



**New Jersey
Alternate Proficiency Assessment (APA)
Science**

**2015
Technical Report**

February 2016

PTM # 1510.64

**Copyright © 2015 by New Jersey Department of Education
All rights reserved**

STATE BOARD OF EDUCATION

- MARK W. BIERDRONHunterdon
President
- JOSEPH FISICAROBurlington
Vice President
- ARCELIO APONTE.....Middlesex
- RONALD K. BUTCHERGloucester
- JACK FORNARO.....Warren
- EDITHE FULTONOcean
- ERNEST P. LEPOREHudson
- ANDREW J. MULVIHILLSussex
- J. PETER SIMONMorris
- DOROTHY S. STRICKLANDEssex

David C. Hesse, Commissioner
Secretary, State Board of Education

It is a policy of the New Jersey State Board of Education and the State Department of Education that no person, on the basis of race, creed, national origin, age, sex, handicap, or marital status, shall be subjected to discrimination in employment or be excluded from or denied benefits in any activity, program, or service for which the department has responsibility. The department will comply with all state and federal laws and regulations concerning nondiscrimination.

TABLE OF CONTENTS

| | |
|--|----|
| PART 1: INTRODUCTION | 1 |
| 1.1 Purpose of the Assessment..... | 1 |
| 1.2 Overview of the Assessment..... | 3 |
| Background | 3 |
| Portfolio Assessment | 3 |
| Uses of Assessment Results..... | 4 |
| 1.3 Organizational Support | 5 |
| PART 2: TEST DESIGN AND TEST DEVELOPMENT | 7 |
| 2.1 Design History | 7 |
| 2.2 Test Design | 10 |
| 2.3 Test Specifications | 12 |
| 2.4 Alignment | 13 |
| PART 3: TEST ADMINISTRATION AND TRAINING | 17 |
| 3.1 Participation in the Alternate Proficiency Assessment..... | 17 |
| Personnel Responsibilities | 18 |
| 3.2 Test Administration Procedures..... | 18 |
| 3.3 Pre-Administration Training | 20 |
| 3.4 Test Security Procedures..... | 20 |
| 3.5 Portfolio Construction..... | 21 |
| Developing an APA Portfolio Entry | 21 |
| Scoring a Piece of Evidence | 25 |
| Scoring for Accuracy | 25 |
| Scoring for Independence | 25 |
| PART 4: SCORING..... | 27 |
| 4.1 Scorer Selection | 30 |
| Security at the Scoring Site..... | 31 |
| 4.2 Rangefinding | 32 |
| 4.3 Scorer Training | 35 |
| 4.4 Scoring Procedures | 36 |
| 4.5 Quality Control of Scoring..... | 41 |
| 4.6 Task Examination | 42 |
| PART 5: RELIABILITY AND VALIDITY..... | 48 |
| 5.1 Reliability..... | 48 |
| Inter-rater Reliability | 48 |
| Decision Consistency..... | 50 |
| 5.2 Validity | 51 |
| Appropriateness of Content Definition | 51 |
| Adequacy of Content Representation | 53 |
| Convergent Validity..... | 55 |
| Consequential Validity..... | 56 |
| Other Validity Studies..... | 59 |
| PART 6: STANDARD SETTING..... | 60 |
| 6.1 Overview of the Process | 60 |
| 6.2 Procedures..... | 61 |

| | |
|---|-----|
| Performance Level Descriptors (PLDs)..... | 61 |
| Standard Setting Process..... | 63 |
| 6.3 Summary of Results..... | 66 |
| PART 7: REPORTING..... | 69 |
| 7.1 Interpreting Reports..... | 71 |
| Student Demographic Information..... | 71 |
| Score Information..... | 71 |
| Student Sticker and Individual Student Report (ISR)..... | 75 |
| All Subjects Roster..... | 79 |
| Student Roster..... | 81 |
| Summary of Performance – School, District..... | 83 |
| Performance by Demographic Groups..... | 85 |
| District Student Data..... | 87 |
| State Summary..... | 87 |
| 7.2 Parent Letter..... | 87 |
| 7.3 Quality Control of Reporting..... | 89 |
| Scanning and Scoring..... | 90 |
| Quality Control of Score Reporting..... | 90 |
| APPENDIX A: Development of the CPI Links..... | 92 |
| APPENDIX B: Performance Level Descriptors..... | 95 |
| APPENDIX C: Consistency between APA Portfolio Scorers by Individual Grade and Entry..... | 101 |
| APPENDIX D: 2015 Executive Summary..... | 104 |
| APPENDIX E: 2015 Frequency Tables of Proficiency Levels by Disability Category.. | 114 |
| References..... | 121 |

TABLES

| | |
|--|----|
| Table 1.1 2015 APA Number of Valid Scores and Percent of Students at Each Proficiency Level..... | 2 |
| Table 2.1 APA Proficiency Classification (2003–2007) | 8 |
| Table 2.2 Number of Valid Scores 2004–2005 through 2014–2015 Administrations for Science..... | 10 |
| Table 3.1 2014–2015 Calendar for the APA | 19 |
| Table 3.2 2014–2015 Training Modules..... | 20 |
| Table 3.3 Scoring of Items for Accuracy and Independence..... | 26 |
| Table 4.1 Total Number of Entries for the APA Science Portfolios | 27 |
| Table 4.2 Summary of the Scorers’ Characteristics..... | 31 |
| Table 4.3 Distribution of Condition Codes by Grade | 44 |
| Table 4.4 Distribution of Unscorable Error Codes and Scores – Science | 45 |
| Table 4.5 Distribution of Unscorable Error Codes by Grade and Entry – Science | 46 |
| Table 4.6 Distribution of Unscorable Error Codes by Subcode – Science | 47 |
| Table 5.1 Consistency between APA Portfolio Scorers | 49 |
| Table 5.2 Links for Academic Learning (LAL) Alignment Criteria | 55 |
| Table 5.3 2015 APA Combined Grade Proficiency Level Frequencies by Disability Category..... | 57 |
| Table 5.4 Combined Grade Proficiency Level Frequencies by School Type..... | 58 |
| Table 6.1 Demographic Background of Standard Setting Panelists | 64 |
| Table 6.2 Cut Scores After Rangefinding and Pinpointing Rounds | 67 |
| Table 6.3 Approved 2009 Cut Scores..... | 68 |
| Table 7.1 Distribution of the APA Reports | 70 |
| Table 7.2 2015 APA Dimension Scoring | 72 |

FIGURES

| | |
|--|----|
| Figure 1.1 Linkage | 2 |
| Figure 2.1 APA Structure | 11 |
| Figure 2.2 Test Specifications..... | 13 |
| Figure 3.1 Choosing a CPI Link for the APA..... | 23 |
| Figure 3.2 Administering and Scoring an Activity for the APA | 24 |
| Figure 4.1 APA Scoring Rubric..... | 28 |
| Figure 4.2 Unscorable Codes..... | 29 |
| Figure 6.1 Graph for Reasoned Judgment Warm-Up Task | 66 |
| Figure 7.1 Sample Student Stickers | 75 |
| Figure 7.2 Sample Individual Student Report (Grade 12 Front) | 77 |
| Figure 7.3 Sample Individual Student Report (Back)..... | 78 |
| Figure 7.4 Sample All Subjects Roster..... | 80 |
| Figure 7.5 Sample Student Roster | 82 |
| Figure 7.6 Sample Summary of District Performance..... | 84 |
| Figure 7.7 Sample District Performance by Demographic Groups | 86 |
| Figure 7.8 Sample Parent/Guardian Letter | 88 |

PART 1: INTRODUCTION

The purpose of this technical report is to provide information about the New Jersey Science Alternate Proficiency Assessment (APA) administered in 2014–2015. This report is intended for use by those who evaluate tests, interpret scores, or use test results for making educational decisions. It consists of the following sections: test design and test development, test administration and training, scoring, reliability and validity, standard setting, and reporting. It includes references to additional reports, documents, and websites related to the APA.

The 2015 Science APA was assessed in grades 4 and 8 and in grade 9, 10, 11, or 12 depending on the grade in which a student received Biology instruction. A total of 3,957 students were evaluated by the 2015 Science APA. Of these, 3,656 students had valid Science scores. Table 1.1 presents the overall performance of students on the 2015 Science APA. The table shows the number of valid scores and the percent of students at each proficiency level for students assessed.

1.1 Purpose of the Assessment

The APA was developed for two purposes:

- To measure the progress of a small percentage of students with the most significant cognitive disabilities who cannot participate in the regular statewide assessments even with accommodations
- To ensure that the educational results for all students are included in the statewide accountability system at the individual, school, district, and state levels

Accountability through assessment provides equity in program and educational opportunities for all students. Alternate assessment ensures an inclusive statewide assessment system and student accountability linked to the common core of learning within the general curriculum in New Jersey.

The New Jersey APA represents a cohesive approach where curriculum, instruction, and assessment work together to build a comprehensive educational program. Curriculum drives instruction and assessment. Assessment and instruction inform the curriculum as well as each other.

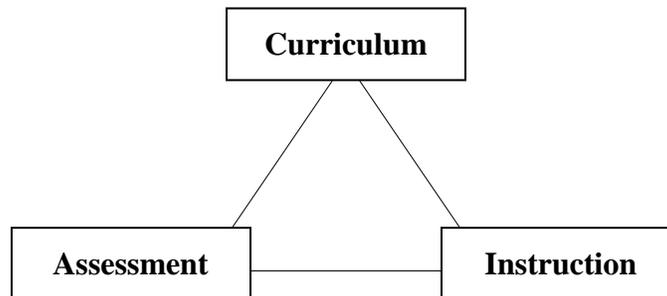
Table 1.1 2015 APA Number of Valid Scores and Percent of Students at Each Proficiency Level

| Grade | Number of Portfolios Processed | Science | | | |
|------------|--------------------------------|------------------------|---------------|---------|--------------|
| | | Number of Valid Scores | % Part. Prof. | % Prof. | % Adv. Prof. |
| 4 | 1,299 | 1,193 | 47.9 | 51.7 | 0.4 |
| 8 | 1,383 | 1,273 | 47.8 | 41.1 | 11.1 |
| 9* | 188 | 176 | 65.3 | 31.8 | 2.8 |
| 10* | 282 | 260 | 56.5 | 39.6 | 3.8 |
| 11* | 691 | 668 | 49.1 | 48.8 | 2.1 |
| 12* | 114 | 86 | 53.5 | 44.2 | 2.3 |
| All Grades | 3,957 | 3,656 | 49.7 | 45.5 | 4.8 |

*In 2014–2015, the APA assessed Science in grade 9, 10, 11, or 12 depending on the grade in which a student received Biology instruction.

The triangle in Figure 1.1 highlights the relationship between curriculum, instruction, and assessment.

Figure 1.1 Linkage



High-quality assessment practices provide information upon which to base ongoing development of curriculum that is responsive to individual student needs. Aside from the use of a portfolio to capture student learning, this philosophy considers students with significant cognitive disabilities as valued and contributing members of their schools and communities. This performance-based assessment is designed to measure achievement of knowledge and skills that will prepare students for positive post-school outcomes in education, employment, and independent living.

1.2 Overview of the Assessment

Background

The NJ APA process was developed in response to the *Individuals with Disabilities Education Act of 1997 (IDEA '97)*, which required that states develop and conduct alternate assessments beginning no later than July 1, 2000. With the reauthorization of *IDEA '97* as the *Individuals with Disabilities Education Improvement Act of 2004 (IDEA '04)*, requirements for alternate assessments remain as follows:

ALTERNATE ASSESSMENTS—

- (i) **IN GENERAL**—The State (or, in the case of a district-wide assessment, the local educational agency) has developed and implemented guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in regular assessments under subparagraph (A) with accommodations as indicated in their respective individualized education programs.
- (ii) **REQUIREMENTS FOR ALTERNATE ASSESSMENTS**—The guidelines under clause (i) shall provide for alternate assessments that—
 - (I) are aligned with the State’s challenging academic content standards and challenging student academic achievement standards; and
 - (II) if the State has adopted alternate academic achievement standards permitted under the regulations promulgated to carry out section 1111(b)(1) of the *Elementary and Secondary Education Act of 1965*, measure the achievement of children with disabilities against those standards.
- (iii) **CONDUCT OF ALTERNATE ASSESSMENTS**—the State conducts the alternate assessments described in this subparagraph. (Sec. 612 (a) (16) (C))

In addition, the *No Child Left Behind Act of 2001 (NCLB)* requires that all students, including those with disabilities, participate in the state assessment program. NCLB also requires that the measurement of progress toward meeting state standards include assessment results for all students.

The Science APA fulfills the NCLB requirements and is based on the New Jersey Core Curriculum Content Standards (NJ CCCS) in Science. In this manner, all students in New Jersey are moving toward the same general standards with whatever modifications or supports they need. Including students with disabilities in the assessment and accountability system is critical to ensure appropriate allocation of resources and learning opportunities for these students. The alternate assessment based on alternate achievement standards was designed for a very small percentage of the total school population for whom traditional assessments, even with accommodations, would be inappropriate measures of their progress.

Portfolio Assessment

The Science APA is a portfolio assessment designed to measure progress toward achieving New Jersey’s state educational standards for students with the most significant cognitive disabilities who are unable to participate in the general assessments: the New

Jersey Assessment of Skills and Knowledge in grades 3–8 (NJ ASK), the Partnership for Assessment of Readiness for College and Career (PARCC), and the New Jersey Biology Competency Test (NJBCT).

A portfolio is a collection of student work samples, student demographic data, and instructional information that relates to a student’s progress on the NJ CCCS, strands, grade-level cumulative progress indicators (CPIs), and skill statements called CPI Links. Evidence of student performance as demonstrated in the student portfolio was collected twice during instructional activities over the school year. To score the portfolios, trained expert scorers used a scoring rubric designed to measure student performance on the skill, the level of independence when performing the skill, and the relationship of the skill to the grade-level cumulative progress indicator.

Uses of Assessment Results

The Science APA measures the student’s achievement of the NJ CCCS. APA results should not be used as the sole basis for instructional decisions.

The three proficiency levels assessed for Science are:

- **Advanced Proficient:** exceeds the state level of proficiency
- **Proficient:** meets the state level of proficiency
- **Partially Proficient:** is below the state minimum level of proficiency.

The proficiency level classification allows the APA results to be combined with the results from general assessment for accountability purposes for state and federal reports. For accountability purposes, the APA is both a student assessment and a school/district program assessment.

It is important to recognize that the APA system does not report scale scores. The data provided are the key components when interpreting the portfolio results. The APA scores are based solely on the information provided in the portfolio submitted; therefore, it is inappropriate to compare these scores to other APA students and students taking the general assessments. Scale scores are not appropriate for use for the APA system as there are no issues of equating involved. Because there are no sets of test items, there are no item difficulties, nor is there a need to equate test scores from year to year.

For additional information about the 2014–2015 APA, the standards on which the APA is based, or information regarding the participation of students with disabilities in the statewide assessment system, see these documents published by the New Jersey Department of Education (NJDOE):

New Jersey Alternate Proficiency Assessment Procedures Manual – Science at https://nj-servicepoint.questarai.com/NJxx01_Documentation_2014_2015

New Jersey Core Curriculum Content Standards at <http://www.nj.gov/njded/cccs>

1.3 Organizational Support

New Jersey Department of Education (NJDOE). The APA is administered by the Office of Assessments (OS) within the NJDOE. The NJDOE coordinates the development and implementation of New Jersey’s statewide assessment program, which is designed to measure student attainment of the NJ CCCS. The OS works collaboratively within the department and with school districts to collect and report information about student academic achievement in order to inform instruction, increase student learning, and help parents and the public assess the effectiveness of their schools.

The staff of the NJDOE plans, schedules, and directs all APA activities. They are extensively involved in the APA development, training, document review, assessment security and authenticity, and quality-control procedures.

Questar Assessment, Inc. The prime contract for developing, administering, and scoring the APA was awarded to Questar Assessment, Inc. (Questar) in August 2012. Beginning in the 2014–2015 school year, the scope of work was changed to Science only. In partnership with Inclusive Large Scale Standards and Assessment (ILSSA) and Keystone Assessment, Questar presents extensive administrator training materials, sample activities, forms templates, planning tools, instructional materials, and resources for APA educators at https://nj-servicepoint.questarai.com/NJxx01_Documentation.aspx.

Documentation for 2014–2015 can be found at https://nj-servicepoint.questarai.com/NJxx01_Documentation_2014_2015.

Major Questar activities include the following:

- Creating and monitoring the schedule for the APA administration, all tasks, subtasks, and activities to be conducted;
- Developing all APA reports, programs, committee communications, training materials, etc., in consultation with NJDOE staff;
- Designing, constructing, proofing, and printing assessment materials, forms, and documents;
- Packaging, distributing, and retrieving all assessment documents;
- Processing and scoring the student portfolios;
- Providing electronic data management and documentation;
- Establishing and implementing psychometric reporting.

Inclusive Large Scale Standards and Assessment (ILSSA) and Keystone Assessment assist the NJDOE and Questar with content development, planning, and execution, including training and scoring support for the APA. Both entities consist of a group of educators dedicated to improving educational opportunities for all students, especially those with significant cognitive disabilities.

Since 2001, ILSSA has worked with the NJDOE to implement the APA. During their years of partnership with the NJDOE, ILSSA has provided technical assistance and professional development on a range of topics, from all aspects of implementation of the

APA, to research-based practices and access to the general curriculum. Beginning in the summer of 2007, ILSSA worked closely with the NJDOE on revisions of the APA through the development of an upfront alignment design, redesign of the scoring rubric, standard setting, and increasing the standardization of the assessment items. They also worked closely with New Jersey educators to provide training and support for teachers with examples of standards-based instruction for better meeting requirements of the revised portfolio assessment.

ILSSA was formed in August 1998 in response to states' and school districts' need to respond to the assessment and other requirements of *IDEA '97* and the *Elementary and Secondary Education Act*.

New Jersey APA Educators. Due to the nature of the APA, educators are more extensively involved with the APA administration than the other NJ statewide assessments. For that reason, the NJDOE developed the APA with the important assistance of several APA educator committees. The committees included representatives of various groups that are knowledgeable about educating students with significant cognitive disabilities and have an interest in alternate assessment. The committees consisted of panels of special education teachers, child study team members, general education teachers, and administrators. Participants were chosen because of their qualifications and educational expertise. Selection criteria included number of years teaching, student population served, district factor group (DFG), type of educational facility, and regional location. Special care was taken to ensure gender and racial/ethnic representation on the committees. Committee meetings supporting the 2014–2015 APA were as follows:

- APA Curriculum (Created Sample Items) Committee:
 - July 14–17, 2014
 - July 21–25, 2014
 - August 25–28, 2014
- APA Rangelinding Committee: March 17–20, 2015

PART 2: TEST DESIGN AND TEST DEVELOPMENT

2.1 Design History

The NJ APA was first administered during the 2001–2002 school year in two content areas: Language Arts Literacy and Mathematics at grades 4, 8, and 11. During the 2004–2005 school year, the APA was expanded to include Science in grades 4, 8, and 11 and the assessment of Language Arts Literacy and Mathematics in grade 3.

Since the 2006–2007 administration, Language Arts Literacy and Mathematics have been assessed in grades 3–8 and 11, and Science has been assessed in grades 4, 8, and 11. With the implementation of the High School End of Course Biology Exam in 2009, however, Science expanded to grades 9 and 10 depending on when a student was enrolled in Biology. In 2010, eligible students who were not assessed in Language Arts Literacy, Mathematics, or Science in grade 11 were required to assess in grade 12 (including students who did not take a Biology course until grade 12). Starting with the 2011 administration, the High School End of Course Biology Exam has been renamed to the New Jersey Biology Competency Test.

Since 2002–2003, APA student performance results have been combined with the results of the general assessment for state and federal accountability reporting. The APA proficiency levels were designed to parallel the general education assessment. Up through 2007, portfolios were scored based on six dimensions: student progress, connection to standards, social interaction, independence, self-determination, and generalization. For each content area, student performance was classified into one of three proficiency levels based on progress and program:

- Advanced Proficient
- Proficient
- Partially Proficient

A student's progress score for each content area was classified into one of three levels:

- Substantial Progress
- Considerable Progress
- Minimal Progress

A student's program score was also classified into one of three levels:

- Commendable
- Satisfactory
- Needs Improvement

The program score was derived by adding the scores of the remaining five dimensions: Connection to Standards, Social Interaction, Independence, Self-Determination, and Generalization. A holistic sorting method was used to determine the cut scores for the three program levels.

The student progress level and the program level were combined to derive the three proficiency levels. At the recommendation of the APA Advisory Committee, the performance classification weights the program level more than the student progress level due to the use of state assessment results for school and district accountability. Table 2.1 prescribes how the proficiency was classified.

Table 2.1 APA Proficiency Classification (2003–2007)

| Proficiency Levels | | Student Progress Levels | | |
|--------------------|-------------------|-------------------------|----------------------|----------------------|
| | | Substantial | Considerable | Minimal |
| Program Levels | Commendable | Advanced Proficient | Advanced Proficient | Proficient |
| | Satisfactory | Proficient | Proficient | Proficient |
| | Needs Improvement | Proficient | Partially Proficient | Partially Proficient |

A standard setting was conducted in January and February 2003 in order to determine the cut scores for the program level. These cut scores were applied to all grade levels for both Mathematics and Language Arts Literacy. When Science was added to the APA in the 2004–2005 administration, the same program-level cut scores were applied.

For the 2006–2007 administration, in preparation for the transition to a new test design, the weight of program score determined by the Social Interaction, Independence, and Generalization dimensions was reduced by half. The scoring rubrics were revised to reflect the changes.

The APA underwent significant changes between 2007–2008 and 2008–2009, including changes to the test specifications, assessable content, and scoring dimensions. Prior to the 2007–2008 administration, peer reviewers from the U.S. Department of Education (USED) provided the NJDOE test design and administration recommendations for the new version of the APA (administered in 2008–2009). These recommendations included the following:

- APA students must be assessed on a subset of skills from the general assessment. The skills must be mapped to the general assessment specifications and address the breadth and depth of skills tested across grade levels.
- The skills assessed must link to the CPIs of the student’s assigned grade level.
- Students in the same grade must be assessed on the same content; teachers choose from a limited selection of standards and strands to assess their students.
- Strengthen the alignment of the APA program design to grade-level academic content and progress indicators.

In light of these recommendations, 2007–2008 was an interim year of change prior to full implementation of the new APA test design in 2008–2009. Based on the USED peer review, skills assessed on the APA were required to be academic in nature and linked to a

grade-level CPI. Therefore in 2008, for the purpose of Adequate Yearly Progress reporting, only the dimensions of Student Progress and Connection to Standards were assessed. The dimensions of Social Interaction, Independence, Self-Determination, and Generalization assessed in previous years were not evaluated in 2008. In addition, the connection to standards score replaced the previous program dimension score. An interim standard setting was conducted in April 2008. The interim standard setting was to ease the further transition of additional changes for the redesigned APA.

The 2008 APA proficiency level for each content area was based on the total score, calculated as the sum of the Connection to Standards and Student Progress scores. These two score dimensions are described below:

- **Student Progress** – to evaluate student progress toward achieving the targeted skills related to the NJ CCCS
- **Connection to Standards** – to determine the extent to which the portfolio content is linked to the NJ CCCS

Each content area assessed received a proficiency classification – Advanced Proficient, Proficient, or Partially Proficient – which allowed the APA results to be combined with New Jersey’s general assessment results for accountability purposes as required by USED.

In 2008–2009, the fully redesigned APA became operational. As a result, new performance level descriptors (PLDs) and a new standard setting were required. The new design, described in Section 2.2, was scored on the three dimensions: Complexity, Independence, and Performance, which are combined to determine a total score. A new standard setting was held and the cut scores that resulted were used for reporting in 2009 and onward. Longitudinal analyses and comparisons across or including the 2008–2009 assessment year are not recommended, nor are they likely to be interpretable.

- The **Complexity** Dimension is used to evaluate the CPI Link assessed and how closely the complexity and difficulty (Matched, Near, Far) links to the NJ CCCS and grade-level cumulative progress indicators (CPIs).
- The **Independence** Dimension is used to evaluate the extent to which the student completed the assessment items independently.
- The **Performance** Dimension is used to evaluate the student’s accuracy when performing skills represented in the CPI Links.

Table 2.2 shows the number of portfolios with valid scores for Science by grade level for the APA test administrations from 2004–2005 through 2014–2015.

Table 2.2 Number of Valid Scores 2004–2005 through 2014–2015 Administrations for Science

| Grade | 2004–2005 | 2005–2006 | 2006–2007 | 2007–2008 | 2008–2009 | 2009–2010 | 2010–2011 |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 4 | 710 | 794 | 894 | 958 | 1,009 | 1,140 | 1,278 |
| 8 | 723 | 871 | 989 | 892 | 1,011 | 1,069 | 1,054 |
| 9* | --- | --- | --- | --- | 55 | 130 | 95 |
| 10* | --- | --- | --- | --- | 109 | 210 | 170 |
| 11* | 554 | 596 | 885 | 66 | 503 | 756 | 711 |
| 12* | --- | --- | --- | --- | --- | 83 | 129 |
| All Grades | 1,987 | 2,261 | 2,768 | 1,916 | 2,687 | 3,388 | 3,437 |

| Grade | 2011–2012 | 2012–2013 | 2013–2014 | 2014–2015 |
|------------|-----------|-----------|-----------|-----------|
| 4 | 1,299 | 1,386 | 1,310 | 1,193 |
| 8 | 1,127 | 1,205 | 1,280 | 1,273 |
| 9* | 103 | 166 | 137 | 176 |
| 10* | 222 | 247 | 230 | 260 |
| 11* | 704 | 751 | 726 | 668 |
| 12* | 105 | 96 | 106 | 86 |
| All Grades | 3,560 | 3,851 | 3,789 | 3,656 |

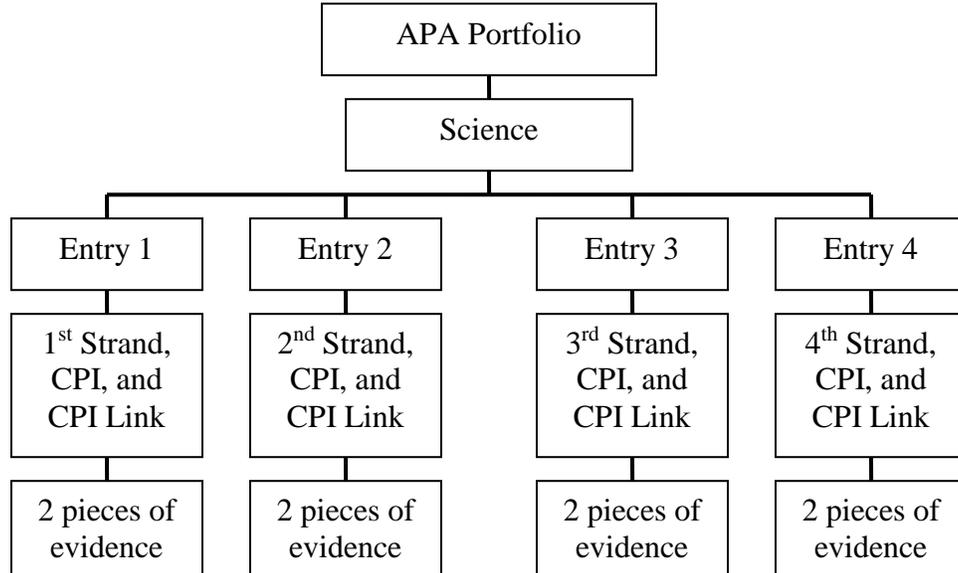
*From 2009–2015, the APA assessed Science in grade 9, 10, 11, or 12 depending on the grade in which a student received Biology instruction.

2.2 Test Design

The design of the Science APA consists of four strands from the NJ CCCS. For each strand, a CPI from the NJ CCCS and an associated CPI Link must be identified for measurement. The CPI Links and their associated CPIs and strands for 2014–2015 are available online at [https://nj-servicepoint.questarai.com/NJxx01_Documentation 2014 2015](https://nj-servicepoint.questarai.com/NJxx01_Documentation_2014_2015). To assess student mastery of the CPI Link, the teacher uses data collected from classroom learning and assessment activities.

The student’s ability to complete the tasks in the activities is measured once early in the assessment window, providing the first piece of evidence. The student is then measured later in the assessment window on the same targeted skill to see the extent to which their performance has improved, providing the second piece of evidence. A graphic representing the structure of the Science APA is presented in Figure 2.1.

Figure 2.1 APA Structure



- Each entry is scored on 3 dimensions: **Performance**, **Complexity** and **Independence** by a minimum of two scorers.
- **Performance** is worth twice as many points as Complexity or Independence.
- **Performance** is the largest contributor to total score.
- Total Score = Entry 1 + Entry 2 + Entry 3 + Entry 4
- An Entry = (**Performance**_{scorer1} + **Performance**_{scorer2}) + **Complexity**_{average} + **Independence**_{average}

Each entry in a student’s portfolio is scored on the three dimensions defined previously: Complexity, Independence, and Performance. These dimensions are evaluated using the two pieces of evidence submitted for each entry. One piece of representative evidence is collected early in the year as a baseline score; another piece of representative evidence is collected near the end of the year. The difference in student performance exemplified on the two is a measure of the student’s performance. Scores are combined across entries to determine the student’s proficiency level. This scoring is described in greater detail in Part 4.

2.3 Test Specifications

The APA has test specifications by grade and content area that prescribe the standards and strands that must be assessed. Test specifications were written in order to provide more guidance on how to link to grade-level CPIs and to address the federal requirement of linkage to the skills tested in the general assessments. Specifying the requirements increases standardization of the assessment for students with significant cognitive disabilities. Students may not be assessed in functional, behavioral, or access (social, motor, etc.) skills. Functional activities and materials might be used to promote understanding during instruction, but the evidence and activities demonstrating student achievement for assessment must be academically focused and represent the entire grade-level CPI Link.

The Science APA portfolio in grades 4, 8, and high school must have four entries. The test specifications below identify the standards, strands, and CPIs that must be assessed.

- Four entries based on different Science standards from the NJ CCCS
 - Grade 4
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.5 (Characteristics of Life)
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.6 (Physical Science – Chemistry)
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.8 (Earth Science)
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.9 (Astronomy and Space Science)
 - Grade 8
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.5 (Characteristics of Life)
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.6 (Physical Science – Chemistry)
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.7 (Physical Science – Physics)
 - One entry based on a specified strand, CPI, and CPI Link from Standard 5.9 (Astronomy and Space Science)
 - High School (Grade 9, 10, 11, or 12)

- Two entries based on two different strands, CPIs, and CPI Links from standard 5.5 (Characteristics of Life)
- Two entries based on two different strands, CPIs, and CPI Links from standard 5.10 (Environmental Studies)

Figure 2.2 illustrates the required components for each APA Science portfolio, including the standards, strands, and CPIs that must be assessed by the APA.

Figure 2.2 Test Specifications

| 2014–2015 NJ APA Science Test Specifications | | Entry | | |
|--|---|-------|----|----|
| Standard (NJ CCCS) | Strand | G4 | G8 | HS |
| 5.5 Characteristics of Life | A: Matter, Energy, and Organization in Living Systems | X | | X |
| | B: Diversity and Biological Evolution | | X | X |
| 5.6 Chemistry | A: Structure and Properties of Matter | X | | |
| | B: Chemical Reactions | | X | |
| 5.7 Physics | B: Energy Transformations | | X | |
| 5.8 Earth Science | B: Atmosphere and Water | X | | |
| 5.9 Astronomy and Space Science | A: Earth, Moon, and Sun System | X | | |
| | B: Solar System | | X | |
| 5.10 Environmental Studies | A: Natural Systems and Interactions | | | X |
| | B: Human Interactions and Impact | | | X |

2.4 Alignment

Federal peer review guidance indicates that a state’s academic achievement standards must be aligned with the state’s academic content standards and capture the full range and depth of knowledge and skills defined in the state’s academic content standards (USED, 2007). For the APA this was achieved by the development of grade-level specific PLDs and proficiency levels that cover the full range of knowledge and skills articulated in the CPI Links. The process for developing the descriptors and setting the proficiency levels is fully described in Section 6. This section details the development of the CPI Links and their alignment to the state’s content standards.

Prior to the development of essence statements and CPI Links, a subset of the NJ CCCS was prioritized for measurement on the APA. In 2007, the NJDOE worked with ILSSA and New Jersey educators to identify appropriate standards and associated CPIs for the APA population. The identified standards and CPIs differed across grades to ensure the

broadest coverage of the NJ CCCS. Subsequently, the essence associated with each identified CPI from the NJ CCCS was established by a committee of New Jersey educators, facilitated by ILSSA. A flowchart explaining this process is attached as Appendix A.

The CPI Links are skill statements that directly link to the critical essence of CPIs from the NJ CCCS. Providing these statements removes the need for educators to determine an appropriate instructional link to the CPIs as the CPI Links have already been vetted using criteria developed in New Jersey based on the peer-reviewed work of special education researchers and the National Alternate Assessment Center (NAAC). The criteria used as guiding principles for test development and alignment processes are excerpted below from page 24 of the *2014–2015 NJ APA Procedures Manual – Science*.

| |
|--|
| <p>Table 1: <i>Criteria for Instruction and Assessment that Links to Grade Level Content</i></p> |
| <ol style="list-style-type: none"> 1. The content is academic and includes the major domains/strands of the content area as reflected in state and national standards (e.g., science). 2. The content is referenced to the student’s assigned grade level. 3. The achievement expectation is linked to the grade-level content but differs in depth or complexity; it is not grade level achievement. 4. There is some differentiation in achievement across grade levels or grade bands. 5. The focus of achievement promotes access to the activities, materials and settings typical of the grade level but with the accommodations, adaptations, and supports needed for individualization. 6. The focus of achievement maintains fidelity with the content of the original grade-level Standards (content centrality) and, when possible, the specified performance (category of knowledge). 7. Multiple levels of access to the general curriculum are planned so that students with different levels of symbolic communication can demonstrate learning. |
| <p>Adapted from Browder, D.M., Wakeman, S.Y., Flowers, C.P., Rickelman, R.J., & Pugalee, D. “Creating access to the general curriculum with links to grade-level content for students with significant cognitive disabilities: An explication of the concept.” <i>Journal of Special Education</i>. v41 n1 p2–16 Spr 2007.</p> |

As a result of the development of the essences and the CPI Links, educators no longer need to develop appropriate targeted skills and criteria, resulting in increased standardization in the academic content to which APA students are exposed, and in the expectations of performance on that academic content.

Each Link is presented at three different levels of complexity to provide examples of how the essence of grade-level content can be taught to students with the most significant cognitive disabilities who have varied levels of communication and skills. The three levels of connection to each CPI are as follows:

- Matched Link
- Near Link
- Far Link

Each CPI Link maintains fidelity with the grade-level CPI (content centrality), but the complexity and difficulty varies from Matched to Far Link (performance centrality).

Complexity is the expectation level at which the student should perform the skill (remembering, understanding, applying, analyzing, evaluating, and creating). **Difficulty** involves the number of concepts, skills, or ideas on which the student will be working or the type of adaptations and supports in place. Difficulty can be changed by reducing the number of nouns addressed within the CPI, limiting the amount a student has to do, or by using adaptations such as adapted text or a limited number of items.

All CPI Links are aligned with grade-level CPIs; however, they differ in the level of complexity and difficulty at which the student is expected to perform. Matched Links have more complexity and difficulty than the Far Links.

The different levels of the CPI Links do not correspond to a particular communication system, learning style, or disability category of a student. Students may be using a Matched Link in one entry and a Far Link in another.

Matched Link: Contains skill statements that are approximately the *same complexity* level of the CPI expectation but the *level of difficulty is lessened*.

- For instance, if the CPI complexity level is “understanding” then a matched link usually requires the student demonstrate understanding. However, if the CPI expectation is that the student understands similes, metaphors, personification, and alliteration, the matched link *might* only require a few of those concepts, thus modifying the difficulty level.
- Difficulty may also be lessened by providing an adapted text, fewer problems, or other supports.

Near Link: *May be the same or lower complexity* as the CPI expectation but the *difficulty level has been lessened even more*.

- Near Links were developed in two different ways. If the complexity level for the CPI is at the “understanding” level, then the near link *may be* “understanding” but the difficulty level has been modified to include fewer concepts and additional supports.
- Or, a near link may have been developed by modifying the complexity level so that instead of “understanding,” the student is required to demonstrate “remembering.”

Far Link: Contains skill statements that are a *lower complexity* level and *difficulty is lessened even more*.

- For instance, if the CPI expectation is at the “understanding” level, the student is only expected to perform at the “remembering” level.
- Also, the difficulty level has been lessened so that the student is only identifying part of the concept/skill required in the CPI and has additional supports.

Example of a CPI Link

| | | | |
|----------|--|--|---|
| CPI | CPI 5.5.12A2 Explain how plants convert light energy to chemical energy | | |
| | Essence of the CPI: Understand the process of photosynthesis | | Essence of the CPI |
| | Matched Link | Near Link | Far Link |
| | <i>Complexity is the same Difficulty is lessened</i> | <i>Complexity is the same Difficulty is lessened even more</i> OR <i>Complexity is lessened Difficulty is lessened</i> | <i>Complexity is lessened Difficulty is lessened even more</i> |
| CPI Link | <ul style="list-style-type: none"> ◆ Demonstrate and explain the roles different parts of a plant cell have in converting light energy into chemical energy <i>5.1.12A1</i> ◆ Explain the transfer of light energy to chemical energy* | <ul style="list-style-type: none"> ◆ Describe how the products of photosynthesis and cellular respiration are recycled* ◆ Differentiate the role cell parts have in respiration and photosynthesis | <ul style="list-style-type: none"> ◆ Identify the structure and function of plant cells* ◆ Identify the components of photosynthesis* |

PART 3: TEST ADMINISTRATION AND TRAINING

3.1 Participation in the Alternate Proficiency Assessment

All students with disabilities must participate in the state assessment system. Students with disabilities participate in the general assessment for their grade or in the APA. The Individualized Education Program (IEP) team makes decisions about state assessment participation. Decisions regarding participation in the APA must be documented in the student's IEP. A sample of the IEP form with guidance about how to document decisions is shown at http://www.state.nj.us/education/specialed/form/iep/ieptoc_sp.shtml. The IEP team must determine for each content area assessed whether an individual student will participate in the general assessment or the APA. A student may participate in the APA in a content area only if the IEP team determines that the student has not been instructed in the knowledge and skills tested by the assessment and if the student is unable to correctly complete any of the tasks on the general assessment, even with accommodations and modifications [N.J.A.C. 6A: 14-4.10].

Students with disabilities participate in the state assessments during the same grades as their nondisabled peers. Therefore, students with disabilities in grades 3–8 and high school (9, 10, 11, or 12) must participate in the statewide assessment system, regardless of educational placement. The student's assigned grade level determines when a student participates in state assessments. This includes students with disabilities attending the following:

- Local district public schools;
- Local district public schools in another part of town;
- Public schools in other towns;
- Receiving schools, including county special services school district, public educational service commissions, approved private schools for the disabled, college-operated programs, Marie H. Katzenbach School for the Deaf, jointure commissions, and regional day schools;
- Private schools in accordance with a Naples placement;
- Private schools for the disabled out of state (placed there by a New Jersey district or authorized state agency); and
- State educational facilities.

Students on homebound instruction were also required to participate in state assessments.

Guidelines for grade 12 students are as follows:

- If a senior was new to the state and had not participated in either the APA or the PARCC, the IEP team determined which assessment was appropriate and the student participated in that assessment.

Students who were juniors the previous year and should have participated in the APA but did not must participate in the APA.

Students with disabilities who participated in one or more content areas of the PARCC, regardless of whether or not they were required to pass the PARCC in order to meet graduation requirements, were not eligible to participate in the APA in that (those) content area(s).

The document “Guidelines to Determine Which Students Should Participate in the New Jersey Statewide Science Assessment Through the Alternate Proficiency Assessment” appears in the *2014–2015 APA Procedures Manual* beginning on page 11. Also included is a chart that provides the individual determinations that must be made to evaluate student eligibility for participation in the APA.

Personnel Responsibilities

Identifying a student who requires the APA as the state assessment of record requires the input of many personnel. The district’s director of special education, the child study team members, and other educators may be involved in this decision, although the IEP team makes the final decision about participation in the APA. Unlike the general state assessments, the development and submission of the APA assessment also requires the participation of many administrators and educators.

The school administrator, director of special education, and the APA coordinator are responsible for ensuring that the APA is correctly developed for the appropriate students during the prescribed data collection period. These administrators are responsible for the accurate development and timely submission of these assessments. The dissemination of information to the APA student’s educators, oversight of the APA process, and the review of the portfolio are all administrators’ responsibilities. It is also the direct responsibility of the administrators to ensure that these assessments are submitted on time for scoring and that the student demographic information coding on both the general assessment test book/answer folder and the APA Student Demographic Information Form (SDIF) is accurate and complete.

All educators of students who participate in the APA are responsible for reviewing the *APA Procedures Manual* and related training materials, and following all procedures when collecting educational information that will be submitted in a portfolio. All educators should review the scoring guidelines and plan how to include student work in the portfolio that meets these guidelines. In most cases, the evidence contained in the portfolio is submitted by several teachers, though the student’s lead teacher does the coordination of the development and submission of the APA to the coordinator.

3.2 Test Administration Procedures

For each school and district with any student assessed with the APA, the NJDOE required that an administrator (special education director, principal, director of curriculum, child study team members, etc.) be assigned to the role of test coordinator. These individuals were responsible for ensuring that all APA tasks were completed, including the dissemination of information, the completion of all portfolios, the review of the completed portfolios for accuracy and authenticity, and adherence to all APA

deadlines. Table 3.1 displays the calendar shown on the inside front cover of the *2014–2015 APA Procedures Manual – Science*.

Table 3.1 2014–2015 Calendar for the APA

| Event | Date |
|--|---|
| Administrator Training | September 15–19, 2014 |
| Training for APA Teachers | On-line Training – Dates determined by the district beginning September 2, 2014 |
| First Collection Period | September 2, 2014 – November 14, 2014 |
| Second Collection Period | December 8, 2014 – February 13, 2015 |
| Portfolio Completion Date | February 13, 2015 |
| Administrator Review of Portfolio | February 13–24, 2015 |
| Portfolio Collection Materials Sent to Districts/Schools | February 2015 |
| Portfolios Returned to Contractor | February 23 – March 13, 2015 Portfolios shipped after March 13 WILL NOT be accepted for scoring |
| Student Demographic Record Changes | March 23 – April 10, 2015 |
| APA Scoring | Spring 2015 |
| Scores Reported to School Districts | End of June 2015 |
| Portfolios Returned to Districts | August 2015 |

3.3 Pre-Administration Training

For schools with any students participating in the APA, the NJDOE required one administrator and at least one teacher to attend a pre-administration training session held at four regional locations across the state in the fall. The mandatory half-day training session for administrators focused on student participation guidelines for the APA, the administrators' roles and responsibilities, and the APA design. For teachers, online training modules were created that focused on the APA test design, CPI Links, Universal Scoring Rules, the required portfolio components, and scoring rubrics. The training modules also included information on the revisions to the APA. A list of training modules is shown in Table 3.2.

The administrator training for the 2015 assessment was held September 15, 16, 17, 18, and 19, 2014. In addition to the regional training sessions, online training sessions were simulcast via the Internet with an online application called WebEx. The WebEx training sessions enabled districts and schools to facilitate in-district training and reduce the transportation burden of attending the regional training. The WebEx administrator training session was held on September 17, 2014.

Table 3.2 2014–2015 Training Modules

- Module I: NJ APA Teacher Training
- Module II: APA Test Design
- Module III: Acceptable Evidence
- Module IV: Steps to Developing an Entry—Part One
- Module V: Steps to Developing an Entry—Part Two
- Module VI: Teacher Instructional Resources
- Module VII: Universal Scoring Rules
- Module VIII: Complexity
- Module IX: Performance
- Module X: Independence
- Module XI: Sample Entries
- Module XII: Performance Reports and Administrative Topics

Copies of all APA training materials are available on Questar's ServicePoint website at https://nj-servicepoint.questarai.com/NJxx01_Documentation_2014_2015.

3.4 Test Security Procedures

Due to the nature of the APA, educators are more extensively involved in preparing and handling the assessment materials than for other New Jersey statewide assessments. Therefore, it is necessary to provide the statement of professional and ethical responsibility, which describes the role of educators in implementing the APA and appeared on page 5 of the *2014–2015 APA Procedures Manual – Science*.

- **It is the responsibility of all contributors to a student’s portfolio to ensure that any and all data and documentation reflect authentic, accurate, and truthful information.**
- **Any student portfolio that is found to contain inauthentic data and/or documentation may result in professional consequences for staff and financial consequences for the school and district.**

There are several different occurrences that result in a security breach of an APA. As such, it is imperative that all staff involved in the development and submission of an APA adhere to the procedures and guidelines that are defined in this manual.

Evidence submitted in a portfolio must not be fabricated, altered, manipulated, or duplicated across students. Evidence must be dated with the date of the actual occurrence of the production of this evidence. Materials should not reflect score, date, or other changes using white out or other methods. More information on acceptable evidence production can be found in the APA Teacher Training Modules.

District and school administrators, as well as the student’s educators, are responsible for ensuring that the APA reflect a true picture of the student’s acquired knowledge and skills. The test security agreement must be signed which certifies that the assessment was completed in accordance with all directions and requirements.

3.5 Portfolio Construction

Developing an APA Portfolio Entry

An entry is a collection of evidence that documents a student’s knowledge and application of key concepts and skills pertaining to a particular content standard and grade-level CPI. Evidence may include teacher-graded student work samples, captioned photographs, and snapshots of completed student work.

The APA test specifications for each Science grade level delineate four standards and strands that must be assessed. A portfolio entry is produced for each set of standards and strands. In addition, a related cumulative progress indicator (CPI) is selected for assessment from the list in the test specifications.

In addition to the portfolio entries, a completed portfolio contains the following:

Table of Contents – A table of contents helps the teacher and/or student organize the portfolio. It can be adapted to meet the individual needs of each student.

Entry Cover Sheet – The entry cover sheet is used to document the entry type (Science), entry number, standard, strand, CPI, CPI Link type, and the specific CPI Link.

The steps for developing an entry are explained in the *APA Procedures Manual – Science* beginning on p. 48. These eight steps are as follows:

Step 1: Select a CPI and one related CPI Link to be assessed.

Step 2: Plan instruction and assessment concerning the CPI.

Step 3: Design activities that will be used to assess the CPI Link.

Step 4: Assess the student to get an initial piece of evidence for APA purposes.

Step 5: Implement instruction.

- Ensure instruction reflects the essence of the strand and standard.
- Ensure instruction is age- and grade-level appropriate.
- Retain a working folder of instructional activities and classroom-based assessments implemented between the activities that generate the initial and final pieces of evidence.

Step 6: Determine when evidence can be collected to document the final instructional assessment of the CPI Link for APA purposes.

Step 7: Based on the student’s accuracy score and level of prompt information on the “final” activity, determine if additional instruction and collection of evidence needs to occur for the entry.

Step 8: Review evidence to ensure that all information related to test design requirements are included.

For teachers preparing to administer the APA, extensive instructions appeared in the procedures manual on the teacher training slides, as well as on the website https://nj-servicepoint.questarai.com/NJxx01_Documentation_2014_2015. A number of annotated examples of acceptable evidence and unacceptable evidence were pictured in the *APA Procedures Manual*. The instructions also listed acceptable and unacceptable work samples.

To begin development of an APA portfolio entry, teachers selected a CPI and one related CPI Link to be assessed. Figure 3.1 summarizes how decisions for choosing CPI Links should and should not be made. CPI Links for each grade level and content area appear in Appendix E of the *2014–2015 APA Procedures Manual – Science*.

“Use of Prompting and Scoring Evidence,” Chapter 5 in the *2014–2015 APA Procedures Manual – Science*, describes the types of supports, prompts, and activity

formats that are acceptable for instruction and those that are acceptable for assessment. Beginning on p. 39, the *2014–2015 APA Procedures Manual – Science* provides teachers with information about task directions, prompts, and instructional supports.

Additionally, Appendix B of the *2014–2015 APA Procedures Manual – Science* shows the “Planning Tool” form with instructions. On page 1 of the “Planning Tool,” teachers documented their planned instructional lessons/unit of study needed to teach the skills and concepts of the CPI and the CPI Link. Also on page 1, teachers listed the supports by answering:

1. How will the student *access* instruction?
2. How will the student *interact* with instruction and materials?
3. How will the student *demonstrate knowledge, skills, and concepts* acquired?

After selecting the CPI and related CPI Link, teachers assessed students to obtain the initial pieces of evidence. Figure 3.2 summarizes the important points that teachers had to consider as they prepared to administer and score the initial entry.

Figure 3.1 Choosing a CPI Link for the APA

| How Do You Choose a CPI Link? Think About a Student | |
|--|--|
| <p>Decisions Are Based On:</p> <ul style="list-style-type: none"> • The student’s grade • What the student already knows • How quickly the student learns new information • High expectations for students • Initial level of prompts (<i>if any</i>) needed for the student to succeed • How well the student performs on the initial activity | <p>Decisions Are Not Based On:</p> <ul style="list-style-type: none"> • Student’s mode of communication • The student’s disability category • Low expectations for students • Supports needed by the student to participate and perform in the curriculum |

Figure 3.2 Administering and Scoring an Activity for the APA

Scoring the activity correctly for assessment purposes is important. The evidence must include scoring information (percent scores) about

- a student's accuracy when performing the skill, and
- the number of items/questions/task elements that the student performed independently.

Teachers must understand the difference between:

- providing *task directions*,
- providing *supports*,
- providing *indirect prompts* (verbal, model, and gestural),
- providing *physical prompts*, and
- providing the answer (*directly prompting the student with the answer to the question*)

To ensure that scoring information on the evidence is accurate.

Scoring an activity for APA requires documentation of how well the student performed the skill.

- Accurate performance

And documentation of how many of the items/questions/task elements were done independently.

- Independence level

Scoring for APA separates these two concepts.

Scoring the activity for accuracy requires a consistent understanding of when to mark an answer right or wrong.

- Certainly, if the student performed the skill independently, the answer is either correct or incorrect.
- But what about when the student receives a prompt? How do you score the item correct or incorrect?

Scoring a Piece of Evidence

When an instructional activity is to be used as evidence in an entry, the activity must be scored by the teacher based on the number of test items the student answered correctly/incorrectly and the number of items the student completed independently.

Each piece of evidence must include two separate scores: one for accuracy and one for independence.

Scoring for Accuracy

Each item on the assessment evidence must be scored as either correct (+) or incorrect (-). The student must perform the skill at least five times by giving responses to at least five items or performing a task of at least five parts for each piece of the assessment. If the student requires a specific prompt level to respond, provide an indirect prompt (V, G, M) or, if necessary, a physical prompt. If the student requires a physical prompt, the item must be scored as incorrect (-P), because the student was given the correct answer. Accuracy scores are documented on the evidence as a percent score (the number of correct responses divided by the total number of items and multiplied by 100). The total number of test items must always be at least five.

Scoring for Independence

Each item on the assessment will receive a second score based on the level of Independence at which the student performed the skill. If the student responds independently, the item will be marked with an (I). If the student required an indirect prompt level to respond or perform the skill, then the item must be marked with the level of prompt. The typical hierarchy of prompts goes from least to most intrusive as verbal (V), gestural (G), model (M), and physical (P). The level of prompt a student receives is a teacher's decision based on the CPI Link selected, the student's prior knowledge, and other instructional information. Use the first letter of the prompt level to mark the answer, or if the student completes all of the items independently, mark each item with an (I). Independence is scored based on the prompt level or independence level at which the student performed the skill. Calculate the percentage of independent student responses by dividing the number of independent responses by the total number of items then multiplying by 100. This percentage must be calculated and documented on every piece of evidence.

Table 3.3 summarizes the correct and incorrect scoring of items for accuracy and independence.

Table 3.3 Scoring of Items for Accuracy and Independence

| An item is scored as correct (+) when | An item is scored incorrect (–) when |
|--|--|
| The student performs the skill independently and correctly | The student performs the skill independently but incorrectly |
| An indirect verbal prompt is provided and the student performs the skill correctly | An indirect verbal prompt is provided and the student performs the skill incorrectly |
| An indirect gestural prompt is provided and the student performs the skill correctly | An indirect gestural prompt is provided and the student performs the skill incorrectly |
| An indirect model prompt is provided and the student performs the skill correctly | An indirect model prompt is provided and the student performs the skill incorrectly |
| <i>You may never mark an item correct when using a physical prompt.</i> | A physical prompt is provided (e.g., the teacher moves the student’s hand, wrist, elbow, etc., to place the sticker in the correct place on the coordinate grid) |

PART 4: SCORING¹

From late April to early June 2015, Questar scored the Science APA portfolios. Each portfolio included four entries.

Each entry in a portfolio was scored independently by *at least* two readers for each dimension of the scoring rubric. Table 4.1 shows the total number of Science entries across grade levels.

Table 4.1 Total Number of Entries for the APA Science Portfolios

| | Grade 4 | Grade 8 | Grade 9 | Grade 10 | Grade 11 | Grade 12 |
|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|
| Science | | | | | | |
| Entry 1 | 1,209 | 1,287 | 179 | 262 | 677 | 88 |
| Entry 2 | 1,209 | 1,287 | 179 | 262 | 677 | 88 |
| Entry 3 | 1,209 | 1,287 | 179 | 262 | 677 | 88 |
| Entry 4 | 1,209 | 1,287 | 179 | 262 | 677 | 88 |
| Total | 4,836 | 5,148 | 716 | 1,048 | 2,708 | 352 |

As part of operational scoring, each entry of a portfolio was reviewed and given a rating of 0 to 4 for Complexity, 0 to 4 for Performance, and 0 to 4 for Independence. If the entry was found to merit a 0 in Complexity, an unscorable code that applied to all dimensions was assigned. The scoring rubric shown in Figure 4.1 presents the criteria used to score each APA entry, and Figure 4.2 lists all unscorable codes and their reason descriptions.

Each entry is scored independently by at least two readers for each dimension of the rubric. An entry score is derived from two scores, one from each reader. If the scores given by the two readers are not equal, a third reader scores the “discrepant” dimension(s). The third reader’s score is then combined with the equivalent score. If a fourth reading was necessary, it was performed by the scoring director.

¹ All tables in Chapter 4 exclude students who are not required to take the APA or students who took the general assessment.

Figure 4.1 APA Scoring Rubric

| Dimension | 0 | 1 | 2 | 3 | 4 |
|---------------------|--|---|--|--|---|
| Complexity | Evidence provided is unscorable; all dimensions will receive a score of zero | CPI Link was assessed but there are major flaws in the evidence | CPI Link is a Far Link to the grade-level indicator | CPI Link is a Near Link to the grade-level indicator | CPI Link is a Matched Link to the grade-level indicator |
| Performance | Evidence is not clear or all items are not marked as correct / incorrect | Accuracy of work is 0–39% based on the last activity. OR Second activity includes more intrusive prompt | Accuracy of work is 40–59% based on the last activity | Accuracy of work is 60–80% based on the last activity | Accuracy of work is 81–100% based on the last activity |
| Independence | Evidence is not clear or all items are not marked for Independence/ prompt level | Student completed items/tasks independently 0–39% of the time | Student completed items/tasks independently 40–59% of the time | Student completed items/tasks independently 60–80% of the time | Student completed items/tasks independently 81–100% of the time |

Figure 4.2 Unscorable Codes

| Unscorable Codes | Subcodes | Reason Descriptions |
|--------------------------------|-----------------|--|
| Entry Errors (EN) | 1. EN-A | Entry is missing from the portfolio. |
| | 2. EN-B | Evidence was submitted for only one collection period (less than two pieces of evidence). |
| | 3. EN-C | Entry contains more than four (4) pieces of evidence. |
| | 4. EN-D | Entry cover sheet is missing from the portfolio and there is insufficient information for scoring. |
| | 5. EN-E | Entry Cover Sheet is incomplete and there is insufficient information for scoring. |
| Test Specification Errors (TS) | 6. TS-A | CPI Standard/strand/link was not allowable for student’s assigned grade level. |
| | 7. TS-B | Link referenced does not exist in current test specifications. |
| | 8. TS-C | Same strand/link is used in more than one entry. |
| Documentation Error (DC) | 9. DC-A | Student name is missing from one or more pieces of evidence. |
| | 10. DC-B | Date(s) on evidence are missing or incomplete (month/day/year). |
| | 11. DC-C | Date on evidence is outside the Collection Period(s). |
| Evidence Error (EV) | 12. EV-A | Type of evidence is not acceptable (data charts, observations). |
| | 13. EV-B | Photo(s) submitted do not meet evidence requirements. |
| | 14. EV-C | Yes/No and True/False question formatting is not acceptable. |
| | 15. EV-D | The balance of representation has not been met. |
| | 16. EV-E | One or both pieces of evidence do not have at least five test items. |
| | 17. EV-F | Student responses are unclear/unreadable or not evident for a minimum of 5 items. |
| | 18. EV-G | Evidence is not presented in the student’s mode of communication (based on documentation provided by teacher). |
| Link Assessment Error (LA) | 19. LA-A | Evidence does not assess the CPI link. |
| | 20. LA-B | Evidence does not connect to the essence of the CPI/Strand/Standard. |
| | 21. LA-C | Evidence assesses two different CPIs or CPI links. |
| | 22. LA-D | The CPI link includes multiple skills, but each piece of evidence assessed a different skill from the link. |
| | 23. LA-E | Evidence assesses more than the CPI link skill(s). |
| | 24. LA-F | One or more items indicate that the concept was incorrectly assessed. |

Major milestones and meetings for the **2014–2015 APA** portfolio scoring included the following:

- Rangefinding preparation..... February 26 – March 16, 2015
- Rangefinding meeting March 16–19, 2015
- Scoring preparation March 19 – April 14, 2015
- Questar and ILSSA meet to finalize training process March 25–31, 2015
- Training..... April 15–22, 2015
- Scoring begins..... April 27, 2015
- Scoring ends May 13, 2015

4.1 Scorer Selection

All scorers selected for the APA had at least a bachelor's degree. Preference was given to candidates with the following credentials:

- educational background, teaching experience, and/or certification in special education
- experience in scoring alternate assessment portfolios
- experience in scoring large-scale educational assessments

For Spring 2015, Questar hired 85 scorers: 55 females and 30 males. Thirteen readers hired for the project did not show up on the first day of scoring, leaving 72 to begin the training and certification process. All scorers had a minimum of a bachelor's degree, and 31 readers had graduate degrees. The scorer degrees included 14 education majors, 19 English/journalism majors, 12 science majors, 19 business majors, and 18 social and behavioral science majors.

All scorers received rigorous training prior to scoring and then received continuous training and monitoring throughout the scoring process. There were 72 scorers present on the first day of training: 72 scorers took the qualification tests, 14 scorers were either unable to qualify or resigned during the qualifying window, and 58 scorers met the qualifying criterion. Scorers' characteristics are summarized in Table 4.2.

After completion of scorer training and qualification, 10 table leaders and 6 feedback supervisors were selected based on their qualification scores and ability to oversee a team.

Table 4.2 Summary of the Scorers' Characteristics

| Scorers' Characteristics | Number |
|---|---------------|
| Number of Scorers Hired | 85 |
| Number of Scorers Trained | 72 |
| Experience | |
| Rehires | 49 |
| Previously Scored NJ APA | 25 |
| New Hires | 9 |
| Education: by Degree | |
| Business | 19 |
| Education/Special Education | 14 |
| English/Journalism | 19 |
| Science | 12 |
| Social and Behavioral Science | 18 |
| Post-Graduate Degrees | 31 |
| Qualification | |
| Scorers Present for Qualification Test | 72 |
| Scorers met criterion | 58 |
| Scorers not meeting criterion or left project | 14 |

Security at the Scoring Site

Providing an environment that promotes the security of test items, student responses, data, and employees is of utmost concern to Questar. Therefore, throughout the NJ APA operational scoring, Questar employed the following standard safeguards for security at the Bloomington, Minn., scoring site:

- A security guard and two site personnel were stationed at the entrance to verify that only employees or authorized visitors were permitted access.
- Entrance to the building was limited to scoring staff with picture ID badges.
- No materials were allowed outside the facility during the project without the permission of a person or persons designated by the NJDOE.
- Scoring personnel signed a nondisclosure and confidentiality form in which they agreed not to use or divulge any information concerning tests, scoring guides, or individual student responses.
- All staff displayed Questar identification badges at all times while in the scoring facility.

4.2 Rangefinding

Rangefinding is the process by which a wide range of portfolios are reviewed by a committee of New Jersey Special Education teachers for the purpose of selecting exemplars to use in the training, monitoring, and qualification of scorers and for establishing/revising the scoring guidelines. To the extent possible, these portfolios represent the range of abilities and characteristics in the population tested as well as a range of student work sample types.

Preparation for the 2015 rangefinding began March 9–13, 2015, in Apple Valley, Minn., to identify portfolios for New Jersey teachers and administrators to score during rangefinding. Participants in this work included the following:

- ILSSA and Keystone content specialists who collaborated with Questar staff to develop the scoring training materials and share the training responsibility
- Questar scoring directors who oversaw and monitored the scoring
- Questar program management team members who directed the day-to-day operations for the APA by working with NJDOE staff members and New Jersey educators

At this meeting, ILSSA, Keystone, and Questar staff reviewed training materials from the rangefinding of the previous assessment year and made necessary revisions. ILSSA and Keystone drew upon their several years of experience scoring the APA, and Questar used its experience with other alternate assessment programs to prepare and revise the training materials for rangefinding. Revised materials for rangefinding were reviewed and approved by the NJDOE.

To provide portfolios for rangefinding, Questar followed a selection process that would result in the early return sample including approximately 8% of the tested population. Primary factors for selection were public versus private schools, district factor group (DFG), and region.

Selection Requirements:

- Student representation at all grade levels being assessed
- Public and private each contributes to about 50% of the full sample (try to contain a balance as much as possible)
- Enough districts selected such that DFG classifications are evenly represented, when possible (i.e., A, B, CD, DE, FG, GH, I, J, and O)
 - DFGs S and V are grouped with Private schools in this process
- All regions of the state evenly represented: South, Central, and North (as much as possible)
- Include districts and schools with both large and small student counts.
- Avoid selecting the same districts or private schools two years consecutively, or twice within a three year period. (There may be times that exceptions are necessary in order to obtain the balanced representations of the DFGs.)
- To the extent possible, the sample will be representative of the population gender and ethnicity.

Staff members at ILSSA, Keystone, and Questar pre-screened the early-return portfolios to identify those to use for rangefinding. Portfolios were selected to represent the following:

- range of school districts
- different types of schools
- grade level of students (elementary, middle, high school)
- skill level (access skill, modified expectation)
- severity of disability (severe/profound, moderate, mild-moderate)
- possible score levels (low, medium, high)

Fifty portfolios were selected to go to rangefinding, with the goal of scoring between 40 and 50 portfolios from which entries would be chosen to supplement/replace existing materials.

Fourteen New Jersey teachers and administrators participated in the rangefinding meetings from March 17–20, 2015, at the Mercer County Community College Conference Center in West Windsor, N.J. Rangefinding committee members were certified in special education with appropriate grade-level and content-area expertise.

Staff members from the NJDOE, ILSSA, Keystone, and Questar facilitated the meeting. At the beginning, committee members were introduced to their tasks of reviewing and scoring rangefinding portfolios used to train the scorers. The portfolio components, the scoring handbook, the rangefinding matrix, the sample entries, and the content modules were discussed. The new unscorable codes were introduced and explained.

For the rest of the first day, the rangefinding committee stayed together as one to score two portfolios. After independently scoring the first entry, each committee member voted on the scores for each dimension by holding up the card representing the score to be assigned. Each dimension was discussed and a consensus of scores was reached. This

process was repeated by the committee for all entries in this initial portfolio, as well as for a second complete portfolio.

On the second day, the rangefinding committee was divided into groups of teams (tables) to discuss and score portfolios systematically assigned to each group. A leader was selected for each table whose responsibility was to maintain notes regarding portfolio discussions and record consensus scores. Each table also included a staff member from the NJDOE, ILSSA, Keystone, or Questar to facilitate discussion and answer questions. The table groups scored through two phases described as follows:

- Phase I – Three members of a team independently scored a portfolio. After the portfolio was scored, the table leader guided the reconciliation discussion. If there were differences among the three scores, the group reached agreement through discussion and review of the rubric. The group then noted specific details for their scoring of the portfolio on the rangefinding matrix. The scoring worksheets and the rangefinding matrix were placed in an envelope. When the process was complete, the portfolio was transferred to another table to be scored by a member of another table team.
- Phase II – After the portfolio was scored the fourth time by another table, staff members from NJDOE/ILSSA/Keystone/Questar compared the group score sheet with the fourth score sheet. This provided a check for consistency across the table groups. If scores were not consistent, a scorer from the original team and the fourth scorer from the different table discussed the scores to determine a consensus score.

A Scoring Services team member was responsible for facilitating the flow of the portfolios and maintaining a log detailing the scoring for each portfolio. Security of the rangefinding material was maintained throughout the meeting. While the meetings were in session, a staff member from the NJDOE, ILSSA, Keystone, or Questar was present in the meeting room at all times. The meeting rooms were locked when the meeting was not in session.

The NJDOE received a copy of the official rangefinding record from Questar, including the consensus scores and the teachers' comments.

Immediately following the rangefinding meeting, staff members from the NJDOE, ILSSA, Keystone, and Questar met to finalize and approve the consensus scores. Questar, ILSSA, and Keystone staff met to select rangefinding entries to replace some of the entries in the current training and qualifying sets. Questar compiled all final scores and produced a spreadsheet including content and CPI-specific decisions, scoring rationales (including reasons for dimension scores as well as unscorable codes), and notes outlining the purpose for which the entries could be used. The notes also included recommendations and rationale as to which entries should be used for training sets, which should be used for qualifying sets, which should not be used at all, and which could be exemplars, both "good" and "bad," for use in teacher training.

Entries scored by the committee from the current administration were used to supplement existing entries from previous training materials in order to ensure scoring consistency from one administration to another. The entries were specifically chosen to address particular score points and issues, both scorable and unscorable, to ensure that scorers were exposed to and qualified on entries that exemplified a wide variety of scores and codes. The entries were used as follows:

- 3 portfolios/36 entries for practice
- 3 portfolios/36 entries for qualification
- 3 portfolios/36 entries for additional training and qualification

During the weeks following rangefinding, staff members from the NJDOE, ILSSA, Keystone, and Questar reviewed decisions from their home sites. Through this work, the NJDOE, ILSSA, Keystone, and Questar staff continued to discuss the selected portfolios with conference calls and e-mails.

All training sets and qualifying portfolios were submitted to the NJDOE for approval and required sign-off before scorer training began.

4.3 Scorer Training

Training for scoring the APA portfolios was conducted by ILSSA and Keystone content specialists with the guidance of the NJDOE APA Coordinator. The scorers were trained to score Science in all applicable grade levels (grades 4, 8, and high school).

ILSSA and Keystone content specialists began the training with an introduction to the content standards and entry points and how these align to one another. Training included discussion of the training entries, the scores for each dimension, and the rationale behind these scores. ILSSA and Keystone content specialists presented a slide presentation that showed examples and non-examples of each dimension and content area.

Scorers received the *New Jersey Alternate Proficiency Assessment Scoring Handbook 2014–2015* and paper copies of the Content Guide and Five Items Resource for grades 4, 8, and high school, the scoring rubric, the unscorable codes, and the scoring rules by dimension (0's and 1's). The training and qualifying sets were presented to the readers online. The scorers were encouraged to take notes throughout training as well as during the entire scoring process. Scorers had their scoring handbooks available to refer to and were instructed to ask questions regarding specific portfolios throughout scoring. The Content Guide and Five Items Resource documents for grades 4, 8, and high school were explained during the training as content training to help them with their Practice and Qualifying sets.

Scorers worked through the scored rangefinding entries, clarified the scoring criteria, and practiced scoring. Scorers were given the opportunity to score the practice sets based on the training in the scoring handbook and the training set. True scores for these practice sets were then reviewed and justified with the group. Questar scoring directors used the

Cumulative Training Report to assist with the review. Retraining was conducted when indicated by the practice sets.

Qualification sets were then administered. Three qualification rounds (one portfolio per round—36 scores) were administered and scored. A re-qualification round, along with additional training, was available for those who required another round to meet the criteria. A reader's scores for the three qualification rounds and re-qualification (if necessary) rounds were averaged.

During qualification, Questar scoring directors, ILSSA and Keystone staff, and the NJDOE APA Coordinator reviewed and analyzed several reports including the Qualifying Reports by Set and the Cumulative Qualifying Reports

To qualify, scorers were required to attain a total of 75% exact agreement and 86.1% exact plus adjacent agreement (summative) across all portfolios and dimensions. Also, a minimum of 83.3% of exact and adjacent agreement scores (summative) was required for the Complexity dimension in order to qualify. Potential scorers who did not meet these requirements but were statistically close (would qualify if successful on two more portfolios) were retrained.

If an entry does not meet the test design requirements, a score of zero may be applied to all dimensions or individual dimensions as defined by the scoring rules. Because the zero score rules were very important to APA scoring, all scorers received additional training as necessary on the entries with zero rules.

After qualification, scorers were given additional content training. The Content Guide and Five Items Resource documents for grades 3, 5, 6, and 7 were given to the scorers and trained as those grades came onto the scoring floor during live scoring. Grades 4, 8, and high school were reviewed again as those grades came onto the scoring floor.

The NJDOE APA Coordinator participated in the final qualification round and the beginning of scoring. The feedback supervisors and team leaders were given additional training by ILSSA and Keystone staff.

4.4 Scoring Procedures

The purpose of scoring is to measure whether the evidence submitted for each CPI Link demonstrates that (1) the student has attained the conditions required for independent and accurate performance and (2) the degree the evidence is aligned to the New Jersey Content Standards. Participants during scoring included the Questar scoring directors, supervisors, and trained scorers; ILSSA and Keystone content specialists; and NJDOE and teacher representatives.

The Questar scoring directors and supervisors ensured that scoring was conducted independently by trained and qualified scorers without discussion between or among scorers. Team leaders monitored scorers under close supervision of the scoring directors.

Scorers were required to bring questions about scoring a particular portfolio and rubric interpretation to their team leaders and/or scoring director.

Scorers worked at tables of 8 to 10 people under the supervision of a team leader. Portfolios to score were placed on large carts labeled as follows:

- Needs First Reading
- Needs Second Reading
- Needs Third Reading
- Reading Complete

Each scorer began by selecting a portfolio from the Needs First Reading cart. The scorer removed the portfolio from its envelope and verified that the portfolio number on the envelope matched the portfolio number on the binder. The scorer first checked for correspondence from the district in the form of letters and/or attendance sheets. If found, the scorer delivered the portfolio to a scoring director.

Scorers followed the detailed instructions in the *New Jersey Alternate Proficiency Assessment Scoring Handbook 2014–2015* to score the portfolios. Scorers began their work using the “Universal Scoring Rules for Each Entry” shown in Figure 4.3. Critical points included checking that the appropriate standards, strands, and CPIs were assessed for the grade level; verifying that the dates fell within the appropriate collection period; confirming that the first piece of evidence had an accuracy score of 39% or lower; replicating the percent score for independence; identifying at least five test items; and determining that only the specified CPI Link was assessed.

Instructions for the scoring rubric in the scoring handbook provided several pages of detailed information for each dimension. These instructions extensively expanded the scoring rubric to include a definition of terms, flowcharts, scoring rules/clarifications, and scoring notes. The instructions for the dimension scoring are shown in the *APA Scoring Handbook 2014–2015* beginning with the rubric on page 10. In addition, the scorers were provided instructions regarding the procedures for missing entries, alerts, possible security breach, and unscorable codes.

The score assigned for one dimension was not to influence the score assigned for another dimension. Each dimension of the rubric was reviewed and scored separately. Also, each content area was scored independently. No information from one content area was to influence the scoring of another.

Three monitor codes were used for scoring the APA entries. A scoring director was assigned the task of verifying and assigning void codes for security breach, insufficient evidence due to extended sick leave (medical emergency), or no evidence (excessive absence non-medical).

Entries that did not meet the test design requirements were assigned a score of zero for all dimensions (by assignment of an unscorable code) or individual dimensions depending

on the type of error. Additional training was provided to scorers to enable them to better identify issues that would result in a zero score. Scorers were authorized to assign zeros pertaining to less complex scoring issues.

Scorers escalated portfolios with more complex scoring issues that required assistance from their team leaders. The team leaders answered the questions as appropriate or escalated the portfolio to the feedback supervisors, scoring directors, or ILSSA and Keystone, depending on the issue identified. The portfolios were shelved in the appropriate area to await review. The portfolio was reviewed and an Explanation Sheet was completed through the online scoring system, and the appropriate scores were assigned. The explanation sheets were printed by the clerking staff and inserted into the portfolio. Explanation sheets were written for 1,635 portfolios out of 3,957 portfolios.

When scoring was completed, the scorer placed the portfolio on the Reading Complete cart and selected another portfolio from carts of portfolios needing first or second readings.

APA REQUIREMENTS

In order to begin scoring an entry, both pieces of evidence in that entry must adhere to the APA Universal Scoring Rules and all test design requirements. If any of the criteria has not been met for two pieces of evidence within the entry, the entry will be unscorable for all three dimensions.

1. Evidence must include the **student’s name**.
2. Evidence must include the **complete date**. (Month/day/year)
 - Date on initial evidence must fall within the collection period of
Sept. 2, 2014 – Nov. 14, 2014
 - Date on final evidence must fall within the collection period of
Dec. 8, 2014 – Feb. 13, 2015

NOTE: If the final piece of evidence is dated **January 2nd, 5th, 6th, or 7th with 2014** as the year, this is acceptable.
3. Evidence must be presented in the **appropriate format**.
 - Actual student work that meets the evidence requirements must be submitted
 - **No data sheets, observations, interviews, etc.**
4. Evidence must reflect the student’s **mode of communication**.
5. Evidence must include **at least 5 items** that assessed the CPI Link.
6. **Student response** must be evident **on at least 5 items**.
7. Evidence must **assess the entire link** while connecting to the essence of the CPI, standard, and strand.
8. Both pieces of evidence must assess the **same CPI Link and skills**.
9. The evidence must **NOT** include **more** than the skills contained within the CPI Link.

**Scoring Rules by Dimension:
Assigning Score of “0” (in one dimension only) and Score of “1”**

| Dimension | Score | Verify |
|---------------------|---|--------------------------------|
| Complexity | Complexity Score of “0” | |
| | A single score of “0” cannot be assigned for Complexity; if the entry is unscorable it will receive an Unscorable Code. | Not Allowable No single “0” |
| | Complexity Score of “1” | |
| | Major flaw: Same activity is used for both pieces of evidence (same context AND application). | Verify with Team Leader |
| | Major flaw: Only part of the CPI link has been assessed (the same part of the link has been assessed in both pieces of evidence). | Verify with Team Leader |
| Performance | Performance Score of “0” | |
| | Initial piece of evidence has a score of 40% or higher. | Reader Assigns |
| | Each item is not individually marked as correct (+) or incorrect (-). | Reader Assigns |
| | Writing rubric is provided (3.2. link) but there is no feedback on student work that corresponds to the rubric. | Verify with Team Leader |
| | First/initial activity is clearly more difficult than the second/final activity. | Verify with Team Leader |
| | Performance Score of “1” | |
| | Accuracy of student work on final activity is 0-39% | Reader Assigns |
| | A more intrusive prompt was given for the final activity than was given for the initial activity. | Verify with Team Leader |
| Independence | Independence Score of “0” | |
| | Each item is not individually marked for Independence/prompt level. | Reader Assigns |
| | Independence Score of “1” | |
| | Student completed 0-39% of tasks on final activity independently. | Reader Assigns |

4.5 Quality Control of Scoring

A team leader monitored 8 to 10 scorers under close supervision of the scoring director. Scorers were required to bring questions about scoring a particular portfolio and rubric interpretation to their team leaders or scoring director in every instance.

- Performance Assessment Reports – The scoring directors had access to reports that documented individual and group performance such as inter-rater reliability, frequency distribution, project completion, and validity. Scoring directors reviewed reports daily to ensure that all items were scored within acceptable parameters and within the scheduled timeframe.
 - **Reader Reliability Report:** scoring directors reviewed inter-rater reliability reports daily to assess how accurately scorers assigned scores. The reader reliability report was available in either daily or cumulative format.

This report showed the exact, adjacent, and non-adjacent agreement for each scorer. Scoring directors used this report to evaluate individual scorer, team, and room totals and determine if any retraining was needed. If a scorer, team, or the room as a whole had an average agreement below the acceptable level predetermined by the NJDOE, it indicated that there was a misconception held by a portion of the scorers that needed to be addressed.

- **Score Point Distribution Report:** this report documents the percentage of scores assigned to each score point (0-4) and unscorable code by each scorer. This report was reviewed by the scoring directors and was produced both on a daily and cumulative basis.
- Read Behinds – In conjunction with the statistics provided by the reader performance reports, team leaders and scoring directors read behind between 5 and 10% of the portfolios already scored. This helped identify individual trends and tendencies that were the foundation for individual feedback and retraining indicators.
- Validity – Scorers were required to score student portfolios that had a pre-assigned “true score.” Statistics from the scoring of validity portfolios showed how often scorers agreed with the true score and was an indication of problem scorers or scoring trends. Each scorer was required to attain a percentage agreement with the true scores as established by the NJDOE. Any scorer who fell below this validity requirement was retrained and placed on probation. If a scorer fell below the established percentage on two consecutive validities, they could be released from the project.

Additionally, the NJDOE monitored scoring. Reports available during scoring for the NJDOE review included the following:

- Cumulative Reader Reliability Report
- Cumulative Score Point Distribution Report

4.6 Task Examination

Before the portfolios were scored, condition codes were assigned as follows:

| | |
|---|--|
| 6 | Security Breach |
| A | Insufficient evidence due to extended sick leave (illness) |
| B | No evidence (not ill) |

Table 4.3 provides the number of total portfolios processed² (across dimensions) that were assigned a condition code for Science within a grade and the number and percentage of condition codes associated with each of the three code categories (i.e., 6, A, and B). For example, 16 of the 1,299 total portfolios processed in grade 4 Science (approximately 1.2%) resulted in a condition code. Of those 16 processed, four (or 25.0%) were due to security breach, six (or 37.5%) were due to insufficient evidence related to illness, and six (or 37.5.0%) were due to no evidence being provided. The “% Assigned a Code” column shows the percentage of the condition codes out of the total number of portfolios. For example, four of the 1,299 portfolios processed in grade 4 Science (approximately 0.3%) received a security breach condition code. “No evidence” was the most frequent condition code assigned, followed by “security breach” and then “insufficient evidence due to illness.”

Prior to 2012, the number of unscorable codes was calculated based on the number of reads, where one student’s portfolio is counted multiple times based on how many readers scored each entry. Since 2013, the scoring rule has been changed so that the number of unscorable codes is now calculated based on the entry so that one student’s portfolio is counted just one time. Tables 4.4 through 4.6 were prepared based on the number of entries read.

Table 4.4, based on the number of entries read and broken down by dimension, shows the distribution of assigned unscorable error codes and scores by grade and dimension for Science. The greatest percentage of codes assigned to portfolio entries was at grade 9 (31.6%). Grades 4, 8, 10, 11, and 12 showed percentages of 18.8%, 20.3%, 28.6%, 15.6%, and 26.4%, respectively.

Overall, students did better on the Performance and Independence dimensions than the Complexity dimension across all three content areas. For example, at grade 4 for Science, 44.7% of the entries received a score of 4 on the Performance dimension, 55.6% of the entries received a score of 4 on the Independence dimension, and only 2.5% of the entries received a score of 4 on the Complexity dimension.

Table 4.5, based on the number of entries read, provide the overall number of unscorable error codes (listed below) assigned to each entry across grades during the handscoring process for Science.

² Portfolios that received a condition code were not scored and are therefore not included in the total number of entries.

- EN : Entry Error
- TS : Test Specification Error
- DC : Documentation Error
- EV : Evidence Error
- LA : Link Assessment Error

Across all grades, the most common error code was Link Assessment (LA), which indicates that the CPI Link, strand, and/or standard was not properly assessed.

Table 4.6 break the unscorable codes down into subcodes assigned to all entries combined by grade for Science. (See Figure 4.2 on page 31 of this technical report for a list of each unscorable code's subcodes and their descriptions.) For all grades, the most frequent unscorable error subcode was in Link Assessment Error, LA-A (i.e., "Evidence/rubric does not assess the CPI link").

Table 4.3 Distribution of Condition Codes by Grade

| Grade | Content Area | Total Portfolios Processed | Portfolios Resulting in a Condition Code | | 6 - Security Breach | | A - Insufficient Evidence due to Illness | | B - No Evidence | |
|-------|--------------|----------------------------|--|------------------|---------------------|-------------------|--|-------------------|-----------------|-------------------|
| | | | # | % of Total Proc. | # | % Assigned a Code | # | % Assigned a Code | # | % Assigned a Code |
| 4 | Sci | 1,299 | 16 | 1.2 | 4 | 0.3 | 6 | 0.5 | 6 | 0.5 |
| 8 | Sci | 1,383 | 14 | 1.0 | 6 | 0.4 | 2 | 0.1 | 6 | 0.4 |
| 9 | Sci | 188 | 3 | 1.6 | 1 | 0.5 | -- | -- | 2 | 1.1 |
| 10 | Sci | 282 | 2 | 0.7 | -- | -- | -- | -- | 2 | 0.7 |
| 11 | Sci | 691 | 9 | 1.3 | 3 | 0.4 | 2 | 0.3 | 4 | 0.6 |
| 12 | Sci | 114 | 2 | 1.8 | -- | -- | -- | -- | 2 | 1.8 |

Table 4.4 Distribution of Unscorable Error Codes and Scores – Science

| | Scores # Entries Read | CODES | | 0 | | 1 | | 2 | | 3 | | 4 | |
|-----------------|-----------------------------|-------|------|-----|-----|-----|------|-------|------|-------|------|-------|------|
| | | # | % | # | % | # | % | # | % | # | % | # | % |
| Grade 4 | | | | | | | | | | | | | |
| Complexity | 4,836 | 908 | 18.8 | 12 | 0.2 | 79 | 1.6 | 2,267 | 46.9 | 1,447 | 29.9 | 123 | 2.5 |
| Performance | 4,836 | 908 | 18.8 | 146 | 3.0 | 299 | 6.2 | 309 | 6.4 | 1,013 | 20.9 | 2,161 | 44.7 |
| Independence | 4,836 | 908 | 18.8 | 82 | 1.7 | 385 | 8.0 | 231 | 4.8 | 540 | 11.2 | 2,690 | 55.6 |
| Grade 8 | | | | | | | | | | | | | |
| Complexity | 5,148 | 1,043 | 20.3 | 23 | 0.4 | 35 | 0.7 | 1,890 | 36.7 | 1,647 | 32.0 | 510 | 9.9 |
| Performance | 5,148 | 1,043 | 20.3 | 177 | 3.4 | 201 | 3.9 | 356 | 6.9 | 1,136 | 22.1 | 2,235 | 43.4 |
| Independence | 5,148 | 1,043 | 20.3 | 135 | 2.6 | 280 | 5.4 | 177 | 3.4 | 532 | 10.3 | 2,981 | 57.9 |
| Grade 9 | | | | | | | | | | | | | |
| Complexity | 716 | 226 | 31.6 | 1 | 0.1 | 13 | 1.8 | 372 | 52.0 | 64 | 8.9 | 40 | 5.6 |
| Performance | 716 | 226 | 31.6 | 18 | 2.5 | 40 | 5.6 | 50 | 7.0 | 100 | 14.0 | 282 | 39.4 |
| Independence | 716 | 226 | 31.6 | 10 | 1.4 | 47 | 6.6 | 15 | 2.1 | 51 | 7.1 | 367 | 51.3 |
| Grade 10 | | | | | | | | | | | | | |
| Complexity | 1,048 | 300 | 28.6 | -- | -- | 9 | 0.9 | 563 | 53.7 | 93 | 8.9 | 83 | 7.9 |
| Performance | 1,048 | 300 | 28.6 | 15 | 1.4 | 55 | 5.2 | 57 | 5.4 | 162 | 15.5 | 459 | 43.8 |
| Independence | 1,048 | 300 | 28.6 | 17 | 1.6 | 39 | 3.7 | 23 | 2.2 | 59 | 5.6 | 610 | 58.2 |
| Grade 11 | | | | | | | | | | | | | |
| Complexity | 2,708 | 422 | 15.6 | 10 | 0.4 | 31 | 1.1 | 1,829 | 67.5 | 176 | 6.5 | 240 | 8.9 |
| Performance | 2,708 | 422 | 15.6 | 80 | 3.0 | 206 | 7.6 | 231 | 8.5 | 675 | 24.9 | 1,094 | 40.4 |
| Independence | 2,708 | 422 | 15.6 | 29 | 1.1 | 342 | 12.6 | 137 | 5.1 | 315 | 11.6 | 1,463 | 54.0 |
| Grade 12 | | | | | | | | | | | | | |
| Complexity | 352 | 93 | 26.4 | -- | -- | -- | -- | 183 | 52.0 | 39 | 11.1 | 37 | 10.5 |
| Performance | 352 | 93 | 26.4 | 5 | 1.4 | 13 | 3.7 | 13 | 3.7 | 77 | 21.9 | 151 | 42.9 |
| Independence | 352 | 93 | 26.4 | 3 | 0.9 | 20 | 5.7 | 5 | 1.4 | 36 | 10.2 | 195 | 55.4 |

Table 4.5 Distribution of Unscorable Error Codes by Grade and Entry – Science

| | Total # Entries Read | Total # of Error Codes | | EN: Entry Error | | TS: Test Specification Error | | DC: Documentation Error | | EV: Evidence Error | | LA: Link Assessment Error | |
|-----------------|----------------------|------------------------|-------------|-----------------|------------|------------------------------|------------|-------------------------|------------|--------------------|------------|---------------------------|-------------|
| | | # | % | # | % | # | % | # | % | # | % | # | % |
| Grade 4 | | | | | | | | | | | | | |
| Entry 1 | 1,209 | 158 | 13.1 | 29 | 2.4 | -- | -- | 25 | 2.1 | 8 | 0.7 | 96 | 7.9 |
| Entry 2 | 1,209 | 212 | 17.5 | 33 | 2.7 | -- | -- | 24 | 2.0 | 12 | 1.0 | 143 | 11.8 |
| Entry 3 | 1,209 | 245 | 20.3 | 35 | 2.9 | 1 | 0.1 | 27 | 2.2 | 13 | 1.1 | 169 | 14.0 |
| Entry 4 | 1,209 | 293 | 24.2 | 34 | 2.8 | | | 37 | 3.1 | 12 | 1.0 | 210 | 17.4 |
| Total | 4,836 | 908 | 18.8 | 131 | 2.7 | 1 | 0.0 | 113 | 2.3 | 45 | 0.9 | 618 | 12.8 |
| Grade 8 | | | | | | | | | | | | | |
| Entry 1 | 1,287 | 235 | 18.3 | 29 | 2.3 | -- | -- | 19 | 1.5 | 32 | 2.5 | 155 | 12.0 |
| Entry 2 | 1,287 | 356 | 27.7 | 31 | 2.4 | -- | -- | 23 | 1.8 | 118 | 9.2 | 184 | 14.3 |
| Entry 3 | 1,287 | 168 | 13.1 | 30 | 2.3 | -- | -- | 28 | 2.2 | 25 | 1.9 | 85 | 6.6 |
| Entry 4 | 1,287 | 284 | 22.1 | 30 | 2.3 | -- | -- | 49 | 3.8 | 10 | 0.8 | 195 | 15.2 |
| Total | 5,148 | 1,043 | 20.3 | 120 | 2.3 | -- | -- | 119 | 2.3 | 185 | 3.6 | 619 | 12.0 |
| Grade 9 | | | | | | | | | | | | | |
| Entry 1 | 179 | 49 | 27.4 | 10 | 5.6 | -- | -- | 5 | 2.8 | 1 | 0.6 | 33 | 18.4 |
| Entry 2 | 179 | 37 | 20.7 | 10 | 5.6 | -- | -- | 7 | 3.9 | 2 | 1.1 | 18 | 10.1 |
| Entry 3 | 179 | 87 | 48.6 | 10 | 5.6 | -- | -- | 4 | 2.2 | 17 | 9.5 | 56 | 31.3 |
| Entry 4 | 179 | 53 | 29.6 | 11 | 6.1 | 2 | 1.1 | 4 | 2.2 | 2 | 1.1 | 34 | 19.0 |
| Total | 716 | 226 | 31.6 | 41 | 5.7 | 2 | 0.3 | 20 | 2.8 | 22 | 3.1 | 141 | 19.7 |
| Grade 10 | | | | | | | | | | | | | |
| Entry 1 | 262 | 62 | 23.7 | 5 | 1.9 | -- | -- | 13 | 5.0 | 2 | 0.8 | 42 | 16.0 |
| Entry 2 | 262 | 56 | 21.4 | 6 | 2.3 | -- | -- | 7 | 2.7 | 10 | 3.8 | 33 | 12.6 |
| Entry 3 | 262 | 90 | 34.4 | 4 | 1.5 | -- | -- | 8 | 3.1 | 16 | 6.1 | 62 | 23.7 |
| Entry 4 | 262 | 92 | 35.1 | 5 | 1.9 | -- | -- | 15 | 5.7 | | | 72 | 27.5 |
| Total | 1,048 | 300 | 28.6 | 20 | 1.9 | -- | -- | 43 | 4.1 | 28 | 2.7 | 209 | 19.9 |
| Grade 11 | | | | | | | | | | | | | |
| Entry 1 | 677 | 113 | 16.7 | 9 | 1.3 | -- | -- | 15 | 2.2 | 11 | 1.6 | 78 | 11.5 |
| Entry 2 | 677 | 77 | 11.4 | 9 | 1.3 | 1 | 0.1 | 15 | 2.2 | 6 | 0.9 | 46 | 6.8 |
| Entry 3 | 677 | 141 | 20.8 | 9 | 1.3 | -- | -- | 13 | 1.9 | 34 | 5.0 | 85 | 12.6 |
| Entry 4 | 677 | 91 | 13.4 | 10 | 1.5 | -- | -- | 17 | 2.5 | 4 | 0.6 | 60 | 8.9 |
| Total | 2,708 | 422 | 15.6 | 37 | 1.4 | 1 | 0.0 | 60 | 2.2 | 55 | 2.0 | 269 | 9.9 |
| Grade 12 | | | | | | | | | | | | | |
| Entry 1 | 88 | 24 | 27.3 | 3 | 3.4 | -- | -- | 3 | 3.4 | -- | -- | 18 | 20.5 |
| Entry 2 | 88 | 15 | 17.0 | 3 | 3.4 | -- | -- | 1 | 1.1 | -- | -- | 11 | 12.5 |
| Entry 3 | 88 | 38 | 43.2 | 3 | 3.4 | -- | -- | -- | -- | 12 | 13.6 | 23 | 26.1 |
| Entry 4 | 88 | 16 | 18.2 | 3 | 3.4 | -- | -- | -- | -- | 2 | 2.3 | 11 | 12.5 |
| Total | 352 | 93 | 26.4 | 12 | 3.4 | -- | -- | 4 | 1.1 | 14 | 4.0 | 63 | 17.9 |
| Total | 14,808 | 2,992 | 20.2 | 361 | 2.4 | 4 | 0.0 | 359 | 2.4 | 349 | 2.4 | 1,919 | 13.0 |

Table 4.6 Distribution of Unscorable Error Codes by Subcode – Science³

| | Grade 4 | | Grade 8 | | Grade 9 | | Grade 10 | | Grade 11 | | Grade 12 | |
|-------------------------------------|---------|------|---------|------|---------|------|----------|------|----------|------|----------|------|
| | # | % | # | % | # | % | # | % | # | % | # | % |
| Total # of Entries Read | 4,836 | -- | 5,148 | -- | 716 | -- | 1,048 | -- | 2,708 | -- | 352 | -- |
| Total # of Error Codes | 908 | 18.8 | 1,043 | 20.3 | 226 | 31.6 | 300 | 28.6 | 422 | 15.6 | 93 | 26.4 |
| EN: Entry Error | | | | | | | | | | | | |
| EN-A | 106 | 2.2 | 64 | 1.2 | 39 | 5.4 | 12 | 1.1 | 32 | 1.2 | 8 | 2.3 |
| EN-B | 25 | 0.5 | 54 | 1.0 | 2 | 0.3 | 8 | 0.8 | 5 | 0.2 | 4 | 1.1 |
| EN-C | -- | -- | 1 | 0.0 | -- | -- | -- | -- | -- | -- | -- | -- |
| EN-D | -- | -- | 1 | 0.0 | -- | -- | -- | -- | -- | -- | -- | -- |
| EN-E | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| TS: Test Specification Error | | | | | | | | | | | | |
| TS-A | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| TS-B | 1 | 0.0 | -- | -- | -- | -- | -- | -- | 1 | 0.0 | -- | -- |
| TS-C | -- | -- | -- | -- | 2 | 0.3 | -- | -- | -- | -- | -- | -- |
| DC: Documentation Error | | | | | | | | | | | | |
| DC-A | 1 | 0.0 | 3 | 0.1 | -- | -- | -- | -- | -- | -- | -- | -- |
| DC-B | 28 | 0.6 | 35 | 0.7 | 2 | 0.3 | 8 | 0.8 | 11 | 0.4 | 1 | 0.3 |
| DC-C | 84 | 1.7 | 81 | 1.6 | 18 | 2.5 | 35 | 3.3 | 49 | 1.8 | 3 | 0.9 |
| EV: Evidence Error | | | | | | | | | | | | |
| EV-A | 1 | 0.0 | 1 | 0.0 | -- | -- | -- | -- | -- | -- | -- | -- |
| EV-B | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| EV-C | 11 | 0.2 | 93 | 1.8 | 2 | 0.3 | 12 | 1.1 | 28 | 1.0 | 9 | 2.6 |
| EV-D | -- | -- | 46 | 0.9 | 5 | 0.7 | 5 | 0.5 | 12 | 0.4 | -- | -- |
| EV-E | 10 | 0.2 | 6 | 0.1 | -- | -- | 4 | 0.4 | 2 | 0.1 | 1 | 0.3 |
| EV-F | 23 | 0.5 | 39 | 0.8 | 15 | 2.1 | 7 | 0.7 | 12 | 0.4 | 4 | 1.1 |
| EV-G | -- | -- | -- | -- | -- | -- | -- | -- | 1 | 0.0 | -- | -- |
| LA: Link Assessment Error | | | | | | | | | | | | |
| LA-A | 245 | 5.1 | 255 | 5.0 | 101 | 14.1 | 116 | 11.1 | 160 | 5.9 | 36 | 10.2 |
| LA-B | 113 | 2.3 | 11 | 0.2 | -- | -- | 2 | 0.2 | 4 | 0.1 | -- | -- |
| LA-C | 33 | 0.7 | 41 | 0.8 | 8 | 1.1 | 13 | 1.2 | 11 | 0.4 | 2 | 0.6 |
| LA-D | 4 | 0.1 | -- | -- | 10 | 1.4 | 14 | 1.3 | 16 | 0.6 | 4 | 1.1 |
| LA-E | 86 | 1.8 | 115 | 2.2 | 15 | 2.1 | 38 | 3.6 | 29 | 1.1 | 11 | 3.1 |
| LA-F | 137 | 2.8 | 197 | 3.8 | 7 | 1.0 | 26 | 2.5 | 49 | 1.8 | 10 | 2.8 |

³ All percentages are out of the Total # of Entries for each grade.

PART 5: RELIABILITY AND VALIDITY

5.1 Reliability

Many traditional measures of reliability are not appropriate for portfolio-based alternate assessments because they do not offer opportunities for test-retest or provide internal standardized items or tasks as a sample of a domain that can be used for all students. These limitations do not prohibit applying the concept of reliability to portfolio-type alternate assessments. Instead of trying to apply traditional statistics, we need instead to look for opportunities to look for sources of consistency in student performance and opportunities in which sources of error external to the students and their abilities may be impacting student scores. For sources of error, we can look to inter-rater reliability and decision accuracy.

Inter-rater Reliability

Inter-rater reliability investigates the extent to which examinees would obtain the same performance level if the portfolio had been scored by different scorers. Inter-rater reliability is calculated as the percent agreement between raters. The metrics tracked and reported are “exact agreement” and “adjacent agreement.” Exact agreement is when the two independent scorers assign the same score to the same student work. Adjacent agreement is when the two independent scorers assign adjacent scores to the same work.

Table 5.1 shows the percent of portfolio entries scored with exact agreement and adjacent agreement as well as the percent of scores that required resolution. All entries were scored for each of the three dimensions: Complexity, Performance, and Independence. A third scorer must score if the first two scores are not equal.

Table 5.1 shows, for example, that scores for the grade 4 Science entries on the Complexity dimension were in exact agreement for 98.0% of the entries. A third reader was required for scoring 2.0% of the entries. For the grade 4 Science entries on the Performance and Independence dimensions, scores were in exact agreement for 97.1% and 98.2% of the entries, respectively. A third reader was required for scoring 3.0% of the entries on the Performance dimension and 1.8% of the entries on the Independence dimension.

The percentage of entries requiring a third reader (not including grade 9, 10, and 12 due to the smaller number of examinees in those grades) for resolution ranged from 1.3% to 3.4% in Science. Resolution rates were highest in grade 8 in Science. A high inter-rater reliability coefficient indicates that subjectivity and differences between scorer’s estimates of student work was not a source of significant error in the students’ scores.

Appendix C shows the consistency between APA scorers for each entry for every grade.

Table 5.1 Consistency between APA Portfolio Scorers

| | GRADE 4 | | | GRADE 8 | | | GRADE 9 | | |
|----------------|------------|---------------|------------|------------|---------------|------------|------------|---------------|------------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | |
| Complexity | 98.0 | 1.8 | 2.0 | 98.4 | 1.2 | 1.6 | 97.7 | 1.7 | 2.4 |
| Performance | 97.1 | 1.6 | 3.0 | 96.6 | 1.9 | 3.4 | 97.3 | 1.3 | 2.7 |
| Independence | 98.2 | 1.1 | 1.8 | 98.2 | 0.8 | 1.8 | 98.6 | 0.9 | 1.4 |
| | GRADE 10 | | | GRADE 11 | | | GRADE 12 | | |
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Complexity | 98.8 | 0.8 | 1.7 | 98.7 | 0.8 | 1.3 | 99.7 | 1.2 | 1.2 |
| Performance | 97.9 | 1.3 | 2.1 | 97.3 | 1.3 | 2.7 | 98.0 | 2.9 | 2.7 |
| Independence | 99.1 | 0.8 | 0.9 | 98.8 | 0.6 | 1.3 | 99.1 | 1.2 | 1.2 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension for resolution.

Decision Consistency

Decision consistency is a psychometric term that refers to "the agreement between the classifications based on two non-overlapping, equally difficult forms of the test" (Livingston & Lewis, 1995). In some cases the classification agreement is assessed between a test score and a reliable second source. Regardless of the source of the second measure, a high consistency between the two classifications is desirable because they confirm each other's decision. In the APA case, a decision consistency study is designed to compare the performance level assigned through APA test scores with the performance level given by teachers. Teachers were asked to indicate the performance level they expect students to achieve based on their classroom experience with the students.

A study by Meisels, Bickel, Nicholson, Xue, and Atkin-Burnett (2001) confirmed that teacher rating has a high correlation with standardized battery when the rating is based on a curriculum-embedded performance assessment. The curriculum-embedded performance assessment is similar to the APA in that content standards, instruction, and test are designed as one interlocked components of an educational program. The similarity suggested that teacher rating will be a reliable second source.

In 2010–2011, teacher rating was collected through the convergent validity study (see Appendix K of the *2010–2011 APA Technical Report*, which is located online at <http://www.state.nj.us/education/assessment/apa/info/APA11TechReport.pdf>). This study shared the same design as the decision consistency; therefore, the results are used here to interpret decision consistency. The study showed that the overall exact agreement rate between teacher rating and performance level assigned by test scores were around 50% for the three content areas. The study provided the following plausible causes of the less than optimal outcomes:

- teachers might not fully understand the PLDs and factors other than grade-level academic skills might be considered by teachers,
- materials submitted might not adhere to the APA rules,
- teachers might not be as comfortable with standard-based teaching and assessment as they are with behavioral-based based teaching and assessment, and
- evidence submitted might not fully capture student performance.

5.2 Validity

The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) states that “Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees” (p. 17).

This section presents efforts to document and gather evidence to support the interpretation of APA performance scores. Efforts focus on documenting content aspects of evidence and gathering consequential aspects of evidence. While this section summarizes evidence supporting claims as to the validity of the APA performance scores, many parts of this technical report provide appropriate evidence for validity. Given the procedural and empirical evidence available and rationale presented below, valid performance standards-based interpretations and uses of the scores are generally supported.

The process implemented by the NJDOE for developing and implementing the APA is an example of the content aspect of validity. The content aspect includes evidence of construct relevance, representativeness, and technical quality. Baker and Linn (2002) suggest that “Two questions are central in the evaluation of content aspects of validity. Is the definition of the content domain to be assessed adequate and appropriate? Does the test provide an adequate representation of the content domain the test is intended to measure?” (p. 6).

The following sections help answer these two very important questions and also address Standard 1.6 of *the Standards for Educational Psychological Testing*.

Standard 1.6 When the validation rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified in reference to the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

Appropriateness of Content Definition

In 1996, the New Jersey State Board of Education adopted the New Jersey Core Curriculum Content Standards (NJ CCCS), an ambitious framework for educational reform in the State’s public schools. New Jersey’s standards were created to improve student achievement by clearly defining what all students should know and be able to do at the end of 13 years of public education. The NJDOE was conscientious in involving content specialists, alternate assessment specialists, policy experts, and measurement experts to ensure that the program was designed and implemented appropriately given the population of students being assessed and the federal requirements that the program must meet. New Jersey educators, NJDOE staff, special education directors, and other state

stakeholders were involved in the process throughout and provided feedback and guidance on all stages of APA development. Such stakeholder involvement helps to ensure that the results of the APA portfolios are viewed as meaningful and important to teachers and parents.

Since the adoption of those standards, the NJDOE has continuously engaged in discussion with educators, business representatives, and national experts about the impact of the standards on classroom practices. To assist teachers and curriculum specialists in aligning curriculum with the standards, the department provided local school districts with a curriculum framework for each content area. The frameworks provided classroom teachers and curriculum specialists with sample teaching strategies, adaptations, and background information relevant to each of the content areas. In addition, the statewide assessments were aligned to the NJ CCCS. This alignment of standards, instruction, and assessment was unprecedented.

The State Board required that the standards be reviewed and revised every five years. The review process, begun in May 2001, involved teachers, school administrators, students, parents, and representatives from business, higher education, and the community. In addition, several content areas were reviewed by Achieve, Inc., and the Council of Chief State School Officers (CCSSO). In response to this unprecedented review, the 2004 NJ CCCS provide the level of specificity and depth of content that will better prepare students for post-secondary education and employment. The standards are based on the latest research in each of the content areas and identify the essential core of learning for all students.

The Language Arts Literacy, Mathematics, and Science standards were adopted by the State Board of Education in July 2002. In April 2004, the Language Arts Literacy standards were revised to comply with the requirements of the No Child Left Behind Act of 2001 (NCLB) and readopted by the Board. Five content areas including the visual and performing arts, comprehensive health and physical education, world languages, career education and consumer, family and life skills, and technological literacy were also adopted by the Board in April 2004. To complete the revision process, the Social Studies standards were adopted in October 2004. The 2004 standards in all nine content areas replace the 1996 standards. Local school districts must align their curriculum and instructional program with the 2004 NJ CCCS. As required by regulation, the next five-year revision process began during the 2008–2009 school year for all nine content areas.

Since the adoption of the original 1996 NJ CCCS, the State Board approved administrative code that implements all aspects of standards-based reform. N.J.A.C. 6A:8 requires districts to: align all curriculum to the standards; ensure that teachers provide instruction according to the standards; ensure student performance is assessed in each content area; and provide teachers with opportunities for professional development that focuses on the standards.

In January 2008, the NJDOE Office of Academic Standards released Phase One of a standards clarification project. The purpose of this project is to provide materials in each of the nine content areas that convey an understanding of the priorities in the current NJ CCCS and how to capture those priorities in designing local curriculum and assessments, as well as in managing local instruction across content areas.

Phase One contained guidance framed as Areas of Focus for state assessment of Language Arts Literacy, Mathematics, and Science in grades 5–8. Developed by the Office of Academic Standards working with teams of field-based educators, the Areas of Focus included exemplars of how cumulative progress indicators may be assessed on state assessments.

In January 2008, the NJ CCCS in Mathematics were readopted with the following revisions:

- The new standards are more specific and clearer than the previous standards;
- The new standards are organized into a smaller number of standards that correspond to the content clusters of the statewide assessments;
- The new standards are intended to serve as clear guides to the assessment development committees so that there should be no gaps between the standards and the test specifications; and
- The new standards include expectations at grades 2, 3, 5, 6, and 7, as well as at grades 4, 8, and 11.

In preparing its recommendations, the Mathematics panel considered the *Principles and Standards for School Mathematics* published by National Council of Teachers of Mathematics (NCTM, 2000); the review of New Jersey’s 1996 standards by Achieve, Inc.; and other states’ standards.

Similarly, the NJ CCCS in Language Arts Literacy were influenced by the national standards developed by the National Council of Teachers of English and the International Reading Association, the Achieve review of the 1996 standards, and research by the National Reading Panel. Standards for the end of grade 12 were adopted in January 2008.

The NJ CCCS in Science were adopted in 2002 and published in 2004. Revised standards were adopted in June 2009. The projects and publications of the American Association for the Advancement of Science, the National Research Council, the National Science Teachers Association, and the National Assessment of Educational Progress were considered by the Science panel during the development of the standards.

Adequacy of Content Representation

Adequacy of the content representation of the APA is critically important because the test must provide an indication of student progress toward achieving the knowledge and skills identified in the NJ CCCS, and the test must fulfill the requirements under NCLB.

In December 2007, January 2008, and February 2008, the APA Advisory Committee met with a number of special education and content specialists to develop the APA test specifications. The APA test specifications delineate the standards and strands that must be assessed for each grade level and content area. ILSSA content specialists, NJDOE special education and content specialists, and special and general education teachers selected the Cumulative Progress Indicators (CPIs) available for the APA assessment. Then, skill statements that directly link the critical essence of the CPIs were developed. Documents used during this process included the NJ CCCS, scope and sequence for each content area, and the Areas of Focus from the Standards Clarification Project.

The work of the APA committees was influenced by the “Links for Academic Learning” developed and validated by Flowers, Wakeman, Browder, and Karvonen (2009). Initially, the “Criteria for Instruction and Assessment that Links to Grade Level Content” by Browder, Wakeman, Flowers, Rickelman, Pugalee, Karvonen (2007) and shown in Part 2 of this technical report consisted of eight criteria developed from the recommendations of a panel of alignment experts.

Flowers et al. (2009) described modifications to reflect both current federal policy and needs identified by special educators, measurement experts, and general education experts. The criteria were field tested in three states using varied alternate assessment formats, revised following review by measurement and special education experts and 20 state directors of alternate assessments, and field tested a second time with three additional states.

The revised eight criteria are shown in Table 5.2. Three of the earlier eight criteria are numbered 1, 2, and 3 in Table 5.2. During the work of the APA test development committees and the additional APA committees that followed, the eight criteria and these Standards were addressed:

Standard 3.11 Test developers should document the extent to which the content domain of a test represents the defined domain and test specifications.

Standard 10.1 In testing individuals with disabilities, test developers, test administrators, and test users should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disabilities and their associated characteristics extraneous to the intent of the measurement.

Evidence to support the APA alignment is given in this technical report in the test development and design sections of Part 2, the portfolio construction section of Part 3, the scoring rubric and procedures sections of Part 4, and the proficiency level descriptor and standard setting sections of Part 6 and the Appendices. APA committee groups included curriculum, range-finding, performance level descriptor, and standard setting committees.

Inherent in the portfolio design of the APA is instruction. Parts 2 and 3 describe the teachers' scoring and instruction that occurs between the initial and final collection for the portfolios. Sample activities developed by teachers are available on the APA website. Score reporting for instructional purposes is explained in Part 7.

Table 5.2 Links for Academic Learning (LAL) Alignment Criteria

| |
|---|
| <ol style="list-style-type: none"> 1. The content is academic and includes the major domains/strands of the content area as reflected in state and national standards (e.g., reading, math, science). 2. The content is referenced to the student's assigned grade level (based on chronological age). 3. The focus of achievement maintains fidelity with the content of the original grade level standards (content centrality) and when possible, the specified performance. 4. The content differs from grade level in range, balance, and DOK, but matches high expectations set for students with significant cognitive disabilities. 5. There is some differentiation in content across grade levels or grade bands. 6. The expected achievement for students is for the students to show learning of grade referenced academic content. 7. The potential barriers to demonstrating what students know and can do are minimized in the assessment. 8. The instructional program promotes learning in the general curriculum. |
| <p>Flowers, C., Wakeman, S.Y., Browder, D.M., & Karvonen, M. (2009). Links for academic learning (LAL): A conceptual model for investigating alignment of alternate assessments based on alternate achievement standards. <i>Educational Measurement: Issues and Practice</i>. 28(1), 25–37.</p> |

With information from teachers and scorers from the 2008–2009 APA administration, the following modifications will be made for future administrations:

- Some CPI Links will be revised and a few will be added.
- CPI Links related to assessment of spelling words will be deleted since these did not link to the other assessment specifications.
- Teachers must mark every item/question with an “T” when an item is performed independently, even if 100% of the test items were completed in this manner.
- When a teacher assesses a writing skill that requires a rubric for scoring, the student's writing sample must have editing/scoring notations that correspond with the rubric scores.

Convergent Validity

Convergent validity is one of the methods that examine the relationship of test scores with external variables (AERA, APA, & NCME, 1999). A special study of APA convergent validity was conducted in 2011. This study examined the correlation between proficiency levels assigned by two sources of information; teachers and test scores. As can be seen the design of convergent validity study is the same as the design of decision

consistency study. Therefore, the outcomes of this study can be used in decision consistency as well as convergent validity study. Information has been provided in the "Decision Consistency" section and is not repeated here. Instead, a brief summary of the study outcome is provided below. Details of the convergent validity study can be found in Appendix K of the *2010–2011 APA Technical Report*, which is located online at <http://www.state.nj.us/education/assessment/apa/info/APA11TechReport.pdf>.

The study showed that the exact agreement rate (the APA teacher's expectations were the same as the examinee's proficiency level) was approximately 50%. Various explanations were suggested to interpret the exact agreement outcome. Additionally, several possible next steps for the APA to increase the level of exact agreement such as including more detailed PLDs, increased training, and potential standardization of certain aspects of the APA were proposed.

Consequential Validity

Additional important validity evidence comes from the intended and unintended consequences of test use (AERA, APA, & NCME, 1999). The concept of consequential validity was introduced by Messick in 1989. Messick (1995) defined consequential validity as "evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long-term." There are two sources of consequential validity evidence provided in this section: (1) an analysis of performances of different sub-groups as suggested in the AERA, APA, and NCME *Standards* (1999) and (2) a survey study based on Messick's definition.

The consequences of test use can be investigated by looking at distributions of scores across subgroups in the tested population. The number and percent of students from various subgroups who achieve each of the three proficiency levels have been calculated separately by grade and content area. The subgroups addressed are disability category and public versus private school attendance.

For the disability category analysis, frequencies were computed to investigate the number of students from each disability category categorized into each of the three proficiency levels. These frequencies were looked at separately for each content area with all grades combined as well as within each content area at each grade.

In the body of the report, only the combined grades frequencies of disability category by proficiency level are presented in Table 5.3 for Science. The tables for each grade separately are included in Appendix E.

The frequencies provide an indication of whether there are differences with respect to disability category and/or proficiency level. The frequency tables provide an indication that in almost all grades there is some relationship between the indicated disability category and the proficiency level into which a student is categorized. However, the relationship seems weak and is not consistent enough across grades to indicate bias. Additionally, while all students with significant cognitive disabilities are likely able to make progress on academic content, and all deserve the opportunity to be exposed to

academic content, there is also likely some relationship between the types and significance of students’ disabilities and their ability to reach proficiency as defined for accountability purposes under the No Child Left Behind regulations.

The relationship between proficiency level private and public school attendance was also investigated by content area; sample sizes were too small to interpret when looked at by grade. The combined, across-grade frequencies for each proficiency level are provided by school type in Table 5.4. Similar to the results of proficiency level by disability categories analyses, there is a relationship between students’ placements in public or private school and their proficiency level. However, it is difficult to interpret these numbers or to conclude bias due to the nature of private school placements of students with significant cognitive disabilities in New Jersey.

Table 5.3 2015 APA Combined Grade Proficiency Level Frequencies by Disability Category

| Combined Grade Table of Disability⁴ | | | | |
|---|-------------------|--------------|--------------------|--------------|
| | Science | | | |
| Disability Category | Adv. Prof. | Prof. | Part. Prof. | Total |
| Auditorily Impaired | 2 | 4 | 5 | 11 |
| Autistic | 69 | 657 | 686 | 1,412 |
| Cognitively Impaired | 29 | 215 | 317 | 561 |
| Communication Impaired | 14 | 102 | 74 | 190 |
| Emotionally Disturbed | -- | 7 | 7 | 14 |
| Multiply Disabled | 46 | 551 | 593 | 1,190 |
| Deaf-Blindness | 1 | -- | 2 | 3 |
| Orthopedically Impaired | -- | 1 | 4 | 5 |
| Other Health Impaired | 5 | 62 | 54 | 121 |
| Specific Learning Disability | 7 | 40 | 45 | 92 |
| Traumatic Brain Injury | 3 | 22 | 20 | 45 |
| Visually Impaired | 1 | 1 | 3 | 5 |
| Blank or Multiple Grid | | 1 | 6 | 7 |

⁴ Beginning in 2013–2014, the “Social Maladjustment” disability category was no longer reported and is therefore not included in this table.

Table 5.4 Combined Grade Proficiency Level Frequencies by School Type

| | Science | | | |
|----------------|------------|-------|-------------|-------|
| | Adv. Prof. | Prof. | Part. Prof. | Total |
| Private School | 17 | 432 | 427 | 876 |
| Public School | 160 | 1,230 | 1,380 | 2,770 |
| Total | 177 | 1,662 | 1,807 | 3,646 |

A consequential validity special study was conducted in the end of 2011. The focus of this consequential validity study surrounding the implementation of the APA. Surveys were given to parents, teachers, and administrators. In addition, focus group interviews were conducted to collect information from a subset of administrators. The parent and teacher surveys were administered online. For the administrators, both the survey and the focus groups occurred during mandatory administrator training sessions to maximize the number of potential participants (participation was voluntary). Details of the study, such as survey questions, focus group questions, sample sizes, and findings can be found in Appendix K of the *2010–2011 APA Technical Report*, which is located online at <http://www.state.nj.us/education/assessment/apa/info/APA11TechReport.pdf>. Below is the conclusion of the consequential validity study.

The survey and focus group results provided information concerning the effect the introduction of the APA has had on the education process in New Jersey from three different groups of stakeholders. Regarding the first group, responding administrators, they are marginally positive about the APA. From the administrators view, the APA has been successful in more closely integrating curriculum, instruction, and assessment for APA students. Additionally, the APA has some positive unintended consequences (more collaboration within and across schools, more interaction between APA students and the general education population, and an increase in parental involvement in special education programs) and some negative consequences (increased teacher concerns about lost instruction time and increased special education teacher turnover). Lastly, two suggestions for improving the APA emerged from the administrator focus groups. First, improved perceptions in the consistency of scoring would help teachers and administrators overall understanding of the APA process. Second, some sort of standardization of tasks, which are pre-approved for use in the APA portfolio, could potentially improve the APA process.

In terms of the second group, responding parents, they are generally neutral to negative about the APA. Roughly 70–80% of responding parents are neutral to negative concerning their views about the APA. Only 20–30% of responding parents believe either that the APA has benefited their child or that the APA is a good measure of their child’s educational strengths or challenges. It does not appear overall that there is strong support

from those parents responding for the APA in its current form. The third group, responding teachers, has generally neutral to negative feelings about the APA. Roughly 65% of the responding teachers are neutral to very negative about the APA, with 30% negative to very negative. Twenty percent of responding teachers have both negative and positive opinions of the APA, depending on the issue. The last group is roughly 15% of responding teachers that have neutral to very positive opinions of the APA. Teacher respondents generally have a negative view of the APA and are neither pleased with nor supportive of this instrument. When asked about allocation of time for functional skills and content standards, responding teachers indicated that insufficient time is spent on functional skills. They also expressed that even before the implementation of the APA, not enough time was spent on functional skills, but that since the implementation of the APA even less time is spent on functional skills. Additionally, they feel that even before the implementation of the APA too much time was spent on content standards, and that since the implementation of the APA even more time has been spent on content standards.

In conclusion, of those responding, the teachers and parents have fairly negative views of the APA. The responding administrators have more balanced views of the APA. It is possible this is due to the levels of training and information concerning the APA that is provided to the three groups of stakeholders and the amount of direct contact with the APA that each group has. The administrators are provided with quite a bit of training and information and have substantial participation with the APA. Teachers have quite a bit of participation, but less training and information concerning the APA. Parents have minimal training and information and not much contact with the APA process.

Other Validity Studies

The NJDOE conducted two extra validity studies that related to test scores in 2011. The first study of these additional reports deals with scoring patterns (see Appendix K of the *2010–2011 APA Technical Report*, which is located online at <http://www.state.nj.us/education/assessment/apa/info/APA11TechReport.pdf>). It looked at the examinee scores for APA to determine the relative importance of the three scoring dimensions, complexity, performance, and independence. It found that almost all APA teachers were choosing appropriate levels of complexity and independence for their students to maximize their students' performance and proficiency levels. Additionally, it showed that an examinee's overall performance on the NJ APA is primarily based on an examinee's performance subtotals and not on the complexity or independence subtotals.

The second study of these additional reports deals with Pearson's Performance Scoring Center (PSC) (see Appendix K of the *2010–2011 APA Technical Report*, which is located online at <http://www.state.nj.us/education/assessment/apa/info/APA11TechReport.pdf>). It analyzed explanation sheets provided by scoring staff members at PSC, which were created when an examinee received a zero score for any scoring dimension. Through this analysis, the most common errors associated with the APA were identified. It recommended focusing training resources on those areas, primarily content alignment, to reduce the incidence of zero scores in the administration of the APA.

PART 6: STANDARD SETTING

6.1 Overview of the Process

New performance level descriptors should be created and new standards should be set whenever a testing procedure is adopted that is judged to be meaningfully different than previous testing procedures or whenever the assessed content meaningfully changes due to new test specifications or new content standards. The APA underwent significant changes between the 2007–2008 and 2008–2009 school years, including changes to the test specifications, assessable content, and scoring dimensions. As a result, both new performance level descriptors (PLDs) and a new standard setting were required.

PLDs are a required component of all assessments under Title I of the *Elementary and Secondary Education Act* (Federal Register, Volume 67, Number 129, 34CFR, Part 200, August, 2002). PLDs are descriptions of what students should know and be able to do academically to achieve a certain proficiency level given the range of skills assessed. The PLDs outline expectations for student performance at each proficiency level given the assessed components of the curriculum. In February 2009, the standard setting process began with the development of specific PLDs for each grade and content area for the APA administered in 2008–2009.

A standard setting was conducted June 9–12, 2009, to describe and delineate the thresholds of performance that are indicative of APA Partially Proficient, Proficient, and Advanced Proficient performances for Language Arts Literacy and Mathematics in grades 3–8 and 11 and for Science in grades 4, 8, and high school. Results of these studies were used to formulate recommendations to the Commissioner of Education and the New Jersey State Board of Education for the adoption of the cut scores (i.e., proficiency levels). In late June and early July, the standard setting panelists' recommendations were reviewed by senior staff in the Office of State Assessments and the Office of Special Education Programs, the Assistant Commissioner for the Division of Student Services, the Deputy Commissioner, and the Commissioner. The review led to some modifications to the panels' recommended cut scores, chiefly affecting the Advanced Proficient cut points. These cut scores were presented to the State Board of Education on July 15, 2009, and approved unanimously by resolution.

Both the PLD development meeting and the standard setting meeting were conducted by the staff from the NJDOE, Pearson, and ILSSA. Appendix B of this document provides a listing of the final PLDs, and an overview of the standard setting process is provided in the following section. A comprehensive report describing the PLD development process and participants is provided in Appendix G of the *2008–2009 APA Technical Report*. Similarly, an abbreviated version of the standard setting technical report, which summarizes the participant background information, outlines applied methodology, and presents some resulting tables, is provided in Appendix H of the *2008–2009 APA Technical Report*, which is located online at <http://www.state.nj.us/education/assessment/apa/APA09TechReport.pdf>

The full standard setting report, available from the NJDOE, provides complete descriptions of the standard setting planning, presentation documents and scripts, demographic information of the panelists, panelists' ratings from one round to the next, and their responses on the evaluation forms. The final cut scores approval by the State Board of Education is also presented.

Educators with extensive knowledge and experience in special education served as panelists for both the PLDs and the standard setting meetings. The expert judgments of panelists are most important for developing the PLDs and determining the standard setting cut scores. Nominations were solicited from school districts for teachers and administrators representing excellence in the teaching profession in terms of knowledge and experience in special education. Qualifications considered for the selection of panelists included the following:

- Current Position Description
- Years Teaching Special Education in New Jersey
- Years Teaching Regular Students in New Jersey
- APA Experience
- Type of Program
- Grade Level/Age of Current Students
- Type of Certification
- Highest Degree

6.2 Procedures

Performance Level Descriptors (PLDs)

In February 2009, 24 PLD panelists met for the purpose of writing the PLDs for Partially Proficient, Proficient, and Advanced Proficient performance. The PLDs are statements of what a student should know and be able to do at each proficiency level given the content standards assessed.

Dr. Kelly Burling served as primary meeting facilitator and facilitated the Language Arts Literacy group. Dr. Jason Meyers facilitated the Mathematics group, and Dr. Paul Nichols facilitated the Science group. Additional expertise in each content area was contributed by a content specialist in Mathematics and Science from the NJDOE as well as specialists from the Office of Special Education.

Tables 1–5 in the report present the panelists' gender and ethnicity, the geographic location of their districts, and the panelists' instructional experience by grade ranges. Panelists attended from 18 different districts in New Jersey and several private school settings. The panelists' years of experience ranged from 1 to 33 years with a median of 7.5 years. Seventeen of the 24 participants worked in special education. Their positions included social workers, teachers in self-contained classrooms, curriculum directors for students with disabilities, assessment coordinators, academic teachers, and administrators.

Panelists received training to ensure a common understanding of the APA, the target population, and the scoring dimensions. Extensive training and discussion were provided about the purpose and development of PLDs, including activities designed to familiarize the participants with elements of successful PLDs. Panelists were given copies of PLDs from the New Jersey Assessment of Knowledge and Skills (NJ ASK) Grade 4 Mathematics. Pearson facilitators led discussions of the following questions:

1. What language in the NJ ASK PLDs distinguishes each level from the others?
2. How are the definitions of student performance different from one another?
3. How is language used to convey meaning?
4. Would that language be useful to describe student performance on the APA?

The process was then repeated with the NJ ASK Grade 8 Mathematics PLDs. The ensuing discussions included the following:

1. What language is the same or similar?
2. Is the content (knowledge and skills) different from grade 4? How?
3. Do the PLDs reflect qualitative differences in student expectations from one level to the next and one grade to the next?
4. Do they show progression with respect to specific skills students should know and be able to do and not just list the same skills at different levels with the only defining factor being the degree of consistency with which the skills are displayed?
5. Are there times when the degree of consistency is an appropriate defining difference?

Notes taken by the facilitators during this discussion were given to all panelists as a resource for the PLD development within their content area groups.

The PLD analysis activities also established a basic format for the content area groups to use. Panelists identified the format used in the NJ ASK Grade 8 Mathematics as one they would like to follow for creating the APA PLDs. This format included an introductory statement followed with a bulleted list of knowledge and skills from the NJ CCCS.

Additional training was provided about the purpose and development of the CPI Links, which were developed to provide the test specification structure for the APA. Panelists were given (1) a copy of the *APA Procedures Manual* with tabs marking CPI Links and scoring rubrics; (2) a worksheet designed to help the participants review the CPI Links and identify language, knowledge, and skills to be used in the PLDs; and (3) a list of PLD evaluation criteria.

The content area groups were initially tasked with reviewing the CPI Links for the lowest assessed grade in their content area and beginning to draft statements and sentences that would comprise draft statements for that grade. Panelists continued working through the grades within their content area. Detailed descriptions of the procedures and discussions for developing the PLDS are included with the PLDs in Appendix B.

Standard Setting Process

Following the assessment administration and the creation of the PLDs, the standard setting panelists met in June 2009 to recommend cut scores. Approximately two-thirds of the operationally scored portfolios were available for standard setting examples. In addition, distributions of scores from the operational 2008–2009 administration were available to serve as impact data. The use of impact data provided panelists an additional frame of reference for their decision making.

Panelists were asked to recommend cut scores distinguishing between:

- Partially Proficient and Proficient
- Proficient and Advanced Proficient

Panelists recommended cut scores for Language Arts Literacy and Mathematics in grades 3–8 and 11 and for Science in grades 4, 8, and high school.

The panelists for standard setting consisted of 81 committee members including special education teachers, child study team members, general education teachers, and administrators. Committee members worked in seven panels based on content and grade. Pearson research scientists served as facilitators for the following groups:

- Mathematics grades 3, 4, and 5
- Mathematics grades 6, 7, and 8
- Mathematics and Science grade 11
- Language Arts Literacy grades 3, 4, and 5
- Language Arts Literacy grades 6, 7, and 8
- Language Arts Literacy grade 11
- Science grades 4 and 8

The demographic background by grade and content panel is presented for current grade taught, position type, and current content area in Table 6.1. Additional tables for grade and content panel are included in Appendix H of the *2008–2009 APA Technical Report* for gender, school location, ethnicity, and region.

Similar to the PLD development meeting, the standard setting meeting began with an introduction and extensive training leading to standard setting. Dr. Paul Nichols from Pearson served as the primary meeting facilitator. Dr. Debbie Traub from ILSSA presented the history of the APA and explained how the APA portfolios were constructed and scored. Dr. Nichols described the Body of Work standard setting method.

Dr. Traub recounted the regulatory history behind the APA and the purpose of IDEA and NCLB. She defined the population of students that participate in the APA. She also defined an alternate assessment and alternate achievement standards. Federal regulations requiring all students to be exposed to grade-level content were explained. Students with the most significant cognitive disabilities must be provided with challenging academic content that is clearly linked to grade-level standards. The content is determined by the

student's grade level based on assigned grade, not on functional level. Across all grades, students must be assessed on the full breadth and depth of the curriculum.

Table 6.1 Demographic Background of Standard Setting Panelists

| | | Current Grade Taught | | | | | |
|-----------------------|------------|-----------------------------|---------|-------------|----------|----------|-----------------|
| Content Area | Grade Band | K-2 | 3-5 | 6-8 | 9-12 | Multiple | Missing |
| LAL | 3-5 | 1 | 5 | 0 | 0 | 6 | 1 |
| LAL | 6-8 | 0 | 0 | 5 | 0 | 3 | 3 |
| LAL | 11 | 0 | 0 | 0 | 6 | 5 | 2 |
| Mathematics | 3-5 | 1 | 5 | 1 | 0 | 6 | 0 |
| Mathematics | 6-8 | 0 | 0 | 5 | 3 | 3 | 1 |
| Mathematics & Science | 11 | 0 | 0 | 0 | 8 | 3 | 1 |
| Science | 4 & 8 | 0 | 2 | 3 | 1 | 4 | 2 |
| | | Position Type | | | | | |
| Content Area | Grade Band | Special Ed. | Admin. | Curr. Spec. | Reg. Ed. | Other | Missing |
| LAL | 3-5 | 10 | 2 | 1 | 0 | 0 | 0 |
| LAL | 6-8 | 4 | 2 | 2 | 0 | 2 | 3 |
| LAL | 11 | 3 | 2 | 2 | 0 | 2 | 3 |
| Mathematics | 3-5 | 9 | 2 | 1 | 0 | 1 | 0 |
| Mathematics | 6-8 | 9 | 0 | 1 | 2 | 0 | 0 |
| Mathematics & Science | 11 | 7 | 2 | 1 | 0 | 0 | 2 |
| Science | 4 & 8 | 8 | 0 | 0 | 2 | 0 | 2 |
| | | Current Content Area Taught | | | | | |
| Content Area | Grade Band | Math | Science | LAL | Multiple | Missing | Not Applicable* |
| LAL | 3-5 | 0 | 0 | 0 | 10 | 1 | 2 |
| LAL | 6-8 | 0 | 0 | 0 | 3 | 3 | 5 |
| LAL | 11 | 0 | 0 | 1 | 6 | 4 | 2 |
| Mathematics | 3-5 | 1 | 0 | 1 | 7 | 1 | 3 |
| Mathematics | 6-8 | 2 | 1 | 0 | 6 | 2 | 1 |
| Mathematics & Science | 11 | 4 | 1 | 1 | 3 | 2 | 1 |
| Science | 4 & 8 | 0 | 2 | 0 | 8 | 2 | 0 |

*Not Applicable: The panelist was not currently in the classroom (e.g., administration).

This introduction was followed with a review of the portfolio process. The portfolio design, scoring of the three dimensions (Performance, Complexity, and Independence), links to the NJ CCCS, and grade-level CPIs were described. The review included examples of portfolio entries and evidence, and an extensive explanation of the role of the CPI Links was provided.

A reasoned judgment step was a warm-up task for the subsequent Body of Work procedure. This warm-up task had two goals:

1. Help panelists become familiar with the three scored dimensions
2. Encourage panelists to think about how the scored dimensions can be combined into total scores

Prior to the reasoned judgment task, panelists were introduced to the scoring rubrics for each score dimension and the descriptions of the dimensions. Panelists became familiar with the three scored dimensions (Performance, Independence, and Complexity) and the ways the dimensions can be combined into total scores. Then, panelists were asked to recommend what combinations of scores would be categorized as Partially Proficient, Proficient, and Advanced Proficient. Panelists were asked to consider a sample of score combinations and were presented the graph shown in Figure 6.1.

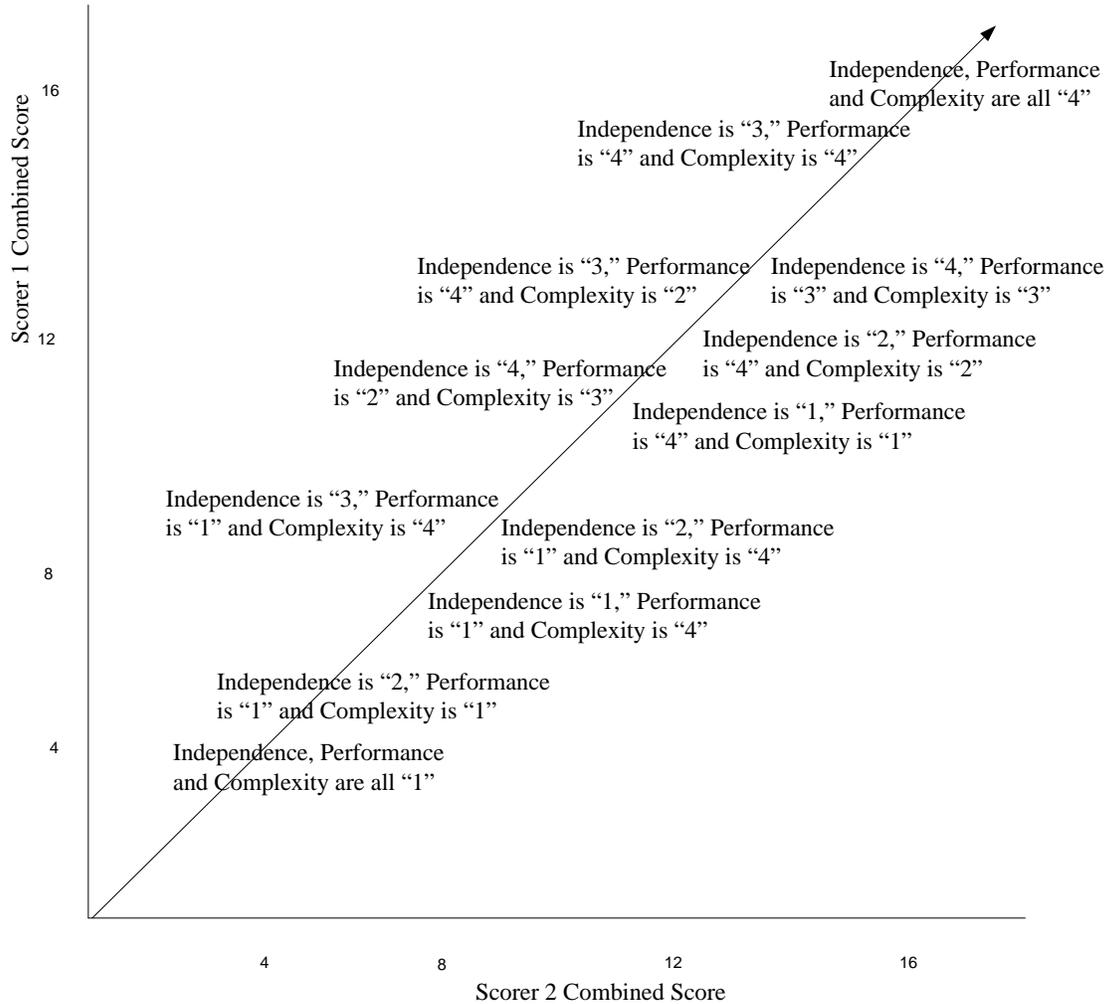
Panelists examined the figure showing the different score combinations and were reminded that each score was rated 0-4, but that entries receiving a 0 for either Performance or Complexity receive a 0 for the entire entry. Panelists were given a ratings sheet listing a progression of score combinations from Independence 0, Performance 1, and Complexity 1 to Independence 4, Performance 4, and Complexity 4. Panelists wrote Partially Proficient, Proficient, or Advanced Proficient next to each score combination on the ratings sheet.

The Body of Work method is intended for use with evidence of student learning displayed in a format other than a multiple-choice assessment. For the APA, the portfolio submitted comprises a “body of work.”

The Body of Work method uses portfolios in a number of different ways. For a student, a portfolio comprises a complete “body of work.” A student’s portfolio is double-scored to increase accuracy. Students whose body of work is of uneven quality were excluded. Only students whose scores were consistent were included. By including only students whose work is consistent, panelists were presented with an easier to understand example of a “Proficient” student or an “Advanced Proficient” student.

Panelists set standards in three steps: training, rangefinding, and pinpointing. Refer to the Procedures section of the standard setting report for the grade sequence used by each panel, the steps followed by each facilitator as they worked through the standard setting rounds, and the presentation of impact data. The next section in the report, Panelists, shows that 11 to 13 people served on each of the panels.

Figure 6.1 Graph for Reasoned Judgment Warm-Up Task



6.3 Summary of Results

The results summary in the standard setting report is organized into five sections: cut score, evaluations, decision factors, reliability, and vertical articulation.

In the standard setting report, Table 24 shows the summary of recommended cut scores and impact data for Language Arts Literacy. Table 25 presents the summary recommended APA cut scores and impact data for Mathematics and Science.

Cut scores computed following rangefinding round 1, rangefinding round 2, and the pinpointing rounds for Language Arts Literacy, Mathematics, and Science are shown in Table 6.2. Note that values are multiplied by 10.

Table 6.2 Cut Scores After Rangefinding and Pinpointing Rounds

| Grade | Content Area | Rangefinding Round 1 | | Rangefinding Round 2 | | Pinpointing Rounds | |
|-------|--------------|----------------------|-------|----------------------|-------|--------------------|-------|
| | | Cut 1 | Cut 2 | Cut 1 | Cut 2 | Cut 1 | Cut 2 |
| 3 | LAL | 356 | 506 | 356 | 518 | 368 | 518 |
| 4 | LAL | 423 | 525 | 409 | 531 | 403 | 542 |
| 5 | LAL | 419 | 534 | 410 | 538 | 426 | 546 |
| 6 | LAL | 377 | 511 | 366 | 517 | 379 | 520 |
| 7 | LAL | 391 | 529 | 386 | 529 | 397 | 532 |
| 8 | LAL | 283 | 527 | 398 | 529 | 404 | 531 |
| 11 | LAL | 433 | 527 | 424 | 537 | 415 | 529 |
| 3 | Mathematics | 370 | 499 | 356 | 509 | 374 | 510 |
| 4 | Mathematics | 422 | 533 | 414 | 534 | 426 | 532 |
| 5 | Mathematics | 380 | 520 | 377 | 517 | 373 | 502 |
| 6 | Mathematics | 381 | 502 | 371 | 514 | 384 | 517 |
| 7 | Mathematics | 401 | 526 | 400 | 532 | 405 | 522 |
| 8 | Mathematics | 393 | 515 | 389 | 520 | 389 | 520 |
| 11 | Mathematics | 287 | 528 | 416 | 531 | 416 | 531 |
| 4 | Science | 295 | 538 | 301 | 547 | 453 | 561 |
| 8 | Science | 422 | 551 | 429 | 564 | 429 | 564 |
| 11 | Science | 412 | 516 | 404 | 528 | 422 | 537 |

*Note that values are multiplied by 10.

New Jersey’s normal standard setting process for all assessment programs includes two additional steps: (1) a senior staff level review of standard setting panel recommendations to assure articulation with state education policy and priorities – this review may result in modifications to the panelists recommendations; and (2) the presentation of the final cut scores to the State Board for formal adoption by resolution.

The APA panelists recommendations were reviewed over several days by directors, managers, and associated staff from both the Office of State Assessments and the Office of Special Education Programs, and then by the Assistant Commissioner responsible for Special Education, the Deputy Commissioner, and the Commissioner. These consultations led to some modifications to the panels’ recommended cut scores, chiefly affecting the Advanced Proficient cut points. The final set of APA cut scores approved by the State Board is shown in Table 6.3.

Table 6.3 Approved 2009 Cut Scores

| | | 2009 APA Impact Percentages (2008 in Parentheses) <i>All Rounded. May Not =100%</i> | | | | |
|-------|--------------|---|-------------------------------------|------------------------------|-----------------|-----------------------------|
| | | Raw Scores 0-64 | | | | |
| Grade | Content Area | Proficient Cut Score | Advanced Proficient Cut Score | % Partially Proficient | % Proficient | % Advanced Proficient |
| 3 | LAL | 36.8 | 56.2 | 27 (22) | 47 (49) | 25 (29) |
| 4 | LAL | 40.3 | 60.0 | 33 (26) | 58 (49) | 8 (26) |
| 5 | LAL | 41.6 | 60.5 | 37 (29) | 55 (47) | 8 (24) |
| 6 | LAL | 37.9 | 58.1 | 32 (27) | 57 (49) | 11 (25) |
| 7 | LAL | 39.7 | 58.2 | 35 (30) | 51 (42) | 14 (28) |
| 8 | LAL | 40.4 | 59.3 | 35 (39) | 52 (40) | 12 (22) |
| 11 | LAL | 41.5 | 56.2 | 33 (36) | 36 (46) | 30 (19) |
| 3 | Mathematics | 37.4 | 57.5 | 35 (17) | 42 (52) | 23 (31) |
| 4 | Mathematics | 41.6 | 56.6 | 40 (22) | 33 (47) | 27 (31) |
| 5 | Mathematics | 37.3 | 55.0 | 34 (27) | 39 (47) | 27 (26) |
| 6 | Mathematics | 38.4 | 57.3 | 40 (29) | 46 (45) | 15 (26) |
| 7 | Mathematics | 40.5 | 58.3 | 36 (35) | 49 (39) | 15 (26) |
| 8 | Mathematics | 38.9 | 58.9 | 32 (46) | 51 (34) | 17 (20) |
| 11 | Mathematics | 41.6 | 57.9 | 40 (56) | 36 (30) | 24 (14) |
| 4 | Science | 43.0 | 62.1 | 46 (23) | 52 (50) | 3 (27) |
| 8 | Science | 42.9 | 58.3 | 35 (32) | 46 (41) | 19 (28) |
| 11 | Science | 42.2 | 60.6 | 40 (26) | 51 (56) | 10 (18) |

*Cut scores approved by the New Jersey State Board of Education on July 15, 2009.

PART 7: REPORTING

Questar was awarded the APA contract in August 2012. Beginning with the 2012–2013 assessment, the APA reports are delivered online by Questar.

The scored portfolios are returned to the schools from Questar after reporting. The portfolios are confidential pupil records. School and district staff must maintain the confidentiality of the portfolio contents. The portfolio contents are to be shared with parents and others in accordance with pupil records regulations.

The NJ APA provides a variety of reports to the school districts. Score reports are designed to display student identification and score information that can help identify student strengths and weaknesses and recognize weaknesses in instructional programs of the curriculum content standards. Information regarding student progress can assist Individualized Education Program (IEP) teams in selecting appropriate goals and objectives and evaluation criteria for individual students.

Both attending and sending districts receive score reports. Table 7.1 lists the distribution of the specific APA reports. On the APA rosters the instruction and assessment status for APA students is indicated to assist districts review and identify the performance of their students:

Status 1 = students are assessed at the school of residence

Status 2 = students are sent outside school of residence for instruction and assessment

Status 3 = students are received from another school for instruction and assessment

Statuses 2 and 3 actually describe the same student; therefore, status 3 students are not included in the summary of performance reports so that the same student is not counted twice.

Districts are required to report test results to their boards of education and to the public within 60 days of receiving test results. However, any report that contains data for less than 11 students may not be publicly reported due to the need to protect student confidentiality.

For teachers and administrators who need to discuss score reports with others, the NJDOE publishes the *Alternate Proficiency Assessment (APA) Score Interpretation Manual* available via ServicePoint at https://nj-servicepoint.questarai.com/NJxx01_Documentation.aspx. The manual provides a broad range of information to assist in the analysis, interpretation, and use of the different APA reports.

In late fall after reporting is complete, a state summary is produced and posted to the NJDOE website at www.state.nj.us/njded/schools/achievement/index.html. The state

summary is a data file, available in text and Excel formats, containing the same type of results as in the performance by demographics report at the state level.

Table 7.1 Distribution of the APA Reports

| <i>Report Distribution Overview</i> |
|---|
| <i>District Reports for Students Educated In and Out of the District</i> (* Receiving Districts, including Private Schools, will receive only the All Subjects Roster) |
| All Subjects Roster |
| Summary of Performance - District |
| Summary of Performance - School |
| Performance by Demographic Groups - District |
| Performance by Demographic Groups - School |
| Performance by Demographic Groups - State |
| |
| <i>School Reports for Students who Attend a Receiving School (if applicable)</i> |
| <u>Receiving School the Student Attends</u> will receive: |
| Individual Student Reports |
| Explanation Sheets (delivered only when student has non-score codes) |
| All Subjects Roster |
| Student Roster |
| <u>Sending School</u> will receive: |
| Student Stickers |
| Individual Student Reports |
| Explanation Sheets (delivered only when student has non-score codes) |
| All Subjects Roster |
| Student Roster |
| Summary of Performance - School |
| Performance by Demographic Groups - School |
| Performance by Demographic Groups - State |
| |
| <i>School Reports for Students who Attend a School in their District of Residence (if applicable)</i> |
| <u>Sending School</u> will receive: |
| Student Stickers |
| Individual Student Reports |
| Explanation Sheets (delivered only when student has non-score codes) |
| All Subjects Roster |
| Student Roster |
| Summary of Performance - School |
| Performance by Demographic Groups - School |
| Performance by Demographic Groups - State |
| |
| Note for ISR: If a student attends an out-of-state facility, the sending school should provide a copy of the ISR to the student's parents and to the out-of state attending facility as feedback. |

7.1 Interpreting Reports

Student Demographic Information

APA teachers included a scan sheet with student demographic information in the inside front cover pocket of the binder for each APA portfolio. The scan sheet information was used to prepare score reports and attach APA scores to the proper schools and districts. Also, the information was used to produce federal reports, including the Adequate Yearly Progress report.

Beginning with the 2006–2007 APA, New Jersey schools had the opportunity to provide student demographic information on a “student pre-ID” file. If a pre-ID file was provided, each student’s demographic information was preprinted on the front side of the SDIF. If any information was found to be missing or incorrect, it could be provided/corrected by the districts gridding the appropriate section on the scannable SDIF.

After the portfolios were submitted and demographic information scanned, districts were given access to an online Record Change process via Questar’s ServicePoint website. The Record Change application displayed each student’s demographic information collected on the SDIFs. A record change period allows the districts an opportunity to review and correct inaccurate student demographic information that the district provided for the assessment. Record changes are completed before reporting. Corrections to the student information are reflected in the reports. For the APA, the sending district is responsible for making all student demographic data changes. Both sending and receiving (attending) districts have access, but only the sending district can update. The sending district is also responsible for making all student data changes requested by the district where a student attends. If the receiving district identifies any errors, they must contact the sending district promptly, allowing time to have the corrections applied.

Terms and definitions used across the APA reports are listed in the *2014–2015 APA Score Interpretation Manual* beginning on page 29.

Score Information

Scores are reported by content area. A full description of the scoring rubric used for rating the APA dimensions is presented in Part 4 of this technical report. Proficiency level is assigned based on the student’s total earned score; a combination of the Complexity, Performance, and Independence scores for entries within the content area. The scores are based solely on the information provided in the portfolio; therefore, it is inappropriate to compare these results to other APA students and students taking the general assessments.

Each content area assessed receives a proficiency level. Table 7.2 summarizes the dimension scores.

Table 7.2 2015 APA Dimension Scoring

| Dimension Scoring (2009–2015) | | | | | |
|--|------------------------|----------------------------------|-----------------------|-----------------------------------|--|
| Dimension | Score Range per Reader | Calculation of Two Reader Scores | Score Range per entry | Entries Required Per Content Area | Maximum Possible Points By Content Area (Across Entries) |
| Complexity | 0–4 | average | 0–4 | 4 | 16 |
| Performance | 0–4 | add | 0–8 | 4 | 32 |
| Independence | 0–4 | average | 0–4 | 4 | 16 |
| Maximum Possible Score per Content Area | | | | | 64 |

Of the required four entries, only one scorable entry is required to assign a proficiency level. If the “subject portfolio” contains only one scorable entry, the total score and proficiency level are reported based on the dimension scores of that entry.

Some scoring related improvements were made in 2010–2011 based on feedback from the field. The rules on assigning zero scores for all three dimensions were relaxed so that some violations will result in zero score for only individual dimension instead of all three dimensions of the entry. In addition, some violations were scored.

Unscorable Entry Errors (zero scores for all three dimensions):

The term unscorable means that an entry error results in zero scores for all three dimensions (a score point of 0 for Complexity, 0 for Performance, and 0 for Independence), and an unscorable code is assigned that describes the error. An Explanation Sheet with the unscorable code, description, and a typed explanation of the error (as required) is placed inside the front of the scored portfolio to indicate that a basic test design requirement was not followed. (A copy of the Explanation Sheet also accompanies the ISR on ServicePoint.) If any of the following requirements are missing from a piece of evidence, the entire entry will be considered unscorable:

1. Student’s name
2. Complete dates (month/day/year) within the specified collection periods
3. A piece of evidence must include at least 5 test items that assess the CPI Link.
4. A writing rubric with each piece of evidence in the entry, when specified in a Writing CPI Link
 - Writing rubric must have at least 5 elements that assessed the Link
5. Evidence presented in the appropriate amount and format

An unscorable code may also result when the evidence/rubric presented in the entry does not align to the CPI/Strand/Standard. An Explanation Sheet is placed inside the front cover of a scored portfolio to provide additional information on these types of errors:

1. Evidence must assess the link while connecting to the essence of the standard and strand.
2. The same CPI Link must be assessed in both pieces of evidence.
3. Evidence must not include more than the skills contained within the CPI Link. (This is true for both the student work and a writing rubric.)

Evidence Errors (zero score for one dimension):

A zero score is assigned to an individual dimension, instead of receiving zero score codes for all 3 dimensions in the entry. This allows the other two dimensions to receive score points. An individual dimension receives a zero when the following violations occur:

- Some or all test items are not marked for accuracy (Performance)
- Accuracy score for initial evidence is higher than 39% (Performance)
- No editing marks related to the scoring rubric appear on the student writing response (Performance)
- Some or all test items are not marked for Independence/prompting (Independence)
- The first activity in the entry is clearly more difficult than the second activity (Performance)

Errors (receives score different than may be expected):

Rather than assigning an unscorable code for the entry, the scorers are allowed to recalculate percentages or reassign the appropriate performance score. The recalculation may result in a different final score point (1-4) than may be expected. The following violations are reviewed and the accuracy/independence scores recalculated by the scorers:

- One or more items are marked as physically prompted and correct (P+)
- Items are marked correct/incorrect but no percentage provided
- Items are marked Independent/prompted but no percentage provided
- One or more of the percentage scores provided are inaccurate
- One or more test items are not correctly graded (marked) for accuracy

Other terms and definitions useful for interpreting the score reports include the following:

No Proficiency Rating:

There are times that a student will not receive a proficiency classification in a content area. This occurs only when all entries are deemed unscorable.

Unscorable:

Scores are reported by content area. Entries that do not meet the APA requirements or are missing are reported as “0’s” along with an unscorable code. If all entries within a content area are unscorable, a student will receive a void for the proficiency level as long as there is documentation accompanying the portfolio indicating why the student should receive a void code. The void code

will be displayed in the sub-total of each dimension and total score for the content area.

Valid scores:

There is at least one scorable entry in a content area.

Void:

This indicates that a student’s assessment result is coded void. One or more content areas can be voided. The proficiency level in a content area is voided if all entries of that content area are unscorable. Instead of a proficiency level, one of the following notations is displayed in the reports:

| Entry Deemed Unscorable | Void Code | Proficiency Display |
|--|------------------|----------------------------|
| Insufficient evidence collected due to extended sick leave | V1/ME | Medical Emergency |
| No evidence provided in entry | V4 | Void 4 |
| Student took general assessment in a content area | V4 | Took General Assessment |
| Security breach occurred | V5 | Security Breach |

Void 1. Medical Emergency (ME):

When a student is out of school for an extended amount of time and not receiving instruction due to extensive sick leave or hospitalization, the portfolio may be eligible to receive a Void code 1 (medical emergency). The portfolio will be voided due to extended illness during the collection period. The student will receive a Void code 1 for each dimension and a “Medical Emergency” for the proficiency level will be displayed on the reports. Eligibility is based only on the following:

- If the student is receiving instruction for 10 days or less during a collection period, and
- The student has an extended hospitalization or leave due to illness and is not receiving instruction, and
- An official record documenting the student absences.

Void 4. No Evidence:

No entry evidence is provided in the portfolio. When entries are unscorable due to missing portfolio components, students will receive a Void 4 for their proficiency level.

A student transferred to New Jersey from out-of-state after October 31, 2013, is not required to submit portfolio evidence for scoring. These students will receive a Void 4 for their proficiency level. A Void 4 is also assigned if the student has excessive absences for non-medical reasons but meets the requirements as outlined for Void 1.

Void 4. Took General Assessment (NJ ASK, PARCC)

A student may not participate in both the APA and the statewide general assessment in the same content area. A student may participate in the APA in one or more content area(s) and the general assessment with accommodations in the other content area(s) or the APA in all content areas assessed. Students who took the general assessment in a content area will receive a Void code 4 and the result of the general assessment will be used for accountability reporting.

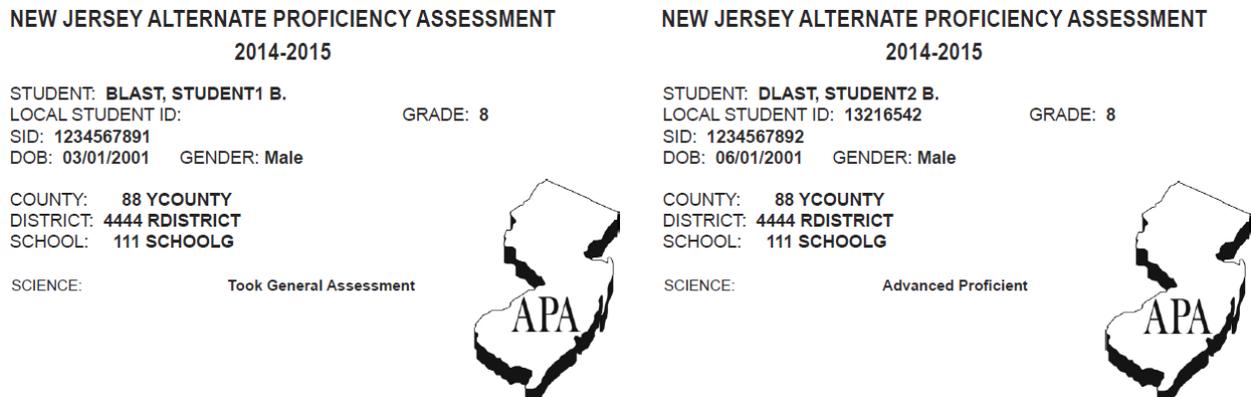
Void 5. Security Breach:

Breach of test security by a school or district. In this case the student report will reflect a Void code 5 for each dimension of the entry and a “Security Breach” for the proficiency level. If a security breach is detected in one content area, the entire portfolio (all content areas) is treated as a security breach and all results voided.

Student Sticker and Individual Student Report (ISR)

The Student Sticker (Figure 7.1) is produced by student name alphabetically. One sticker for each student within the school is provided. It is a peel-off label designed to be easily attached to the student’s permanent record. The Student Sticker is sent in printed format only to the Sending District or the School/District of Residence only. Receiving Districts do not receive Student Stickers.

Figure 7.1 Sample Student Stickers



The Individual Student Report (ISR) is a two-sided report showing specific student score information on the front of the ISR. A description of the APA and an interpretation of the scores are printed on the back. The school the student attends receives two printed copies of the ISR, whether it is a receiving school (private school for the disabled, special services school district, jointure commission, educational services commission, college-operated program, or state facility) or a school in the district of residence. Both the sending and attending schools will have the ability to download and print student ISRs.

It is the responsibility of the school the student attends to send a copy of the ISR to the child’s parent/guardian within 60 days of receipt. The sending school, if applicable,

receives one copy of the ISR. The district of residence also receives a copy of the ISR for review by the director of special education and the case manager.

Figure 7.2 presents the front of a student's sample report with demographic information and APA results. The proficiency levels in Science are shown in the top section. The scores for the Complexity, Performance, and Independence dimensions for every entry of the student's APA portfolio are provided on the lower half of the ISR. In addition, the maximum number of points obtainable per entry, for each dimension, is displayed in the parentheses below the dimension name for reference. The score data included for each rubric dimension assist in the identification of students' strengths and weaknesses.

Figure 7.3 shows the back of the ISR printed for all students. Information provided assists parents and educators with score interpretation.

Figure 7.2 Sample Individual Student Report (Grade 12 Front)



**New Jersey Statewide
Alternate Proficiency Assessment
2014–2015
Individual Student Report
Grade 12**

School Student Attends
CDS: 88-4444-333
County: YCOUNTY
District: RDISTRICT
School: SCHOOLT

Student Name: HLAST, STUDENTS

Title 1: S
SE: 02
LEP:
HB:

Date of Birth: 05/12/1997
Gender: Female
Local Student ID: 13216549
SID: 1234567552

Sending District
CDS: 88-4444-111
County: YCOUNTY
District: RDISTRICT
School: SCHOOLG

| Subject | Proficiency Level |
|---------|-------------------|
| Science | Proficient |

| Science | | Points Student Obtained | | |
|-----------------|--|-------------------------|---------------------|---------------------|
| | Strand | Complexity | Performance | Independence |
| Entry 1 | 5.5.A: Matter, Energy and Organization in Living Systems | 4 | 8 | 4 |
| Entry 2 | 5.5.B: Diversity and Biological Evolution | 3 | 7 | 3 |
| Entry 3 | 5.10.A: Natural Systems and Interactions | 4 | 7 | 4 |
| Entry 4 | 5.10.B: Human Interaction and Impact | 0 ^{EV-A} | 0 ^{EV-A} | 0 ^{EV-A} |
| Subtotal | | 11 out of 16 | 22 out of 32 | 11 out of 16 |
| Total | | 44 out of 64 | | |

EV-A – Type of evidence is not acceptable (media, data charts, observations)

Run Date 11/21/2014

Figure 7.3 Sample Individual Student Report (Back)

The **Alternate Proficiency Assessment (APA)** is a portfolio-based alternate assessment administered to students with the most significant cognitive disabilities who are unable to take the General Assessment. The APA is administered at every grade level at which a general statewide assessment is administered. The portfolio is a collection of student work samples that demonstrates how well students have learned the knowledge and skills covered by the New Jersey Core Curriculum Content Standards (CCCS), based on alternate achievement standards for their grade level in the content area of Science. The assessment measures a student's progress related to the CCCS, strands, grade-level Cumulative Progress Indicators (CPIs), and skill statements (called CPI Links).

The 2014-2015 APA was administered in Science to approximately 3948 students in grades 4 and 8, and in grades 9, 10, 11, or 12, in whichever year the student receives instruction in Biology.

HOW TO READ THIS REPORT

This **Individual Student Report (ISR)** represents the score results for the student. Scores are based solely on the information provided in each individual student portfolio; therefore, it may not be possible to compare the results earned by the student to other APA students or to students taking the general assessment. The ISR report is available only to parents, guardians, students, and authorized school personnel.

Additional information about score interpretation can be found on the New Jersey ServicePoint site at <https://nj-servicepoint.questarai.com>, click on *Documentation*, and refer to the Score Interpretation Manual. If you have any questions about the report or how to interpret the results, please contact the student's teacher, principal, or case manager.

The Student **Proficiency Level** (Advanced Proficient, Proficient, Partially Proficient), reported on the reverse side, is assigned based upon the student's total score earned across the four entries in each content area. The scores earned represent student knowledge and skills in each of the identified strands. Student work is scored for the following dimensions:

Complexity: The complexity dimension evaluates how closely the CPI Link assesses the CCCS CPI. The CPI Links vary by complexity and/or difficulty in relation to the CPI (Matched, Near, or Far).

Performance: The performance dimension evaluates the student's accuracy performing the skills represented in the CPI Links. The student's performance is documented by evidence of the student working on the CPI Link within the two collection periods in a school year (September 2, 2014–November 14, 2014, and December 8, 2014–February 13, 2015).

Independence: The independence dimension evaluates the extent to which the student completed the assessment tasks independently.

Portfolio requirements: A portfolio contains four entries per content area. Each entry is based on a specified standard and strand from the CCCS, and selected CPI and CPI Links. Requirements by content are:

- Science: Four entries
 - Grade 4: One strand each from standards 5.5 Characteristics of Life, 5.6 Chemistry, 5.8 Earth Science, and 5.9 Astronomy and Space Science
 - Grade 8: One strand each from standards 5.5 Characteristics of Life, 5.6 Chemistry, 5.7 Physics, and 5.9 Astronomy and Space Science
 - High School (Grade 9, 10, 11, or 12): Two different strands each from standards 5.5 Characteristics of Life and 5.10 Environmental Studies

Explanation of Zero Scores and Unscorable Entries:

An entry may be deemed unscorable or receive a score of "0" for a dimension in the event that certain errors occur in the production of evidence. Should the entry be unscorable across all three scoring dimensions, an Explanation Code will be assigned representative of the specific error that occurred; a "0" score will be used for calculation purposes in all fields containing an Explanation Code. Additional information about specific conditions resulting in unscorable entries is provided in the APA Score Interpretation Manual which can be obtained from the student's teacher or on the NJ ServicePoint website (<https://nj-servicepoint.questarai.com>).

An entry may also be deemed unscorable if there is a security breach, no evidence is provided, insufficient evidence is collected due to student on extended medical leave, or the student took the general assessment. Instead of a Proficiency Level, one of the following notations will appear:

- Medical emergency (indicating insufficient evidence due to extended sick leave)
- Void 4 (indicating no evidence for this subject or took General Assessment)
- Void 5 (indicating security breach occurred)

Run Date 04/16/2015

All Subjects Roster

The All Subjects Roster, as shown in Figure 7.4, provides a convenient method for reviewing students' complete APA results. An All Subjects Roster is generated for each grade level. Each report displays student names in alphabetical order (last name first) sorted by status. Users of this report can quickly determine how a particular student performed in Science (when applicable).

Receiving schools receive an All Subjects Rosters for all of the students who participated in the APA who are educated in that school. District schools receive an All Subjects Roster that includes the APA participant students who attend the school, those who live in the catchment area of the school but attend schools out of district, and those who attend a program within the school but reside in another school district.

Figure 7.4 Sample All Subjects Roster

CDS: 88-444-111
 County: YCOUNTY
 District: RDISTRICT
 School: SCHOOLG
 Page: 1 OF 1

**New Jersey Statewide Assessment System
 Alternate Proficiency Assessment
 2014-2015
 All Subject Roster
 Grade 8**



| STUDENT NAME DATE OF BIRTH | SID | Status ^a | Ethnicity | TITLE 1 | ED | Migrant | SE | LEP | TIS | TID | Gender | PROFICIENCY LEVEL |
|----------------------------------|------------|---------------------|-----------|---------|----|---------|----|-----|-----|-----|--------|-------------------------|
| BLAST, STUDENT1 B. 03/01/1999 | 1234567891 | 1 | W | | | | 06 | | | | M | Took General Assessment |
| DLAST, STUDENT2 B. 06/01/1999 | 1234567892 | 1 | W | S | | | 02 | | | | M | Advanced Proficient |
| FLAST, STUDENT3 K. 07/01/2000 | 1234567893 | 1 | H | | | | 08 | | | | M | Proficient |
| GLAST, STUDENT4 N. 12/01/1999 | 1234567894 | 1 | W/P | | | | 03 | | | | M | Partially Proficient |
| ILAST, STUDENT6 M. 11/01/1999 | 1234567896 | 1 | W | | | | 08 | | | | M | Proficient |
| RLAST, STUDENT7 J. 06/01/1999 | 1234567897 | 2 | W | S | | | 02 | | | | M | Security Breach |
| MLAST, STUDENT5 A. 10/01/1999 | 1234567895 | 3 | W | S | | | 08 | | | | F | Took General Assessment |
| SLAST, STUDENT8 M. 07/01/1999 | 1234567898 | 3 | W | | F | | 08 | | | | F | Advanced Proficient |

^a

| | |
|---|--|
| } | 1 = Student was assessed at school of residence. |
| | 2 = Student was sent outside school of residence for instruction and assessment. |
| | 3 = Student was received from another school for instruction and assessment. |
| | 4 = No scorable evidence. |

Note: All names and data are fictional and are for illustrative purposes only.

Run Date: 11/12/2014

Student Roster

Student Rosters for Science are produced for each grade level assessed. Students' names are listed in descending order by proficiency level. Figure 7.5 shows an example of the Student Roster – Science for Grade 8. Following a student's identification information, the student's proficiency level is given. These scores enable the program staff to identify strengths and weaknesses across students within the content area. Voided portfolio content areas are noted, where applicable.

Receiving schools receive Student Rosters that include all of the students who participated in the APA who attend that school.

Sending schools or the Schools of Residence receive Student Rosters that include the students participating in the APA who attend that school, those who live in the school catchment but attend a school out of district, and those who attend a program within the school but reside in another school district.

Summary of Performance – School, District

General Information. Two types of summary performance reports are generated, one at the district level and one at the school level. For each grade, a Summary of District Performance report is generated for each district. Within the district, for each grade level, a Summary of School Performance report is generated for each school. These reports provide summary statistics for each subject assessed. Summary reports are produced for public schools and districts only. Summary reports reflect data for students who were sent out of district, as well as students remaining in the district.

Summary reports are not available for receiving districts. A sample Summary of District Performance report is shown in Figure 7.6.

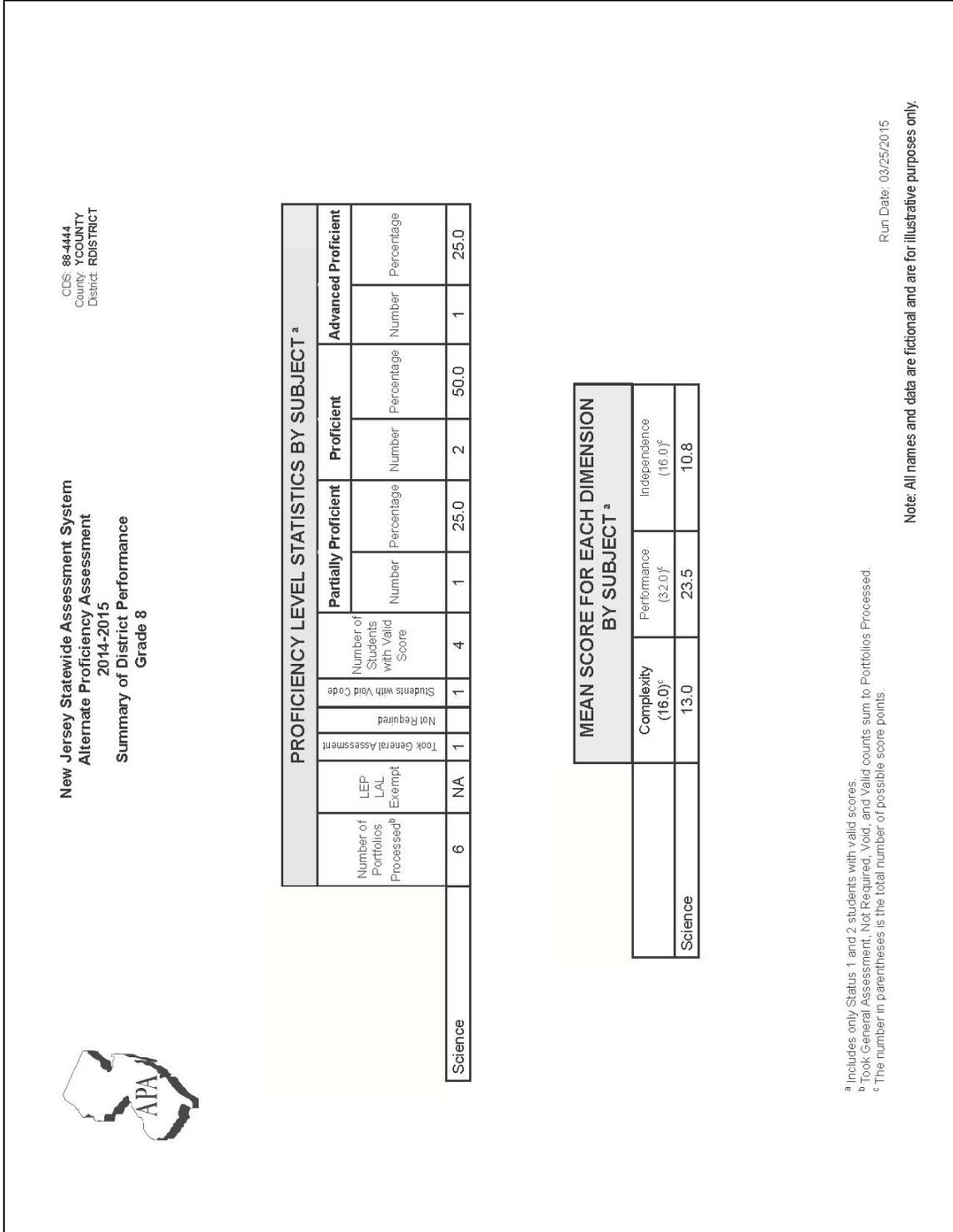
District and School Identification Information. This part of the report includes the names and code numbers of the county and district (and school for school summary).

Proficiency Level Statistics by Subject. This part of the report provides the number and percent of students in each proficiency level for Science at either the district or school level. The following summary is provided for each subgroup shown on the report:

- Number of portfolios processed (sum of GA, NR, Void, and Valid)
-
- Number of students that took the general assessment (NJ ASK or PARCC)
- Number of students not required to submit entries for the content area (also includes high school students who did not take the Biology course)
- Number of students with Void Codes (Security Breach, Medical Emergency, and V4 due to no content evidence in the portfolio).
- Number of students with valid scores
- Number of students in each proficiency level (number is based on students with valid scores.)
- Percent of students at each proficiency level (number is based on students with valid scores.)

Mean Score for Each Dimension by Subject. School or district means for each dimension are provided by subject (content area) based on students with valid scores.

Figure 7.6 Sample Summary of District Performance



Performance by Demographic Groups

The Performance by Demographic Groups report summarizes overall performance by student demographic subgroups: Total, LEP Status, Gender, Ethnicity, Economic Status (Disadvantaged vs. Not Disadvantaged), and Migrant Status. These group reports provide additional performance information that can be used to make adjustments to curricula that may better serve these students. Both sending and attending districts and schools will receive this report.

The Performance by Demographic Groups reports are produced at school and district levels by grade for reporting. The district level report presents aggregated data for the district. The school-level report shows school-level data. They are distinguished by report title. If a district has only one school in which the test was administered, the summary data will be identical in both the district report and the school report. State-level data is produced with the state summary reports which will be posted on the New Jersey Department of Education website. At the state level, reports are also produced by District Factor Groups, Charter Schools (DFG-R), Non-Special Needs Districts, and Special Needs Districts and are distinguished by report title.

This one-page report includes performance data for Science (where administered). The percentage of students who fall into each of the three proficiency levels is based on the number of valid scores. This report does not disaggregate the data at the dimension level. Figure 7.7 shows an example of a District Performance by Demographic Groups report.

Figure 7.7 Sample District Performance by Demographic Groups

CDS: 88-4444
 County: YCOUNTY
 District: RDISTRICT

**New Jersey Statewide Assessment System
 Alternate Proficiency Assessment
 2014-2015
 District Performance by Demographic Groups
 Grade 8**



| | Number of Portfolios Processed | Took General Assessment | | Number of Students with Valid Score | | Science ^a | | |
|------------------------------------|--------------------------------|-------------------------|----------|-------------------------------------|------------------------|----------------------|------------------|--|
| | | Not Required | Required | Number of Students with Valid Score | % Partially Proficient | % Proficient | % Adv Proficient | |
| TOTAL | 6 | 1 | 1 | 4 | 25.0 | 50.0 | 25.0 | |
| LEP Status^b | | | | | | | | |
| LEP (Current & Former) | 0 | | | | | | | |
| Current LEP | 0 | | | | | | | |
| Former LEP | 0 | | | | | | | |
| Non-LEP | 6 | 1 | 1 | 4 | 25.0 | 50.0 | 25.0 | |
| Gender^d | | | | | | | | |
| Female | 0 | | | | | | | |
| Male | 6 | 1 | 1 | 4 | 25.0 | 50.0 | 25.0 | |
| Ethnicity^e | | | | | | | | |
| White | 4 | 1 | 1 | 2 | 0.0 | 50.0 | 50.0 | |
| Black | 0 | | | | | | | |
| Asian | 0 | | | | | | | |
| Pacific Islander | 0 | | | | | | | |
| Hispanic ^f | 1 | 0 | 0 | 1 | 0.0 | 100.0 | 0.0 | |
| American Indian/Alaska Native | 0 | | | | | | | |
| Other ^g | 1 | 0 | 0 | 1 | 100.0 | 100.0 | 0.0 | |
| Economic Status^h | | | | | | | | |
| Disadvantaged | 0 | | | | | | | |
| Non-Disadvantaged | 6 | 1 | 1 | 4 | 25.0 | 50.0 | 25.0 | |
| Migrant Statusⁱ | | | | | | | | |
| Migrant | 0 | | | | | | | |
| Non-Migrant | 6 | 1 | 1 | 4 | 25.0 | 50.0 | 25.0 | |

^aExcludes Status 3 students. Students are included in Total only once, but they appear in each applicable category. Percentages may not total 100 due to rounding.
^bIncludes students coded Medical Emergency.
^cDifferences in totals among demographic categories resulted from gridding errors or missing data in materials received from districts.
^dIncludes students coded Hispanic with or without other ethnic affiliations.
^eIncludes students coded with more than one ethnicity (none of these ethnicities were Hispanic), or their ethnicity was not provided by district.
^fFor the LEP category in this report, Total Portfolios processed is the sum of Non-LEP and Current LEP Portfolios.

Run Date: 02/12/2015
Note: All names and data are fictional and are for illustrative purposes only.

District Student Data

Districts of residence and sending districts will receive student level data files of their students electronically. Files in fixed-width ASCII and Excel formats are created for districts and schools with ten or more students. When necessary, a CD-ROM may be requested.

State Summary

A state summary data file and an Executive Summary will be completed based on the reporting data and posted on the NJDOE website before the end of the calendar year (<http://www.nj.gov/education/schools/achievement/>). The data file, available in text and Excel formats, contains the same type of test results as in the Performance by Demographic report. Due to the small size of the APA population, the APA reports the state summary at the state level only. The Executive Summary is included in Appendix D.

7.2 Parent Letter

To help explain to parents and guardians both the purpose of the APA and the information provided on the Individual Student Report (ISR), a sample form letter is included (Figure 7.8) that can be adapted, signed, photocopied, and sent home with each student along with his/her ISR.

Additionally, a Parent Information Guide was produced in Fall 2014 that addresses common questions about the APA. It is available on Questar's website (https://nj-servicepoint.questarai.com/NJxx01_Documentation_2014_2015).

Figure 7.8 Sample Parent/Guardian Letter

Dear Parent/Guardian:

Your child's Individual Student Report for the New Jersey Alternate Proficiency Assessment (APA) in Science is attached. The APA is a portfolio assessment that consists of a collection of student work which was gathered by your child's teachers during instructional activities. Your child participated in the APA between September 2, 2014, and February 13, 2015. Your child's APA portfolio was then submitted to the New Jersey Department of Education and scored by trained readers during the spring of 2015. The attached report provides your child's APA scores in the content area assessed. The 2014-2015 APA assesses Science only. The alternate assessment for English Language Arts and Mathematics transitioned to a new assessment system called Dynamic Learning Maps (DLM) under a different contractor.

The report tells you the proficiency levels your child achieved on the skills assessed for Science. A level of "proficient" or "advanced proficient" is considered meeting the state standard for the APA. The boxes below the proficiency levels show the scores for each "dimension" scored for each Science standard assessed by the portfolio. Please refer to the back of the Individual Student Report for further information regarding these boxes.

APA results should not be used as the sole basis for instructional decisions. It is important that districts consider multiple measures on all students before making decisions about the student's instructional placement.

This report is available only to parents, guardians, students, and authorized school officials. If your child attends a school outside of this district, reports are sent to the home school district, your child's neighborhood school, and the school your child attends. All reports are kept confidential. If you have any questions about the report, please contact _____ (district contact name / case manager / teacher / the principal of the school) at _____ (phone number) for assistance.

7.3 Quality Control of Reporting

Quality control procedures are an integral part of Questar's Software Development Lifecycle (SDLC). Questar's SDLC, which is employed for software and application development, involves software project planning and tracking, requirements management, software development, software quality assurance, and software configuration and release management. A few examples of Questar's documentation include the Statement of Work, Master Schedule, Project Plan, Functional Specifications, Design Review Document, Quality Assurance (QA) Test Plan, Requirements Traceability Matrix, and Release Management Plan. Questar's SDLC is influenced by Software Engineering Institute's (SEI) Capability Maturity Model (CMM) for software development process management and control, and Questar continues to strive to continually revisit the process and compare it against industry best practices to ensure quality and accuracy of our products.

After software requirements have been identified, the project team led by the Business Analyst prepares the project plan and functional specifications. Then, the development team reviews the functional specifications and prepares design documents. In addition, unit test plans are created by the development team. A unit test plan is a list of specific modular tests they run to make sure the building blocks of their code are working according to specifications. The purpose of the code-unit test process is to ensure that software is developed, maintained, documented, and verified to meet the project requirements for coding and unit testing. As such, the process provides the mechanisms that are necessary to implement the software requirements and design in order to catch errors early on the process. This approach firmly exemplifies our commitment to quality through the whole SDLC. At the same time, the QA Team reviews the requirements and creates their QA Test Plans and Requirements Traceability Matrix. The QA Test Plan focuses on System and Integrated Testing. Also, when necessary, test plans are created for Performance and Load Testing.

After all modules (units) are developed and unit tested, the application requires system and integration tests. These tests ensure that all the application modules work together and that outputs from one module match up to the proper inputs for the next module in the system. These types of tests validate data quality and correctness across all the modules. Through the traceability matrix, it ensures each requirement is tested and expected results to ensure that all requirements have been met. As the need arises, we also conduct performance and load tests to ensure that the application conforms to performance requirements, which are defined by benchmarks under specific load conditions. It is important that these tests be performed by a group that is independent of the software development team. Also, note that these tests are performed in an environment different from the development domain. These processes allow independent verification and interpretation of the requirements. Once the independent QA group has completed the test and given its approval, the system is moved into production mode.

Scanning and Scoring

Before actual documents are machine-scanned, a comprehensive check of the scanning and scoring system is performed. The software development tester creates test decks of gridded scanned documents with specific test criteria. The test decks are designed and gridded to cover all response ranges, ID ranges, blanks, and multiple grids, as well as any other responses used by the APA. A file containing the scanned responses is then compared to the expected test results for each document to ensure the scanner is operating correctly. The test decks are processed through the programs for scanning and editing scanned, and packetizing and printing scoring monitors. The second check involves processing and quality-checking the first actual scanned documents received.

As described in the rangefinding section of Part 4, the NJDOE Office of State Assessments asked districts to return their portfolios early following testing so actual portfolios could be used for rangefinding. Some early return portfolios and additional portfolios received during the scheduled return served a quality-control purpose beginning with hand checking and following with periodical checking throughout scoring.

For both the rangefinding and quality-control purposes, portfolios were selected to represent the following:

- range of school districts
- different types of schools
- grade level of students (elementary, middle, high school)
- skill level (access skill, modified expectation)
- severity of disability (severe/profound, moderate, mild-moderate)
- possible score levels (low, medium, high)

Quality Control of Score Reporting

NJDOE Office of State Assessments conducted a quality control of score reporting in June 2015. Questar scored a sample of portfolios from a variety of students across grades and content areas.

Questar printed all applicable reports for 8–10 districts that met requirements specified by the Office of State Assessments for quality control. Requirements for the selected districts included the following:

- All grades in at least 2 districts
- Each grade represented at least 4 times across the districts
- 3 urban districts, at least 1 private school
- 4–6 public districts (non-specialized districts)
- 4 private districts such as the Department of Children and Families (DCF) districts
- No more than 50 students in a district (multiple schools)

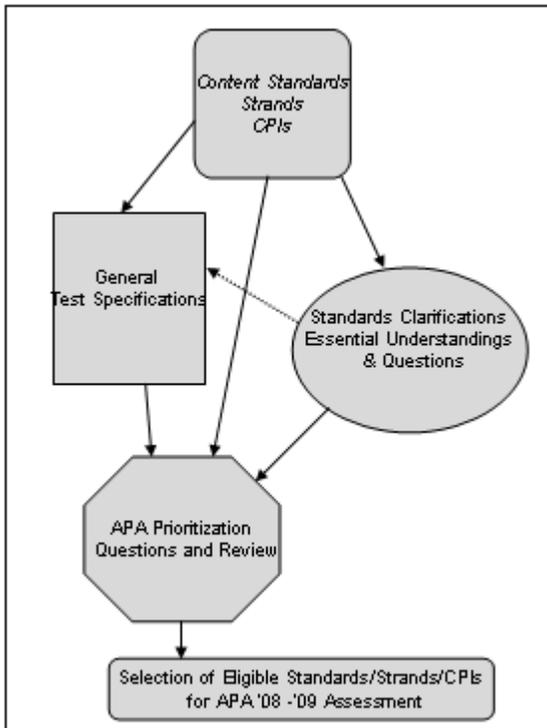
- Sending/receiving relationship and Status: some related districts through sending/receiving relationship (e.g., at least, Status 2 and Status 3), minimum of 3 sets. A minimum of 2 districts should be “independent” (e.g., with Status 1 only)

Additionally, the quality-control requirements included these student demographics:

- Migrant: 3–4 students
- SE: As many different codes as possible (including N-unknown or multiple).
- T-I: 3–4 cases
- Economically Disadvantaged: 3-4 students
- LEP: 3–4 cases of each code (<, 1, 2, 3, F1, F2, and Y).
- Home: 3–4 homebound students
- Homeless: 3–4 homeless students
- Ethnicity: 3–4 cases (of all codes, including multiple-codes)
- TIS/TID: 3 cases at minimum of TID only and both TIS and TID.
- Void: At least 3 cases per code (V1, V4, V5); some must have dimension scores for one entry
- Report Footnote: Every case of each footnote (including unscorable codes)
- General Assessment: Several cases of students whose scan sheet indicated they took the general assessment
- 4th Rater: Several cases requiring a fourth reader, with resolution information provided.
- Breach: 3–4 cases.
- Grade Changes: 3–4 cases.
- Student Not Assessed: 3–4 cases.

For the NJDOE quality control, Questar provided the demographic scan sheets, scoring monitors, record changes printout, and school names with CDS codes.

APPENDIX A: Development of the CPI Links



Content Standards Strands CPIs

- Source document for instruction and assessment
- Describes what all students should know and be able to do, including students with disabilities
- A scope and sequence document is available to assist in planning for instruction

January 14, 2008 APA Redesign Flow
Charland Process 2

APA Prioritization Questions and Review

- The process of defining the eligible standards/strands/CPIs for APA assessment requires
 - A review of the intersection of the standards from the test specs and standards clarifications and
 - A prioritization of the remaining available standards based on the APA student population and a series of questions.
 - *The questions should help to define what is most important to assess. This process should not exclude strands based on the belief of what may not need to be or is not currently taught.*

January 14, 2008 APA Redesign Flow
Charland Process 5

Educators will consider these questions and statements:

1. APA students must receive standards-based instruction that is linked to grade-level and must be held to high expectations.
2. Which of the strands and CPIs are essential for students to master?
3. Which of the strands and CPIs are very desirable for students to master?
4. Which of the strands and CPIs are desirable for students to master?
5. Which strands and CPIs support learning of higher level skills?
6. Which strands and CPIs promote instruction of foundational skills that will prepare students for future learning?

January 14, 2008 APA Redesign Flow
Charland Process 6

Selection of Eligible Standards/Strands/CPs
for APA '08 -'09 Assessment

- The ILSSA group has produced a first draft of the eligible standards, strands, and CPs eligible for APA assessment.
- The Advisory committee will review the draft considering the prioritization questions, content standards, scope and sequence, and other reference documents. Committee will revise draft if necessary and document their justification of revisions.
- DOE content experts will review the committee's product and revise if necessary.
- A committee of educators will review the final draft before publication.

January 14, 2008

APA Redesign Flow
Charland Process

7

APPENDIX B: Performance Level Descriptors

Performance Level Descriptors Science

Grade 4 Science

Partially Proficient

Fourth grade students performing at the partially proficient level may require prompting, modifications and/or additional supports while recalling knowledge and demonstrate emerging skills in characteristics of life, chemistry, earth science and astronomy with inconsistent performance. Partially proficient students will typically use fewer categories to:

- Identify matter, energy and organization in living systems
- Identify physical properties and changes of matter
- Identify components of the water cycle and states of water in the Earth's system
- Identify components and their sequence within the Earth, Moon and Sun system

Partially proficient students will sometimes demonstrate the ability to identify vocabulary, collect and record data and make a few connections to their real-life experiences.

Proficient

Fourth grade students performing at the proficient level may require some prompting, modifications and/or additional supports while recalling knowledge and demonstrating skills in characteristics of life, chemistry, earth science and astronomy with increased performance. Proficient students will typically be able to:

- Classify and/or sequence matter, energy and organization in living systems
- Classify, compare, and/or describe physical properties and changes of matter
- Sequence and/or order the water cycle, describe states of water in the Earth's system
- Describe, illustrate and/or demonstrate an understanding of the sequence and order within the Earth, Moon and Sun system

Proficient students will frequently demonstrate the ability to comprehend vocabulary, use data to draw conclusions and make connections to the real-world.

Advanced Proficient

Fourth grade students performing at the advanced proficient level will demonstrate the qualities outlined for the proficient student. They may require minimal prompting, modifications and/or additional supports while applying vocabulary, knowledge and skills to explain the characteristics of life, chemistry, earth science and astronomy with a high-level of performance. Advanced proficient students will typically be able to perform skills such as: make predictions, observe, collect data, draw conclusions and make inferences relating to the real-world.

Grade 8 Science

Partially Proficient

Eighth grade students performing at the partially proficient level may require prompting, modifications and/or additional supports while recalling knowledge and demonstrate emerging skills in characteristics of life, chemistry, physics and astronomy with inconsistent performance. Partially proficient students will typically use fewer categories to:

- Identify organisms based upon the diversity of their characteristics. Identify characteristics best suited for survival in a particular environment.
- Identify physical changes and chemical reactions
- Identify types of energy and types of energy transformations
- Identify objects and/or the physical characteristics of the planets and other objects within the Solar system

Partially proficient students will sometimes demonstrate the ability to identify vocabulary, collect and record data and make a few connections to their real-life experiences.

Proficient

Eighth grade students performing at the proficient level may require some prompting, modifications and/or additional supports while recalling knowledge and demonstrating skills in characteristics of life, chemistry, physics and astronomy with increased performance. Proficient students will typically be able to:

- Classify organisms based upon the diversity of their characteristics. Describe the biological evolution of organisms.
- Classify, compare, and/or describe examples of physical changes and chemical reactions
- Classify, illustrate and/or describe types of energy and types of energy transformations
- Compare and/or classify the physical characteristics of the planets and other objects within the Solar system

Proficient students will frequently demonstrate the ability to comprehend vocabulary, use data to draw conclusions and make connections to the real-world.

Advanced Proficient

Eighth grade students performing at the advanced proficient level will demonstrate the qualities outlined for the proficient student. They may require minimal prompting, modifications and/or additional supports while applying vocabulary, knowledge and skills to explain the characteristics of life, chemistry, physics and astronomy with a high-level of performance. Advanced proficient students will typically be able to perform skills such as: make predictions, observe, collect and analyze data, draw conclusions and make inferences relating to the real-world.

High School NJBCT

Partially Proficient

High School Biology students performing at the partially proficient level may require prompting, modifications and/or additional supports while recalling knowledge and demonstrate emerging skills in characteristics of life and environmental studies with inconsistent performance. Partially proficient students will typically use fewer categories to:

- Identify the components involved in photosynthesis and their role in the energy cycle of life
- Identify the process of evolution by natural selection. Identify the impact of inherited traits and the environment on natural selection.
- Identify the impact of human actions and/or naturally occurring processes on the environment
- Identify the ways human actions impact the ecosystems

Partially proficient students will sometimes demonstrate the ability to identify vocabulary, collect and record data and make a few connections to their real-life experiences.

Proficient

High School Biology students performing at the proficient level may require some prompting, modifications and/or additional supports while recalling knowledge and demonstrating skills in characteristics of life and environmental studies with increased performance. Proficient students will typically be able to:

- Describe the process of photosynthesis and its role in the energy cycle of life.
- Describe the process of evolution by natural selection. Describe the impact of inherited traits and the environment on natural selection.
- Describe, compare and/or contrast the impact of human actions versus naturally occurring processes on the environment
- Use data to assess the impact of human actions on the ecosystems

Proficient students will frequently demonstrate the ability to comprehend vocabulary, use data to draw conclusions and make connections to the real-world.

Advanced Proficient

High School Biology students performing at the advanced proficient level will demonstrate the qualities outlined for the proficient student. They may require minimal prompting, modifications and/or additional supports while applying vocabulary, knowledge and skills to explain the characteristics of life and topics in environmental studies with a high-level of performance. Advanced proficient students will typically be able to perform skills such as: make predictions, observe, collect and analyze data, support conclusions and make inferences relating to the real-world.

**APPENDIX C: Consistency between APA Portfolio Scorers by Individual Grade
and Entry**

Consistency between APA Portfolio Scorers by Entry – Grade 4

| | Entry 1 | | | Entry 2 | | | Entry 3 | | | Entry 4 | | |
|----------------|---------|------------|---------|---------|------------|---------|---------|------------|---------|---------|------------|---------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | | | | |
| Complexity | 97.8 | 1.8 | 2.2 | 98.3 | 1.6 | 1.7 | 96.7 | 2.8 | 3.4 | 99.1 | 0.8 | 0.9 |
| Performance | 96.1 | 2.2 | 3.9 | 98.0 | 1.1 | 2.0 | 97.6 | 1.1 | 2.5 | 96.6 | 2.2 | 3.4 |
| Independence | 98.1 | 1.3 | 1.9 | 98.7 | 0.8 | 1.3 | 98.4 | 1.0 | 1.8 | 97.7 | 1.3 | 2.3 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension.

Consistency between APA Portfolio Scorers by Entry – Grade 8

| | Entry 1 | | | Entry 2 | | | Entry 3 | | | Entry 4 | | |
|----------------|---------|------------|---------|---------|------------|---------|---------|------------|---------|---------|------------|---------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | | | | |
| Complexity | 98.4 | 0.9 | 1.6 | 98.8 | 1.0 | 1.2 | 99.1 | 0.7 | 0.9 | 97.4 | 2.0 | 2.6 |
| Performance | 96.9 | 1.6 | 3.1 | 95.7 | 2.1 | 4.3 | 96.9 | 2.2 | 3.2 | 97.1 | 1.7 | 2.9 |
| Independence | 97.9 | 0.9 | 2.1 | 97.5 | 1.2 | 2.5 | 98.7 | 0.5 | 1.3 | 98.7 | 0.5 | 1.3 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension.

Consistency between APA Portfolio Scorers by Entry – Grade 9

| | Entry 1 | | | Entry 2 | | | Entry 3 | | | Entry 4 | | |
|----------------|---------|------------|---------|---------|------------|---------|---------|------------|---------|---------|------------|---------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | | | | |
| Complexity | 98.3 | 1.7 | 1.7 | 97.2 | 1.1 | 2.8 | 98.3 | 1.7 | 2.3 | 97.2 | 2.3 | 2.8 |
| Performance | 98.3 | -- | 1.7 | 98.9 | 0.6 | 1.1 | 97.7 | 1.1 | 2.3 | 94.3 | 2.3 | 5.7 |
| Independence | 98.3 | -- | 1.7 | 99.4 | 0.6 | 0.6 | 98.9 | 0.6 | 1.1 | 97.7 | 1.7 | 2.3 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension.

Consistency between APA Portfolio Scorers by Entry – Grade 10

| | Entry 1 | | | Entry 2 | | | Entry 3 | | | Entry 4 | | |
|----------------|---------|------------|---------|---------|------------|---------|---------|------------|---------|---------|------------|---------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | | | | |
| Complexity | 98.8 | 0.8 | 1.2 | 97.3 | 0.4 | 2.7 | 100.0 | -- | -- | 98.8 | 1.2 | 1.2 |
| Performance | 98.5 | 1.5 | 1.5 | 97.7 | 0.8 | 2.3 | 97.3 | 1.5 | 2.7 | 98.1 | 1.2 | 1.9 |
| Independence | 98.8 | 0.4 | 1.2 | 98.8 | 1.2 | 1.2 | 99.2 | -- | 0.8 | 99.6 | -- | 0.4 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension.

Consistency between APA Portfolio Scorers by Entry – Grade 11

| | Entry 1 | | | Entry 2 | | | Entry 3 | | | Entry 4 | | |
|----------------|---------|------------|---------|---------|------------|---------|---------|------------|---------|---------|------------|---------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | | | | |
| Complexity | 98.7 | 1.0 | 1.5 | 97.9 | 0.7 | 2.2 | 99.4 | 0.6 | 0.6 | 99.0 | 0.9 | 1.0 |
| Performance | 96.1 | 1.8 | 4.0 | 97.9 | 1.0 | 2.1 | 96.9 | 1.5 | 3.1 | 98.5 | 0.9 | 1.5 |
| Independence | 98.4 | 0.6 | 1.8 | 98.8 | 0.6 | 1.3 | 98.7 | 0.9 | 1.3 | 99.4 | 0.4 | 0.6 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension.

Consistency between APA Portfolio Scorers by Entry – Grade 12

| | Entry 1 | | | Entry 2 | | | Entry 3 | | | Entry 4 | | |
|----------------|---------|------------|---------|---------|------------|---------|---------|------------|---------|---------|------------|---------|
| | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* | % Exact | % Adjacent | % Res.* |
| Science | | | | | | | | | | | | |
| Complexity | 98.8 | 1.2 | 1.2 | 100.0 | -- | -- | 100.0 | -- | -- | 100.0 | -- | -- |
| Performance | 98.8 | -- | 1.2 | 96.5 | 3.5 | 3.5 | 100.0 | -- | -- | 96.5 | 2.3 | 3.5 |
| Independence | 98.8 | -- | 1.2 | 98.8 | 1.2 | 1.2 | 100.0 | -- | -- | 98.8 | 1.2 | 1.2 |

*Complexity, Performance, and Independence Dimensions—If the first two scores are not equal, then a third reader must score the dimension.

APPENDIX D: 2015 Executive Summary

2015 New Jersey Alternate Proficiency Assessment

Executive Summary

The New Jersey Alternate Proficiency Assessment (APA) is a portfolio assessment designed to measure progress toward achieving New Jersey's state educational standards for students with the most significant cognitive disabilities who are unable to participate in the general assessments: New Jersey Assessment of Skills and Knowledge (NJASK), the Partnership for Assessment of Readiness for College and Career (PARCC), or New Jersey Biology Competency Test (NJBCT).

The APA was developed for two purposes:

- To measure the progress of a small percentage of students with the most significant cognitive disabilities who cannot participate in the regular statewide assessments even with accommodations.
- To ensure that the educational results for all students are included in the statewide accountability system at the individual, school, district, and state levels.

Accountability through assessment provides equity in program and educational opportunities for all students. Alternate assessment ensures an inclusive statewide assessment system and student accountability linked to the common core of learning within the general curriculum in New Jersey.

The APA was designed and developed to meet the requirements of the *Individuals with Disabilities Education Act of 1997 (IDEA '97)*, *Individuals with Disabilities Education Improvement Act of 2004 (IDEA '04)*, and the *No Child Left Behind Act of 2001 (NCLB)*.

NCLB requires that all students, including those with disabilities, participate in the state assessment program. NCLB also requires that the measurement of progress toward meeting state standards includes assessment results for all students.

The APA fulfills these requirements and is based on the New Jersey Core Curriculum Content Standards (NJ CCCS) in the content areas of Language Arts Literacy, Mathematics, and Science. In this manner, all students in New Jersey are moving toward the same general standards with whatever modifications or supports they need.

The 2014–2015 APA was administered in Science in grades 4 and 8 and in grade 9, 10, 11, or 12 depending on the grade in which a student received Biology instruction. Evidence of student performance as demonstrated in the student portfolio was collected during two collection periods from September 2, 2014, through November 14, 2014, and December 8, 2014, through February 13, 2015. A portfolio is a collection of student work samples that measure a student's progress related to the NJ CCCS, strands, grade-level cumulative progress indicators (CPIs), and skill statements called CPI Links.

Extensive APA information is available at the ServicePoint website provided by Questar Assessment, Inc. (Questar), the current APA vendor, at https://nj-servicepoint.questarai.com/NJxx01_Documentation.aspx.

For the *New Jersey Core Curriculum Content Standards*, see <http://www.nj.gov/njded/cccs>.

The 2015 APA state summary reports appear at <http://www.state.nj.us/education/schools/achievement/>.

Test Design

Peer reviewers from the U.S. Department of Education (USED) assisted the New Jersey Department of Education (NJDOE) in designing the current version of the APA by providing test design and administration recommendations. These recommendations included the following:

- APA students must be assessed on a subset of skills from the general assessment. The skills must be mapped to the general assessment specifications and address the breadth and depth of skills tested across grade levels.
- The skills assessed must link to the CPIs of the student's assigned grade level.
- Students in the same grade must be assessed on the same content; teachers choose from a limited selection of standards and strands to assess their students.
- Strengthen the alignment of the APA program design to grade-level academic content and progress indicators.

In accordance with these recommendations, the APA is developed using test specifications, by grade and content area, that prescribe the standards and strands that must be assessed. Test specifications were written in order to provide more guidance on how to link to grade-level CPIs and to address the federal requirement of linkage to the skills tested in the general assessments. Specifying the requirements increases standardization of the assessment for students with significant cognitive disabilities. Students may not be assessed in functional, behavioral, or access (social, motor, etc.) skills. Functional activities and materials might be used to promote understanding during instruction, but the evidence and activities demonstrating student achievement for assessment must be academically focused and represent the entire grade-level CPI Link.

Test specifications for the 2014–2015 Science APA administration are provided below. The specific standards to be assessed differ by grade.

- **Science:** Four entries
 - Grade 4: One strand each from standards 5.5, 5.6, 5.8, and 5.9
 - Grade 8: One strand each from standards 5.5, 5.6, 5.7, and 5.9
 - High School Biology (grade 9, 10, 11, or 12): Two different strands (A and B) each from standards 5.5 and 5.10

The CPI Links were developed from a subset of the NJ CCCS, strands, and CPIs. The subset was prioritized for assessment on the APA by Inclusive Large Scale Standards for

Assessment (ILSSA) content specialists, NJDOE content specialists, New Jersey special education teachers and general education teachers, and the APA advisory committee. Individuals from each of these areas were also involved in drafting the content in the CPI Links and ensuring its alignment to the NJ CCCS. Each CPI Link offers three levels of connection to each CPI: Matched Link, Near Link, and Far Link. Educators choose one CPI Link per entry and use that as the basis for developing portfolio entries for assessment within the APA.

New test standards should be set whenever a testing procedure is adopted that is judged to be meaningfully different from previous testing procedures. A standard setting for the redesigned APA, administered operationally for the first time in 2008–2009, was conducted from June 9–12, 2009, to describe and delineate the thresholds of performance that are indicative of APA Partially Proficient, Proficient, and Advanced Proficient performance for Language Arts Literacy and Mathematics in grades 3–8 and 11, and for Science in grades 4, 8, and high school (grade 9, 10, 11, or 12). Results from the standard setting studies were used to formulate recommendations to the Commissioner of Education and the New Jersey State Board of Education for the adoption of the cut scores (i.e., proficiency levels). Subsequently, in late June and early July of 2009, the standard setting panels' recommendations were reviewed by the senior staff in the Office of State Assessments and the Office of Special Education Programs, the Assistant Commissioner for the Division of Student Services, the Deputy Commissioner, and the Commissioner. The review led to some modifications to the panels' recommended cut scores, chiefly affecting the advanced proficient cut points. These cut scores were presented to the State Board of Education on July 15, 2009, and approved unanimously.

Scoring Process

The entries of the APA portfolio are scored based on three dimensions:

- The **Complexity** Dimension is used to evaluate the CPI Link assessed and how closely the complexity and difficulty (Matched, Near, Far) links to the NJ CCCS and grade-level CPI.
- The **Independence** Dimension is used to evaluate the extent to which the student completed the assessment items independently.
- The **Performance** Dimension is used to evaluate the student's accuracy when performing skills represented in the CPI Links.

Complexity is the expectation level at which the student should perform the skill (remembering, understanding, applying, analyzing, evaluating and creating). Difficulty involves the number of concepts, skills, or ideas on which the student will be working or the type of adaptations and supports in place. Performance measures how well the student has demonstrated the skill specified in the CPI Link within the collection periods.

To score the portfolios, trained expert scorers used a scoring rubric designed to measure student performance on the skill, the level of independence when performing the skill, and the relationship of the skill to the grade level cumulative progress indicator.

A proficiency classification for each content area is derived by combining the scores of the three dimensions. Performance contributes twice as many points as Complexity and Independence to the total score. Each content area assessed receives a proficiency level. The three proficiency levels are:

Advanced Proficient exceeded the level of proficiency

Proficient met the state level of proficiency

Partially Proficient is below the state minimum level of proficiency.

Scores are reported by content area. Entries that do not meet the APA requirements are reported as “0’s” along with an unscorable code. Of the required four entries for a content area, only one scorable entry is required to assign a proficiency level. If the portfolio contains only one scorable entry within a content area, the total score and proficiency level for that content area are reported based on the dimension scores of that entry.

The proficiency level classification allows the APA results to be combined with the general assessment results for accountability purposes as required by USED.

It is important to recognize that the APA system does not report scale scores. The data provided are the key components to interpreting the portfolio results. The APA scores are based solely on the information provided in the individual portfolio submitted. Therefore, it may not be possible to compare these scores to other APA students and students taking the general assessments. Scale scores are not appropriate for use for the APA system so there are no issues of equating involved. There are no sets of test items; therefore, there are no item difficulties, nor is there a need to equate test scores from year to year.

This executive summary includes four tables derived from the statewide summary for the 2015 APA in Science. The state summary data file and the state level Performance by Demographic Group reports are produced and posted on the NJDOE website. The Performance by Demographic Group reports show additional columns including the number of portfolios processed and the percentages of students who scored at the Partially Proficient, Proficient, and Advanced Proficient level. Values are suppressed and an asterisk is printed when the number of students with valid scores for a particular group is greater than zero but 10 or less, or when the percentage for Partially Proficient in a group is greater than 90%.

Table 1 in this executive summary provides the number of participating APA students with valid scores in Science and the percent of students at each APA proficiency level. The percentages may not total to 100.0% due to rounding.

As seen in the Table 1 summary data, 3,957 students were evaluated by the 2015 APA in Science. Of these, 3,656 students had valid Science scores. Science was assessed in grades 4 and 8 and for high school in grade 9, 10, 11, or 12 if the student was enrolled in a Biology course.

Table 2 presents the grade level performance by demographic groups for Science. Results are presented for the total student group and the following demographic variables: limited English proficient (LEP) status, gender, ethnicity, economic status, and migrant status. These tables show the number of students with valid scores and the percentage of students who scored at or above Proficient on their portfolios. This percentage, the students in Proficient or Advanced Proficient, was calculated by subtracting the percentage of students in Partially Proficient from 100.

Students are counted in the state total only once but are counted in as many other categories that apply. Some students might not be included in a gender group because of incomplete or missing information. Students with only one ethnic code are reported in the appropriate ethnic group. Examiners were asked to code all categories applicable to indicate a student's ethnicity. Students with multiple ethnic codes or no ethnic code (unspecified) are counted in the category called "Other." LEP is reported as LEP (Current plus Former) with two subcategories: Current LEP and Former LEP.

The demographic information originates from the data collected on the APA Student Demographic Information Forms (SDIFs) submitted for the students by school districts. Demographic information was reviewed by the school district personnel prior to reporting, allowing them an opportunity to correct any errors.

Highlights from the 2015 APA Performance Results

Tables 2 presents the number of students with valid scores and the percentage of Science APA students who scored at or above Proficient on their portfolios in the tested grade levels. Total results are summarized as follows:

Science:

- Grade 4 – 52.1
- Grade 8 – 52.2
- Grade 9 – 34.7
- Grade 10 – 43.5
- Grade 11 – 50.9
- Grade 12 – 46.5

For high school, Science was assessed in grade 9, 10, 11, or 12 depending on the grade in which a student received Biology instruction. The greatest number of high school students with valid scores was 691 students in grade 11 (as shown in Table 1). Since much smaller numbers of students took Science in grades 9, 10, and 12, the discussion is limited to the grade 11 group. If there were no students associated with a particular subgroup, an n-count of 0 is provided and % At or Above Proficient is left blank.

LEP Status Approximately 1.5% of the APA test-taking population was classified as Limited English Proficient (LEP). For the following summary of LEP students' performance, LEP is defined as current and former LEP students combined. In Science, only grade 4 had more than 10 students

taking the APA. Most LEP students were current LEP students rather than former LEP students. Of the 18 grade 4 LEP students, 61.1% were classified as at or above proficient.

Gender

Approximately 2 to 2.5 times as many male students took the Science APA as female students. This ratio had a decreasing trend from grades 4 and 8 to grade 11, decreasing from 70.5% in grade 4 to 65.0% in grade 11.

Across grades 4, 8, and 11, the percentages of female students and male students scoring at or above Proficient were similar. The largest difference was at grade 11 with 53.6% of female students and 49.4% of male students scoring at or above Proficient. In grades 4 and 8, the percentage of students scoring at or above Proficient was greater for male students compared to female students. In grade 11, percentages were higher for female students.

Ethnicity

The highest and lowest n-counts, in consideration of valid portfolios, associated with Science varied as follows:

| | |
|-----------------|---|
| White | 554 students in grade 8 to 319 students in grade 11 |
| Black | 322 students in grade 8 to 160 students in grade 11 |
| Asian | 109 students in grade 4 to 52 students in grade 11 |
| Hispanic | 315 students in grade 4 to 125 students in grade 11 |

Since 10 or fewer students were associated with the Pacific Islander, American Indian/Alaskan Native, and other ethnic groups (some grades had more than 10 students for this category, but the n-counts were small), data for these groups were not reported. (Values are suppressed and an asterisk is printed when the number of students with valid scores for a particular group is greater than zero but 10 or less.) If there were no students associated with a particular sub-group, an n-count of 0 is provided and % At or Above Proficient is left blank.

Within a given grade level for Science, moderate differences in ethnic group performance were observed. The difference between the highest and lowest performing ethnic group, in terms of percentage of students Proficient or above, ranged from 11.4% in grade 4 to 16.1% in grade 11. The average difference across grades 4, 8, and 11 was approximately 13%.

White students did not have the highest or lowest percentage of students classified as Proficient or above for any grade. Black students had the highest percentage of students classified as Proficient or above for grade 11 and the lowest percentage of students classified as

Proficient or above for grade 4. Asian students had the highest percentage of students classified as Proficient or above for grades 4 and 8. Hispanic students had the lowest percentage of students classified as Proficient or above for grades 8 and 11.

For grade 4, the percentage of students scoring at or above the Proficient level ranged from 46.4% of Black students to 57.8% of Asian students. For grade 8, the percentage ranged from 45.9% of Hispanic students to 58.7% of Asian students. For grade 11, the percentage ranged from 40.8% of Hispanic students to 56.9% of Black students.

Economic Status The number of students with valid scores indicates that the average percentage of economically disadvantaged for grades 4, 8, and 11 was approximately 37%.

The percentage of economically disadvantaged students scoring at or above Proficient was greater than the percentage of non-disadvantaged students scoring at or above Proficient for grade 4. The greatest difference was at grade 11 with 44.0% of the economically disadvantaged and 54.4% of the non-economically disadvantaged students scoring at or above Proficient. The smallest difference in performance was observed in grade 8 with 50.8% of economically disadvantaged students and 53.0% of the non-economically disadvantaged students scoring at or above Proficient. The average difference in performance across grades, with respect to the percentage of students classified as Proficient or above, was approximately 6%.

Migrant Status Only non-migrant data appear in this report. Since 10 or fewer migrant students took the Science APA in each grade, data are suppressed for student confidentiality. If there were no students associated with a particular sub-group, an n-count of 0 is provided and % At or Above Proficient is left blank.

Reporting Rules for APA State Summary

In order to safeguard student confidentiality, certain information is suppressed in the state summary files according to the following reporting rules:

- Data are not reported where the number of students with valid scores for a particular group is greater than zero but 10 or less.
- Data are not reported when it is otherwise possible to identify individual student performance.

Table 1
2015 New Jersey Alternate Proficiency Assessment
Number of Valid Scores and Percent of Students at Each APA Proficiency Level

| Grade | Number of Portfolios Processed | Science | | | |
|------------|--------------------------------|------------------------|---------------|---------|--------------|
| | | Number of Valid Scores | % Part. Prof. | % Prof. | % Adv. Prof. |
| 4 | 1,299 | 1,193 | 47.9 | 51.7 | 0.4 |
| 8 | 1,383 | 1,273 | 47.8 | 41.1 | 11.1 |
| 9* | 188 | 176 | 65.3 | 31.8 | 2.8 |
| 10* | 282 | 260 | 56.5 | 39.6 | 3.8 |
| 11* | 691 | 668 | 49.1 | 48.8 | 2.1 |
| 12* | 114 | 86 | 53.5 | 44.2 | 2.3 |
| All Grades | 3,957 | 3,656 | 49.7 | 45.5 | 4.8 |

*In 2014–2015, the APA assessed Science in grades 9, 10, 11, or 12 depending on the grade in which a student received Biology instruction.

Table 2
2015 New Jersey Alternate Proficiency Assessment
Statewide Performance by Demographic Groups
Science

| | Grade 4 | | Grade 8 | | Grade 9 | | Grade 10 | | Grade 11 | | Grade 12 | |
|------------------------|--------------------------------------|--------------------------|--------------------------------------|--------------------------|--------------------------------------|--------------------------|--------------------------------------|--------------------------|--------------------------------------|--------------------------|--------------------------------------|--------------------------|
| | Number of Students with Valid Scores | % At or Above Proficient | Number of Students with Valid Scores | % At or Above Proficient | Number of Students with Valid Scores | % At or Above Proficient | Number of Students with Valid Scores | % At or Above Proficient | Number of Students with Valid Scores | % At or Above Proficient | Number of Students with Valid Scores | % At or Above Proficient |
| STATE TOTAL | 1,193 | 52.1 | 1,273 | 52.2 | 176 | 34.7 | 260 | 43.5 | 668 | 50.9 | 86 | 46.5 |
| LEP STATUS | | | | | | | | | | | | |
| LEP (Current & Former) | 18 | 61.1 | * | * | * | * | * | * | * | * | * | * |
| Current LEP | 14 | 50.0 | * | * | * | * | * | * | * | * | * | * |
| Former LEP | * | * | * | * | 0 | | 0 | | 0 | | 0 | |
| Non-LEP | 1,179 | 52.2 | 1,270 | 52.2 | 175 | 34.3 | 256 | 43.0 | 667 | 51.0 | 85 | 47.1 |
| GENDER | | | | | | | | | | | | |
| Female | 352 | 52.0 | 378 | 50.5 | 53 | 28.3 | 77 | 40.3 | 233 | 53.6 | 31 | 38.7 |
| Male | 841 | 52.2 | 895 | 52.8 | 123 | 37.4 | 183 | 44.8 | 435 | 49.4 | 55 | 50.9 |
| ETHNICITY | | | | | | | | | | | | |
| White | 495 | 54.9 | 554 | 55.1 | 83 | 42.2 | 112 | 34.8 | 319 | 53.6 | 35 | 54.3 |
| Black | 250 | 46.4 | 322 | 50.9 | 39 | 17.9 | 65 | 52.3 | 160 | 56.9 | 16 | 43.7 |
| Asian | 109 | 57.8 | 80 | 58.7 | * | * | 20 | 65.0 | 52 | 46.2 | * | * |
| Pacific Islander | * | * | * | * | 0 | | 0 | | * | * | 0 | |
| Hispanic | 315 | 51.4 | 279 | 45.9 | 46 | 37.0 | 59 | 42.4 | 125 | 40.8 | 29 | 37.9 |
| Amer.Indian/AK Native | 0 | | * | * | 0 | | * | * | * | * | 0 | |
| Other | 20 | 35.0 | 33 | 51.5 | 0 | | * | * | * | * | * | * |
| ECONOMIC STATUS | | | | | | | | | | | | |
| Disadvantaged | 508 | 55.3 | 461 | 50.8 | 74 | 31.1 | 101 | 47.5 | 225 | 44.0 | 33 | 39.4 |
| Non-Disadvantaged | 685 | 49.8 | 812 | 53.0 | 102 | 37.3 | 159 | 40.9 | 443 | 54.4 | 53 | 50.9 |
| MIGRANT STATUS | | | | | | | | | | | | |
| Migrant | 0 | | * | * | * | * | 0 | | 0 | | 0 | |
| Non-Migrant | 1,193 | 52.1 | 1,272 | 52.2 | 173 | 34.1 | 260 | 43.5 | 668 | 50.9 | 86 | 46.5 |

*Values are suppressed for student counts greater than zero but 10 or less, or when the percentage for Partially Proficient in a group is greater than 90%.

APPENDIX E: 2015 Frequency Tables of Proficiency Levels by Disability Category⁵

⁵ Beginning in 2013–2014, the “Social Maladjustment” disability category was no longer reported and is therefore not included in these tables.

Proficiency Level Distribution by Disability Category – Grade 4

| | Science | | | |
|-------------------------------------|---------------------|------------|----------------------|-------|
| | Advanced Proficient | Proficient | Partially Proficient | Total |
| Auditorily Impaired | -- | 2 | 2 | 4 |
| Autistic | 4 | 279 | 253 | 536 |
| Cognitively Impaired | 1 | 58 | 85 | 144 |
| Communication Impaired | -- | 61 | 33 | 94 |
| Deaf-Blindness | -- | -- | 1 | 1 |
| Emotionally Disturbed | -- | 3 | -- | 3 |
| Multiply Disabled | -- | 158 | 159 | 317 |
| Orthopedically Impaired | -- | -- | -- | -- |
| Other Health Impaired | -- | 31 | 20 | 51 |
| Specific Learning Disability | -- | 18 | 10 | 28 |
| Traumatic Brain Injury | -- | 5 | 5 | 10 |
| Visually Impaired | -- | 1 | 1 | 2 |
| Blank or Multiple Grid | -- | 1 | 2 | 3 |
| Total | 5 | 617 | 571 | 1,193 |

Proficiency Level Distribution by Disability Category – Grade 8

| | Science | | | |
|-------------------------------------|---------------------|------------|----------------------|-------|
| | Advanced Proficient | Proficient | Partially Proficient | Total |
| Auditorily Impaired | 1 | 2 | -- | 3 |
| Autistic | 57 | 216 | 252 | 525 |
| Cognitively Impaired | 20 | 77 | 94 | 191 |
| Communication Impaired | 13 | 25 | 11 | 49 |
| Deaf-Blindness | 1 | -- | -- | 1 |
| Emotionally Disturbed | -- | 3 | 4 | 7 |
| Multiply Disabled | 39 | 165 | 213 | 417 |
| Orthopedically Impaired | -- | -- | 1 | 1 |
| Other Health Impaired | 3 | 16 | 15 | 34 |
| Specific Learning Disability | 6 | 11 | 7 | 24 |
| Traumatic Brain Injury | -- | 8 | 10 | 18 |
| Visually Impaired | 1 | -- | 2 | 3 |
| Blank or Multiple Grid | -- | -- | -- | -- |
| Total | 141 | 523 | 609 | 1,273 |

Proficiency Level Distribution by Disability Category – Grade 9

| | Science | | | |
|-------------------------------------|---------------------|------------|----------------------|-------|
| | Advanced Proficient | Proficient | Partially Proficient | Total |
| Auditorily Impaired | -- | -- | -- | -- |
| Autistic | 1 | 17 | 25 | 43 |
| Cognitively Impaired | 1 | 6 | 37 | 44 |
| Communication Impaired | -- | 4 | 8 | 12 |
| Deaf-Blindness | -- | -- | -- | -- |
| Emotionally Disturbed | -- | 1 | 2 | 3 |
| Multiply Disabled | 2 | 20 | 33 | 55 |
| Orthopedically Impaired | -- | -- | -- | -- |
| Other Health Impaired | -- | 3 | 5 | 8 |
| Specific Learning Disability | -- | 2 | 5 | 7 |
| Traumatic Brain Injury | 1 | 3 | -- | 4 |
| Visually Impaired | -- | -- | -- | -- |
| Blank or Multiple Grid | -- | -- | -- | -- |
| Total | 5 | 56 | 115 | 176 |

Proficiency Level Distribution by Disability Category – Grade 10

| | Science | | | |
|-------------------------------------|---------------------|------------|----------------------|-------|
| | Advanced Proficient | Proficient | Partially Proficient | Total |
| Auditorily Impaired | -- | -- | -- | -- |
| Autistic | 2 | 30 | 38 | 70 |
| Cognitively Impaired | 3 | 16 | 27 | 46 |
| Communication Impaired | -- | 5 | 15 | 20 |
| Deaf-Blindness | -- | -- | 1 | 1 |
| Emotionally Disturbed | -- | -- | 1 | 1 |
| Multiply Disabled | 2 | 42 | 38 | 82 |
| Orthopedically Impaired | -- | -- | 3 | 3 |
| Other Health Impaired | 1 | 3 | 7 | 11 |
| Specific Learning Disability | 1 | 4 | 15 | 20 |
| Traumatic Brain Injury | 1 | 3 | 2 | 6 |
| Visually Impaired | -- | -- | -- | -- |
| Blank or Multiple Grid | -- | -- | -- | -- |
| Total | 10 | 103 | 147 | 260 |

Proficiency Level Distribution by Disability Category – Grade 11

| | Science | | | |
|-------------------------------------|---------------------|------------|----------------------|-------|
| | Advanced Proficient | Proficient | Partially Proficient | Total |
| Auditorily Impaired | 1 | -- | 3 | 4 |
| Autistic | 3 | 102 | 110 | 215 |
| Cognitively Impaired | 4 | 47 | 57 | 108 |
| Communication Impaired | 1 | 7 | 5 | 13 |
| Deaf-Blindness | -- | -- | -- | -- |
| Emotionally Disturbed | -- | -- | -- | -- |
| Multiply Disabled | 3 | 155 | 133 | 291 |
| Orthopedically Impaired | -- | -- | -- | -- |
| Other Health Impaired | 1 | 8 | 6 | 15 |
| Specific Learning Disability | -- | 4 | 8 | 12 |
| Traumatic Brain Injury | 1 | 3 | 2 | 6 |
| Visually Impaired | -- | -- | -- | -- |
| Blank or Multiple Grid | -- | -- | 4 | 4 |
| Total | 14 | 326 | 328 | 668 |

Proficiency Level Distribution by Disability Category – Grade 12

| | Science | | | Total |
|-------------------------------------|---------------------|------------|----------------------|-------|
| | Advanced Proficient | Proficient | Partially Proficient | |
| Auditorily Impaired | -- | -- | -- | |
| Autistic | 2 | 13 | 8 | 23 |
| Cognitively Impaired | -- | 11 | 17 | 28 |
| Communication Impaired | -- | -- | 2 | 2 |
| Deaf-Blindness | -- | -- | -- | -- |
| Emotionally Disturbed | -- | -- | -- | -- |
| Multiply Disabled | -- | 11 | 17 | 28 |
| Orthopedically Impaired | -- | 1 | -- | 1 |
| Other Health Impaired | -- | 1 | 1 | 2 |
| Specific Learning Disability | -- | 1 | -- | 1 |
| Traumatic Brain Injury | -- | -- | 1 | 1 |
| Visually Impaired | -- | -- | -- | -- |
| Blank or Multiple Grid | -- | -- | -- | -- |
| Total | 2 | 38 | 46 | 86 |

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Baker, E.L., & Linn, R.L. (2002) Validity issues for accountability systems. Center for the Study of Evaluation. Technical Report 585, Los Angeles, CA.
- Browder, D.M., & Spooner, F. (2006). Teaching language arts, math, and science to students with significant cognitive disabilities. Baltimore, MD: Paul H. Brookes Publishing Co.
- Browder, D.M., Wakeman, S.Y., Flowers, C., Rickelman, R.J., Pugalee, D., & Karvonen, M (2007). Creating access to the general curriculum with links to grade-level content for students with significant cognitive disabilities: An explication of the concept. *The Journal of Special Education*, 41(1), 2–16.
- Clayton, J., Burdge, M., Denham, A., Kleinert, H.L., & Kearns, J. (2006). A four-step process for accessing the general curriculum for students with cognitive disabilities. *Teaching Exceptional Children*, 38(5), 20–27.
- Flowers, C., Wakeman, S.Y., Browder, D.M., & Karvonen, M. (2009). Links for academic learning (LAL): A conceptual model for investigating alignment of alternate assessments based on alternate achievement standards. *Educational Measurement: Issues and Practice*. 28(1), 25–37.
- Kleinert, H.L., & Kearns, J.F. (2001) *Alternate assessment: Measuring outcomes and supports for students with disabilities*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5–8.
- New Jersey Department of Education (2008). *New Jersey Alternate Proficiency Assessment (APA) 2008–2009 Procedures Manual*. New Jersey.
- U.S. Department of Education. (Revised December 21, 2007, to include modified academic achievement standards. Revised with technical edits January 12, 2009.) *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: Author. www.ed.gov/policy/elsec/guid/saaprguidance.pdf.