devices; Braille or miniatures; sign language; and communication boards, booklets or other devices. Other accommodations involved changes in the setting, in timing, or in scheduling of the assessment. In addition to these accommodations, teachers were permitted to substitute entirely different materials without penalty, provided that the essential characteristics of the task were maintained, as judged by the scorer. This option is utilized rarely.

**Colored versions** of printed test materials were made available as an accommodation for students whose vision problems required testing materials with greater contrast. These could be requested at <http://www.pasaassessment.org/administering/coloredmaterials.html>

Items that remained in the test but were not appropriate for students who use American Sign Language as their communication mode were flagged with the designation **D/HOH** on the top right-hand corner of the test item page in the test booklet. Instructions in the Test Administrator Manual and in the front of the Test Booklet directed test administrators to access alternate replacement items. These items could be requested on the *PASA* website at <http://www.pasaassessment.org/administering/dhoh.html>

For hearing students who **cannot produce a spoken response** when presented with **open-ended questions**, the test booklet contained general guidelines for converting questions requiring open-ended responses to selection responses; these suggestions were specifically to use with students who are unable to produce answers without an augmentative and alternative communication device/system. Test items for which a converted version had been designed were designated with an **AAC** in the top right hand corner of the Test Booklet. Again, instructions in the Test Administrator Manual and in the front of the Test Booklet directed test administrators to access replacement items. These items could be requested on the *PASA* website at <http://www.pasaassessment.org/administering/nonverbal.html> . It should be noted that while some teachers of students taking the Level A *PASA* indicated that the AAC version would be needed, these students received the regular Level A assessment, because the items on the Level A assessments do not require verbal responses.

An adapted version of the *PASA* (“**no-vision version**”) was produced for students with no useable/functional vision; these students require the use of tactile and auditory presentations of the test. Most test items needed to be replaced with equivalent items that were suitable to these presentation modes. This version of the *PASA* could be requested on the *PASA* website at <http://www.pasaassessment.org/administering/visualimpairments.html> .

Table 10 presents the counts of students who took one of the adapted versions of the 2012 *PASA Reading, Math, and Science* by grade and test level. Again, teachers whose students took the Level A *PASA* (only) may have indicated that the AAC version of the test was needed, but because verbal responses are not required on the Level A *PASA*, these students received the regular test rather than an adapted version.

Table 10. *Use of Accommodations for the 2012 PASA Math, Reading, and Science as Reported by the Teacher*

Accommodated Version of PASA	Reading			Math			Science		
	Level			Level			Level		
	A	B	C	A	B	C	A	B	C
<b>Grade 3</b>									
Tactile Version	14	2	0	15	1	0	-	-	-
Augmentative Communication Version	0	11	7	0	11	3	-	-	-
Deaf & Hard of Hearing Version	2	6	0	2	3	0	-	-	-
<b>Grade 4</b>									
Tactile Version	9	1	1	11	1	1	8	11	3
Augmentative Communication Version	0	19	4	0	16	2	0	0	0
Deaf & Hard of Hearing Version	6	9	1	3	4	1	-	-	-
<b>Grade 5</b>									
Tactile Version	9	1	2	9	2	1	-	-	-
Augmentative Communication Version	0	18	6	0	18	8	-	-	-
Deaf & Hard of Hearing Version	5	12	1	7	2	0	-	-	-
<b>Grade 6</b>									
Tactile Version	12	2	3	12	1	2	-	-	-
Augmentative Communication Version	0	26	1	0	16	2	-	-	-
Deaf & Hard of Hearing Version	4	12	1	1	1	2	-	-	-
<b>Grade 7</b>									
Tactile Version	15	2	0	15	1	0	-	-	-
Augmentative Communication Version	0	20	7	0	12	3	-	-	-
Deaf & Hard of Hearing Version	3	6	0	4	1	1	-	-	-
<b>Grade 8</b>									
Tactile Version	13	1	1	11	1	1	28	4	2
Augmentative Communication Version	0	15	4	0	6	2	0	0	0
Deaf & Hard of Hearing Version	6	5	0	2	2	0	-	-	-
<b>Grade 11</b>									
Tactile Version	8	2	1	8	3	1	11	8	1
Augmentative Communication Version	0	8	2	0	8	0	0	0	0
Deaf & Hard of Hearing Version	1	6	1	0	0	0	-	-	-
<b>Total</b>									
Tactile Version	80	11	8	81	10	6	47	23	6
Augmentative Communication Version	0	117	31	0	87	20	0	0	0
Deaf & Hard of Hearing Version	27	56	4	19	13	4	-	-	-

## Modifications

Modifications to the task change what is being assessed and alter the difficulty level of the reading- math- or science-related items. Modifications resulted in the student's score being lowered because the targeted skills were not fully observed. Modifications might include changes in the materials that make the correct answer more obvious. Modifications might also have occurred if the teacher provided additional information that made the task easier to complete. Test administrators were encouraged to make modifications if the student was unable to complete a task as specified. These modifications allow for partial participation. If an item is modified during the assessment by the test administrator, the examinee's score on that item drops to a 3 at best. Appendix G contains a complete list of Accommodations and Modifications.

## Operational Test Forms for the 2012 PASA

For the *PASA Reading and Math*, several forms of the Level A and B assessments were administered at each grade span, 3/4, 5/6, 7/8, and 11. For the *PASA Science*, several forms of the Levels A and B assessments were administered in grades 4, 8 and 11. Each of the operational test forms (that were not adapted versions of the *PASA*) contained the same set of 20 scored items and 5 additional, supplemental, 'test' items.

Prior to 2009, the *PASA Reading, Math and Science* contained 20 items that counted towards scoring, and various forms contained between 0 and 5 additional items that were being field tested. From 2009 - 2011, 25 items were administered on the *PASA Math and Reading*, with the plan that 5 items would be repeated for each administration in order to investigate the equivalence of *PASA* assessments across administrations.

For the 2012 administration, each assessment form at the A and B levels has 20 items that are scored, and five additional 'test' items that are items from a higher level of the *PASA* and do not count towards scoring. In order to have the broadest representation of Level B items on the Level A assessments, multiple forms of the Level A and Level B assessments were developed in each content area. The five supplemental items were different on each form. Any students taking the adapted versions of the *PASA* that were described above (e.g., Deaf/Hard of Hearing version, tactile version) received Form 1 of the respective assessment, so that only one version of each grade level/test level form had to be adapted. The final reported score on the 2012 *PASA* is the total score received across the 20 items on the *PASA Math, Reading and Science* that are not the supplemental, test items.

## Chapter 5: Test Administration Procedures

### Materials Required for PASA Administration

Several sets of materials are required for administration of the PASA. These materials are described in the sections that follow.

#### *Test Booklets*

Test booklets contained all the printed materials required for administration of the PASA at each level and grade span. Teacher packets were prepared and included Rosters of all the students enrolled to be assessed on the PASA by that teacher, a blank Supporting Documentation Form for each student on the teacher roster, a set of preprinted Student Labels for each student on the roster, directions to the teacher on PASA administration, Scoring Conference application notices, and Plastic Bags in which to place completed media recordings. Separate packets were assembled by teacher for each subject area. Then packets were sorted by Teacher, School and District and put into plastic bags. Plastic Bags were boxed for shipment to the appropriate Assessment Coordinator. Assessment coordinators in each school, district, or IU were responsible for distributing packets to the appropriate teachers.

#### *Test Kits*

Physical objects needed for test administration were assembled into Test Kits and distributed directly to test administrators. Materials were placed into bins for each level of test. Bins were boxed and return instructions and a UPS return label were included. Kits were distributed in approximately 2 students to 1 test kit ratio. Kits were mailed via UPS directly to the test administrators.

### Preparation of Test Administrators

#### *Training*

The PASA Reading, Math and Science assessments are designed to be individually administered by a student's teacher or someone who knows the student well. The test administrator must be certified to teach in Pennsylvania. The test administrator can exert a large influence on a student's score by the way in which the materials are varied, the way in which questions are asked, and the extent to which accommodations and modifications are provided. In order to train test administrators, on-line training was held. The online training is for both new and veteran test administrators, and features brief video segments with valuable information for before, during, and after testing. Proficiency tests are also available for test administrators to assess their test administration knowledge. A total of 1500 people registered for the test administration training site. A total of 868 attempted at least one question on one of the seven proficiency tests, 678 people passed all seven proficiency tests with a score of 80% or higher with unlimited attempts, and 569 people passed all seven proficiency tests with a score of 80% or higher with maximum two attempts, earning 5 Act 48 credits.

### ***Administrator's Manual***

In addition to training, every test administrator received an Administrator's Manual. The Manual briefly outlines the purpose of the assessment and some commonly asked questions and answers. In addition, it explains in detail how to effectively administer the assessment and how to submit the materials after testing. Information about general accommodations and modifications and those applicable for students with specific kinds of disabilities, a description of the scoring rubric, a description of the difficulty levels (A, B, and C), the performance standards, and a copy of the Supporting Documentation form are also included. Administrator's Manuals are mailed to each teacher with their test booklets. The Administrator's Manual is also available online at [www.pasaassessment.org](http://www.pasaassessment.org), and is found in Appendix G.

## Chapter 6: Processing and Scoring the PASA

### Processing Completed Assessments

UPS return labels are provided for each box of test booklets and each test kit. Test administrators are instructed to repackage all test materials for return to the *PASA* project offices. Boxed materials can be placed with standard outgoing school UPS pickup. If there is no school pickups, UPS can be called to schedule one.

All packages sent from and to the *PASA* project are tracked by UPS Worldship. Once received, all returned packages are opened, inspected, and Teacher Rosters and Media are separated from everything else. Returned media are logged into the *PASA* database, and reboxed for transport to the scoring conferences. Remaining materials are divided between Test Booklets and Kits. Test booklets are destroyed/recycled. Kits are inventoried and/or counted and sent to an approved contractor for disassembly. There, kit materials are separated, counted, cleaned and returned to the *PASA* Project for storage.

For the 2012 administration of the *PASA*, in an increased effort to heighten the test security, all *PASA* test books were barcoded before they were sent out. Books were then scanned in upon return to *PASA* headquarters in an effort to determine how many books that were sent out were returned. A total of 18,210 books were sent out and 16,437 were returned (90.3%). Updated lists of districts for which books were not returned was submitted to PDE with the final list being submitted in June of 2012.

### Scoring Rubric

Each of the reading, mathematics and science items was scored on a scale from 0 to 5. Scores were assigned based on the degree to which the item was performed independently and whether the correct response was ultimately generated. A score of 5 was assigned if an individual completed an item after receiving the initial instruction only. A score of 1 was assigned if the student passively participated as the teacher guided the student through the item. The scoring rubric used to score the 2012 *PASA* is found in Figure 3.

<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>0</b>
Performed target skill  <b>correctly and independently</b> with beginning prompt only	Performed target skill  <b>correctly</b> with beginning prompt and <b>one additional</b> prompt, redirection, or correction	Performed <b>correctly</b>  but the <b>beginning prompt</b> was <b>not heard</b> <i>or</i> a <b>modified</b> (easier) version of the skill <i>or</i> at least <b>one but not all</b> of the <b>steps</b> of a multi-step skill <i>or</i> with the <b>array reduced</b> <i>or</i> with <b>beginning prompt</b> and <b>two or more additional</b> prompts	Performed <b>incorrectly</b> <i>or</i> Demonstrated a skill completely <b>different</b> from targeted skill <i>or</i> <b>Actively engaged</b> and the teacher <b>ensured</b> the correct response <i>or</i> <b>Actively engaged</b> but the <b>response was not seen nor heard</b>	<b>Passively participated</b> and administrator <b>ensured</b> correct response	<b>Entire skill was not recorded</b>  (skipped, not able to be seen nor heard)
<i>Beginning prompt only</i> <i>Lucky guess</i> <i>Self-correction</i>	<i>Beginning prompt + one additional prompt</i> <i>Silent prompt</i> <i>Confirmation</i>	<i>Easier</i> <i>Reduced array</i> <i>Hints,</i> <i>more info,</i> <i>strategies</i> <i>Some steps done; not all</i> <i>Beginning prompt missing</i>	<i>Engaged but...</i> <i>Response wrong</i> <i>Different skill</i> <i>Responses ensured</i>	<i>Not engaged</i> <i>Administrator ensured response</i>	<i>Skill omitted</i> <i>Not recorded</i>

Figure 3. *Rubric for Scoring the 2012 PASA*

A more complete description of scores and decision rules is provided to each scorer for use during the scoring conferences; it is also published in the Administrator Manual so that test administrators will understand the implications of prompting, accommodations, and modifications as they administer the test. A copy of the expanded Scoring Rubric is provided in Appendix G.

## Scoring Conferences

PASA performances are scored at “scoring conferences”, gatherings of volunteer teachers, school administrators, or University faculty who are trained to apply a scoring rubric to each test item on each assessment returned to PASA. Five “scoring conferences” took place in order to evaluate all of the videotapes and narrative notes that had been submitted. A total of 752 people participated as scorers in these conferences, with some serving at more than one conference. At each conference, scorers were assigned to 2-person teams. Across the set of 5 conferences, there were 749 two-person teams. There were 174 new scoring conference attendees, and 580 people had attended a prior scoring conference. Two-hundred sixty-two of the scorers had administered the PASA this year, and all 752 of the scorers were certified teachers.

### *Recruitment and Selection of Scorers*

An invitation to apply for participation in scoring conferences was included in each PASA teacher packet distributed by the assessment coordinators. In addition, anyone who had scored

in the past received a scoring conference application. All other interested parties could find a scoring conference application on the *PASA* website.

Selection was based on criteria such as: proficiency as measured by the *PASA* Scoring, Training and Proficiency Test, type of certification, past and current experience in administering the *PASA* Assessment, past scoring conference participation, foreign/sign language skills, and availability.

### ***Training Scorers***

In 2012, in order to be selected as a scorer, applicants had to participate in an on-line training module and pass the *PASA* Scoring, Training and Proficiency Test. In the training module, the general rules of scoring, videotaped examples, and details pertaining to specific scores were presented, and applicants engaged in practice scoring activities. Upon completion of the training module, applicants completed the *PASA* Scoring, Training, and Proficiency Test. If participants passed the proficiency test (16/20 items scored correctly) in one or two attempts, they were invited to participate in one or more scoring conference. Applicants who did not pass the test were not invited to attend a conference. In order to be selected to score the *PASA* Science (which occurs at the last scoring conference), scorers had to have already participated in a previous scoring conference in 2012. A total of 1237 individuals registered for the training site, and 95 attempted the proficiency test, meaning that they answered at least one question. A total of 927 people passed the proficiency test with a score of 80% or higher.

At the start of the scoring conferences, a short review of the scoring procedures was conducted, and participants scored a reliability tape at the first conference they attended. Participants who met the standard on the reliability tape began scoring. Participants who did not meet the standard met with *PASA* staff who provided feedback on scoring errors and additional training before scoring.

### ***Reliability to Standard***

After the short scoring review, all scoring teams were asked to evaluate one tape to establish initial reliability. Each team worked independently to rate the same videotaped performances. The standards tape contained videotaped performance clips across grades, levels of difficulty, and score ranges. Clips were specifically selected so that they were challenging and offered a reasonable assessment of scorers' understanding of the scoring rules. The ratings by each scoring team were then compared against the scores assigned by the *PASA* staff prior to the conferences. The percentage of items on which there was exact agreement appears in the following table.

Table 11. *Reliability to Standard - Average Percentage of Items with Exact Agreement*

Scoring Conference	Subject	% Exact Agreement
1	Reading	83.1
	Math	89.5
	Total	86.3
2	Reading	82.8
	Math	88.6
	Total	85.7
3	Reading	82.2
	Math	84.9
	Total	83.5
4	Reading	80.3
	Math	89.3
	Total	84.8
5	Science	77.3

***Team-to-Team Reliability***

***Percentage of Tapes Scored Twice.*** In addition to reliability to the standard, a subset of tapes was selected randomly to establish team-to-team reliability at each conference and between conferences. Pairs of scoring teams were not prearranged. Instead, any 2 teams might form a reliability pair for a given tape. The counts and percentages of tapes double scored for reliability appear below.

Table 12. *Team-to-Team Reliability Counts and Percentages*

Grade	Read and Math		Science	
	# tapes scored for reliability	% of tapes scored for reliability	# tapes scored for reliability	% of tapes scored for reliability
3	234	11.3	-	-
4	214	10.4	200	10.1
5	166	7.6	-	-
6	232	10.9	-	-
7	269	12.6	-	-
8	225	10.8	164	8.1
11	288	14.7	171	9.1

Percent Agreement for Item Scores. The percentages of scores that matched exactly, that were within 1 point, or that were 2 points or more were computed. The percentage of exact agreement between scores averaged between 80.7% and 92.0% for all grades for reading, mathematics and science. The specific results appear on the next several tables.

Table 13. *Grades 3 and 4: Percent Agreement for Item Scores*

Percent agreement between scores	3 <sup>rd</sup> grade Reading	3 <sup>rd</sup> grade Math	4 <sup>th</sup> grade Reading	4 <sup>th</sup> grade math	4 <sup>th</sup> grade science
	Average	Average	Average	Average	Average
Exact match	90.0	91.4	90.3	90.0	84.5
Within 1 point	6.9	6.6	5.9	6.5	10.4
Differ by 2 or more points	3.1	2.0	3.8	3.5	5.1

Table 14. *Grades 5 and 6: Percent Agreement for Item Scores*

Percent agreement between scores	5 <sup>th</sup> grade Reading	5 <sup>th</sup> grade Math	6 <sup>th</sup> grade Reading	6 <sup>th</sup> grade math
	Average	Average	Average	Average
Exact match	88.5	89.3	90.4	90.8
Within 1 point	6.7	6.9	5.8	5.3
Differ by 2 or more points	4.8	3.8	3.8	3.9

Table 15. *Grades 7 and 8: Percent Agreement for Item Scores*

Percent agreement between scores	7 <sup>th</sup> grade Reading	7 <sup>th</sup> grade Math	8 <sup>th</sup> grade Reading	8 <sup>th</sup> grade math	8 <sup>th</sup> grade science
	Average	Average	Average	Average	Average
Exact match	86.8	85.2	88.1	88.0	84.0
Within 1 point	8.3	9.8	6.5	7.5	10.0
Differ by 2 or more points	4.9	5.0	5.4	4.5	6.0

Table 16. *Grade 11: Percent Agreement for Item Scores*

Percent agreement between scores	11 <sup>th</sup> grade Reading	11 <sup>th</sup> grade math	11 <sup>th</sup> grade science
	Average	Average	Average
Exact match	91.3	92.0	80.7
Within 1 point	4.7	5.5	12.7
Differ by 2 or more points	4.0	2.5	6.6

## Chapter 7: Score Reporting

Individual student-level scores from the *PASA Reading, Math and Science* are provided to Data Recognition Corporation, the contractor for the *PSSA*. These data are used in the calculation of AYP. In addition, individual student data is summarized in a report that is sent, in duplicate, to schools for placement in the students' permanent records and distributed to parents. The report briefly describes the skills on which the student was tested, the performance levels earned, and an explanation of the meaning of the scores.

Individual score reports are generated and two copies of each student's report are sent to the student's school: one to be placed in the student's permanent record, the other to be sent home to the student's parents/guardian. The score report has undergone minor modification from the 2012 report, in a continuous attempt to improve the feedback provided to students and parents. An example of the score report for the 2012 administration is provided in Appendix H.

A study was conducted to directly examine issues related to the reporting of assessment scores to parents in Pennsylvania. The study examined meaningful ways of reporting alternate assessment results to parents. The study is described in the 'Pennsylvania Alternate System of Assessment Supplemental Report on Validity Studies' and a complete summary of the findings is provided in Appendix D of that supplemental report. Findings from the study will be considered for implementation in future score reports.

## Chapter 8: Statistical Analyses of Item Data

### Descriptive Statistics and Internal Consistency

Descriptive statistics and internal consistency of each level at each grade were computed. Results for reading, mathematics and science are provided in the following tables. Test administrators were asked to mark assessments for which they began to administer the test but stopped. Scorers were asked to verify that the assessment was stopped during scoring. Overall, 450 students in reading, 457 in math, and 143 in science stopped the test after beginning administration. Summary statistics are presented for those students who completed the assessment according to their test administrator. For all analyses in Chapter 8, Chapter 10 and Appendices I, J and K, the subset of examinees with incomplete assessments as marked by the teacher and verified by the scorer is excluded.

Total scores on the test were computed based on 20 items for the *PASA Reading, Math and Science*. As was stated, 20 items were counted towards scoring on the *PASA* for the 2012 administration.

Despite the fact that scores are unweighted by the level of difficulty, students tended to earn the lowest scores, on average, for the most simplistic tasks (Level A) and higher (or comparable) scores on average for the B and C level tasks in reading. Results in mathematics and science were similar. Cronbach's Alpha reliability coefficients were also computed using SPSS, and were at a high level across content areas and test forms.

The standard errors of measurement (SEM) for each test level at each grade level are also provided. The SEMs were computed using the formula:  $SEM = SD * \sqrt{1 - \alpha}$  for the raw scores, where SD is the standard deviation of the raw scores on the test, and is the Cronbach's alpha reliability coefficient. Raw scores on the *PASA* are computed by summing the individual item scores across all items.

### Item Correlations

The inter-item correlations were computed for each item in reading, math and science for each grade span and level. Correlations were also computed for each grade. Inter-item correlations tended to be higher for Level A than for Levels B and C across grades and content areas with few exceptions. Correlations also tended to be higher in grade 11 than in grades the lower grades for all levels of difficulty. The full set of item correlations for each grade span and test level appears in Appendix I.

Table 17. *Descriptive Statistics, Cronbach's Alpha and SEM for the 2012 PASA Reading (20 Items) for All Students Completing the PASA*

Grade/ Level	n	Mean	SD	Cronbach's $\alpha$	SEM
3A	864	82.4	21.2	.967	3.85
3B	761	88.8	11.3	.895	3.66
3C	378	82.6	14.0	.904	4.34
4A	636	82.2	19.2	.959	3.89
4B	954	81.5	19.2	.907	5.86
4C	413	86.4	14.7	.927	3.97
5A	613	78.0	20.2	.957	4.19
5B	939	85.9	13.8	.917	3.98
5C	571	83.1	13.4	.899	4.26
6A	548	78.4	21.1	.961	4.17
6B	928	88.9	12.2	.915	3.56
6C	578	85.1	14.0	.915	4.08
7A	537	80.1	21.6	.967	3.92
7B	915	87.0	13.5	.922	3.77
7C	628	83.7	14.1	.923	3.91
8A	531	81.4	21.0	.969	3.70
8B	879	79.8	20.1	.895	6.51
8C	601	86.2	12.5	.904	3.87
11A	606	82.3	21.8	.970	3.78
11B	654	87.4	15.0	.935	3.82
11C	626	82.6	13.9	.911	4.15

Table 18. *Descriptive Statistics, Cronbach's Alpha and SEM for the 2012 PASA Math for All Students Completing the PASA*

Grade/ Level	n	Mean	SD	Cronbach's Alpha	SEM
3A	867	80.9	19.7	.954	4.23
3B	728	86.7	13.8	.920	3.91
3C	405	82.8	14.3	.917	4.11
4A	713	82.1	17.9	.944	4.24
4B	861	89.1	12.0	.910	3.61
4C	433	87.4	12.5	.912	3.70
5A	646	77.5	19.9	.950	4.45
5B	825	86.8	12.6	.905	3.87
5C	648	83.9	12.6	.892	4.14
6A	566	78.0	20.3	.954	4.36
6B	848	90.2	10.8	.900	3.41
6C	638	86.7	12.5	.902	3.92
7A	626	79.9	20.5	.957	4.24
7B	788	83.4	14.6	.925	3.99
7C	668	83.1	12.6	.894	4.10
8A	613	82.0	17.9	.946	4.16
8B	777	85.5	13.1	.906	4.01
8C	619	84.8	12.0	.884	4.09
11A	621	79.4	22.8	.968	4.07
11B	728	82.3	16.0	.927	4.31
11C	535	81.6	14.3	.908	4.34

Table 19. *Descriptive Statistics, Cronbach's Alpha and SEM for the 2012 PASA Science (20 Items) for All Students Completing the PASA*

Grade/ Level	n	Mean	SD	Cronbach's alpha	SEM
4A	718	80.9	21.0	.963	4.03
4B	830	83.7	13.8	.921	3.87
4C	407	82.6	11.6	.895	3.76
8A	613	77.4	22.1	.966	4.05
8B	765	86.4	12.1	.897	3.87
8C	598	83.4	13.1	.902	4.08
11A	616	72.8	24.1	.966	4.42
11B	693	74.7	13.6	.889	4.54
11C	511	79.4	12.5	.874	4.42

## Item Performance

Item performance varied systematically as a function of level of difficulty, but remained relatively consistent across grades. The mean and standard deviation for each item for each grade span appear in Appendix J. Means and standard deviations for each item by grade were also computed.

## Percentages of Items at Each Score Level

Appendix K shows the percentage of scores (0-5) assigned for each item for each grade and level of difficulty for all students attempting a given task. In reading, mathematics and science, across grades and levels, scores of 5 were the most prevalent, on average. Scores of 0 and 1 are not prevalent.

For field test items, item statistics are used as a means of detecting items that deserve close scrutiny, rather than being a mechanism for automatic rejection or retention of items. Towards this end, a set of criteria was used as a screening tool to identify items that need a closer review by the test development committee. For an item to be flagged, the criteria included any of the following:

- Correlation (Pearson's) of the item score to the total test score of less than .40
- Percent of students scoring 5 greater than 80% or less than 20%
- Percent of students scoring 2 or 3 greater than percent scoring 4 or 5

The tables in Appendix K were also used to evaluate whether teachers (or other individuals) were adequately trained to administer the *PASA*. Scores of 0 indicate that the person

administering the test did not administer the test item. These scores of 0 inevitably lower a student's scores. Additional training to test administrators emphasizing that the full set of items should be attempted in most cases, should reduce instances of 0 scores and improve students' scores.

## **Differential Item Performance**

During the 2006-07 year, differential item performance was studied in students whose visual impairment does not permit using vision as the input mode for the assessment. Performance on items was compared for students with significant visual impairments and those without visual impairments using a matched pairs Wilcoxon test. Items flagged as functioning differently for the groups were further examined by experts in vision to determine the source for the difference in item functioning. A qualitative review of items revealed additional areas of bias: In some cases, students with visual impairments lacked of familiarity with the content (e.g., things that refer to color), or items could not be altered to make them accessible to students with visual impairments (e.g., a complex picture of a playground). A replacement test was created and placed on the *PASA* website for use with students who do not have useable vision.

## Chapter 9: Item Performance on Supplemental ‘Test’ Items

A large scale validity study was designed and implemented during the 2011-12 school year to investigate the performance of students assigned to each level of the *PASA Reading, Math and Science* on items that appeared on test forms at higher levels of the assessment. The study is described in detail in the ‘Pennsylvania Alternate System of Assessment Supplemental Report on Validity Studies’ and is also described briefly here. Appendix B of the Supplemental Report on Validity Studies is replicated as Appendix K of this technical report.

Students assigned to the Level A assessment were also administered a set of 5 items found on the Level B assessment; students assigned to the Level B assessment were also administered a set of 5 items found on the Level C assessment. Students assigned to the Level C assessment were not administered additional test items. Students taking levels A and B of the *PASA* in each content area (Reading, Math and Science) at each assessed grade/grade span (3/4, 5/6, 7/8 and 11 in Reading and Math and 4, 8, and 11 in Science) were administered a set of 5 additional, unscored items.

The study was designed to provide information about the learning progressions underlying the *PASA* levels and the procedure for assigning students to test levels. Test forms were designed such that the more difficult items were interspersed throughout the forms. At each test level, multiple forms were developed so that all of the alternate eligible content assessed on the higher level *PASA* was represented on at least one test form. Different sets of ‘test’ items from the higher level *PASA* were included on different test forms in order to cover the breadth of content assessed on the higher level *PASA*. That is, all of the alternate eligible content assessed on the Level B *PASA* was represented on at least one Level A form; and all of the alternate eligible content assessed on the Level C *PASA* was represented on at least one Level B form. Generally, between 3 and 7 forms were administered across grade levels, test levels and content areas.

For each version of the *PASA* test forms, item frequencies for the five additional test items from the higher level *PASA* were computed. That is, for the Level A form at grade 3, three versions were administered, where each version contained a different set of five ‘test’ Level B items. A total of 15 ‘test’ Level B items were administered to students taking the Level A assessment. Frequency distributions for these items are found in Appendix K. Again, this appendix was also included as Appendix B in the Supplemental Validity Study Report.

Results showed that relatively large percentages of students who were assigned to the Level A *PASA* and were administered the Level B items, received a score of 5 on those items. For instance, for the 15 Level B items that were assessed across the three versions of the grade 3 Level A form, between 16% and 53% of students taking the Level A assessment scored 5 on these items. For students taking the Level B form at grade 3, the percentages of students who scored 5 on the Level C items ranged from 15% to 76%. Across the skills that were assessed, across grade levels, and across content areas, there were some items that only small percentages of students taking the lower level *PASA* could answer correctly. However, on each

form, there were several items that large percentages of students who were assigned to take a lower test level could answer on the higher form. Overall, results suggested that there were several skills on the higher level *PASA* that larger percentages of students who were assigned to the lower level assessment could perform correctly. The percentages of students answering the reading test items were generally lower than the percentages for math. However, for all content areas, a relatively high percentage of students at each performance level on the test items by students taking the lower level of the assessment performed well.

Additional analyses were run to investigate the performance on the items by students who were classified as emerging, novice, proficient and advanced. There was a significant drop in the percentages of scores of 5 for students classified at the performance categories below proficient. These results indicate that additional content-linked analyses would provide additional insight into the learning progressions of the students.

Additional analyses exploring other ways to examine the data to better understand the learning progressions of students with significant cognitive disabilities are under way with this data set. One analysis examines the performance of students on the 'test' items compared to performance on the items from their level test. Analyses are planned to investigate the particular Level B skills and content that students assigned to the Level A assessment scored highest on, and investigate the relationship between student performance on the Level A and Level B items that represent that same alternate eligible content.

## Chapter 10: Setting Standards: Performance Level Descriptions and Cut Scores

### Standard Settings for the PASA Reading, Math and Science

Two previous standard settings for the *PASA Reading and Math* have occurred since the inception of the assessment. The first took place after the first administration of the *PASA* and the second took place using data from the Spring 2006 administration. Due to the time since the previous standard setting, changes to the scoring rubric, and recent literature related to including content into Performance Level Descriptors (PLDs), new standard setting workshops for the *PASA Reading and Math* were conducted in 2009-2010. A detailed summary of the standard setting workshops was provided in the 2010 *PASA* technical manual. The standard setting workshops were held over two weekends, and the Performance Level Descriptors (PLDs) were revised during the first weekend of the standard setting workshop, and cut scores were established during the second weekend.

The first administration of the *PASA Science* took place in 2008, and the standard setting for the *PASA Science* assessment took place in early Fall of the 2008-09 school year. As a result of this process, cut scores were recommended to separate the reported performance levels (Advanced-Level C only, Proficient, Novice and Emerging) for the three levels of tests (A, B, and C) at each grade level (4, 8, and 11). The standard setting for the *PASA Science* took place in 2008, and was summarized in a standard setting report and in the 2009 Technical Manual. The standard setting workshop was held over two weekends, and the Performance Level Descriptors (PLDs) were established during the first weekend of the standard setting workshop, and cut scores were established during the second weekend.

Due to the restructuring of the *PASA* in the 2012 administration, the cut scores had to be revised. The cut scores for the *PASA Reading and Math* had been based on 25 items. All forms of the *PASA* for the 2012 administration were based on 20 items. Upon consultation with the TAC and PDE, new standard setting workshops were not recommended. Instead, mathematical adjustments to the existing cut scores, based on the items that were included in the 2012 administration of the *PASA*, were recommended.

### Brief Overview of the Most Recent Standard Setting Workshops for the PASA Reading, Math, and Science

The standard settings for the *PASA Reading, Math and Science* took place over two consecutive weekends in the Fall of 2009, Spring of 2010, and Fall of 2008, respectively. The purpose of the first weekend of each standard setting was to revise (in the case of reading and math) or write (in the case of science) content-based performance level descriptors (PLDs) for each reported performance level. The purpose of the second weekend of each was to establish cut scores. PLDs and cut scores were established to allow for score reporting for 4 performance categories (Advanced, Proficient, Novice and Emerging) for the Level C tests, and 3 performance categories (Proficient, Novice and Emerging) for the Levels A and B tests. The same methodology was followed for the standard setting workshops, and is briefly described below.

## ***Panelists***

A total of 36 panelists were recruited for the standard setting sessions for math and reading, so that the targeted panels for each grade level (3/4, 5/6, 7/8, and 11) would consist of 9 members. A total of 35 panelists participated in both the standard setting for the PASA Reading and PASA Math due to last minute cancellations. A total of 27 panelists were recruited for the standard setting sessions for science, so that the targeted panels for each grade level (4, 8, and 11) would consist of 9 members. A total of 25 panelists participated due to last minute cancellations. Panelists from several stakeholder groups, including special education and general education teachers at the elementary, middle, and high school levels, content experts, and special education administrators participated. Panelists worked in grade level teams of 8-9 to both develop PLDs and established cut scores for the Level A, B and C assessments for one grade level.

## ***Development of Performance Level Descriptors (PLDs)***

The first weekend of the standard setting workshop was dedicated to developing PLDs for each performance level (Advanced, Proficient, Novice, and Emerging for Level C, and Proficient, Novice, and Emerging for Levels A and B) of the *PASA Reading, Math and Science* for each grade level (3/4, 5/6, 7/8, and 11 in Math and reading; 4, 8 and 11 in science).

For both weekends, panelists worked in grade level panels of 8-9. Each grade level panel of 8-9 individuals worked with all levels of the *PASA* (A, B, and C) at a single grade level (3/4, 5/6, 7/8, OR 11). After the initial training, grade level panels were broken into teams of 2-3 to view approximately 6 individual assessments of the *PASA* that spanned a wide range of ability at each assessment level. Each team at a single grade level viewed a different subset of student assessments to maximize the number and variety of assessments that could be viewed and discussed in the approximately 10 hours allotted to this task. While viewing individual assessments, panelists had a copy of the *PASA* test booklet and a worksheet on which to take individual notes regarding specific items as well as such aspects as the content, skills, and independence that they would expect of students at each performance level. After watching all videos at a single test Level (e.g., Level A), each group of 2-3 panelists filled out a sheet that summarized how the group described the students at that test level. By the end of the day on Saturday, each team had summarized the characteristics of high, medium, and low performing examinees on each test level for their assigned grade level.

On Sunday of Weekend 1, the small teams convened into grade level teams of 8-9 and developed content-based PLDs for each performance level. Drafts were shared and reviewed across grades, and revisions were made based on these reviews. At the end of Weekend 1, grade level panels had a final draft of the PLDs for each level (A, B and C) of their assigned grade level assessment. At various points during the first weekend, panelists were asked to provide documentation of their understanding of the purpose and processes taking place during the weekend, as well as their confidence in results obtained at various points over the weekend.

## ***Establishing Cut Scores***

Weekend 2 of the standard settings was dedicated to establishing cut scores. A combination of the mean estimation method and modified extended Angoff procedure was used for the standard setting. Initially, training on the standard setting procedure was conducted. The training involved summarizing the task, introducing and modeling the standard setting

procedure, and practice by the panelists in implementing the procedure. Panelists were asked to use the test booklet, item characteristics, PLDs they had developed, and scoring rubric to complete the rating task. Panelists were asked to independently estimate the average score that would be received on each of the test items by 100 students who just barely meet the performance standard for each reported performance category. Panelists were asked to carry out this process for these borderline examinees at the Advanced, Proficient, and Novice performance levels for Level C of their grade level test, and at the Proficient and Novice performance levels for Levels A and B of their grade level test. Panelists were also asked to record this information first on paper and then in an electronic file which was double checked by *PASA* staff. The scores for a single panelist were summed across items to get the panelist's estimated cut scores, and then averaged across panelists to get each panel's estimated cut scores.

After the first round of ratings, descriptive statistics on the cut scores were shared with grade level panels. A guided discussion about the general meaning of the ratings, estimated cut scores, and variability in the cut scores took place. Panelist-led, item-level discussion followed. After the grade level group discussions, panelists independently completed a second round of ratings.

After this 2<sup>nd</sup> iteration, the individuals had established new cut scores. Descriptive statistics on the cut scores as well as impact data based on the new cut scores applied to the most recent administration of the *PASA* was provided to facilitate discussion among the panelists. Further, item level frequency distributions based on the most recent administration were shared. After group discussion, the rating process was repeated on an individual level one additional time, and the final cut scores were recommended based on these final ratings.

Several measures of validity evidence were collected during the standard setting weekends. First, the process was documented by surveying panelists at several time points regarding their levels of understanding and confidence in results. Second, an observer was present during the standard setting sessions to evaluate the extent to which the plan for the standard setting is followed. The observer provided a written summary of her evaluation after the second weekend of each standard setting.

## **Establishing Cut Score for the 2012 Administration of the *PASA***

In order to establish the cut scores for the 2012 administration, the item level data obtained at the standard setting workshops was utilized. Many of the items that had been part of the most recent standard setting workshops were found on the 2012 assessment. As was stated, the final cut scores from the standard setting workshops were found by summing across items to obtain a single panelists' estimated cut scores, and then averaging across panelists to get each panel's estimated cut scores. This same process was followed for the subset of 20 items that were administered in 2012. When possible, the final item ratings established by the standard setting panelists for the exact items that were on the 2012 administration of the *PASA* were used to establish the revised cut scores. When exact items from the standard setting did not appear on the 2012 *PASA*, the estimated cut scores from items that had similar characteristics in terms of content, item difficulty, and item format to those from the standard setting were used. One additional adjustment was made to the 3/4 B Math form, where the cut score for proficient from 97 to 93 in order to make the number of student at each of the performance levels more in line with each other. The cut scores for the *PASA Reading, Math and Science* for the 2012 administration are provided in the Table 20. For comparison, the cut scores for the 2011 administration are also provided in Table 21.



Table 20. *Score Ranges for the 2012 PASA Reading, Math and Science*

	<b>Emerging</b>	<b>Novice</b>	<b>Proficient</b>	<b>Advanced</b>
<b>3/4 Reading</b>				
<b>A</b>	0 - 70	71 - 87	88 - 100	-
<b>B</b>	0 - 74	75 - 89	90 - 100	-
<b>C</b>	0 - 71	72 - 84	85 - 93	94 - 100
<b>3/4 Math</b>				
<b>A</b>	0 - 75	76 - 88	89 - 100	-
<b>B</b>	0 - 84	85 - 92	93 - 100	-
<b>C</b>	0 - 77	78 - 86	87 - 94	95 - 100
<b>4 Science</b>				
<b>A</b>	0 - 71	72 - 88	89 - 100	-
<b>B</b>	0 - 70	71 - 87	88 - 100	-
<b>C</b>	0 - 72	73 - 81	82 - 91	92 - 100
<b>5/6 Reading</b>				
<b>A</b>	0 - 71	72 - 87	88 - 100	-
<b>B</b>	0 - 72	73 - 89	90 - 100	-
<b>C</b>	0 - 70	71 - 84	85 - 92	93 - 100
<b>5/6 Math</b>				
<b>A</b>	0 - 77	78 - 89	90 - 100	-
<b>B</b>	0 - 75	76 - 87	88 - 100	-
<b>C</b>	0 - 73	74 - 82	83 - 91	92 - 100
<b>7/8 Reading</b>				
<b>A</b>	0 - 73	74 - 86	87 - 100	-
<b>B</b>	0 - 71	72 - 86	87 - 100	-
<b>C</b>	0 - 70	71 - 84	85 - 93	94 - 100
<b>7/8 Math</b>				
<b>A</b>	0 - 72	73 - 86	87 - 100	-
<b>B</b>	0 - 77	78 - 90	91 - 100	-
<b>C</b>	0 - 74	75 - 85	86 - 92	93 - 100
<b>8 Science</b>				
<b>A</b>	0 - 65	66 - 84	85 - 100	-
<b>B</b>	0 - 68	69 - 86	87 - 100	-
<b>C</b>	0 - 69	70 - 83	84 - 93	94 - 100
<b>11 Reading</b>				
<b>A</b>	0 - 68	69 - 85	86 - 100	-
<b>B</b>	0 - 71	72 - 89	90 - 100	-
<b>C</b>	0 - 65	66 - 82	83 - 92	93 - 100
<b>11 Math</b>				
<b>A</b>	0 - 73	74 - 87	88 - 100	-
<b>B</b>	0 - 72	73 - 87	88 - 100	-
<b>C</b>	0 - 69	70 - 83	84 - 93	94 - 100
<b>11 Science</b>				
<b>A</b>	0 - 85	86 - 95	96 - 105	-
<b>B</b>	0 - 79	80 - 92	93 - 105	-
<b>C</b>	0 - 77	78 - 89	90 - 96	97 - 105

Table 21. Score Ranges for the 2011 PASA Reading, Math and Science

	<b>Emerging</b>	<b>Novice</b>	<b>Proficient</b>	<b>Advanced</b>
<b>3/4 Reading</b>				
<b>A</b>	0 – 89	90 – 110	111 – 125	-
<b>B</b>	0 – 93	94 – 112	113 – 125	-
<b>C</b>	0 – 90	91 – 106	107 – 117	118 – 125
<b>3/4 Math</b>				
<b>A</b>	0 – 91	92 – 110	111 – 125	-
<b>B</b>	0 – 99	100 – 113	114 – 125	-
<b>C</b>	0 – 95	96 – 107	108 – 117	118 – 125
<b>4 Science</b>				
<b>A</b>	0 – 71	72 – 88	89 – 100	-
<b>B</b>	0 – 70	71 – 87	88 – 100	-
<b>C</b>	0 – 72	73 – 81	82 – 91	92 – 100
<b>5/6 Reading</b>				
<b>A</b>	0 – 91	92 – 110	111 – 125	-
<b>B</b>	0 – 91	92 – 111	112 – 125	-
<b>C</b>	0 – 87	88 – 105	106 – 115	116 – 125
<b>5/6 Math</b>				
<b>A</b>	0 – 91	92 – 110	111 – 125	-
<b>B</b>	0 – 95	96 – 111	112 – 125	-
<b>C</b>	0 – 92	93 – 103	104 – 115	116 – 125
<b>7/8 Reading</b>				
<b>A</b>	0 – 92	93 – 109	110 – 125	-
<b>B</b>	0 – 89	90 – 108	109 – 125	-
<b>C</b>	0 – 88	89 – 105	106 – 116	117 – 125
<b>7/8 Math</b>				
<b>A</b>	0 – 91	92 – 108	109 – 125	-
<b>B</b>	0 – 97	98 – 113	114 – 125	-
<b>C</b>	0 – 94	95 – 107	108 – 116	117 – 125
<b>8 Science</b>				
<b>A</b>	0 – 65	66 – 84	85 – 100	-
<b>B</b>	0 – 68	69 – 86	87 – 100	-
<b>C</b>	0 – 69	70 – 83	84 – 93	94 – 100
<b>11 Reading</b>				
<b>A</b>	0 – 87	88 – 108	109 – 125	-
<b>B</b>	0 – 90	91 – 111	112 – 125	-
<b>C</b>	0 – 82	83 – 103	104 – 116	117 – 125
<b>11 Math</b>				
<b>A</b>	0 – 93	94 – 110	111 – 125	-
<b>B</b>	0 – 91	92 – 109	110 – 125	-
<b>C</b>	0 – 88	89 – 105	106 – 116	117 – 125
<b>11 Science</b>				
<b>A</b>	0 – 81	82 – 91	92 – 100	-
<b>B</b>	0 – 76	77 – 88	89 – 100	-
<b>C</b>	0 – 74	75 – 85	86 – 92	93 – 100

Tables 22 to 24 show the percents of students falling at each performance level for the 2012 PASA in the content areas of Reading, Math and Science, respectively.

Table 22. *Percent of PASA Students at Each Performance Levels for the 2012 PASA Reading*

<b>Level</b>	<b>Emerging</b>	<b>Novice</b>	<b>Proficient</b>	<b>Advanced</b>
<b>3A</b>	23.1	21.4	55.4	-
<b>3B</b>	10.2	28.6	61.1	-
<b>3C</b>	15.9	32.0	32.3	19.8
<b>3 ALL</b>	19.4	25.3	51.6	3.6
<b>Grade 4</b>				
<b>4A</b>	21.5	30.3	48.2	-
<b>4B</b>	6.9	23.3	69.8	-
<b>4C</b>	12.1	18.4	33.2	36.3
<b>4 ALL</b>	15.3	23.8	53.7	7.3
<b>Grade 5</b>				
<b>5A</b>	28.9	32.5	38.7	-
<b>5B</b>	11.5	40.2	48.4	-
<b>5C</b>	13.8	31.2	32.2	22.8
<b>5 ALL</b>	19.4	34.5	40.1	6.0
<b>Grade 6</b>				
<b>6A</b>	28.7	28.5	42.7	-
<b>6B</b>	8.7	29.2	62.1	-
<b>6C</b>	9.0	25.2	36.8	29.0
<b>6 ALL</b>	16.8	27.1	48.2	7.9
<b>Grade 7</b>				
<b>7A</b>	29.2	19.9	50.9	-
<b>7B</b>	9.6	23.9	66.4	-
<b>7C</b>	11.8	29.1	38.4	20.7
<b>7 ALL</b>	17.6	23.9	52.5	6.1
<b>Grade 8</b>				
<b>8A</b>	26.3	19.7	53.9	-
<b>8B</b>	6.8	23.7	69.5	-
<b>8C</b>	10.3	22.1	38.0	29.6
<b>8 ALL</b>	15.7	21.5	54.3	8.5
<b>Grade 11</b>				
<b>11A</b>	20.4	20.8	58.8	-
<b>11B</b>	10.8	28.5	60.7	-
<b>11C</b>	7.7	31.4	41.6	19.3
<b>11 ALL</b>	15.8	26.1	51.9	6.2

Table 23. *Percent of PASA Students at Each Performance Levels for the 2012 PASA Math*

<b>Level</b>	<b>Emerging</b>	<b>Novice</b>	<b>Proficient</b>	<b>Advanced</b>
<b>3A</b>	27.7	25.8	46.5	-
<b>3B</b>	31.5	26.6	41.9	-
<b>3C</b>	22.7	29.9	33.8	13.6
<b>3 ALL</b>	30.4	26.1	40.9	2.7
<b>4A</b>	26.9	26.6	46.5	-
<b>4B</b>	23.5	26.6	49.9	-
<b>4C</b>	12.2	20.8	39.7	27.3
<b>4 ALL</b>	24.5	24.6	45.2	5.7
<b>5A</b>	38.3	31.1	30.6	-
<b>5B</b>	15.3	24.8	59.9	-
<b>5C</b>	14.5	23.0	34.0	28.5
<b>5 ALL</b>	24.4	25.4	41.8	8.5
<b>6A</b>	38.4	26.3	35.3	-
<b>6B</b>	8.6	19.2	72.2	-
<b>6C</b>	8.9	14.2	37.6	39.3
<b>6 ALL</b>	19.6	19.0	49.6	11.8
<b>7A</b>	26.6	24.9	48.5	-
<b>7B</b>	26.3	30.0	37.7	-
<b>7C</b>	14.4	37.0	33.4	15.3
<b>7 ALL</b>	24.6	32.1	38.5	4.8
<b>8A</b>	22.1	25.6	52.3	-
<b>8B</b>	21.5	34.6	43.9	-
<b>8C</b>	12.4	32.4	31.5	23.7
<b>8 ALL</b>	21.5	30.2	41.3	7.0
<b>11A</b>	29.4	20.7	49.8	-
<b>11B</b>	19.2	32.2	48.6	-
<b>11C</b>	14.0	34.5	34.1	17.4
<b>11 ALL</b>	23.8	28.0	43.4	4.8

Table 24. *Percent of PASA Students at Each Performance Levels for the 2012 PASA Science*

<b>Level</b>	<b>Emerging</b>	<b>Novice</b>	<b>Proficient</b>	<b>Advanced</b>
<b>4A</b>	22.0	31.8	46.2	-
<b>4B</b>	12.3	40.8	46.9	-
<b>4C</b>	13.0	22.6	45.2	19.2
<b>4 ALL</b>	17.5	33.1	45.5	3.9
<b>8A</b>	23.2	25.6	51.2	-
<b>8B</b>	5.6	34.8	59.6	-
<b>8C</b>	9.5	29.6	44.0	16.9
<b>8 ALL</b>	14.2	29.7	51.1	5.0
<b>11 A</b>	57.6	21.4	20.9	-
<b>11B</b>	50.1	34.5	15.4	-
<b>11 C</b>	26.4	39.1	23.1	11.4
<b>11 ALL</b>	46.0	31.4	19.5	3.2

## Chapter 11: Validity Argument for the *PASA*

### Validity Studies for the Pennsylvania Alternate System of Assessment (*PASA*)

In order to address the validity argument that has been established for the *PASA*, a supplemental report has been developed to accompany this technical manual. The validity study report provides summaries of the validity studies that have been conducted for the *PASA Reading and Math* during the 2009-10, 2010-11, and 2011-12 school years as well as all of the validity studies that have been conducted related to the *PASA Science*, which was administered initially during the 2007-08 school year.

The validity studies present evidence that informs the validity argument being established for the *PASA*. The evidence-based argument links to the intended purposes of the test, which are:

- 1) to measure the degree to which students attain the knowledge and skills described in the statewide alternate eligible content, which is linked to the state academic content standards;
- 2) to provide information regarding district and school accountability; and
- 3) to improve curricular and instructional practice for students with significant cognitive disabilities and increase student learning.

The validity studies that have been conducted for the *PASA* are highlighted in the following tables. Brief summaries of several of these studies are presented in the Supplemental Validity Report and detailed summaries are presented in the Appendices to that report. Additional studies have been summarized in previous technical reports, such as those related to standard setting and PLDs.

Table 25. *Sources of Validity Evidence for the PASA Reading, Math, and Science*

Description	Evidence	Validity Concern	Findings	Uses/Changes
Pilot Proficiency Study	Item frequencies of proficient A students taking B items and proficient B students taking C items	Construct	There were some items from higher level test that students could do. Students performed less well on items from higher test level than on items at their assigned test level	Study provided preliminary information about student performance on higher level items, but only proficient students were assessed. Prompted a large scale study as a follow up.
Proficiency Study 2011-12	Item frequencies of proficient A students taking B items and proficient B students taking C items	Construct	There were several items from each higher level test form that students could do, including some items for which a large percentage of students (e.g., above 60%) assigned to a lower test level could earn a score of 5.	Results will be used in conjunction with results from the 2012-13 proficiency study to examine any needed restructuring of the <i>PASA</i> Additional item level analyses are currently being explored to further examine learning progressions.
Proficiency Study 2012-13	Item frequencies of students at each grade level on items within the same content strand across grades	Construct	Planning Stage, to be conducted in Spring 2013	Results will be used in conjunction with results from the 2011-12 proficiency study to examine any needed restructuring of the <i>PASA</i>
LCI Study	LCI data from teachers who administered <i>PASA</i> in 2010-11. Utilized subset of LCI responses that could be linked by student ID to <i>PASA</i> scores.	Consequential Construct	<i>PASA</i> students have similar learning characteristics to students taking AA-AAS across states. Some differences in learning characteristics of students taking <i>PASA</i> levels related to communication.	Methods for identifying students for the <i>PASA</i> will remain. Results suggest some differences in A, B and C level students. Information will be utilized with evidence from proficiency studies to inform any needed redesign of the <i>PASA</i> .

<b>Description</b>	<b>Evidence</b>	<b>Validity Concern</b>	<b>Findings</b>	<b>Uses/Changes</b>
Parent Score Report Study	Parent focus groups regarding score reports	Consequential	Parents would like to see specific information about tested content	Score report has been modified regularly based on feedback from TAC and earlier focus groups. Results from this study will inform future modifications.
Fidelity of Test Administration Study	Rescoring of 750 student assessments for fidelity of administration	General: Score interpretation and use	No systemic errors in administration were found.	Training of test administrators will be continued. No major changes to the process beyond the continuous improvement model that is implemented were deemed necessary.
Impact of Change to Scoring Rubric Study	Rescoring of 230 2008 student assessments using the revised scoring rubric from 2009	General: Score interpretation and use	Change in scoring rubric resulted in differences that would have caused some students to change performance level classifications. Many classifications remained.	The revised scoring rubric was maintained.
Score of 3 and 4 Study	Rescoring of items on that were assigned scores of 3 and 4 on a subset of student assessments	General: Score interpretation and use	The most common classifications for scores of 4 was because the teacher repeated directions, and for scores of 3 was because the teacher made the task easier.	The scoring rubric remained the same. Training for administrators and scorers is continuously updated and was updated to reflect common errors.

Table 26. *Additional Sources of Validity Evidence for PASA Reading*

<b>Description</b>	<b>Evidence</b>	<b>Validity Concern</b>	<b>Findings</b>	<b>Uses/Changes</b>
Reading Screening Study Appendix E	Supplemental reading test administered to students taking the Level B and C PASA to estimate reading level	Criterion	A wide range of reading ability was observed on the measures of word and passage reading. Children who take the AA-AAS are performing substantially below grade level in word and passage reading.	Results from this study were considered and will inform the amount and degree of difficulty of reading on the PASA.

Table 27. *Additional Sources of Validity Evidence for PASA Science*

Description	Evidence	Validity Concern	Findings	Uses/Changes
Internal Alignment Study	Panelist ratings	Content	Majority of assessment content linked to content standards;	Assessment items related to an alternate assessment anchor that was not found to be linked to the regular education assessment anchors were phased out of the <i>PASA Science</i> .
External Alignment Study	Panelist ratings	Content	Majority of assessment content linked to content standards. Some content at certain grade levels not linked.	Revision to items and item writing procedures

Table 28. *Sources of Evidence Related to Technical Quality of the Assessment Measuring Student Performance*

Description	Evidence	Validity Concern	Findings	Uses/Changes
Technical Quality	PLD development	Construct	Panelists understood and had confidence in the process of PLD development and independent observations supported the procedural fidelity of the process	Used to set cut scores
	Standard setting	Construct	Panelists understood and had confidence in their participation in the standard setting process, across grade and test levels	Periodic recalibration of cut scores to assure accurate and reliable differentiation of groups of students into performance categories
	PLD analysis	Construct	Preliminary evaluation of definition and placement of PLDs within performance categories-overall, students' performance on items matched that predicted by PLDs; Results continue to be tracked to inform whether item level data patterns should impact changes in PLDs	Continuous analysis that informs item development and will inform future standard settings and potentially impact PLDs
	Reliability of Scoring		Overall reliability of scoring from team to team and from team to standard consistently over 85% agreement	Assessment of scoring team performance and performance of individual test items

The tables presented summarize the main aspects of the validity argument for the PASA and show how the information has and will continue to be utilized. Initial findings have impacted subsequent steps of the validation process. For instance, alignment studies suggested that some changes were needed to the science content assessed at one grade level, which expanded the item development process for the assessment program. Development of PLDs and the subsequent and longitudinal analysis and comparison of actual test data to the underlying PLD framework continue to inform the content progressions that have been established for the assessment. Additionally, more expansive studies such as the proficiency study, where students are administered test items from test levels (B or C) that represent more cognitive demand, continue to inform the internal structure of the assessment, and results, when analyzed, may impact the test design.

Aspects related to the administration, scoring and score reporting, such as on-line trainings, proficiency tests, and focus groups of parents continue to improve the administration and scoring aspects of the assessment program. As new technologies have been made available for computer based training, the assessment program has utilized these technologies whenever possible.

With respect to consequential validity, a number of sequential studies were planned and carried out, including a survey research study, artifact study, planned focus group study for science, which resulted in the establishment of a Leadership Institute. This sequence of studies provided several pieces of information. First, the survey and artifact studies provided information about the impact of the assessment program on and the current state of classroom instruction and assessment practice. Follow up focus group studies suggested the need for guidance and information sharing regarding instruction and assessment of students with significant cognitive disabilities, which resulted in the Leadership Institute and corresponding conference, which allowed for the dissemination of content and instructional ideas for teaching sequences of lessons to students with significant cognitive disabilities.

Evidence related to various aspects of the validity argument continue to be planned, collected, and utilized to improve the assessment system. The following chapter highlights some of the studies that are currently being planned.

## Chapter 12: Future Activities

### On-Line Submission of Student Assessments

For the 2012-13 administration of the *PASA*, a pilot study will be conducted to examine the feasibility of online submission of student assessments. The on-line submission system will be piloted in approximately 32 districts and involve approximately 1000 student assessments. Enrollment for these districts will also take place on-line. Additionally, a pilot of scoring these assessments will take place in the Spring of 2013. This may lead to more regionalized scoring sites for future administrations of the *PASA*. PDE and *PASA* are working together to investigate the feasibility of the on-line submission of student assessments.

### On-Line Training of Scorers

On-line training modules for scorers and proficiency tests were developed and utilized for the 2012 administration and are found at <http://www.pasatest.com>. These modules will be updated and will continue to be used to train and select scorers for participation in future scoring conferences. For the 2012 administration, only scorers who passed the proficiency test were asked to participate, and this criteria is also planned for future administrations.

### On-Line Training of Test Administrators

On-line training modules for administering the *PASA* have been developed and are found at <http://www.pasatest.com>. These modules will continue to be used to train and select scorers for participation in future scoring conferences. A proficiency test was developed for the 2012 administration, and additional proficiency test will be developed for 2013. An attempt to increase the number of test administrators who participate in the training will be made. Plans for mandatory training of test administrators are being considered.

### Proficiency Study: Across Grade Learning Progressions

In order to further examine the learning progressions underlying the *PASA* Reading, Math and Science, a large scale study is planned for the 2012-13 school year. This validity study will investigate the performance of students assigned to a particulate test level taking items that appear on different grade level assessments. For instance, students in grades 3 or 4 who are assigned to the Level A assessment will be administered the 20 item Level A assessment as well as a set of supplemental items found on the grades 5/6, 7/8 and 11 Level A assessments. Students in grades 5 and 6 who are assigned to the Level A assessment will be administered the appropriate test form with 20 scored items, and a supplemental test form with items from the grades 3/4, 7/8 and 11 Level A assessments that track the theoretical learning progressions related to similar content. This design allows for investigation of the appropriateness of the grade level learning progressions that underlie the *PASA Reading, Math and Science*. Supplemental test booklets will be utilized, because they will contain between approximately 8

to 17 items. Due to the high number of ‘test’ items on several forms, the items will not be embedded into the *PASA*, but will be included as a supplemental test booklet.

This study involves the administration of a supplemental test form to each student taking the *PASA* Reading, Math and Science. The items on the supplemental forms will assess content strands that span across grade levels. For instance, all students taking the grade 3/4 Level A *PASA* Reading will be administered the appropriate assessment form as well as a supplemental assessment that contains approximately 10 additional items from the Level A assessments on the grades 5/6, 7/8 and 11 assessments. The supplemental assessment forms will be generated such that they track sets of theoretical learning progressions across the grade levels, and such that all learning progressions are assessed on at least one set of forms. Therefore, several supplemental tests will be designed at each test level. Subsets of students in Grades 3/4, 5/6, 7/8 and 11 will receive the same supplemental test booklets. This will allow theoretical learning progressions to be evaluated across all grade levels. In addition, different supplemental test booklets will be generated in order for the breadth of content assessed throughout the *PASA* to be represented. Results will be presented in the 2013 Technical Manual.

## **Enacted Curriculum Study**

### ***Impact Study Phase 1, PASA Math, Reading and Science***

The first phase of the enacted curriculum study involves 450 teachers who will be brought together in one of three locations across Pennsylvania to complete a survey related to the enacted curriculum. Participants will be selected from the subset of teachers who have a student enrolled for the 2012-13 *PASA* science. This criteria was selected in an effort to ensure that participants will be able to address questions about the enacted curriculum for the three content areas of Reading, Math and Science. Representation of teachers across the state and who teach students taking all levels of the *PASA* will be sought. Teachers who do not have a student enrolled for the 2012-13 *PASA Science* may not be able to adequately answer the questions related to science instruction, which serves as the rationale for focusing the sample selection on the subset of teachers who will be administering all three content areas of the *PASA*.

### ***Impact Study Phase 2, PASA Science***

Phase 2 of the enacted curriculum study will involve teachers who will administer the *PASA* Science during the 2012-13 school year and volunteer to participate in this phase of the study. This phase will involve classroom visits to view the teachers while they are instructing, and will provide interview data, teacher logs, classroom instructional and assessment artifacts, IEP reviews, classroom assessments and other data. This complete set of data is meant to provide information about current instructional and assessment practices of teachers across Pennsylvania.

## **APPENDICES**

- Appendix A: Enrollment forms and Example Skills Checklists**
- Appendix B: General skill areas identified for math and reading used in preliminary item development**
- Appendix C: Pennsylvania Alternate Academic Content Standards for Reading and Mathematics**
- Appendix D: Pennsylvania Alternate Anchors and Eligible content for Reading, Mathematics and Science**
- Appendix E: Detailed Grids for Test Development**
- Appendix F: Test Blueprints for the *PASA Reading, Math and Science***
- Appendix G: *PASA Administrator's Manual***
- Appendix H: Example of an Individual Student Score Report**
- Appendix I: Item to Item Correlations**
- Appendix J: Item Means and Standard Deviations**
- Appendix K: Item Frequency Distributions**