New Jersey Assessment of Skills and Knowledge

2015

TECHNICAL REPORT

Science
Grades 4 and 8

February 2016
PTM: 1510.64

STATE BOARD OF EDUCATION

MARK W. BIEDRON ..... Hunterdon
President

JOSEPH FISICARO ..... Burlington
Vice President

ARCELIO APONTE .....Middlesex

RONALD K. BUTCHER ..... Gloucester

JACK FORNARO .....Warren

EDITHE FULTON ..... Ocean

ERNEST P. LEPORE ..... Hudson

ANDREW J. MULVIHILL ..... Sussex

J. PETER SIMON .....Morris

DOROTHY STRICKLAND..... Essex

DAVID C. HESPE, Commissioner of Education

NJ ASK SCIENCE 2015 GRADES 4 and 8 TECHNICAL REPORT

TABLE OF CONTENTS

NJ ASK 2015 GRADES 4&8 TECHNICAL REPORT
TABLES and FIGURES

PART 1: INTRODUCTION

The purpose of this Technical Report is to provide information about the technical characteristics of the 2015 administration of the New Jersey Assessment of Skills and Knowledge (NJ ASK) for Science at grades 4 and 8. This report is intended for use by those who evaluate tests, interpret scores, or use test results for making educational decisions. It includes the following sections: test development, test administration, scoring, standard setting, item and test statistics, equating and scaling, reliability, validity, and score reporting.

This report provides extensive detail about the development and operation of NJ ASK. The traditional concerns with a program are often labeled reliability and validity. The empirical reliability and validity of the assessments are reported explicitly in this document. While reliability (Part 8) is relatively straightforward, the steps in creating the program and putting it into operation are all aspects of validity (Part 9). The validity of any assessment stems from the steps taken in planning it; the processes of developing the tests' content; the processes of consulting with stakeholders; the processes of communicating with users about the test; the processes of scoring and reporting; and the processes of data analysis and appropriate uses of outcomes. Each is an integral part of validity.

Data for the analyses presented in this Technical Report were collected during the spring administration in May 2015. The short time duration between test administration and score reporting necessitated the use of a priority sample for the equating/scaling analyses presented in Part 7 – Equating and Scaling. A priority sample consists of a sub-group (approximately 30%) of the entire state student population that contains a representative sample of students from across the state based on ethnicity, gender and District Factor Group (DFG), a measure of socioeconomic status (see Section 6.5). The answer documents from the selected priority sample are scored and prioritized such that the results from this group are available for score-reporting-timeline-driven-analyses. The entire student population test results were utilized in less time-sensitive analyses such as those reported in Part 6 – Item and Test Statistics and in Part 8 – Reliability. The student N-counts are provided for each analysis in order for the reader to quickly ascertain whether the total student population or a sub-group was used for a given analysis.

In reading this technical report, it is critical to remember that the testing program does not exist in a vacuum; it is not just a test. It is one part of a complex network intended to help schools focus their energies on dramatic improvement in student learning. NJ ASK is an integrated program of testing, accountability, and curricular and instructional support. It can only be evaluated properly within this full context. Detailed descriptions of the NJ ASK 2015 Science are provided in Sections 2.2 and 2.3.

1.1 Description of the Assessment

The NJ ASK Science was administered as an operational assessment in spring 2015 to New Jersey students in grades 4 and 8. These assessments fulfill the requirements under the 2001 No Child Left Behind Act (NCLB) for each state to assess science at least once during grades 3–5 and grade 6–9. (Prior versions of NJ ASK assessed English Language Arts and mathematics. However, in 2015 the NJ DOE switched the assessment of those subjects to the Partnership for Assessment of Readiness for College and Careers (PARCC).)

In 2008, grades 5 through 8 assessments were redesigned as NJ ASK 5–8. Grades 5 through 7 of this new ASK 5–8 replaced the interim ASK 5–7 administered in 2006 and 2007. For grade 8, ASK 8 replaced the Grade Eight Proficiency Assessment (GEPA), marking 2007 as the last GEPA administration; however, the ASK 8 science test design remains unchanged from GEPA. In 2009, ELA and mathematics assessments in grades 3 and 4 were also redesigned.

New Jersey's statewide assessments of science currently include the following components, with versions in both English and Spanish for NJ ASK:

Elementary School:

- Grade 4 New Jersey Assessment of Skills and Knowledge (NJ ASK)

Middle School:

- Grade 8 New Jersey Assessment of Skills and Knowledge (NJ ASK)

High School:

- New Jersey Biology Competency Test (NJBTC)

The NJ ASK Science scores at grades 4 and 8 are reported as scale scores, with score ranges as follows:

- Partially Proficient 100–199
- Proficient 200–249
- Advanced Proficient 250–300

The scores of students who are included in the Partially Proficient level are considered to be below the state minimum of proficiency, and those students may be most in need of instructional support.

1.2 Purpose of the Assessment

As a result of the NCLB requirements, New Jersey established statewide science assessments in grades 4, 8, and high school. The statewide assessments for grades 4 and 8 are administered annually as the New Jersey Assessment of Skills and Knowledge (NJ ASK). High school testing is administered via the New Jersey Biology Competency Test (NJBTC). Testing is conducted in the spring of each year to allow school staff and students the greatest opportunity to achieve the goal of Proficiency.

Schools and districts should use the results to identify strengths and weaknesses in their educational programs. This process is designed to improve instruction and foster better alignment with the New Jersey science standards. The results may also be used, along with other indicators of student progress, to identify those students who may need instructional support in any of the content areas. This support, which could be in the form of individual or programmatic intervention, would be a means to address any identified knowledge or skill gaps.

1.3 Organizational Support

New Jersey's Office of State Assessments (OSA) coordinates the development and implementation of NJ ASK Science. In addition to planning, scheduling, and directing all NJ ASK activities, the staff is extensively involved in numerous test design, item and statistical reviews, security, quality-assurance, and analytical procedures. Measurement Incorporated (MI), the contractor for NJ ASK Grades 4 and 8, is responsible for all aspects of the testing program, including activities such as program management, development of test materials (test items, test booklets, answer documents, and ancillary materials), and psychometric support, including standard setting. MI's other activities include enrollment verification; distribution of all materials; receiving, scanning, editing, and scoring the answer documents; scoring constructed-response items; and creating, generating, and distributing all score reports of test results to students, schools, districts, and the state.

PART 2: TEST DEVELOPMENT

A directory of test specifications and sample items was developed for each science content area. These specifications describe the test, format of the items, and the scores to be generated by the test. The material in the test specifications is designed for use by curriculum specialists and teachers to improve instruction at the district, school, and classroom levels. This document serves as the foundation for all test item development.

2.1 Test Specifications

The 2015 NJ ASK was designed to measure the knowledge and skills identified in the 2004 revision of the New Jersey Core Curriculum Content Standards (CCCS) for science in grades 4 and 8. The following tables provide information about item type, content cluster/standards, and total point value by test section. Table 2.1.1 summarizes the total points possible for each of the content areas of the operational NJ ASK administered in 2015 for grades 4 and 8. Table 2.1.2 shows the number of items by content cluster/standard and skill, where appropriate. An in-depth discussion of the composition of the science assessments can be found in the 2009 NJ ASK Technical Report (PTM 1507-34), Part 2, Section 2.1.

Table 2.1.1: NJ ASK Science 2015 Total Points Possible by Content Area

Science	Grade 4	Grade 8
Total	**39 points**	**54 points**
Life Science	15	20
Physical Science	12	17
Earth Science	12	17
Knowledge	*4*	*6*
Application	*35*	*48*

Table 2.1.2: NJ ASK Science 2015 Number of Items by Content Cluster and Skill

Science**	Grade 4		Grade 8	
Skill	**A**	**K**	**A**	**K**
Life Science	13	2	18	2
Physical Science	9	1	13	2
Earth Science	9	1	13	2
Total	**31**	**4**	**44**	**6**

**K = Knowledge, A = Application

Test Blueprints

The following tables outline the test construction blueprints. The actual test map for each grade and content area for the NJ ASK Science 2015 is included. The NJ ASK Science assessment includes Life, Physical, and Earth Sciences. Each multiple choice item is worth one point; each constructed response item is worth up to three points. Each constructed response item is scored using an item-specific rubric.

Table 2.1.3: Test Construction Blueprint for NJ ASK 4 and 8 Science

		Grade 4	Grade 8
Item Count by Type (does not include field test content)	MC	33	48
	CR	2	2
Total raw score points possible		**39**	**54**
Approximate total testing time (includes field test content)		60 min.	120 min.

Table 2.1.4: Actual Test Map for 2015 Grade 4 Science NJ ASK

Cluster	Cog/Prob	MC (1 pt.)	CR (3 pts.)	# of Items	# of Points
Earth	Application	8	1	9	11
	Knowledge	1	0	1	1
Earth Total		**9**	**1**	**10**	**12**
Life	Application	13	0	13	13
	Knowledge	2	0	2	2
Life Total		**15**	**0**	**15**	**15**
Physical	Application	8	1	9	11
	Knowledge	1	0	1	1
Physical Total		**9**	**1**	**10**	**12**
Grand Total		**33**	**2**	**35**	**39**

Table 2.1.5: Actual Test Map for 2015 Grade 8 Science NJ ASK

Cluster	Cog/Prob	MC (1 pt.)	CR (3 pts.)	# of Items	# of Points
Earth	Application	12	1	13	15
	Knowledge	2	0	2	2
Earth Total		**14**	**1**	**15**	**17**
Life	Application	18	0	18	18
	Knowledge	2	0	2	2
Life Total		**20**	**0**	**20**	**20**
Physical	Application	12	1	13	15
	Knowledge	2	0	2	2
Physical Total		**14**	**1**	**15**	**17**
Grand Total		**48**	**2**	**50**	**54**

2.2 Development of Test Items

The NJ ASK consists of two types of items:

1. Operational items used to determine students' scores.
2. Field-test items evaluated for use as future base test items.

In the item development process, MI developed test and item specifications based upon requirements of the NJ CCCS for science in grades 4 and 8. Details regarding the item development process can be found in the 2009 NJ ASK Technical Report (PTM 1507-34), Part 2, Section 2.2.

Test Form Distribution

Before spring of 2008, the NJ DOE developed items for the NJ ASK using a standalone field-test format. Beginning with the operational administration in spring of 2008, the NJ DOE began embedding field-test items for ELA, mathematics, and science. Thus, twenty-four forms of the NJ ASK 2015 assessments were distributed to New Jersey schools. Each of the 24 test forms at each grade level included identical base test (or operational) items as well as a semi-unique set of field-test items. Note that students earned scores only on operational items. The 24 field-test forms were assigned to school districts such that each district had one and only one test form, except in the case of unusually large districts (i.e., Jersey City, Newark, and Patterson), which received two forms. Moreover, the field-test forms were distributed across ethnic groups and DFG classifications, such that each group or classification was represented across each form. Finally, approximately equal numbers of students (approximately 4,500) were given each test form. Tables showing the final form distribution plan by test form, grade, and DFG classification can be found in Appendix A.

Information regarding the Item Review Process, Item Use, Test Forms Assembly, and Quality Control for Test Construction can also be found in the 2009 NJ ASK Technical Report (PTM 1507-34), Part 2, Sections 2.3 through 2.6, respectively.

PART 3: TEST ADMINISTRATION

Great care is taken to ensure the standard administration of the NJ ASK. Close attention to details is necessary to ensure that students taking the test in different locations have equal opportunities for success. Information about the administration of NJ ASK is available in the Test Coordinator Manual That information is not fully replicated here, but the following elements are of importance to this technical report.

3.1 Participation

State regulations require that all students be included in the statewide assessment program and assessed annually. This includes limited English proficient (LEP) students and students with disabilities. In school year 2001–2002, students with severe cognitive disabilities were administered the Alternative Proficiency Assessment (APA) for the first time statewide.

All public schools, including those without assessed grades, are counted in the state's accountability system. All schools without assessed grades are counted as one unit with their respective receiving schools. This helps ensure closer vertical alignment of instructional services. In addition, special education students served in proprietary schools are counted in the sending schools' accountability results, which ensure that placement decisions are reviewed closely at the school and district level for optimum student academic performance.

New Jersey does not include in the accountability system the results of any student enrolled less than one full academic year in a school for school accountability or in a district for district accountability. This does not exclude from a district's accountability the results of those students who transfer from one school to another within a district.

3.2 Test Security Procedures

The NJ ASK test booklets and their contents are treated as secure materials. Detailed procedures for maintaining the security of test materials while they are in the districts are outlined in the *New Jersey Assessment of Skills & Knowledge Spring 2015 Test Coordinator Manual Grades 4 and 8*. It is the responsibility of the district to guarantee the security of the test materials. Examiners, proctors, and other school personnel are prohibited from copying, reading, discussing, or disclosing any test items before, during, or after test administration. When not being used during a test period, test materials are stored in a secure, locked location that is accessible only to individuals whose access is authorized by the school test coordinator. Inventory forms track test materials as they move from one location to another in districts.

As part of the test development procedures, "breach" test forms and examiner manuals are prepared in the event of a security breach. If the NJ DOE identifies a security breach during the test administration window, MI immediately removes the NJ ASK test materials from the involved district or school. The test booklets for the content area affected are coded with a void code indicating a security breach. If the NJ DOE determines that there was enough time for testing, the breach forms are delivered to the district and the test is administered to the affected students in the content area impacted by the security breach. For students re-tested during the test administration window, scores are reported based on the breach form. If a

security breach is identified after the testing window, the impacted test booklets are coded with a security breach void code and no test results are reported for that content area. However, students receive a score for the content area not impacted by the security breach.

3.3 Test Administration Procedures

Detailed instructions for administering the NJ ASK are provided in the *New Jersey Assessment of Skills & Knowledge Spring 2015 Test Coordinator Manual Grades 4&8.* The NJ ASK 4 and 8 was administered according to the following schedule:

Table 3.3.1: NJ ASK 2015 grades 4 and 8 Science Testing Window

Grade	Test Dates		Testing Time (minutes)*
	Regular testing	Make-up testing	Day 4/Day 5
Grade 4	5/27/15	5/28/15	60
Grade 8	5/27/15	5/28/15	120

*Does not include administrative time but does include field-test time.

Testing was not to be scheduled immediately after an athletic event or an assembly. All test schedules were checked with the appropriate school officials to ensure that other school activities did not interfere with the test administration. Other test administration procedures included:

- All testing had to be scheduled in the morning. Exceptions included homebound and bedside students, as well as students attending out-of-district placements who were tested at that placement by staff from the student's home district.
- The district and school test coordinators (DTCs/STCs) were responsible for scheduling times and places for regular and make-up testing and for ensuring that all testing was completed according to the procedures and schedule described in the *Test Coordinator Manual* and in the *Examiner Manual*.
- Students who were required to test but were absent for the regular test administration had to be tested on the make-up dates.
- Students whose answer folders were voided during testing were considered to have attempted the test section. They were not allowed to retake or resume taking the voided test section during the make-up.
- Students who began a section of the test and did not complete it during the specified testing time were not allowed to complete the test section during the make-up period or any other time unless additional time was specified in their IEP or 504 plan.

3.4 Test Accommodations

To ensure that students are tested under appropriate conditions, the Department of Education has adopted test accommodations and modifications that may be used when testing special populations of students. The content of the test typically remains the same, but administration

procedures, setting, and answer modes may be adapted. Students requiring accommodations must be tested in a separate location from general education students.

General education students receive no special testing accommodations other than the standard room setup and materials distribution described in the examiner's section of the Test Coordinator Manual.

Limited English Proficient (LEP) students who do not take the Spanish form of the test are tested with one or more of these accommodations:

- Additional time up to 150% of the administration times indicated
- Translation of directions only to the student's native language
- Use of a bilingual dictionary, preferably one normally used by the student as part of the instructional program.

Translations of passages, items, prompts, and tasks are NOT permitted.

Students with Disabilities (SE/504) must take the NJ ASK unless their Individualized Education Program (IEP) specifically states that they take the Alternate Proficiency Assessment (APA) and not the NJ ASK.

Students who are eligible under Section 504 of the Rehabilitation Act of 1973 may be tested using modified testing procedures that must be specified in the student's 504 accommodation plan.

Visually impaired students may take either a Braille or large-print version of the test. Specific instructions for administering the Braille and large-print versions of the test are provided in the supplementary instructions for examiners administering these forms.

Students using the Braille test booklets:

- are instructed to bring a Braille ruler and a talking calculator to the test session;
- are instructed to skip some items identified in the Braille instructions—the spaces for these items must be left blank on the student answer folder;
- have answer folders transcribed from Braille version by the examiner;
- dictate their answers to the examiner or use a device that produces Braille.

For dictations and responses recorded in Braille:

- Students must indicate all punctuation and must spell all key words.
- Examiners must transcribe the Brailled responses into the regular answer folder.

Students using the large-print test booklets:

- mark their answers in the large-print answer folders;
- may be instructed to skip some questions—the spaces for these questions must be left blank in the student's large-print answer folder;
- dictate responses on constructed-response items and writing tasks, indicate all punctuation, and spell key words.

Accommodations and modifications of test administration procedures are listed in Appendix C of this report and are included in the Test Coordinator Manual**.**

If a student requires an accommodation or modification that is not listed, district staff are instructed to contact the Office of Assessments, NJ ASK Program Coordinator. Accommodations or modifications are classified as follows:

A = Setting Accommodations
B = Scheduling Accommodations
C = Test Materials/Modifications
D = Test Procedures Modifications

Tables 3.4.1–3.4.2 provide counts and performance results of special education and Section 504 students. Descriptive statistics are also provided for those students classified as special education or Section 504 who were tested *without* accommodations or modifications. Not every special education and Section 504 student is tested with an accommodation or modification. Accommodations and modifications may be used separately or in combination. The use of multiple accommodations for individual students is common.

Table 3.4.1: Statistics for Students Classified as Special Education, NJ ASK Science

Grade	Accommodation	N	Mean	STD	Min	Max	%PP	%P	%AP
4	Yes	12200	218.91	33.00	100	300	24.76	52.39	22.85
	No	3958	238.69	35.31	118	300	12.63	41.03	46.34
8	Yes	13732	195.83	26.94	100	300	54.97	41.20	3.83
	No	1774	202.01	32.39	100	300	48.31	41.88	9.81

Table 3.4.2: Statistics for Students Classified as Section 504, NJ ASK Science

Grade	Accommodation	N	Mean	STD	Min	Max	%PP	%P	%AP
4	Yes	2637	240.06	31.30	118	300	7.55	47.71	44.75
	No	730	244.37	30.99	168	300	5.89	43.01	51.10
8	Yes	2480	219.28	28.51	138	300	22.14	62.22	15.65
	No	956	226.35	30.24	100	300	17.89	58.68	23.43

3.5 Reliability and Validity of Tests for Special Populations[1]

Assessing the reliability and validity of the modifications made for the special populations is as important as assessing these psychometric properties of the operational tests. The reliability of an assessment refers to the consistency of test scores across test administrations. Validity of assessment is the degree to which an assessment measures what it is intended to measure and the extent to which the inferences made and actions taken on the basis of the assessment outcomes are accurate and appropriate. An assessment that is not reliable cannot be valid (AERA, APA, & NCME, 1999).

New Jersey state regulations require that all students be included in the statewide assessment program and assessed annually including limited English proficient (LEP) students and students with disabilities. Beginning in school year 2001–2002, students with severe cognitive disabilities were administered the Alternative Proficiency Assessment (APA) for the first time statewide. All public schools, including those without assessed grades, are counted in the state's accountability system.

Given the high stakes nature of the tests for school accountability, it is important that the tests be reliable and valid. The NJ ASK tests are offered in English, Spanish, Braille, and large-print, and students are allowed various accommodations as determined by the individualized education plan (IEP) and 504 plan. Spanish forms are offered for current limited English proficient (CLEP) students whose dominant language is Spanish, as identified by school districts. Alternate forms of the tests are generated for students who cannot participate in the regular administration of the tests.

The reliability and validity evidence for the 2015 NJ ASK tests for the regular and special populations are documented in various parts of this report. A reliable test is one that produces scores that are expected to be relatively stable if the test is administered repeatedly under similar conditions for the general testing population and across subgroups. For evidence that a test is performing similarly across subgroups, the reliability values for these subgroups can be compared to those of the total population. Note that the reliability measures are impacted by the population distribution and can be lower when the subgroup is homogenous in performance. However, one would expect the subgroup reliabilities to be adequately high for all groups. The test reliabilities measured by Cronbach alpha for the 2015 NJASK tests are described in Part 8. The alphas for overall student responses ranged from 0.83 to 0.88 for science indicating that the tests are highly reliable. The reliability of the tests for Spanish students only is lower than in the general population, ranging from 0.71 to 0.75 for science (see Table 8.1.1), which is still reasonable given the student population. Reliability estimates for special education and limited English proficient students can also be found in Table 8.1.1. The reliabilities for these special populations are quite similar to the general population.

The reliability of the test and test scores is reflected in the evidence of rater consistency (i.e. inter-rater reliability). Although there is no separate inter-rater reliability analysis for CLEP

[1] Sato, E., Worth, P, Gallagher, C., Lagunoff, R., and McKeag, H., (2007). Guidelines for Ensuring the Technical Quality of Assessments Affecting English Language Learners and Students with Disabilities: Development and Implementation of Regulations

students, the rater reliability coefficient for the total population shows relatively high agreement between the raters on the constructed-response items. The scoring processes are described in Part 4, and inter-rater reliability of test scores in constructed-response and writing items are presented in Part 8 of this report.

The 2015 NJ ASK Science validity evidence for special populations is described here in terms of test content, test administration and response process, internal structure, and score reporting as proposed by Sato et. al (2007) for guidelines of evaluating reliability and validity of assessments.

The fact that all tests are constructed under the same blueprint and specifications is evidence of content validity. The NJ ASK tests for special populations including Spanish, Braille, and large-print versions are translated directly from the operational forms. The items are developed to align with and measure the NJ core curriculum standards so that all students can demonstrate the knowledge and skills necessary for the attainment of English language proficiency and the language necessary for achievement in the academic content areas. All standards and assessments are reviewed by specialists from NJ content as well as bias and sensitivity review committees to identify and eliminate elements that may favor one group (e.g., language, culture, ethnicity) over another. Test items are developed under universal test design principles with NJ special student populations in mind so that no student group is disadvantaged. The test development process is described in Part 2 of this technical report.

The test validity is also reflected in the fact that the test is inclusive for all students. In order to minimize or eliminate factors that contribute to assessment ambiguity and inaccuracy such that assessment results accurately reflect student knowledge and ability, various accommodations are provided to the special needs students based on their IEP or 504 plans. A list of acceptable test accommodations or modifications of test administrations is provided in Appendix C.

The test validity further ensures the comparability and interpretation of scores and proficiency standards across different student groups. All NJ ASK item responses for a given grade/content from the general and special populations are combined for item analysis, calibration, and equating. These analyses include all students regardless of the test version taken, i.e., operational, Spanish, Braille, or large-print. An entirely different score conversion table is prepared for tests requiring modifications such that a subset of the total number of items constitutes the total score. However, these special test versions are placed on the same scale as the operational tests; thus, proficiency standards can be applied uniformly to all tests.

The performance of students from various groups—including gender, ethnicity, special education, and LEP—are reported at the school level. Table 6.5.8 presents the mean and standard deviation of scale scores for Braille, large-print, and Spanish test takers. As shown in this table, students from the sub-populations performed less well than the general population in all content areas. The number of students in the sub-groups is small, however.

PART 4: SCORING

4.1 Multiple-Choice Items

The answer keys approved by NJ DOE are used to score the multiple-choice items after the responses have been scanned. Each item has a key associated with the item (A, B, C, or D), which has been supplied and verified by the NJ ASK content specialists. All correct answers are assigned the value of "1" while incorrect answers are assigned the value of "0." At no time in this process is the original scanned answer overwritten, in case the key is determined to be incorrect during the post-scoring quality assurance check. After scoring is completed, simple item statistics are provided to the appropriate NJ ASK content specialist to ensure that the correct keys are being applied. If a key changes, then the process is repeated until the scoring file is correct. The key-check data file contains the following information:

- percent of students getting the question correct (PC);
- correlation of the item to the test as a whole (RPB);
- correlation of each possible response option to the test as a whole (RPBA, RPBB, etc.);
- percentage of students choosing each response option (A, B, C, D or X-omits); and
- flags for items with high difficulty (DFLAG) or low correlations (CFLAG).

4.2 Constructed-Response Items

A discussion of the following topics germane to the scoring of constructed response items can be found in the 2009 NJ ASK Technical Report (PTM 1507-34), Part 4, Section 4.2.

- Scorer Selection
- Range Finding
- Field Test Range Finding
- Scoring Guides
- Team Leader Training and Qualifying
- Scorer Training/Qualifying
- Monitoring Scorer Performance

As the number of scoring personnel varies from year to year, Table 4.2.1 details the levels of staffing for scoring the 2015 NJ ASK. The table shows the numbers of scorers, team leaders and scoring directors at each grade level who participated in scoring.

Table 4.2.1: Scoring Personnel by Grade and Content Area - NJ ASK Science 2015

	Grade	Scorers	Team Leaders	Scoring Director
Constructed Response	4	105	11	4
	8	70	8	3

As shown in Part 8, Reliability, the raters are not in perfect agreement 100% of the time. Thus to ensure that no student is unjustly penalized because a rater may have been a little too stringent, rescoring is conducted automatically for any student who scores within one raw score point of the proficient cut score. MI reviews writing and constructed-response items and verifies the original scores or makes changes where warranted. Scores are never lowered during the automatic rescoring process even if a lower score results. Districts do not need to request rescoring. Table 4.2.2 provides automatic rescoring information for each grade level and content area. All open-ended/constructed response item types were scored by a single rater.

Table 4.2.2: Automatic Rescore Statistics - NJ ASK Science 2015

Grade	Eligible for Automatic Rescore	Score/Proficiency Changes	
		#	%
4	2056	43	2.09
8	2411	23	0.95

4.3 Quality Control[2]

In order to ensure the quality of the testing materials, MI and the NJ DOE work together to rigorously proof all materials prior to printing/production. The steps of the quality control procedures can be found in the 2009 NJ ASK Technical Report (PTM 1507-34), Part 4, Section 4.3.

[2] The NJ DOE checks all test result data for consistency, replicates reported summary data to ensure accuracy, and reviews all printed reporting materials to verify appropriateness. Additionally, the NJ DOE checks the recording and tallying of item scores.

PART 5: STANDARD SETTING

Two separate standard settings have been held for NJ ASK. Both standard settings were relevant to only ELA and math. The first was conducted after the first administration of the new NJ ASK grades 5 through 8 in April–May 2008 and the second occurred after the 2009 administration of the new NJ ASK assessments in grades 3 and 4. Detailed information regarding these two standard settings can be found in the 2009 NJ ASK Technical Report (PTM 1507-34) and the New Jersey Assessment of Skills and Knowledge (NJ ASK) Standard Setting Report from 2008 and 2009. No recent standard settings have been held for science.

PART 6: ITEM and TEST STATISTICS

6.1 Classical Item Statistics

For each administration, classical item analyses were completed prior to item calibration, scaling, and equating. These statistics were calculated again once all of the data were available. These analyses involve computing a set of statistics based on classical test theory for every item in each form. Each statistic was designed to provide some key information about the quality of each item from an empirical perspective. The statistics estimated for the NJ ASK are described below.

- Classical item difficulty ("p-value"):
  This statistic indicates the percentage of examinees in the sample that answered the item correctly. Desired p-values generally fall within the range of 0.20 to 0.90.

- Item discrimination ("r-biserial"):
  This statistic is measured by the poly-serial correlation between the item score and the test criterion score and describes the relationship between performance on the specific item and performance on the entire form. Higher values indicate greater differences in the performance of competent and less competent examinees. Items with negative correlations can indicate serious problems with the item content (e.g., multiple correct answers or unusually complex content) or can indicate that students have not been taught the content. For science, the test criterion score is the total score of all MC and CR items.

- Distractor analyses for MC items:
  This statistic reports the proportion of examinees who select each incorrect response (distractor).

- Percentage of students omitting an item:
  This statistic is useful for identifying problems with test features such as testing time and item/test layout. Typically, we would expect that if students have an adequate amount of testing time, 95% of students should attempt to answer each question.

  When a pattern of omit percentages exceeds 5% for a series of items at the end of a timed section, this may indicate that there was insufficient time for students to complete all items. Alternatively, if the omit percentage is greater than 5% for a single item, this could be an indication of an item/test layout problem. For example, students might accidentally skip an item that follows a lengthy stem.

Item analyses were conducted for the 2015 NJ ASK assessment of science for both grades 4 and 8. In this section, summary information is presented by grade at both the content domain and content cluster level. The information includes mean item scores and discrimination indices, as well as descriptive statistics for number correct raw score and for scale scores. Statistics include N-counts, means, standard deviations, minimum and maximum values, and a variety of data disaggregations, including student demographic group and DFG.

For multiple-choice (MC) items, the mean score is simply the proportion of students who gave a correct response to the item (usually referred to as item difficulty or the p-value), and the discrimination index is the point-biserial correlation between the item score and the total score based on the remaining items.

Descriptive Statistics

Tables 6.1.1 through 6.1.2 summarize by item response format, item difficulty, and discrimination of the items that comprise each content domain and cluster for grades 4 and 8, respectively. For MC items, both the mean and standard deviation are given. The mean value is the average of the p-values of the items in the cluster. For CR items, the mean value is the average item score for the items in the cluster. Item discrimination is the correlation between students' item score and the total score of the remaining items on the test. Both item difficulty and discrimination are expressed in terms of the raw score metric.

Tables 6.1.3 through 6.1.4 summarize frequency distributions for MC item difficulty and discrimination indices of items comprising each content domain and cluster for grades 4 and 8, respectively. The median item difficulty and discrimination is also displayed.

Table 6.1.5 summarizes distractor analyses for MC items by test. The number in each cell indicates the number of items where at least one p-value or discrimination index (point-biserial) for the distractors was higher than the keyed option (answer identified as the correct response).

Table 6.1.1: Grade 4 - Item Difficulty and Discrimination Summary Statistics by Cluster

Test Section/ Cluster	Multiple-Choice				Constructed-Response			
	Item Difficulty			Item Discrimination	Item Difficulty			Item Discrimination
	Nitem	Mean	S.D.	Mean	Nitem	Mean	S.D.	Mean
Science	**33**	**0.67**	**0.10**	**0.32**	**2**	**1.07**	**0.70**	**0.40**
Life Science	15	0.70	0.08	0.35	0	-	-	-
Physical Science	9	0.67	0.07	0.26	1	0.57	-	0.39
Earth Science	9	0.62	0.13	0.33	1	1.56	-	0.41
Knowledge	*4*	*0.74*	*0.10*	*0.33*	*0*	-	-	-
Application	*29*	*0.66*	*0.10*	*0.32*	*2*	*1.07*	*0.70*	*0.40*

Table 6.1.2: Grade 8 - Item Difficulty and Discrimination Summary Statistics by Cluster

Test Section/ Cluster	Multiple-Choice				Constructed-Response			
	Item Difficulty			Item Discrimination	Item Difficulty			Item Discrimination
	Nitem	Mean	S.D.	Mean	Nitem	Mean	S.D.	Mean
Science	**48**	**0.63**	**0.09**	**0.33**	**2**	**0.75**	**0.14**	**0.49**
Life Science	20	0.64	0.10	0.33	0	-	-	-
Physical Science	14	0.64	0.09	0.35	1	0.65	-	0.44
Earth Science	14	0.60	0.09	0.32	1	0.85	-	0.53
Knowledge	*6*	*0.65*	*0.08*	*0.38*	*0*	-	-	-
Application	*42*	*0.62*	*0.09*	*0.33*	*2*	*0.75*	*0.14*	*0.49*

Table 6.1.3: Grade 4 - Difficulty and Discrimination Indices for MC Items by Cluster

		Difficulty						Discrimination					
	Nitem	**Median**	**p < 0.25**	**0.25 <= p < 0.50**	**0.50 <= p< 0.75**	**0.75 <= p< 0.90**	**p >=0.90**	**Median**	**rpb <0.20***	**0.20 <= rpb < 0.30**	**0.30 <= rpb < 0.40**	**0.40 <=rpb< 0.50**	**rpb >= 0.50**
Science	**33**	**0.66**	**0**	**1**	**26**	**6**	**0**	**0.34**	**2**	**10**	**16**	**5**	**0**
Life Science	15	0.69	0	0	12	3	0	0.35	0	5	5	5	0
Physical Science	9	0.69	0	0	8	1	0	0.29	2	3	4	0	0
Earth Science	9	0.63	0	1	6	2	0	0.35	0	2	7	0	0
Knowledge	*4*	*0.78*	*0*	*0*	*1*	*3*	*0*	*0.36*	*0*	*1*	*2*	*1*	*0*
Application	*29*	*0.66*	*0*	*1*	*25*	*3*	*0*	*0.33*	*2*	*9*	*14*	*4*	*0*

* While ideally items should have a point-biserial correlation of at least 0.20, these items had acceptable p-values and were retained to preserve adequate content coverage at the cluster level.

Table 6.1.4: Grade 8 - Difficulty and Discrimination Indices for MC Items by Cluster

		Difficulty						Discrimination					
	Nitem	Median	p < 0.25	0.25 <= p < 0.50	0.50 <= p< 0.75	0.75 <= p< 0.90	p >=0.90	Median	rpb <0.20*	0.20 <= rpb < 0.30	0.30 <= rpb < 0.40	0.40 <=rpb< 0.50	rpb >= 0.50
Science	**48**	**0.62**	**0**	**5**	**38**	**5**	**0**	**0.33**	**2**	**15**	**21**	**10**	**0**
Life Science	20	0.64	0	2	15	3	0	0.33	1	7	7	5	0
Physical Science	14	0.64	0	1	12	1	0	0.35	0	3	8	3	0
Earth Science	14	0.60	0	2	11	1	0	0.32	1	5	6	2	0
Knowledge	*6*	*0.64*	*0*	*0*	*5*	*1*	*0*	*0.39*	*0*	*1*	*2*	*3*	*0*
Application	*42*	*0.62*	*0*	*5*	*33*	*4*	*0*	*0.33*	*2*	*14*	*19*	*7*	*0*

* While ideally items should have a point-biserial correlation of at least 0.20, these items had acceptable p-values and were retained to preserve adequate content coverage at the cluster level.

Table 6.1.5: Number of Multiple-Choice Items Flagged by Distractor Analyses

Test	Grade	Nitems	P-Value*	Point-Biserial*
Science	**4**	**33**	**0**	**1**
	8	**48**	**0**	**0**

* The p-value and point-biserial correlation in this table are calculated in the same way as for a correct answer, except in this case the distractor is used instead of the correct answer.

6.2 Speededness

The consequence of time limits on examinees' scores is called speededness. An examination is "speeded" to the degree that those taking the exam score lower than they would have had the test not been timed. Most speededness statistics are based on the number of items that were not attempted by students. In each separately timed subsection of a test, if a student does not attempt the last item of the test, it can be assumed that the student may have run out of time before reaching the last item. The percentage of students omitting an item provides information about speededness, although it must be kept in mind that students can omit an item for reasons other than speededness (for example, choosing to not put effort into answering a constructed response item). Thus, if the percentage of omits is low, that implies that there is little speededness; if a percentage of omits is high, speededness, as well as other factors, may be the cause.

The NJ ASK was not designed to be a speeded test, but rather a power test. That is, all students are expected to have ample time to finish all items and prompts. As the tests were administered over four days, with multiple sessions each day, students were assumed to have enough time to complete the test. The number of items and item types composing each test, along with the testing time and day of administration, are detailed in Table 6.2.1. Table 6.2.2 presents the percentage of students omitting the last MC item in each test section.

Table 6.2.1: Testing Schedule—Items and Time Allocations

Subject	Grade	Items	Time*
Science	4	33 MC, 2 CR	60
	8	48 MC, 2 CR	120

* Time in minutes

Table 6.2.2: Percent of Students Omitting the Last MC Item in Each Test Section

Grade	Section	Content Area	Location	%
4	1	Science	Item 13	2.46
	2		Item 23	2.64
	3		Item 34	1.38
8	1	Science	Item 18	0.31
	2		Item 33	0.37
	3		Item 49	0.41

6.3 Intercorrelations

The Pearson product-moment correlations among the test sections/clusters are presented in Tables 6.3.1–6.3.2. Generally, the more items a cluster (standard) has, the higher the correlation with the total score. After all, the cluster (standard) makes up more of the points of the total score. For example, the Application total score at grade 4 is highly correlated with the total science test score (0.99) because application items make up 35 of the 39 possible points for science.

Table 6.3.1: Grade 4 Correlation Coefficients among Content Domains and Clusters

	Science	**Life**	**Physical**	**Earth**	***Knowledge***	***Application***
Science	1.00	0.92	0.85	0.88	*0.72*	*0.99*
Life	0.92	1.00	0.68	0.71	*0.70*	*0.91*
Physical	0.85	0.68	1.00	0.62	*0.60*	*0.84*
Earth	0.88	0.71	0.62	1.00	*0.60*	*0.88*
Knowledge	*0.72*	*0.70*	*0.60*	*0.60*	*1.00*	*0.64*
Application	*0.99*	*0.91*	*0.84*	*0.88*	*0.64*	*1.00*

Table 6.3.2: Grade 8 Correlation Coefficients among Content Domains and Clusters

	Science	**Life**	**Physical**	**Earth**	***Knowledge***	***Application***
Science	1.00	0.93	0.91	0.91	*0.80*	*0.99*
Life	0.93	1.00	0.77	0.77	*0.74*	*0.93*
Physical	0.91	0.77	1.00	0.76	*0.75*	*0.91*
Earth	0.91	0.77	0.76	1.00	*0.73*	*0.91*
Knowledge	*0.80*	*0.74*	*0.75*	*0.73*	*1.00*	*0.74*
Application	*0.99*	*0.93*	*0.91*	*0.91*	*0.74*	*1.00*

6.4 DIF Analysis

Using data from the field test items embedded in the 2014 operational tests, Differential Item Functioning (DIF) was examined using the Mantel-Haenszel (1959)[3] procedure for the MC items and the Liu-Agresti cumulative common log odds ratio (Penfield, 2007)[4] for CR items. As all items must be field tested and scrutinized including DIF analyses prior to appearing as an operational item, DIF analyses are not conducted on operational items.

For DIF analyses, all members of the reference group (typically male/majority) are compared against all members of the focal group (typically female/minority). The DIF analyses conducted for NJ ASK 4 and 8 focused on gender and ethnicity. The number of examinees composing the

[3] Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, 22, 719-748.
[4] Penfield, R. (2007). An approach for categorizing DIF in polytomous items. *Applied Measurement In Education*, 20, 335-355.

reference and focal groups differ dependent upon the year in which a given item was field tested. In general, appropriately 4,500 examinees respond to each field test item.

The Mantel-Haenszel (MH) method is a non-parametric approach to DIF. In the MH procedure, total raw scores are held constant while the odds ratio is estimated. The ETS categorization is applied to flag the significance of DIF effects (Dorans & Holland, 1993)[5]. The Liu-Agresti cumulative common log odds ratio allows for the ETS categorization to be applied to polytomous items. DIF analyses are detailed in Section 2.2 - Development of Test Items. The letters A, B, and C are used to denote the ETS categorizations. A indicates a smaller degree of DIF, B indicates moderated DIF, and C indicates larger differences in the performance of the reference and focal groups on a given item. Slightly different categorizations were used for the constructed response items. A or NS indicates a smaller degree of DIF, B or S indicates moderated DIF, and C- indicates larger differences. Table 6.4.1 represents the ETS categorization of each of the items used in the 2015 NJ ASK operational test when they were field tested.

Table 6.4.1: 2015 NJ ASK Operational Items - DIF Categories by Item Type and Grade

Test	Grade	Group	Multiple Choice*			Constructed-Response*		
			A**	B**	C**	A/NS+	B/S+	C-+
Science	4	M/F	31	2	0	2	0	0
		W/B	26	6	1	2	0	0
		W/H	28	5	0	2	0	0
	8	M/F	43	4	0	2	0	0
		W/B	43	4	0	1	1	0
		W/H	44	3	0	1	1	0

* The Mantel-Haenszel procedure is applied to MC and CR items.
** DIF categories for MC items: A, negligible; B, slight to moderate; and C, moderate to severe.
*** DIF contrast groups: M/F, Male versus Female; W/B, White versus Black; and W/H, White versus Hispanic.
\+ DIF categories for CR items: A/NS, negligible; B/S, moderate to severe; and C-, severe.

[5] Dorans, N. J. & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

6.5 Summary Statistics

Descriptive Statistics for Total Raw Score

Descriptive statistics of total scores for NJ ASK 2015 are summarized in Table 6.5.1 by test content, form, and grade level. A total of 200,043 students participated in the science tests in grades 4 and 8.

Table 6.5.1: Descriptive Statistics for Total Raw Score by Content Area and Grade Level

Test	Grade	Form	N	Mean	STD	Min	Max	Nitem	Max Possible
Science	4	OP	98,073	24.35	6.70	0	39	35	39
		BR	2	15.50	3.54	13	18	35	39
		LP	66	20.97	7.01	6	33	35	39
		SP	833	15.93	5.62	4	34	35	39
	8	OP	99,814	31.47	9.64	0	54	50	54
		BR	5	28.80	10.66	12	38	49	54
		LP	60	25.83	11.07	7	50	50	54
		SP	1,190	21.21	6.21	4	43	50	54

*OP: Operational Test; BR: Braille; LP: Large Print; SP: Spanish Version.

Descriptive Statistics for Total Raw Score by Cluster

Tables 6.5.2 through 6.5.3 summarize the means and standard deviations for raw score attained by cluster for the 2015 NJ ASK operational test forms.

Table 6.5.2: Grade 4 Means and Standard Deviations for Raw Score

	Number of Items		Number of	Raw Score		Mean % of Points
	MC	CR	Possible Points	Mean	STD	Available
Science	**33**	**2**	**39**	**24.35**	**6.70**	**62.45%**
Life Science	15	0	15	10.56	3.05	70.40%
Physical Science	9	1	12	6.65	2.19	55.39%
Earth Science	9	1	12	7.15	2.55	59.56%
Knowledge	*4*	*0*	*4*	*2.97*	*1.01*	*74.37%*
Application	*29*	*2*	*35*	*21.38*	*6.09*	*61.08%*

Table 6.5.3: Grade 8 Means and Standard Deviations for Raw Score

	Number of Items		Number of	Raw Score		Mean % of Points
	MC	CR	Possible Points	Mean	STD	Available
Science	**48**	**2**	**54**	**31.47**	**9.64**	**58.27%**
Life Science	20	0	20	12.65	3.92	63.26%
Physical Science	14	1	17	9.59	3.46	56.42%
Earth Science	14	1	17	9.22	3.35	54.26%
Knowledge	*6*	*0*	*6*	*3.87*	*1.56*	*64.44%*
Application	*42*	*2*	*48*	*27.60*	*8.50*	*57.50%*

Scale Score Distributions by Content Area and Grade

Descriptive statistics for scale scores and percentage distributions of students' performance levels are summarized in Table 6.5.4 by content area and grade. Science student records flagged as void, not present, or missing were removed. For all test forms, scale scores have a range of 100 to 300. A student is classified as Partially Proficient (PP) if his/her scale score is lower than 200. A student is classified as Advanced Proficient (AP) if his/her scale score is 250 or higher. All other students are classified as Proficient (P).

Table 6.5.4: Descriptive Statistics of Students' Performance Levels by Content and Grade

Test	Grade	Form	N+	Mean	STD	Min	Max	%PP	%P	%AP
Science	4	OP	98073	240.97	32.95	100	300	9.06	43.40	47.55
		BR	2	199.00	15.56	188	210	50.00	50.00	0.00
		LP	66	224.15	33.54	148	288	21.21	48.48	30.30
		SP	833	200.97	26.59	130	296	44.54	49.34	6.12
	8	OP	99814	221.08	30.64	100	300	22.46	57.99	19.54
		BR	5	212.80	32.77	161	241	20.00	80.00	0.00
		LP	60	204.15	36.24	138	298	46.67	43.33	10.00
		SP	1190	189.77	18.90	116	257	66.97	32.69	0.34

* OP: Operational Test; BR: Braille; LP: Large Print; SP: Spanish Version

+ Reflects N counts

Scale Score Distributions by Demographic Group

Descriptive statistics of scale scores and percentage distributions of students' Performance by Demographic Groups can be found at http://www.nj.gov/education/schools/achievement. Scale score cumulative frequency distributions are attached as Appendix F. Note that Alternate and Braille forms are excluded from the cumulative frequency distributions.

Scale Score Distributions by District Factor Groups (DFG)

New Jersey has an established history of applying DFGs[6] in the analysis and reporting of assessment results. DFG is an indicator of the socioeconomic status of citizens in each district and has been useful for the comparative reporting of test results from New Jersey's statewide testing programs. The measure was first developed in 1974 using demographic variables from the 1970 United States Census. A revision was made in 1984 to take into account new data from the 1980 United States Census. The DFG designations were updated again in 1992 after the 1990 census. The current DFG designations are based upon the 2000 census. The DFGs are labeled from A (lowest) to J (highest). Additional DFGs are designated for special groups that are not defined geographically. For example N is used to designate districts with a percentage of students in public schools too low for a DFG value to be assigned; O and S indicate schools receiving special populations and are not included in the tables; R represents charter schools; and V denotes vocational schools

Descriptive statistics of scale scores and percentage distributions of student performance by DFG for General Education group are summarized in Tables 6.5.5 by content area and grade. For each of the content areas, students who were flagged as "void" or "not present" were removed.

[6] For more information on DFGs, see the following link: http://www.state.nj.us/education/finance/rda/dfg.shtml

Table 6.5.5: Descriptive Statistics for Science Performance Levels by DFG

Grade	DFG[+]	N	Mean	STD	Min	Max	%PP	%P	%AP
4	A	17,523	219.50	31.14	102	300	22.74	55.51	21.75
	B	10,387	230.47	31.23	100	300	13.11	53.85	33.04
	CD	9,341	235.78	30.70	102	300	9.70	50.36	39.94
	DE	11,947	241.64	30.30	100	300	6.59	45.58	47.83
	FG	11,348	245.80	30.21	130	300	5.17	41.16	53.67
	GH	13,172	249.93	30.73	100	300	4.48	36.60	58.92
	I	17,311	256.16	28.95	118	300	2.61	29.96	67.43
	J	4,230	261.11	28.64	102	300	2.01	24.35	73.64
	N	521	213.41	29.15	140	300	29.56	56.24	14.20
	0	7	197.71	19.20	173	219	42.86	57.14	0.00
	R	3,187	236.19	32.47	140	300	11.26	47.91	40.82
8	A	16,122	199.96	26.89	100	300	48.28	47.20	4.52
	B	10,091	209.71	28.25	100	300	33.91	56.90	9.19
	CD	9,334	215.16	28.09	100	300	26.11	61.96	11.93
	DE	12,732	220.50	27.92	125	300	19.99	63.63	16.38
	FG	12,824	224.85	28.37	100	300	16.56	62.48	20.96
	GH	13,211	229.16	29.84	116	300	14.52	58.52	26.96
	I	19,470	234.68	27.87	132	300	9.35	58.20	32.46
	J	4,286	241.23	27.67	100	300	5.86	53.03	41.11
	N	376	196.80	26.66	132	266	52.39	45.21	2.39
	0	21	183.10	24.62	152	242	80.95	19.05	0.00
	R	2,584	214.43	28.06	132	300	28.10	59.52	12.38
	V	18	219.56	26.50	175	266	22.22	61.11	16.67

[+]N = majority of students in private schools; R = charter schools; V = vocational schools

PART 7: EQUATING AND SCALING

This section details the equating and scaling procedures applied to the NJ ASK 2015 operational tests[7]. Equating and scaling procedures were applied to the grades 4 and 8 Science assessments.

7.1 Descriptive Statistics for Equating Data

In 2008, data from approximately 35,000 students were used as impact data for standard setting. These data were used to establish new Proficient and Advanced Proficient cut scores for ELA and mathematics in grades 5 through 8. Thus, 2008 became the new base year to which future ELA and mathematics grade 5-8 assessments are equated. Likewise, standard settings were conducted for grades 3-4 ELA and mathematics in 2009; thus, making 2009 the year to which future grades 3-4 ELA and mathematics assessments are equated. The base years for science grades 4 and 8 are 2005 and 2000, respectively. The 2014 assessments were placed on the corresponding base-year scale using common item non-equivalent group with anchor test equating design based on an equating sample of approximately 30% of the total student population of science examinees. The NJ ASK 2015 equating samples are summarized in Table 7.1.1.

Table 7.1.1: N-Counts for the Equating Samples by Content and Grade[2]

Test	Grade	Total[3]	Percent	Valid	Invalid[1]
Science	4	39,576	39.09	38,703	873
	8	42,474	40.88	41,192	1,282

[1]Invalidation occurs when void codes are applied or a non-attempt flag is present
[2]Please Note: All Tables in Part 7are based on the equating sample.
[3]Minor data differences across tables reflect small amounts of unreported information.

The 2015 equating sample was selected using a stratified random sampling methodology with DFG as a stratum. In addition, the samples were representative of the total student population in terms of demographic variables such as gender ethnicity, economic status, and Current Limited English Proficiency (CLEP). Comparisons between data from the 2015 Form Distribution Plan and the sample data used for equating and scaling are presented in Tables 7.1.2 to 7.1.3. These tables show the differences between the 2015 Form Distribution Plan and the equating sample.

[7] All equating results are verified by two external reviewers.

Table 7.1.2: Comparison of the Equating Sample and the Statewide DFGs—Grade 4

	Statewide	Science	
DFG[+]	Distribution	Obs(%)	Diff
A	17.66	17.98	-0.33
B	10.47	11.05	-0.58
CD	9.45	9.48	-0.03
DE	12.13	12.02	0.11
FG	11.58	11.65	-0.07
GH	13.30	14.37	-1.07
I	17.45	16.51	0.94
J	4.25	3.90	0.35
N	0.54	0.06	0.48
O	0.01	0.02	-0.01
R	3.17	2.97	0.20
S	0.00	0.00	0.00
V	0.00	0.00	0.00

* Indicates the maximum difference between statewide distribution and the sample.
[+] N = majority of students in private schools, O and S = schools receiving special populations; R = charter schools; V = vocational schools

Table 7.1.3: Comparison of the Equating Sample and the Statewide DFGs—Grade 8

	Statewide	Science	
DFG[+]	Distribution	Obs(%)	Diff
A	16.07	16.44	-0.37
B	9.94	9.32	0.62
CD	9.24	9.46	-0.21
DE	12.64	10.83	1.81
FG	12.72	11.35	1.37
GH	13.08	16.12	-3.04
I	19.18	17.79	1.39
J	4.19	6.76	-2.57
N	0.38	0.05	0.33
O	0.02	0.01	0.01
R	2.52	1.83	0.69
S	0.00	0.00	-0.00
V	0.02	0.04	-0.03

* Indicates the maximum difference between statewide distribution and the sample.
[+] N = majority of students in private schools, O and S = schools receiving special populations; R = charter schools; V = vocational schools

Table 7.1.4 presents the N-counts for the 2015 equating samples by DFG, gender, and ethnicity. Note that the sum for males and females does not equal the total in Table 7.1.1 as some examinees did not identify their gender. Similarly, some examinees did not identify ethnicity or marked

multiple ethnicities, therefore the sum over ethnic groups does not equal the total number of students. Also reported in Table 7.1.4 are the numbers of economically disadvantaged students as well as CLEP students.

Table 7.1.4: Equating Sample N-Counts by Gender and Ethnicity: Science

Grade	DFG[+]	Male	Female	Asian	Black	Hispanic	Indian Alaska	Hawaii Pacific	White	EconDis[2]	LEP[1]
4	**A**	3,551	3,404	118	2,090	4,031	13	5	557	5,437	1,324
	B	2,151	2,120	158	698	1,982	3	13	1,366	2,855	614
	CD	1,888	1,778	225	687	1,106	2	4	1,600	1,834	338
	DE	2,408	2,237	311	729	981	7	8	2,446	1,550	163
	FG	2,294	2,211	484	476	797	7	14	2,620	1,163	186
	GH	2,885	2,675	698	486	758	9	12	3,494	926	226
	I	3,242	3,146	1,138	229	430	8	9	4,368	321	120
	J	775	735	248	17	52	2	5	1,152	12	22
	N	13	10	0	4	10	0	0	9	12	7
	O	6	1	0	5	0	0	0	2	5	0
	R	540	606	76	460	449	0	2	136	694	33
	S	0	0	0	0	0	0	0	0	0	0
	V	0	0	0	0	0	0	0	0	0	0
	Total	**19,753**	**18,923**	**3,456**	**5,881**	**10,596**	**51**	**72**	**17,750**	**14,809**	**3,033**
8	**A**	3,468	3,292	138	2,443	3,609	4	3	463	4,609	646
	B	1,955	1,879	213	672	1,761	4	12	1,143	2,269	350
	CD	2,034	1,860	228	759	1,070	0	3	1,799	1,846	133
	DE	2,274	2,176	246	515	862	6	14	2,519	1,252	88
	FG	2,396	2,278	419	588	738	6	17	2,823	1,150	101
	GH	3,435	3,204	1,061	775	800	11	5	3,876	1,096	102
	I	3,692	3,630	1,135	270	401	4	7	5,232	404	90
	J	1,445	1,339	661	47	101	1	3	1,943	41	33
	N	15	5	0	2	3	0	0	14	1	0
	O	1	2	0	2	0	0	0	1	1	0
	R	362	392	53	313	275	0	5	101	464	11
	S	1	0	0	0	0	0	0	1	0	0
	V	10	8	3	4	6	0	0	5	11	0
	Total	**21,088**	**20,065**	**4,157**	**6,390**	**9,626**	**36**	**69**	**19,920**	**13,144**	**1,554**

[1] Current LEP status in 2015
[2] Economically Disadvantaged
[+] N = majority of students in private schools, O and S = schools receiving special populations; R = charter schools; V = vocational schools

Table 7.1.5 displays descriptive statistics of the raw scores for the equating samples by grade and test content. Tables 7.1.6 and 7.1.7 summarize descriptive statistics for raw scores for the equating samples by gender. Table 7.1.8 summarizes descriptive statistics for raw scores for the samples by DFG. Note, the maximum possible score was achieved at all grade levels in science.

Table 7.1.5: Descriptive Statistics for Raw Scores by Grade and Test Content

Test	Grade	N	Mean	STD	Min	Max	Max Possible
Science	4	38,703	24.14	6.77	0	39	39
	8	41,192	31.40	9.76	0	54	54

Table 7.1.6: Descriptive Statistics for Raw Scores by Gender—Male

Test	Grade	N	Mean	STD	Min	Max
Science	4	19,753	24.29	7.03	0	39
	8	21,088	31.66	10.22	0	54

Table 7.1.7: Descriptive Statistics for Raw Scores by Gender—Female

Test	Grade	N	Mean	STD	Min	Max
Science	4	18,923	23.99	6.49	0	39
	8	20,065	31.15	9.25	0	54

Table 7.1.8: Descriptive Statistics for Raw Scores by District Factor Group: Science

Grade	DFG[+]	N	Mean	STD	Min	Max
4	A	6,960	19.78	6.62	0	38
	B	4,276	22.46	6.45	0	39
	CD	3,670	23.28	6.43	2	39
	DE	4,651	24.41	6.28	0	39
	FG	4,507	24.80	6.41	0	39
	GH	5,561	26.02	6.13	0	39
	I	6,388	27.45	5.60	0	39
	J	1,510	27.93	5.81	0	39
	N	23	19.78	6.91	8	31
	O	7	15.14	4.26	10	20
	R	1,150	23.48	6.59	0	39
	S	0	0.00	0.00	0	0
	V	0	0.00	0.00	0	0
8	A	6,770	24.33	8.95	0	53
	B	3,839	28.15	9.10	0	53
	CD	3,896	29.54	9.07	0	54
	DE	4,460	31.70	9.13	0	53
	FG	4,676	31.84	9.13	0	53
	GH	6,641	33.74	9.17	5	53
	I	7,327	35.91	8.19	6	53
	J	2,786	37.60	7.97	2	54
	N	20	35.20	7.92	18	45
	O	3	34.67	5.13	29	39
	R	755	29.30	9.24	6	52
	S	1	16.00	0.00	16	16
	V	18	31.11	8.67	16	45

[+] N = majority of students in private schools, O and S = schools receiving special populations; R = charter schools; V = vocational schools

7.2 Equating and Scaling Procedures

Item Calibration

In order to accomplish equating and scaling for science in grades 4 and 8, the NJ ASK 2015 operational tests were calibrated using Winsteps (Linacre, 2006)[8]. Winsteps is designed to produce a single scale by jointly analyzing data resulting from students' responses to both multiple-choice and open-ended items. Multiple-choice items were calibrated using the Rasch model (Rasch, 1960[9], Wright & Stone, 1979[10]; Anderich, 1978[11]), while the partial credit model (Masters, 1982)[12] was used for open-ended items.

Rasch scaling is "a method for obtaining objective, fundamental, linear measures from stochastic observations of ordered category responses" (Linacre, 2006, p.10). In the Rasch model, the probability of a correct response to item *i* given θ is:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

where θ = latent trait or ability level and
b_i = the difficulty parameter for item *i*.

Similar to other IRT models (Hambleton, 1989[13]; Hambleton & Swaminathan, 1985[14]), the Rasch model requires an assumption of unidimensionality (Smith, Jr., 2004)[15]. Unidimensionality means that all items measure a single construct. If the data fit the model, the measurement units (logits) have the desirable property of maintaining the same size over the whole continuum. These interval measures may then be used in subsequent statistical analyses that assume an interval scale (Smith, Jr., 2004). Also, like other IRT models, the Rasch model allows for separability of parameter estimates (Hambleton, Swaminathan, & Rogers, 1991[16]; van der Linden & Hambleton, 1997[17]). That is, the ability estimates of persons are freed from the distributional properties of the specific items attempted. Likewise, the estimated difficulties of items are freed from the distributional

[8] Linacre, J. M. (2006). A *User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs*. Chicago
[9] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
[10] Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
[11] Anderich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
[12] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.
[13] Hambleton, R. K (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Washington, DC: American Council on Education.
[14] Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory. Principles and Applications*. Boston: Kluwer.
[15] Smith, Jr. E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith, Jr. & R. M. Smith, Introduction to Rasch measurement: Theory, models and applications. Maple Grove, MN: JAM Press.
[16] Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). Fundamentals of Items Response Theory. Newbury Park, CA: Sage Publications.
[17] van der Linden, W. J. & Hambleton, R. K. (1997). Handbook of Modern Item Response Theory. New York: Springer-verlagVerlag.

properties of specific examinees used in the calibration. This property was useful for the Braille and large-print test score scaling described below in Section 7.4.

The following steps detail the procedure used to equate the NJ ASK 2015 tests to the base scale.

(1) Calibrate the 2015 assessment without constraint

The first step in equating the NJ ASK 2015 tests to the base scale was to create data files for each test by grade and content area. These data were imported into Winsteps where an unconstrained, or free, calibration was conducted. This free calibration allowed Winsteps to calculate the 2015 Rasch values based strictly on how the examinees and items performed without regard to previous performance.

(2) Examine the stability of the common items

A set of items largely from the 2014 NJ ASK Operational tests calibrated to the base scale were selected as the potential anchor items for the NJ ASK 2015 Operational tests in science. These anchor items were internal – contributing to the students' total score. These items were sound in statistical characteristics and representative of the test contents. Given that these tests were not released, these anchors can still be considered secure. The anchor sets included both multiple-choice and constructed response items.

Assessing the stability of the common items was accomplished through comparing the constrained Rasch values from prior usage with the unconstrained 2015 Rasch values of the common items for all content area and grade combinations. The stability of common items refers to the expectation that common items function the same way for the groups involved in an equating study. It is recommended that the stability of common items be examined visually and statistically (Kolen and Brennan, 2004)[18]. In the NJ ASK 2015 anchor evaluation, both visual and analytical methods were applied. Anchor items were evaluated using both the 0.3 Criterion and the Delta Plot. In order for an anchor item to be considered for removal from the anchor set, the absolute logit difference between the adjusted 2015 "free" calibrations and the 2012 "base" calibrations has to be greater than 0.3 logits (Miller, Rotou, & Twing. 2004)[19] and more than two standard deviations away from the line of best fit fitted to the base year and current year normalized inverse p-values in the delta plot. Figures 7.2.1 to 7.2.2 present scatter plots by content area and grade that were used for visual examination. Tables supporting the analytical examination are presented in the appendices of the 2015 Equating Report, Equating of NJ ASK Regular, Braille, large-print, and Alternate Test Forms.

[18] Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practice*. NY: Springer.

[19] Miller, G.E., Rotou, O., & Twing, J.S. (2004). Evaluation of the 0.3 logits screening criterion in common item equating. *Journal of Applied Measurement*, 5(2), 172-177.

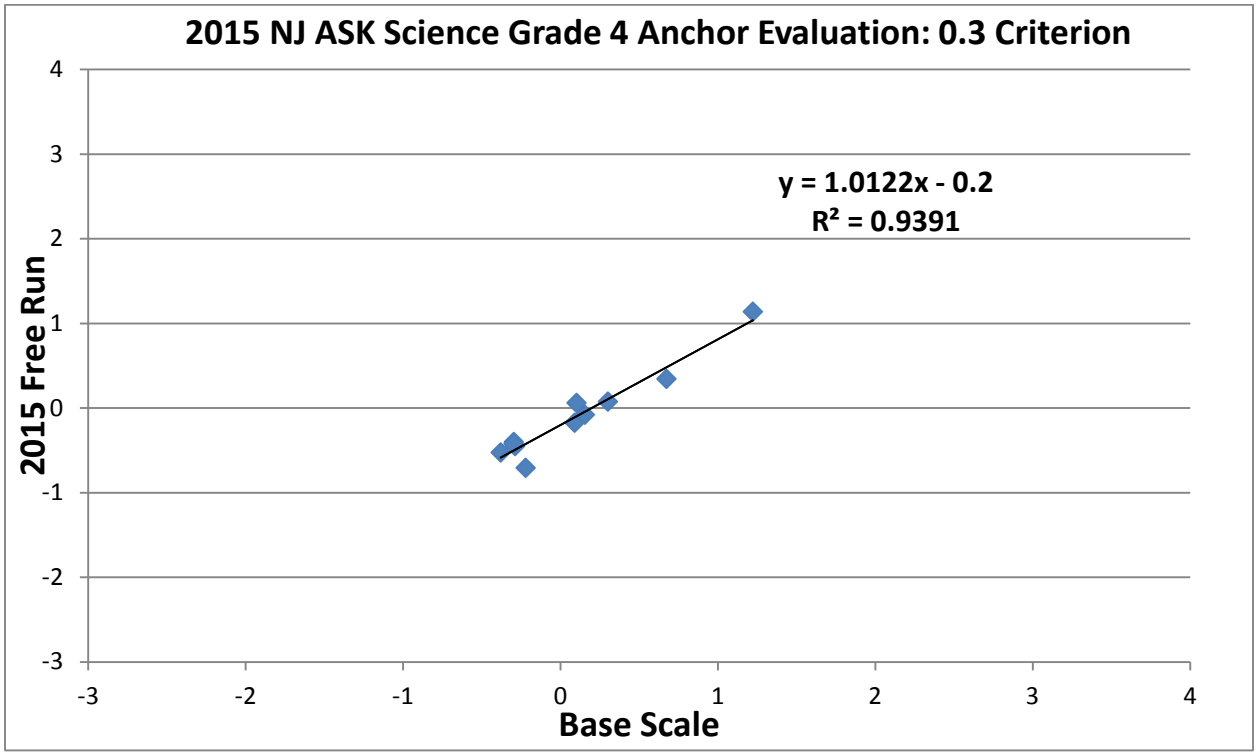

Figure 7.2.1: Scatter Plot of Anchor Items – Science Grade 4

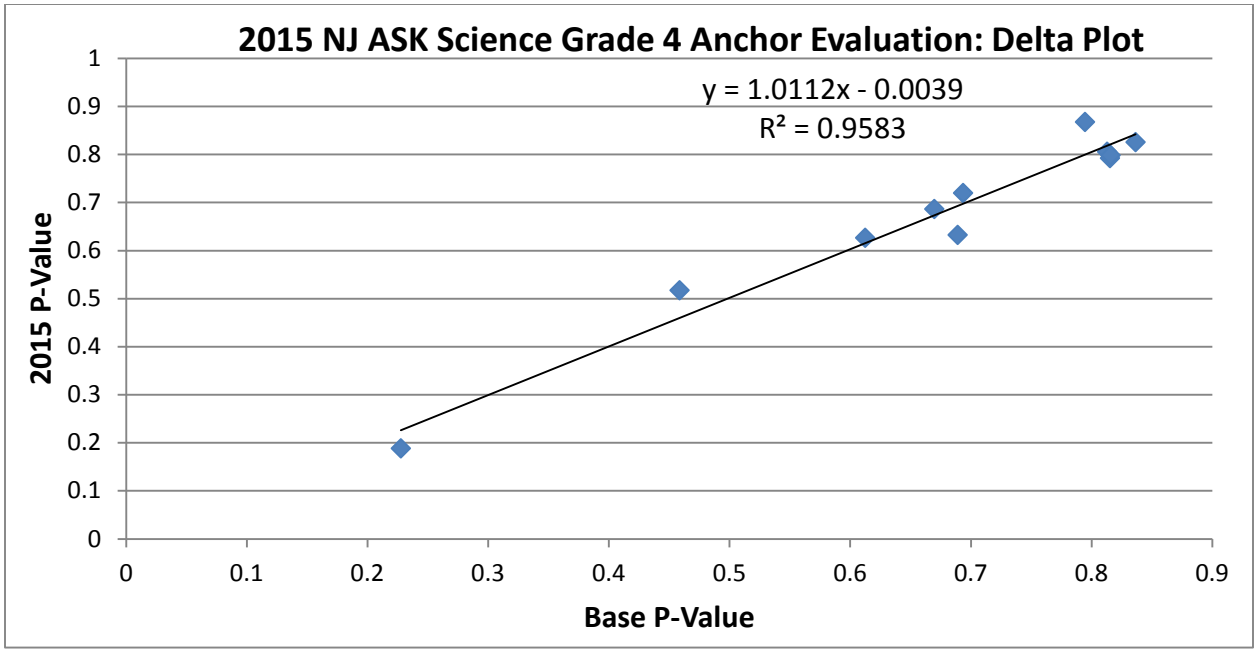

Figure 7.2.2: Scatter Plot of Anchor Items – Science Grade 4 Delta Plot

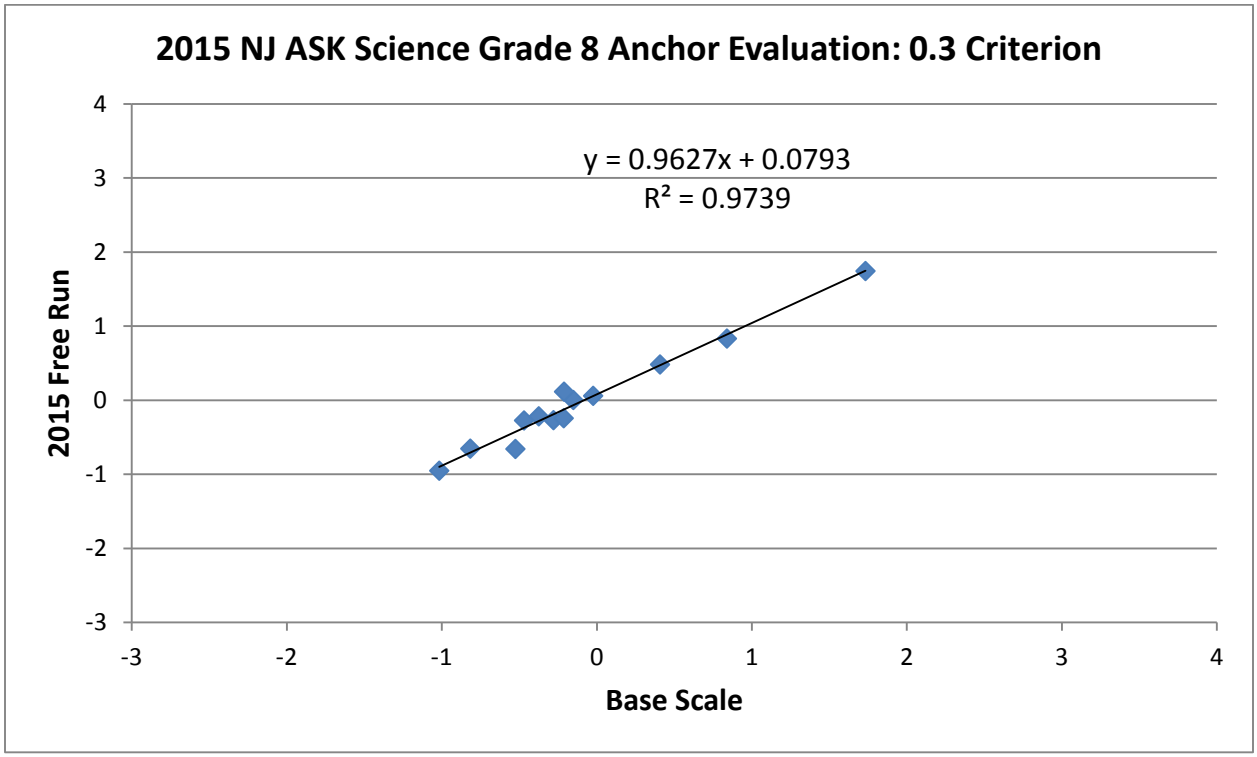

Figure 7.2.3: Scatter Plot of Anchor Items – Science Grade 8

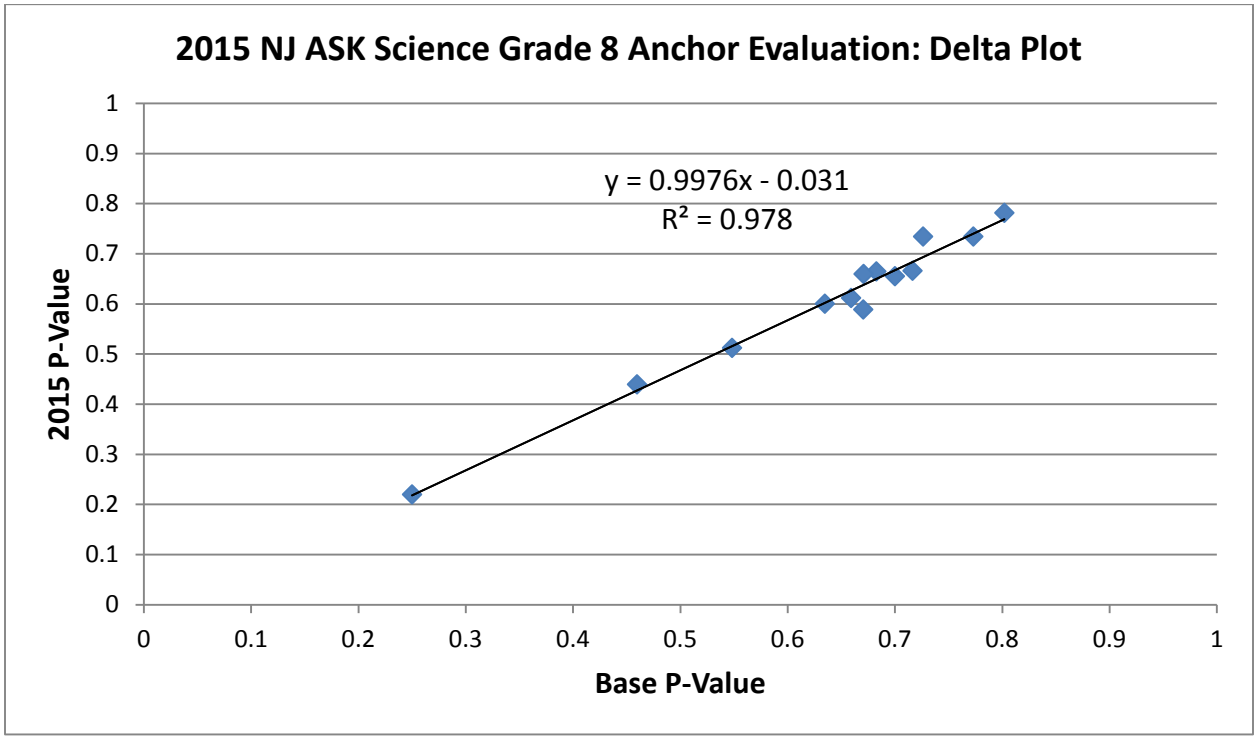

Figure 7.2.4: Scatter Plot of Anchor Items – Science Grade 8 Delta Plot

(3) Equate the 2015 assessments to the "base" scale

It was assumed that the latent traits measured by the 2015 operational tests and the "base" tests were the same. Note that all anchor items for 2015 operational assessments were selected from previous administrations where the items were already calibrated to the base scale. Given the fact that common anchor items were used and the blueprint and item specifications were the same, it appears reasonable to assume that the underlying latent trait or construct measured by each assessment was the same. To equate the 2015 assessments to the "base" scale, the Rasch values (difficulties and Rasch-Anderich thresholds for the open-ended items) of the common items were fixed to the "base" scale. This resulted in a raw score to theta conversion on the "base" scale for the 2015 assessment (i.e., the 2015 assessments were scaled to the "base" metric).

(4) Assess the model fit

Winsteps was able to produce an ability estimate (theta) for every possible number correct, raw score total as one or more examinees obtained a perfect score on each CR item in science. Table 7.1.1 shows the number of examinees used for the calibrations by grade and content area.

Table 7.2.1 summarizes Infit and Outfit statistics for the NJ ASK 2015 tests. The Infit statistic is more sensitive to unexpected behavior affecting responses near an examinee's ability level while the Outfit statistic is more sensitive to unexpected behavior by examinees far from their ability level (see Winsteps Manual, pp.199-202). Infit and Outfit can be expressed as a mean square (MNSQ) statistic or on a standardized metric (ZSTD). MNSQ values are more oriented toward practical significance, whereas Z values are more closely related to statistical significance. As a rule of thumb, the Rasch model fits the data well when the item mean square ("Infit") indices are within the range of 0.70 to 1.30. The tables indicate that the majority of Infit indices are in the range of 0.70 to 1.30. The Infit statistics for science were within the recommended range. The Rasch model fits the data very well with an average Infit of approximately 1.0.

Table 7.2.1: Summary of the Infit and Outfit Statistics by Grade—Science

				INFIT		OUTFIT	
Grade		**Measure**	**Model Error**	**MNSQ**	**ZSTD**	**MNSQ**	**ZSTD**
4	**Mean**	0.19	0.01	0.99	-1.26	0.99	-1.52
	SD	0.33	0.00	0.08	7.75	0.14	8.20
	Max	1.22	0.01	1.21	9.90	1.42	9.90
	Min	-0.38	0.00	0.78	-9.90	0.74	-9.90
8	**Mean**	-0.08	0.01	1.00	0.80	1.00	0.79
	SD	0.59	0.00	0.07	8.30	0.11	8.14
	Max	1.73	0.01	1.15	9.90	1.27	9.90
	Min	-1.02	0.01	0.85	-9.90	0.73	-9.90

The Item Parameter tables located in Appendix G contain the displacement statistics for the common items generated from the anchor calibrations. The displacement statistic is a measure of the size of the change in the parameter estimate that would be observed in the next iteration if the targeted parameter were unconstrained and all other parameter estimates were held constant at current values. A large displacement value indicates lack of convergence, or the presence of

anchored or targeted values. It is recommended that "random displacements of less than 0.50 logits are unlikely to have much impact in a test instrument" (Linacre, 2006, p. 280).[41] The tables in Appendix G show that all displacement statistics of the common items are smaller than 0.50, indicating the anchored calibrations converged well.

7.3 Summary of Cut Scores

Total scores for NJ ASK 2015 were reported in scale scores with a range of 100–300. Note that scores of 100 and 300 are a theoretical floor and ceiling and may not actually have been observed for some grades and/or content areas. However, for each test, for a perfect raw score, the scale score was set to 300. A scale score of 200 represents the cut point between Partially Proficient (PP) and Proficient (P), while a scale score of 250 represents the cut point between Proficient and Advanced Proficient (AP). The scale score ranges are as following:

Partially Proficient	100 to 199
Proficient	200 to 249
Advanced Proficient	250 to 300

To produce the scale score ranges above, linear transformations were applied to theta estimates and scale scores. The following formula was used to obtain the slopes and intercepts for the transformation functions:

$$sc(y) = \left[\frac{sc(y_2) - sc(y_1)}{\theta_2 - \theta_1}\right] y + \left\{(sc(y_1) - \left[\frac{sc(y_2) - sc(y_1)}{\theta_2 - \theta_1}\right]\theta_1\right\}$$

where θ_1 and θ_2 are person parameter estimates that correspond to the cut score points, and $sc(y_1)$ and $sc(y_2)$ are scale score points. The above formula was adopted from Kolen and Brennan (2004, p. 337)[42]. New standards have been set for various grades and content areas of the NJ ASK assessment at different times. Regardless of when new standards have been set, $sc(y_1)$ has always been 200 and $sc(y_2)$ has always been 250. Slopes and intercepts of the transformation functions are summarized in Table 7.3.1. The following sections specify how these slopes and intercepts were used to generate the scale scores in each content area and grade level. The complete raw to scale score conversion tables can be found in Appendix H.

[41] Linacre, J. M. (2006). A user's guide to Winsteps Ministep Rasch-model computer program. Chicago: MESA Press.

[42] Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practice*. NY: Springer.

Table 7.3.1: Slope and Intercept of Theta to Scale Score Transformation

		Proficient			Advanced Proficient		
Test	**Grade**	**RS**	**Theta**	**SS**	**RS**	**Theta**	**SS**
Science	4	15	-0.0144	200	26	0.6408	250
	8	24	-0.2076	200	41	1.3220	250

After calibrating the 2015 Science assessments in grades 4 and 8 to the base scales, the raw score to theta conversion table produced by Winsteps was used to develop the raw to scale score tables. Using the slopes and intercepts shown in Table 7.3.1, linear transformations of the Winsteps theta estimates were conducted to produce the final science scaled scores for grades 4 and 8.

NJDOE policy requires that scaled scores below 100 are rounded up to 100 and scaled scores above 300 are rounded down to 300. Additionally, NJDOE requires that the following rules apply:

1. If a raw score maps to an unrounded scaled score that is greater than 199.499 and less than or equal to 200.000, it will serve as the proficient cut score. Otherwise, the highest raw score that maps to a scaled score less than or equal to 199.499 will serve as the cut score. The selected cut score will be assigned a value of exactly 200.
2. If a raw score maps to an unrounded scaled score that is greater than 249.499 and less than or equal to 250.000, it will serve as the advanced cut score. Otherwise, the highest raw score that maps to a scaled score less than or equal to 249.499 will serve as the cut score. The selected cut score will be assigned a value of exactly 250.
3. In the unlikely event that two scores fall >199.499 and <200.000 or >249.999 and <250.000, the lower of these two scores would become the cut score.
4. When the implementation of the above rounding rules results in two raw scores mapping to a rounded scaled score of 200 or two raw scores mapping to a rounded scaled score 250, the scaled score associated with the higher of the two raw scores will be adjusted upwards by one (1) scaled score. Thus, the higher of the two raw scores at the proficient or advanced proficient cut point will be mapped to a rounded scaled score of 201 or 251, respectively.

7.4 Equating and Scaling for Braille, Large-Print, and Alternate forms

This section describes the equating procedures for scores from the Braille, large-print, and Alternate forms of the NJ ASK 2015. Items that the Commission of the Blind deemed inappropriate were not scored for student with visual impairments. Braille and large-print test forms were constructed by removing the inappropriate items from the corresponding regular test forms. No items were removed from the 2015 NJ ASK large-print forms for any grade. All required modifications are summarized in Table 7.4.1. Results from these "special equatings" appear in Appendix H.

Table 7.4.1: Special Equatings

	Total Items	Items Dropped	Coefficient Alpha	Raw Score Range	Item Mean
Science Grade 4					
Regular	35		0.83	0-39	0.185
Special Administration	35		n/a	0-39	0.121
Science Grade 8					
Regular	50		0.89	0-54	-0.078
Braille	49	#31	n/a	0-53	-0.091
Special Administration	50		n/a	0-54	-0.187

Braille and Large-Print Tests. Several assumptions had to be made in order to equate the scores of the Braille and large-print tests to the scores of the regular test. First, it was assumed that the latent trait measured by the Braille tests and the regular test was the same. Given the fact that the same items were used across the tests within each content area, with the exception of the removed items, it seemed reasonable to assume that changes to item format or item presentation would not greatly change the overall latent trait or construct measured by each assessment.

A second, stronger assumption, however, was that item parameters across the tests within each content area were identical. This of course is a very strong assumption considering the different item formats across the tests. However, this assumption was necessary because sample sizes for the Braille tests were too small to get reliable parameter estimates. Moreover, making these assumptions is considered common and current best practice for these populations. Because the first assumption noted above is reasonable, i.e., for each test the mathematics assessment measures mathematics, the following steps for equating the Braille tests to the regular tests were used:

- Conduct an anchored item calibration. The items in Table 7.4.1 were removed and the parameters and steps of the Braille test items were fixed with the estimates resulting from the corresponding regular test items.
- Transform the theta metric to the scale score metric. Because the theta values obtained from the anchored calibration and those obtained from the regular test score calibration are on the same metric, the transformation functions applied to the regular test scores can be applied to the Braille test scores.

- Create raw score to scale score look-up tables for each Braille test. In cases where no raw score corresponds to the cut scale scores (200 for Proficient and 250 for Advanced Proficient), the raw score point immediately below the cut score was assigned as the cut point scale score.

Alternate Forms. A security violation or a deviation from the standardized administration procedures of the NJ ASK is defined as a testing breach. An alternate form of the test was constructed for each grade of the 2015 NJ ASK tests with items from previous administrations. The alternate form was administered to all students of a given grade affected when a breach occurred. Equating of the alternate forms was conducted in a manner similar to that used with the Braille**.**

PART 8: RELIABILITY

The New Jersey Department of Education is required by federal law to ensure that the instruments it uses to measure student achievement for school accountability provide reliable results. This section shows that results of the NJ ASK Science 2015 grade 4 and 8 measure student achievement in a reliable manner. The size of the measurement error associated with test scores is reasonable and can be taken into account when interpreting the scores for individual students.

8.1 Classical Reliability Estimates of Test Scores

Reliability and Measurement Error

A detailed review of the relationship between reliability and measurement can be found in the 2009 NJ ASK Technical Report (PTM 1507-34), Part 8 , Section 8.1.

Raw Score Internal Consistency

Consistency of individual student performance was estimated using Cronbach's coefficient alpha. Coefficient alpha is conceptualized as the proportion of total raw score variance that may be attributed to a student's true score variance. Ideally, more score variance should be attributable to true test scores than to measurement error. Alpha is an appropriate index of internal consistency for use on untimed tests such as NJ ASK.

Separate analyses were performed for each grade level. Both MC and CR items scores were used in the computations. Coefficient alpha can be interpreted as a lower bound to reliability and was estimated using the following formula:

$$\alpha_{\text{Cronbach}} = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n}\sigma_{Y_i}^2}{\sigma_X^2}\right],$$

where n is the number of items, $\sigma_{Y_i}^2$ is the variance of item i, and σ_X^2 is the variance of total score. SEMs were calculated using the following formula:

$$SEM = S_X\sqrt{1-\alpha_{\text{Cronbach}}},$$

where S_X is the standard deviation of observed total scores.

Table 8.1.1 summarizes coefficient alpha and SEMs by content and form. All groups are included in the content area N counts as well as reported separately in Table 8.1.1. Tables 8.1.2 through 8.1.3 summarize coefficient alpha and SEMs of content clusters by test. Tables 8.1.2a – 8.1.3a summarize coefficient alpha and SEMs by item type at the test and cluster level for

MC items. Reliability coefficients are commonly low when based upon small numbers of items.[43] Note that Spanish test takers are included in Tables 8.1.2 through 8.1.3.

Table 8.1.1: Summary of Coefficient Alpha and SEM by Grade and Content Area

Grade	Test	Form*	N-Count	Alpha	SEM
4	Science	OP	98,073	0.83	2.75
	Science, Spanish	SP	833	0.75	2.83
	Science, Special Education	OP	16,158	0.85	2.83
	Science, Current Limited English Proficient	OP	3,769	0.80	2.85
8	Science	OP	99,814	0.88	3.30
	Science, Spanish	SP	1,190	0.71	3.35
	Science, Special Education	OP	15,506	0.86	3.33
	Science, Current Limited English Proficient	OP	2,903	0.78	3.34

* OP: Operational Test; SP: Spanish Version; N-counts were insufficient to produce values for Braille and large-print.

Table 8.1.2: Grade 4 Coefficient Alpha and SEM for Clusters

	Number of Items				
	MC	OE	Max Points	Alpha	SEM
Science	**33**	**2**	**39**	**0.83**	**2.75**
Life Science	15	0	15	0.73	1.61
Physical Science	9	1	12	0.50	1.56
Earth Science	9	1	12	0.61	1.60
Knowledge	*4*	*0*	*4*	*0.38*	*0.80*
Application	*29*	*2*	*35*	*0.81*	*2.63*

Table 8.1.2.a: Grade 4 Coefficient Alpha and SEM for MC Clusters*

	Number of Items	Alpha	SEM
Science MC	**33**	**0.82**	**2.47**
Life Science	15	0.73	1.61
Physical Science	9	0.45	1.33
Earth Science	9	0.60	1.31
Knowledge	*4*	*0.38*	*0.80*
Application	*29*	*0.80*	*2.34*

* Except where a cluster contains no constructed responses, the statistics apply to item types that comprise parts of tests or parts of clusters.

[43] See the following for a further discussion of the relationship between measures of reliability and numbers of items: Traub, R. E. and Rowley, G. L. (2008). Understanding reliability. *Instructional topics in educational measurement.* Madison, WI: National Council on Measurement and Education 176-177.

Table 8.1.3: Grade 8 Coefficient Alpha and SEM for Clusters

	Number of Items				
	MC	**OE**	**Max Points**	**Alpha**	**SEM**
Science	**48**	**2**	**54**	**0.88**	**3.30**
Life Science	20	0	20	0.75	1.98
Physical Science	14	1	17	0.71	1.86
Earth Science	14	1	17	0.69	1.86
Knowledge	*6*	*0*	*6*	*0.54*	*1.06*
Application	*42*	*2*	*48*	*0.87*	*3.12*

Table 8.1.3.a: Grade 8 Coefficient Alpha and SEM for MC Clusters*

	Number of Items	**Alpha**	**SEM**
Science MC	**48**	**0.88**	**3.08**
Life Science	20	0.75	1.98
Physical Science	14	0.70	1.64
Earth Science	14	0.65	1.70
Knowledge	*6*	*0.54*	*1.06*
Application	*42*	*0.85*	*2.89*

* Except where a cluster contains no constructed responses, the statistics apply to item types that comprise parts of tests or parts of clusters.

8.2 Reliability of Performance Classifications

Two measures of reliability are presented below in Table 8.2.1. Stratified Alpha is used to assess the reliability of the different item types, e.g., multiple choice and constructed response. Stratified Cronbach Alpha can be calculated using the following formula:

$$\text{Stratified } \alpha = 1 - \frac{\sum \sigma_i^2 \left(1 - \rho_{ii'}\right)}{\sigma_t^2}$$ [44]

where

σ_i^2 = variance of score on cluster *i*,

σ_t^2 = variance of total score, and

$\rho_{ii'}$ = reliability coefficient of score on cluster *i*.

[44] Maryland school assessment – Reading: Grades 3 through 8 (2004). http://www.marylandpublicschools.org/NR/rdonlyres/26BD65BE-6F27-4F35-8699-139BC98BF99F/8812/2004_MDTech_Reading_Report_3.pdf

Reliability index for proficiency classifications (kappa) is an estimate of how reliably the test classifies students into the performance categories (Partially Proficient, Proficient, and Advanced Proficient). Kappa was computed with the BB-CLASS program (Brennan, 2004)[45] that is based on the beta-binomial model. Coefficient kappa is given by:

$$\kappa = \frac{\varphi - \varphi_C}{1 - \varphi_C},$$

where φ is the probability of a consistent classification and φ_C is the probability of a consistent classification by chance. A classification consistency index can be regarded as the percentage of examinees that would hypothetically be assigned to the same achievement level if the same test was administered a second time or an equivalent test was administered under the same conditions.

Table 8.2.1 displays two cut scores for each grade. The lower cut score is the minimum raw score required to be classified as proficient and the higher cut score is the minimum raw score required for classification as advanced proficient.

Table 8.2.1: Consistency Indices for Performance Levels—Science

Grade	Alpha	SEM	Stratified Alpha	Cut Score	Kappa	φ
4	0.83	2.73	0.84	15, 26	0.56	0.74
8	0.88	3.26	0.89	24, 41	0.61	0.77

Item Maps and Test Information Functions

Item maps for science are presented in Appendix I. These figures indicate how well the item difficulties and person ability levels match.

The test information function is another method of assessing the reliability or the precision of a test. The reliability of a test, however, is not uniform across the entire range of test scores. The highest and lowest scores typically have more measurement error than do scores in the middle of the range because more examinees tend to score in the middle of the score range. With item response theory (IRT), the item and test information functions can assess test reliability across the range of scores. The item information function is the probability of a correct response multiplied by the probability of an incorrect response. Item information functions (I_{ij}) for every item (j) at every level of student ability (i) can be calculated for each item using the following equation:

$$\boldsymbol{I_{ij}\left(\theta_i, \delta_j\right) = P_{ij} * \left(1 - P_{ij}\right)}$$

[45] Brennan, R. L. (2004). Manual for BB-CLASS: A computer program that uses the beta-binomial model for classification consistency and accuracy (version 1). CASMA Research Report 9. Iowa City, IA.

The total test information function for a given ability level is simply the sum of all the item information functions for that ability level (Lord & Novick, 1968[46]; Hambleton, 1989[47]). Computing an item information function for each ability level and summing these functions to derive test information functions for each ability level, one can plot the total information function for a test, as shown in Figures 8.2.2–8.2.15. Each item yields the greatest amount of information at the point at which the difficulty of the item (δ_j) is equal to the ability of the student (θ_i).

These figures illustrate the level of information at theta values ranging from -4 to $+4$. As shown, the information or reliability of the test scores is lower at the extremes and higher in the middle. More information implies less measurement error. Ideally, the Proficient cut score would occur at the peak of the information function where the most information and the least measurement error occur. Thus, scores in this area yield the most error free measurements. Two arrows appear in each TIF shown below. The arrow with the lower value on the *x*-axis (measure) represents the proficient cut and the arrow with the higher value represents the advance proficient cut. As depicted in these figures, the Proficient cut scores for science occur near the peak of information.

[46] Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley.
[47] Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: American Council on Education and Macmillan.

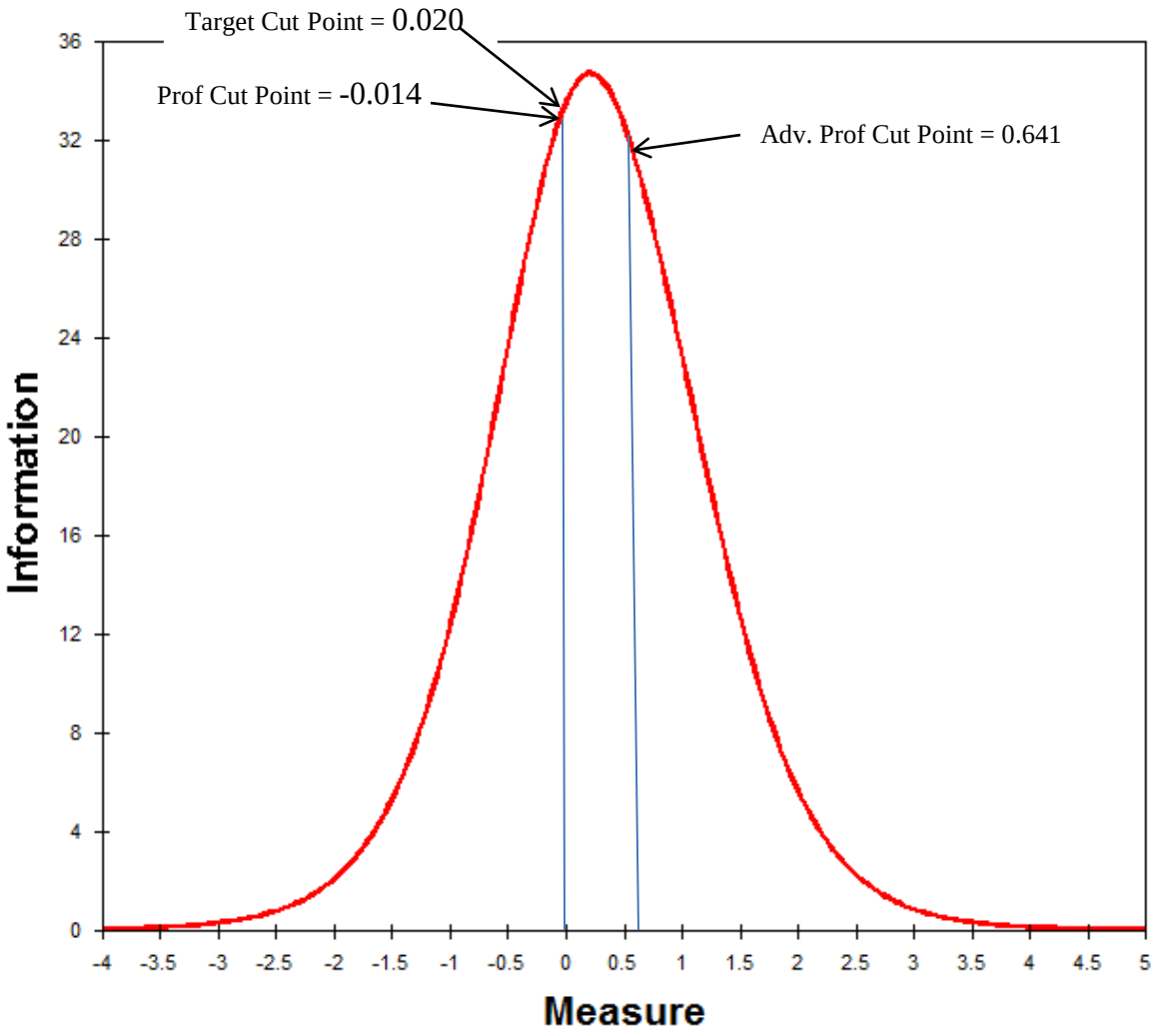

Figure 8.2.1: Grade 4 Science Test Information Function

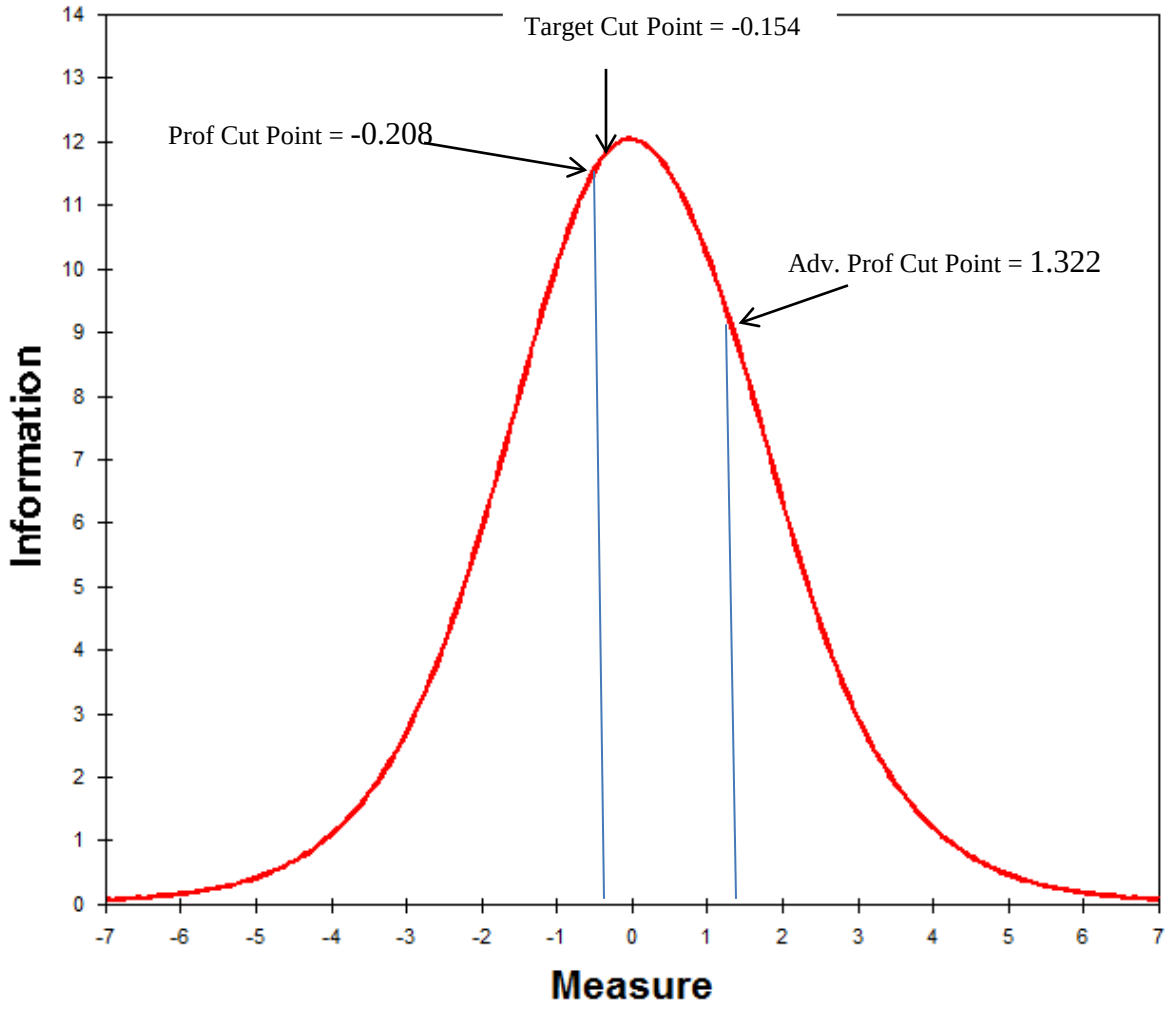

Figure 8.2.2: Grade 8 Science Test Information Function

8.3 Conditional Estimate of Error at Each Cut-Score

The NJ ASK 2015 Science 4 and 8 raw cut scores and the corresponding conditional standard error of measurement (CSEM) are summarized in Table 8.3.1. WINSTEPS calculates the standard error at each score point using item response theory and the information function. The equation for the standard error at each value of theta (ability) is given by

$$SE\left(\hat{\theta}\right) = \frac{1}{\sqrt{I\left(\theta\right)}}$$

where $I(\theta)$ is the information function for a test at θ.

For the Rasch model, the information provided by a test at θ is the sum of the item information functions at θ. Interpolation of the raw cut scores were used to derive the CSEM from the standard error associated with the theta at each cut score.

Table 8.3.1: Raw Score Cut Scores with Conditional Standard Error of Measurement

Grade		Proficient	Advanced Proficient
4	Cut score	15	26
	(CSEM)	(2.87)	(2.74)
8	Cut score	24	41
	(CSEM)	(3.44)	(3.01)

8.4 Rater Reliability

Tables 8.4.1 and 8.4.2 show the percentages of constructed-response items scored with exact agreement, adjacent agreement, and resolution needed by grade. Raters used scoring rubrics with a score range of 0 to 3. There were no half points assigned for any of the CR items. Only 10% of the constructed-response items were read by a second rater. The purpose of the second reading for the constructed-response items was to investigate the consistency between raters for the NJ ASK Science 2015. As shown in the tables below, the exact agreement rates ranged from 69.0% to 92.6% at the test level. An adjacent score is a score assigned by the second rater that is no more than ±1 score point from the score assigned by the first rater. The adjacent agreement rates ranged from 7.2% to 26.4%. On average, approximately 1.5% of the scores require resolution by a third rater.

Table 8.4.1: Grade 4 Scoring Consistency of Constructed-Response Items

	% Raters in Exact Agreement	% Raters in Adjacent Agreement	% Resolution Needed
Science Total	**80.8**	**16.8**	**2.2**
CR 1	69.0	26.4	4.2
CR 2	92.6	7.2	0.2

Table 8.4.2: Grade 8 Scoring Consistency of Constructed-Response Items

	% Raters in Exact Agreement	% Raters in Adjacent Agreement	% Resolution Needed
Science Total	**82.5**	**16.4**	**0.8**
CR 1	77.4	21.2	1.2
CR 2	87.6	11.6	0.4

PART 9: VALIDITY

The *Standards for Educational and Psychological Testing* states, "Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing program. 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).[48] While this section summarizes evidence supporting claims as to the validity of NJ ASK performance scores, many parts of this technical report provide appropriate evidence for validity. Given the procedural and empirical evidence available and the rationale presented below, valid performance standards-based interpretations and uses of the scores are generally supported.

The following section begins with a review of important federal statutes requiring the NJ ASK Science 4 and 8 and explains the purposes and intended uses of performance test scores, suggesting the value implications of performance scores for schools, teachers, students, and parents. Content-related evidence supporting validity is presented in terms of the adequacy and appropriateness of the state content standards and the representation of the content standards on the tests. Then, validity evidence based on the internal structure of NJ ASK is provided through a correlational analysis of NJ ASK content clusters with each other. Reference to specific Standards within the *Standards for Educational and Psychological Testing* are provided where appropriate.

9.1 Content and Curricular Validity[49]

Content validity of a test refers to the degree to which the content of a test is congruent with the purpose of testing, as defined by the curriculum. Baker and Linn (2002)[50] 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 two sections help answer these two very important questions and also address Standard 1.6 of the *Standards for Educational and Psychological Testing*.

Appropriateness of Content Definition

In 1996, the New Jersey State Board of Education adopted the New Jersey Core Curriculum Content Standards, 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 thirteen years of public education. Since

[48] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington: APA.

[49] 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 (page 18).

[50] Baker, E. L., & Linn, R. L. (2002). Validity Issues for Accountability Systems. Center for the Study of Evaluation. Technical Report 585, Los Angeles, CA.

the adoption of those standards, the NJ DOE has continuously engaged in discussions 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 NJ DOE 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.

The review process required by the State Board 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 New Jersey Core Curriculum Content Standards provide the level of specificity and depth of content that better prepares 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.

Since the adoption of the original 1996 New Jersey Core Curriculum Content Standards (CCCS), the New Jersey State Board of Education 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.

Adequacy of Content Representation

The content-related evidence of validity includes the extent to which the test items represent these specified content domains and cognitive dimensions. Adequacy of the content representation of the NJ ASK is critical because the tests must provide an indication of student progress toward achieving the knowledge and skills identified in the CCCS, and the tests must fulfill the requirements under NCLB.

Adequate representation of the content domains defined in the CCCS is assured through use of a test blueprint and a responsible test construction process. New Jersey performance standards, as well as the CCCS, are taken into consideration in the writing of multiple-choice and constructed-response items and constructed-response rubric development. Each test must align with and proportionally represent the sub-domains of the test blueprint. Evidence to support the above is described in Part 2, Test Development Process, and Part 6, Item and Test Statistics. Part 2 provides evidence that the NJ test specifications were followed in the development of test items; alignment of items with the CCCS; and the review of items by NJ content experts, teachers, and Sensitivity committee. Item writers were recruited with specific qualifications and were trained and validated before they started writing items. Tables 2.1.3 through 2.1.5 in Part 2 provide a comparison of target test construction maps to actual test maps for science. The tables indicate that the target blueprint representation in terms of number of items and score points for each sub-domain was adequately met.

The CCCS are represented on each test by balancing sub-domain coverage on each test, by proportionally representing items corresponding to Partially Proficient, Proficient, and Advanced Proficient performance categories on each test, and by matching item format to the requirements of the content and standards descriptions.

Adherence to Test Specification

MI followed statistical and content specifications to make sure that the 2015 NJ ASK assessments are valid. The statistical specification described the psychometric characteristics of the items included in the 2015 assessments. The primary statistical targets used for NJ ASK test assembly were the p-value estimates also called proportion correct or item difficulty, the point bi-serial correlation which is a measure of how well the items discriminate among test takers and is related to the overall reliability of the test, and proportion correct value which is an indication of test difficulty. Similarly, the minimum target value for a proportion-correct was set at 0.25 and maximum was set at 0.95. In addition, content experts made sure that the items selected for the 2015 NJ ASK tests were free from poor model fit and differential item functioning when they were first field tested.

Content specification pertains to the adherence to content representation across content standards and sub-domains. MI developed all test items to conform to the NJ ASK content standards and test blueprints. Part 2 of this document describes test development activities of the NJ ASK assessments. The actual test maps for NJ ASK are shown in Tables 2.1.3 through 2.1.5 and the 2015 tests configurations are shown in Tables 2.1.1 and 2.1.2. The Tables indicate that the 2015 assessments adequately adhere to the test blueprints.

Qualified item writers who were familiar with the NJ state specifications and populations were recruited and trained for item writing. Detail procedures are described in Part 2 of this document. The items were reviewed by NJ's content review committee and sensitivity review committee comprised of NJ teachers making sure that the items align with the state standards and are free from bias for a specific group of student population. Only items accepted from the committees were added to the bank for possible use in future operational tests.

Test Administration and Accommodations

Part 3 of this report describes the test administration process. In order to securely administer the tests the test administrator were trained on the process and test co-coordinator manuals were produced for the step by step process. A test form distribution list was prepared for the proportional representation of DFG prior to the test administration. Tests were administered under standard condition. For the case where the standard condition was compromised or breached a separate alternate form was developed for each grade and content area. The alternate test forms matched test blueprint and difficulty of the regular tests.

The tests were also translated into large-print and Braille, and a separate Spanish version of the test was developed for state approved accommodations for LEP. Similarly, various accommodations (see Appendix C for the list of accommodations) were offered for students identified by IEP and 504 plans to minimize assessment ambiguity and inaccuracy.

9.2 Construct Validity[51]

Because the NJ ASK testing program assesses student performance in several content areas using a variety of testing methods, it is important to study the pattern of relationships among the content areas and testing methods. Therefore, this section addresses evidence based on responses and internal structure. One method for studying patterns of relationships to provide evidence supporting the inferences made from test scores is the multi-trait matrix. Tables 6.3.1 through 6.3.2 summarize Pearson correlation coefficients among test content domains and clusters by grade level. The correlations between clusters within a content area were generally found to be higher than the correlations between clusters across the content areas.

Scaling and Performance Score Reporting

The NJ ASK Science grade 4 and 8 are scaled in several ways: raw score points, Item Response Theory (IRT), and performance standard level (based on scale-score cuts). New Jersey actively promotes the use of performance level results, reporting them annually on each content test at the student, school, district and state levels. Individual student and average scale scores are also used, but should play a secondary role, generally interpreted with reference to their distance from performance-score cut points. Test results are reported for students as a whole as well as by student group including sex, ethnicity, disability, English language proficiency, migrant status, and DFG. Scores are reported to schools and districts in the annually published reports (see Part 10: Reporting).

NJ ASK performance scores indicate that an individual student performs at the Partially Proficient, Proficient, and Advanced Proficient level in a content area. Performance standard descriptions associated with each level provide details of the performance that students have met or exceeded. No stakes for students or teachers are attached by the state to student-level scores. Teachers are counseled to interpret individual student scores only in the context of other assessment results and their own experience.

[51] Standard 1.11 – If the rationale for a test use or interpretation depends on premises about the relationships among parts of the test, evidence concerning the internal structure of the test should be provided.
Standard 1.12 – When interpretation of subscores, score differences, or profiles is suggested, the rationale and relative evidence in support of such interpretation should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.

9.3 Criterion-Related Validity

Validity evidence related to other Standards is listed below:

Standard 1.5[52]

- The composition of the sample of examinees from which validity evidence was obtained is described in detail in Part 6 – Item and Test Statistics, including major relevant sociodemographic characteristics. This information is imbedded within the Tables of Part 6. These tables also provide descriptive statistics for number correct raw score and for scale scores. Statistics include N-counts, means, standard deviations, minimum and maximum values, and a variety of data disaggregations, including student demographic group and DFG.

Standard 1.7[53]

- Standard setting procedures, including the selection process and the characteristics of judges, is described in detail in Part 5.
- The NJ ASK Science 2015 4 and 8 constructed-response items required hand scoring. The processes of selecting and training scorers, reading and scoring papers, and monitoring scoring are described in detail in Part 4.

Standard 1.13[54]

- The conditions under which the data were collected are described in Part 2. Information about the administration of NJ ASK is available in the *New Jersey Assessment of Skills & Knowledge Spring 2015 Test Coordinator Manual Grades 4 & 8* which can be found at https://www.measinc.com/nj/NJASK/Default.aspx

[52] Standard 1.5 - The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics.

[53] Standard 1.7 – When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications, and experience, of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

[54] Standard 1.13 - When validity evidence includes statistical analyses of test results, either alone or together with data on other variables, the conditions under which the data were collected should be described in enough detail that users can judge the relevance of the statistical findings to local conditions. Attention should be drawn to any features of a validation data collection that are likely to differ from typical operational testing conditions and that could plausibly influence test performance.

PART 10: REPORTING

Previously, scores were reported in two cycles. Data for Cycle I reporting were produced after districts submit record changes. Data for Cycle II reporting were produced after the completion of automatic rescoring of the constructed-response items and writing tasks.

Beginning in 2011, only one reporting cycle was used. Reports were produced after districts submitted record changes and the automatic rescoring of the constructed response items and writing tasks were completed.

10.1 Reports

While there is only one reporting cycle currently, the same reports were produced as in previous years, with one exception. The Preliminary Performance by Demographic Group—School and District are no longer produced.

The following reports were produced separately for each grade.

- Student Sticker (1 per student)
- Individual Student Report (ISR) (2 per student)
- Student Roster – Science (Grade 8 only)
- All Sections Roster
- Performance by Demographic Group –School
- Performance by Demographic Group –District
- Cluster Means Report

Brief descriptions of each report and the score reporting categories can be found in the *2015 NJ ASK Score Intrepretation Manual* (PTM 1510.52), Part 3, at http://www.state.nj.us/education/assessment/njask/ref/SIM15.pdf.

10.2 State Summary Reporting

The state summary data file contains the same type of test results based on the performance by demographics reports at the state, district, and school levels. This data file is available in text and in Excel formats.

APPENDIX A
Field Test Form Distribution Plan

Table A.1: NJ ASK 2015 Grade 4 Test Form Distribution Plan

Form	DFG													Grand
Grade 4	A	B	CD	DE	FG	GH	I	J	N	O	R	S	V	Total
A	1,375	450	516	493	489	526	1844	45		9	58			**5,805**
B	1,647	276	738	365	534	1223	560	218		22	165	22		**5,770**
C	1,032	503	294	1664	622	761	713	98			27			**5,714**
D	953	548	445	930	748	721	1212	102		27	54			**5,740**
E	422	1,082	360	422	1089	818	1033	369		106	170	40		**5,911**
F	765	484	1148	499	681	582	1170	89		44	281			**5,743**
G	427	788	770	436	988	476	1332	160	22	93	182	67		**5,741**
H	739	566	529	509	703	668	1736			88	191	18		**5,747**
J	1,682	423	365	374	338	931	1264	240		18	129			**5,764**
K	1,184	231	361	1010	1015	530	890	80	458	0	85			**5,844**
L	769	588	419	720	611	1357	836	169		85	280	13		**5,847**
M	1,059	623	699	352	762	983	1028	151		18	151			**5,826**
N	695	614	383	1156	561	538	1180	734		0	53			**5,914**
O	560	757	276	1064	796	543	867	526		35	326	9		**5,759**
P	1,139	992	578	520	743	543	618	543		27	205			**5,908**
R	1,474	992	401	485	493	570	1308	71		9	67	9		**5,879**
S	912	254	257	694	668	1816	583	472		44	76			**5,776**
T	645	339	1064	1024	792	530	744	414		80	98	58		**5,788**
Grand Total	**17,479**	**10,510**	**9,603**	**12,717**	**12,633**	**14,116**	**18,918**	**4,481**	**480**	**705**	**2,598**	**236**	**0**	**104,476**

[1] DFG, or district factor group, is a district-level socioeconomic measure based upon 2000 U.S. Census data, with A referring to districts at the lowest end and J at the highest end. N = districts with too low a percentage of students in public schools for a DFG value to be assigned. O and S = schools receiving special populations. R = charter school. V= vocational school

Table A.2: NJ ASK 2015 Grade 8 Test Form Distribution Plan

Form	DFG													Grand
Grade 8	A	B	CD	DE	FG	GH	I	J	N	O	R	S	V	Total
A	1,495	467	521	561	494	517	1,868			13	45			**5,981**
B	1,349	290	796	450	743	1,287	596	182		31	187	22		**5,933**
C	752	561	272	1,709	587	841	1,157	93			22			**5,994**
D	1,121	214	473	1,011	748	832	1,389	134		13	54			**5,989**
E	361	925	325	471	1,081	788	1,039	378		142	169	76		**5,755**
F	770	855	1,162	471	708	459	1,247	120		48	134			**5,974**
G	339	890	890	445	1,108	557	1,309	76	13	195	49	102	22	**5,995**
H	650	468	566	593	743	792	1,891			222	27	27		**5,979**
J	1,540	449	196	401	1,152	605	1,282	196		18	120			**5,959**
K	970	245	338	975	1,126	499	1,236	89	294	13				**5,785**
L	730	512	530	761	551	1,317	758	169		102	299	18		**5,747**
M	912	712	904	152	765	948	1,290	187		58	22			**5,950**
N	636	668	418	1,295	373	565	1,174	743		22				**5,894**
O	414	711	285	1,215	845	347	913	538		75	473	36		**5,852**
P	1,113	814	512	569	667	534	752	627		22	196			**5,806**
R	1,202	1,095	396	503	547	627	1,256	76		9	71	13		**5,795**
S	779	258	266	721	721	1,754	458	516		58	263			**5,794**
T	454	321	1073	1,117	828	561	828	383		208	107	75		**5,955**
Grand Total	**15,587**	**10,455**	**9,923**	**13,420**	**13,787**	**13,830**	**20,443**	**4,507**	**307**	**1,249**	**2,238**	**369**	**22**	**106,137**

[1] DFG, or district factor group, is a district-level socioeconomic measure based upon 2000 U.S. Census data, with A referring to districts at the lowest end and J at the highest end. N = districts with too low a percentage of students in public schools for a DFG value to be assigned. O and S = schools receiving special populations. R = charter school. V= vocational school

APPENDIX B
CHECKLIST FOR FORMS DEVELOPMENT

Table B-1: Checklist for Forms Development

Item Data
Average Target Rasch value based on standard setting year*
As many items as possible have a p-value above 0.35 and below 0.90
As many items as possible have a pt. bis above 0.25
No item was used as a sample item.

* Targets are the theta cuts from the standard setting year.

Item Distribution
Item standards are distributed equally throughout the test
There are a variety of indicators assessed in each standard
MC items are generally in passage order, and OE items are at the end of the passage sets. WT items are in the appropriate places.
Answer key distribution is nearly equal between answer choices: A B C D
Having more than 2 MC items in a row with the same answer is avoided.

Name, Gender, and Ethnicity Distributions
Check gender distribution (number of passages or prompts which have a male and/or female): Male Female Both
Check ethnicity distribution (number of passages or prompts): Caucasian_Hispanic Asian_African American Other_____
There are NOT two or more items in the same session that have similar contexts.
There are NOT two or more items with similar answers or answer choices.
Sample items and test items do NOT clue each other.
Items do NOT have any fairness or sensitivity related to the names and contexts of the items.

APPENDIX C

MODIFICATIONS OF TEST ADMINISTRATION PROCEDURES FOR LIMITED ENGLISH PROFICIENT, SPECIAL EDUCATION STUDENTS, AND STUDENTS ELIGIBLE UNDER SECTION 504 OF THE REHABILITATION ACT OF 1973

Accommodations for Limited English Proficient (LEP) Students

NCLB prohibits exemptions from testing based on limited English proficient (LEP) status. As permitted by NCLB, Spanish forms of the test were available for LEP students whose dominant language was Spanish, as identified by school districts. For those LEP students who were tested in English, one or more of the following accommodations were permitted.

- Additional time up to 150% of the administration times indicated
- Translation of directions only to the student's native language.
- Translations of passages, items, prompts, and tasks are NOT permitted
- Use of a bilingual dictionary, preferably one normally used by the student as part of the instructional program.

Accommodations for Special Education students, and students eligible under section 504

In accordance with the Individuals with Disabilities Education Act (IDEA), students who are receiving special education services must participate in each subject area of the age-appropriate statewide assessment with the following exception:

Students with disabilities shall participate in the Alternate Proficiency Assessment in each content area where the nature of the student's disability is so severe that the student is not receiving instruction in any of the knowledge and skills measured by the general statewide assessment and the student cannot complete any of the types of questions on the assessment content area(s) even with accommodation and modifications. (New Jersey Administrative Code Chapter 6A:14-4.11[a]2)

Districts may use modifications of test administration procedures when administering the NJ ASK to special education students or to students eligible under Section 504 of the Rehabilitation Act of 1973. Decisions about participation and accommodations/modifications are made by the Individualized Education Program (IEP) or 504 team. Information about test content and item types from the test specifications booklets can be used to make this determination. Modifications in the areas listed below may be used separately or in combination.

Any accommodations or modifications of test administration procedures for students eligible for special education under the IDEA or eligible under Section 504 of the Rehabilitation Act of 1973 must be specified in the student's IEP or 504 accommodation plan. Accommodations or modifications must be consistent with the instruction and assessment procedures used in the student's classroom. Students eligible for modifications under Section 504 may not be classified but do have a permanent or temporary impairment in a major life function (for example: performing manual tasks, walking, seeing, hearing, speaking, etc.).

ACCEPTABLE ACCOMMODATIONS OR MODIFICATIONS

Code

A. Setting Accommodations

1. Administering the assessment:
    a. individually in a separate room
    b. in a small group in a separate room
    c. in the resource room
    d. in a special education classroom
    e. using carrels
    f. at home or in a hospital (this depends on the nature of the assessment task)
2. Seating the student in the front of the room near the examiner or proctor
3. Seating the student facing the examiner or proctor
4. Providing special lighting
5. Providing special furniture (e.g., desks, trays, carrels)

B. Scheduling Accommodations

1. Adding time as needed
2. Providing frequent breaks
3. Terminating a section of the test when a student has indicated that he/she has completed all the items he/she can. The test examiner must ensure that the student has attempted all items in a section since items are not ordered by difficulty. When this accommodation is used, the test must be administered in a small group or individually to avoid distraction.

C. Test Materials Modifications

1. Administering the large-print version of test materials
2. Administering the Braille version of test materials

D. Test Procedures Modifications

1. Administration modifications
    a. reading directions aloud
    b. reading test items (not reading passages) aloud
    c. providing and ensuring that amplification (hearing aid and/or FM system) is in working order
    d. using a sign language or cued speech interpreter for administration of directions or items but not reading passages
    e. masking a portion of the test booklet and/or answer folder to eliminate visual distractors or providing reading windows
    f. repeating, clarifying, or rewording directions
    g. providing written directions on a separate sheet or transparency
    h. using an examiner who is familiar with the student

i. using an examiner who can communicate fluently in sign language (American Sign Language or a form of Manually Coded English)
j. providing manipulatives for math items
k. using graph paper for math section
l. using a Braille ruler and talking calculator
m using tactile or visual cues for deaf or hard of hearing students to indicate time to begin, time remaining, and time to end a particular part of the test

2. Response modifications

a. having an examiner record the student's identifying information on the answer folder, or grid corrections to the pre-ID label
b. dictating oral responses to a scribe (person who writes from dictation) – student must indicate all punctuation and must spell all key words
c. using a Braille writer to record responses
d. signing responses to a sign language interpreter (student must indicate all punctuation and must spell all key words)
e. recording responses on a word processor
f. using large-face calculators
g. using talking calculators
h. providing an Augmentative Communication device
i. using a larger diameter or modified special grip #2 pencil
j. masking portions of the answer folder to eliminate visual distractors
k. marking answers in the test booklet (an examiner would transfer the answers to an answer folder)
l. Allowing separate additional continuation pages for writing tasks. These pages MUST be properly marked to link them to the correct student for credit.

OTHER CONSIDERATIONS

Ensure that:

a. any medication has been appropriately adjusted to prevent interference with the student's functioning.
b. eyeglasses are used, if needed.
c. hearing aids, FM systems, Augmentative Communication devices, word processors, or other equipment are functioning properly.
d. source and strength of light are appropriate.
e. all students can clearly see and hear the examiner.
f. all deaf or hard of hearing students who communicate aurally/orally are watching the examiner when instructions are given.
g. responses to CR items and writing tasks which are written or typed on separate sheets of paper by students eligible for this accommodation are labeled with student data paper-clipped to the front of the answer folder, and placed in the fluorescent orange envelope provided. Copies of these pages

should be made and retained on file by the school district until scores are received.

h. students using the large-print test booklets
    1. mark their answers in the large-print answer folder. All responses must be transcribed into the regular answer folder provided in the large print kit.
    2. may be instructed to skip items identified in the LP instructions. The spaces for these items must be left blank on the student's answer folder (included in the large-print kit).
    3. who dictate responses on CR items and writing tasks indicate all punctuation and spell all key words.

i. students using the Braille test booklets
    1. are instructed to bring a Braille ruler and a talking calculator to the test session.
    2. are instructed to skip dropped items identified in the Braille instructions. The spaces for these items must be left blank on the student transcription answer folder (included in the Braille kit).
    3. have answer folders transcribed from the Braille version by the examiner.
    4. dictate their answers to the examiner or use a device that produces Braille. For dictations and responses recorded in Braille:
        - Students must indicate all punctuation and must spell all key words.
        - Examiners must transcribe the Braille responses into the regular answer folder included in the Braille kit.

j. students who communicate in sign language
    1. have an interpreter to translate oral directions and test items (but not the Reading passages in the English Language Arts section of the test). The interpreter should be able to communicate in the mode used by the student, American Sign Language or a form of Manually Coded English. The interpreter should be instructed to interpret so as not to give the answer to the student through the use of a particular sign or finger spelling.
    2. using American Sign Language for CR and writing task responses will sign the responses to the interpreter who will interpret them into spoken English and a scribe will record the responses in the answer folder.
    3. using Signed English or cued speech will sign/cue to the interpreter who will transliterate (word for word) into spoken English and a scribe will record the responses.

.

APPENDIX D
SCORING RUBRICS

Scoring Rubrics

Table D.1: NJ ASK Generic Science Rubric

3-Point Response	Student response is reasonably complete, clear, and satisfactory.
2-Point Response	Student response has minor omissions and/or some incorrect or non-relevant information.
1-Point Response	Student response includes some correct information, but most information included in the response is either incorrect or not relevant.
0-Point Response	Student attempts the task but the response is incorrect, irrelevant, or inappropriate.

APPENDIX E
STANDARD SETTING
PERFORMANCE LEVEL DESCRIPTORS

Standard Setting

Demographic background information of PLD panelists and standard setting participants from the 2008 and 2009 meetings can be found in Appendix E1 of the 2009 NJ ASK Technical Report (PTM 1507-34),

New Jersey Assessment of Skills and Knowledge (NJ ASK)
and Grade Eight Proficiency Assessment (GEPA)
Performance Level Descriptors
Science
Grade 8

Proficient

The Proficient student can recognize the structural levels of living things. This student knows that some traits of organisms are beneficial and some detrimental. This student can interpret visual and textual data to understand the relationship within a food web and the interdependence of living and nonliving systems.

The proficient student can recognize the effect force has on an object, trace the flow of energy through a system, and use the properties of matter to identify and separate materials. This student can understand different types of energy and use information from data charts to interpret relationships and predict outcomes.

The proficient student can recognize the existence of a relationship between the moon and tides, recognize the different characteristics of the planets in the solar system, and understand the natural forces that change the surface of the Earth, including chemical and physical weathering.

Advanced Proficient

The advanced proficient student can support scientific conclusions with valid contextual and visual data and make predictions based on the interactions of living things. This student is able to use interpretive skills to analyze visual and textual data in order to solve problems dealing with the application of force and energy.

The advanced proficient student understands the difference between types of energy waves and can recognize and apply experimental principles and empirical data.

The advanced proficient student can recognize the nature of the tides' relationship to Earth, Sun, and moon; interpret topographical maps; and identify the steps in the process of weathering and erosion.

APPENDIX F
SCALE SCORE CUMULATIVE FREQUENCY DISTRIBUTIONS

Table F.1: Science Grade 4

Science Grade 4								
		All Students* Cumulative		Female Cumul.	Male Cumul.	Afr. A. Cumul.	Hispanic Cumul.	White Cumul.
Raw Score	Scale Score	#	%	%	%	%	%	%
1	100	3	0.00	0.00	0.00	0.00	0.00	0.00
2	102	9	0.01	0.01	0.01	0.02	0.00	0.01
3	118	28	0.03	0.03	0.03	0.07	0.03	0.02
4	130	75	0.08	0.06	0.09	0.19	0.10	0.03
5	140	167	0.17	0.13	0.21	0.44	0.24	0.06
6	148	369	0.37	0.32	0.42	0.92	0.56	0.14
7	155	696	0.70	0.62	0.79	1.70	1.09	0.26
8	162	1,197	1.21	1.06	1.35	2.93	1.91	0.45
9	168	1,929	1.95	1.69	2.19	4.56	3.15	0.70
10	173	2,862	2.89	2.49	3.26	6.65	4.76	1.04
11	179	4,021	4.06	3.57	4.51	9.06	6.79	1.51
12	184	5,526	5.58	5.06	6.06	12.24	9.39	2.11
13	188	7,257	7.33	6.75	7.87	15.92	12.36	2.84
14	193	9,269	9.37	8.74	9.94	19.79	15.71	3.85
15	200	11,671	11.79	11.20	12.33	23.94	19.58	5.23
16	202	14,279	14.43	13.87	14.93	28.60	23.63	6.75
17	206	17,234	17.41	16.96	17.82	33.43	28.12	8.68
18	210	20,579	20.79	20.51	21.03	38.71	32.92	11.05
19	215	24,216	24.47	24.44	24.47	44.06	37.95	13.89
20	219	28,308	28.60	28.78	28.41	49.56	43.38	17.26
21	223	32,579	32.92	33.41	32.42	54.98	48.86	21.06
22	228	37,090	37.47	38.27	36.69	60.10	54.20	25.36
23	232	41,862	42.30	43.42	41.20	65.24	59.54	30.17
24	237	46,946	47.43	48.86	46.05	70.12	65.08	35.47
25	241	52,268	52.81	54.71	50.98	75.09	70.36	41.32
26	250	57,738	58.34	60.51	56.24	79.59	75.20	47.78
27	251	63,372	64.03	66.48	61.67	83.76	79.93	54.57
28	256	68,933	69.65	72.02	67.37	87.54	84.13	61.36
29	262	74,438	75.21	77.62	72.89	90.75	88.16	68.20
30	267	79,669	80.49	82.72	78.35	93.45	91.33	74.99
31	274	84,318	85.19	87.14	83.32	95.49	94.07	80.95
32	280	88,548	89.47	91.02	87.97	96.97	96.31	86.48
33	288	92,087	93.04	94.26	91.86	98.33	97.81	91.22
34	296	94,853	95.84	96.71	94.99	99.12	98.91	94.85
35	300	96,810	97.81	98.34	97.31	99.61	99.49	97.34
36	300	98,038	99.05	99.31	98.81	99.83	99.80	98.90
37	300	98,637	99.66	99.78	99.54	99.93	99.93	99.62
38	300	98,918	99.94	99.96	99.92	100.00	99.99	99.93
39	300	98,974	100.00	100.00	100.00	100.00	100.00	100.00

Table F.2: Science Grade 8

Science Grade 8								
		All Students* Cumulative		Female Cumul.	Male Cumul.	Afr. A. Cumul.	Hispanic Cumul.	White Cumul.
Raw Score	Scale Score	#	%	%	%	%	%	%
1	100	4	0.00	0.00	0.00	0.01	0.00	0.01
2	100	6	0.01	0.00	0.01	0.00	0.00	0.01
3	106	12	0.01	0.01	0.02	0.03	0.01	0.01
4	116	18	0.02	0.01	0.03	0.04	0.03	0.00
5	125	36	0.04	0.02	0.05	0.10	0.04	0.02
6	132	88	0.09	0.04	0.13	0.23	0.10	0.06
7	138	196	0.19	0.13	0.25	0.47	0.28	0.09
8	143	387	0.38	0.23	0.53	0.99	0.57	0.16
9	148	701	0.69	0.43	0.95	1.66	1.12	0.29
10	152	1,158	1.15	0.75	1.52	2.88	1.84	0.44
11	157	1,817	1.80	1.25	2.31	4.47	2.91	0.69
12	161	2,671	2.64	1.93	3.31	6.43	4.27	1.04
13	164	3,750	3.71	2.81	4.56	8.78	6.10	1.51
14	168	5,015	4.96	3.86	6.00	11.56	8.23	2.05
15	171	6,456	6.39	5.16	7.54	14.77	10.54	2.70
16	175	8,050	7.96	6.71	9.14	17.88	13.21	3.51
17	178	9,787	9.68	8.31	10.97	21.33	16.13	4.37
18	181	11,703	11.58	10.23	12.85	24.76	19.43	5.36
19	184	13,740	13.59	12.41	14.71	28.34	22.96	6.45
20	187	15,931	15.76	14.76	16.70	32.28	26.34	7.73
21	190	18,214	18.02	17.17	18.81	35.65	30.14	9.16
22	193	20,719	20.50	19.81	21.14	39.63	33.94	10.83
23	195	23,242	23.00	22.55	23.41	43.79	37.60	12.56
24	200	25,900	25.63	25.39	25.83	47.92	41.47	14.47
25	201	28,708	28.40	28.43	28.36	51.77	45.27	16.72
26	204	31,707	31.37	31.73	31.01	55.47	49.10	19.31
27	207	34,816	34.45	35.08	33.83	58.97	53.00	22.12
28	209	38,030	37.63	38.72	36.58	62.62	56.85	25.11
29	212	41,360	40.92	42.28	39.61	66.18	60.56	28.35
30	215	44,828	44.35	46.05	42.73	69.73	64.33	31.82
31	218	48,498	47.99	50.05	46.00	73.30	68.05	35.64
32	221	52,170	51.62	53.89	49.44	76.55	71.54	39.59
33	224	55,953	55.36	57.93	52.91	79.61	74.99	43.84
34	227	59,806	59.17	61.87	56.59	82.46	78.11	48.35
35	230	63,624	62.95	65.76	60.27	84.90	81.03	52.93
36	233	67,357	66.64	69.59	63.83	87.28	83.81	57.36
37	236	71,084	70.33	73.20	67.60	89.33	86.39	62.02
38	239	74,644	73.85	76.77	71.08	91.34	88.63	66.47
39	242	78,162	77.34	80.14	74.67	93.03	90.64	70.92
40	246	81,551	80.69	83.39	78.12	94.59	92.41	75.28
41	250	84,741	83.84	86.32	81.49	95.87	94.06	79.29
42	253	87,657	86.73	89.00	84.57	96.82	95.53	83.08
43	257	90,387	89.43	91.34	87.61	97.71	96.69	86.62
44	261	92,866	91.88	93.49	90.36	98.38	97.66	89.91
45	266	94,891	93.89	95.17	92.67	98.89	98.28	92.52
46	271	96,676	95.65	96.65	94.71	99.30	98.95	94.74
47	276	98,147	97.11	97.75	96.50	99.62	99.36	96.58
48	282	99,237	98.19	98.66	97.74	99.79	99.66	97.91
49	289	100,040	98.98	99.28	98.70	99.89	99.84	98.86
50	298	100,578	99.51	99.64	99.40	99.97	99.93	99.42
51	300	100,884	99.82	99.87	99.77	99.99	99.96	99.75
52	300	101,017	99.95	99.96	99.94	100.00	99.98	99.93
53	300	101,061	99.99	99.99	99.99	100.00	100.00	99.99
54	300	101,069	100.00	100.00	100.00	100.00	100.00	100.00

APPENDIX G
ITEM PARAMETERS TABLES

Table G.1: Science Grade 4: Item Parameters

				INFIT		OUTFIT			Score
Item[1]	Measure	Anchor	Error	MNSQ	ZSTD	MNSQ	ZSTD	Displace	Corr.
1	-0.38069	Yes	0.00726	1.05	5.22	1.09	5.80	0.04	0.26
2	-0.10120	No	0.00635	1.06	9.47	1.13	9.90	0.00	0.22
3	-0.01620	No	0.00615	1.13	9.90	1.29	9.90	0.00	0.15
4	0.08889	Yes	0.00594	0.87	-9.90	0.80	-9.90	-0.08	0.42
5	-0.28795	Yes	0.00692	0.95	-6.94	0.87	-9.90	0.02	0.38
6	-0.00712	No	0.00613	1.21	9.90	1.42	9.90	0.00	0.06
7	0.05014	No	0.00601	0.96	-6.69	0.94	-6.63	0.00	0.36
8	-0.29665	Yes	0.00695	0.98	-2.94	0.85	-9.90	0.08	0.42
9	-0.02310	No	0.00617	0.88	-9.90	0.82	-9.90	0.00	0.45
10	0.19292	No	0.00578	0.95	-9.90	0.92	-9.90	0.00	0.38
11	-0.22222	Yes	0.00670	0.78	-9.90	0.74	-9.90	-0.32	0.23
12	0.20690	No	0.00576	0.97	-6.56	0.94	-8.90	0.00	0.36
13	-0.29951	Yes	0.00696	0.97	-3.19	0.88	-8.91	0.06	0.39
14	0.15393	Yes	0.00584	0.99	-1.65	0.98	-2.74	-0.06	0.30
15	0.01180	No	0.00609	1.00	0.68	1.03	3.40	0.00	0.30
16	0.18370	No	0.00580	1.05	9.75	1.07	9.71	0.00	0.27
17	0.26772	No	0.00569	1.00	-0.61	1.01	1.24	0.00	0.33
18	0.30174	Yes	0.00566	1.07	9.90	1.09	9.90	-0.05	0.23
19	0.47272	No	0.00555	1.00	1.06	1.01	1.36	0.00	0.32
20	0.10277	No	0.00592	1.03	6.43	1.09	9.90	0.00	0.28
21	0.67288	Yes	0.00556	0.98	-5.87	0.98	-3.16	-0.15	0.38
22	0.20342	No	0.00577	0.95	-9.90	0.93	-9.90	0.00	0.39
23	0.75848	No	0.00561	0.95	-9.90	0.98	-4.20	0.00	0.36
24	0.09521	No	0.00593	1.03	5.90	1.03	3.97	0.00	0.28
25	0.40794	No	0.00558	0.96	-9.90	0.94	-9.90	0.00	0.37
26	0.55826	No	0.00554	1.08	9.90	1.11	9.90	0.00	0.23
27	0.32717	No	0.00564	0.97	-6.06	0.95	-7.52	0.00	0.36
28	0.10166	Yes	0.00592	1.10	9.90	1.12	9.90	0.14	0.31
29	0.21259	No	0.00576	0.89	-9.90	0.84	-9.90	0.00	0.45
30	0.21797	No	0.00575	1.01	1.56	1.01	1.40	0.00	0.32
31	0.31506	No	0.00565	0.93	-9.90	0.91	-9.90	0.00	0.40
32	0.41467	No	0.00558	0.99	-1.51	0.99	-1.46	0.00	0.33
33	0.00243	No	0.00611	0.89	-9.90	0.81	-9.90	0.00	0.44
34	1.22248	Yes	0.00351	0.98	-2.68	0.93	-7.01	0.10	0.39
35	0.56999	No	0.00323	1.13	9.90	1.14	9.90	0.00	0.41

[1] Item indicates the order in which items were entered in Winsteps

Table G.2: Science Grade 8: Item Parameters

Item[1]	Measure	Anchor	Error	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	Displace	Score Corr.
1	0.12818	No	0.01084	1.02	5.52	1.01	2.36	0.00	0.33
2	-0.90151	No	0.01240	1.00	-0.46	1.03	2.67	0.00	0.31
3	-0.75143	No	0.01204	0.91	-9.90	0.87	-9.90	0.00	0.42
4	-0.15118	No	0.01107	0.99	-2.04	0.97	-4.32	0.00	0.36
5	-0.02648	No	0.01095	1.09	9.90	1.12	9.90	0.00	0.25
6	-0.29371	No	0.01124	1.00	-0.30	1.00	0.50	0.00	0.34
7	-0.46921	Yes	0.01150	1.00	0.20	1.02	1.93	0.12	0.37
8	-0.27901	Yes	0.01122	0.98	-4.42	1.01	2.02	-0.07	0.34
9	0.23948	No	0.01079	1.03	8.31	1.03	6.40	0.00	0.32
10	-0.09482	No	0.01101	1.04	8.36	1.02	3.22	0.00	0.31
11	0.06402	No	0.01088	1.03	7.95	1.04	6.01	0.00	0.32
12	-0.16693	No	0.01109	1.15	9.90	1.21	9.90	0.00	0.19
13	-0.68573	No	0.01190	0.97	-4.82	0.98	-2.22	0.00	0.35
14	-0.86862	No	0.01232	1.10	9.90	1.27	9.90	0.00	0.19
15	0.31095	No	0.01077	1.09	9.90	1.10	9.90	0.00	0.26
16	-0.61786	No	0.01177	0.89	-9.90	0.84	-9.90	0.00	0.45
17	-0.15108	Yes	0.01107	1.01	2.36	0.99	-2.31	0.08	0.36
18	0.49350	No	0.01075	1.08	9.90	1.10	9.90	0.00	0.26
19	-0.06621	No	0.01099	0.92	-9.90	0.89	-9.90	0.00	0.44
20	0.25928	No	0.01078	1.03	7.86	1.03	5.11	0.00	0.32
21	0.63773	No	0.01077	0.95	-9.90	0.95	-9.90	0.00	0.40
22	0.07493	No	0.01088	1.07	9.90	1.09	9.90	0.00	0.28
23	-0.37497	Yes	0.01135	0.98	-4.81	0.96	-4.80	0.08	0.39
24	-0.81639	Yes	0.01219	1.07	9.90	1.20	9.90	0.09	0.26
25	0.40729	Yes	0.01075	0.98	-4.03	0.99	-2.50	0.00	0.37
26	-1.01600	Yes	0.01271	0.85	-9.90	0.73	-9.90	-0.01	0.48
27	-0.81508	No	0.01219	0.96	-5.92	0.96	-4.42	0.00	0.36
28	-0.52632	Yes	0.01160	0.85	-9.90	0.80	-9.90	-0.21	0.43
29	-0.63364	No	0.01180	0.91	-9.90	0.86	-9.90	0.00	0.43
30	0.17895	No	0.01082	1.03	8.74	1.06	9.90	0.00	0.32
31	0.54794	No	0.01075	1.10	9.90	1.13	9.90	0.00	0.24
32	-0.21340	Yes	0.01114	1.05	9.90	1.06	8.72	-0.11	0.26
33	-0.48392	No	0.01152	1.02	4.31	1.10	9.90	0.00	0.30
34	-0.59378	No	0.01172	0.91	-9.90	0.85	-9.90	0.00	0.44
35	-0.32842	No	0.01129	1.03	5.89	1.02	2.46	0.00	0.31
36	-0.83826	No	0.01224	0.85	-9.90	0.74	-9.90	0.00	0.49
37	0.65914	No	0.01078	1.04	9.34	1.06	9.90	0.00	0.31
38	0.15654	No	0.01083	1.05	9.90	1.06	9.90	0.00	0.30
39	0.83862	Yes	0.01087	1.10	9.90	1.16	9.90	-0.08	0.25
40	-0.74373	No	0.01203	0.95	-9.64	0.90	-9.90	0.00	0.39
41	-0.02480	Yes	0.01095	0.98	-4.92	0.97	-5.51	0.01	0.37
42	0.36977	No	0.01076	1.01	2.49	1.01	1.18	0.00	0.34
43	0.11809	No	0.01085	0.95	-9.90	0.93	-9.90	0.00	0.41
44	0.14566	No	0.01083	0.96	-9.90	0.94	-9.90	0.00	0.40
45	0.49712	No	0.01075	1.05	9.90	1.06	9.90	0.00	0.30
46	0.04698	No	0.01090	1.00	-0.05	1.02	3.10	0.00	0.35
47	-0.24669	No	0.01118	0.93	-9.90	0.88	-9.90	0.00	0.43
48	-0.21248	Yes	0.01114	1.10	9.90	1.13	9.90	0.25	0.30
49	1.58669	No	0.00695	0.93	-9.90	0.90	-9.90	0.00	0.54
50	1.73336	Yes	0.00653	1.11	9.90	1.16	9.90	-0.05	0.44

[1] Item indicates the order in which items were entered in Winsteps

APPENDIX H
RAW SCORE TO SCALE SCORE CONVERSION TABLES

Table H.37: Science Grade 4

SCI Grade 4 OP Theta To Scale Score Table									
Raw Score	Theta	Standard Error	Slope	Intercept	Unrounded	Scale Score	Scale SE	Lower SS	Upper SS
0.0	-2.63866	1.29454	73.98639	198.50548	3.28055	**100**	**38**	100	138
1.0	-1.67439	0.50922	73.98639	198.50548	74.62341	**100**	**38**	100	138
2.0	-1.30904	0.36671	73.98639	198.50548	101.65434	**102**	**27**	100	129
3.0	-1.08718	0.30494	73.98639	198.50548	118.06896	**118**	**23**	100	141
4.0	-0.92385	0.26896	73.98639	198.50548	130.15315	**130**	**20**	110	150
5.0	-0.79239	0.24504	73.98639	198.50548	139.87940	**140**	**18**	122	158
6.0	-0.68092	0.22788	73.98639	198.50548	148.12667	**148**	**17**	131	165
7.0	-0.58308	0.21498	73.98639	198.50548	155.36550	**155**	**16**	139	171
8.0	-0.49507	0.20495	73.98639	198.50548	161.87704	**162**	**15**	147	177
9.0	-0.41440	0.19700	73.98639	198.50548	167.84552	**168**	**15**	153	183
10.0	-0.33936	0.19060	73.98639	198.50548	173.39746	**173**	**14**	159	187
11.0	-0.26873	0.18542	73.98639	198.50548	178.62312	**179**	**14**	165	193
12.0	-0.20158	0.18121	73.98639	198.50548	183.59130	**184**	**13**	171	197
13.0	-0.13717	0.17782	73.98639	198.50548	188.35677	**188**	**13**	175	201
14.0	-0.07492	0.17512	73.98639	198.50548	192.96242	**193**	**13**	180	206
15.0	**-0.01435**	**0.17303**	**73.98639**	**198.50548**	**197.44378**	**200**	**13**	**187**	**213**
16.0	0.04497	0.17148	73.98639	198.50548	201.83265	**202**	**13**	189	215
17.0	0.10339	0.17043	73.98639	198.50548	206.15493	**206**	**13**	193	219
18.0	0.16125	0.16984	73.98639	198.50548	210.43579	**210**	**13**	197	223
19.0	0.21887	0.16970	73.98639	198.50548	214.69888	**215**	**13**	202	228
20.0	0.27654	0.16999	73.98639	198.50548	218.96568	**219**	**13**	206	232
21.0	0.33455	0.17071	73.98639	198.50548	223.25763	**223**	**13**	210	236
22.0	0.39321	0.17188	73.98639	198.50548	227.59767	**228**	**13**	215	241
23.0	0.45282	0.17349	73.98639	198.50548	232.00800	**232**	**13**	219	245
24.0	0.51372	0.17557	73.98639	198.50548	236.51377	**237**	**13**	224	250
25.0	0.57625	0.17816	73.98639	198.50548	241.14014	**241**	**13**	228	254
26.0	**0.64083**	**0.18130**	**73.98639**	**198.50548**	**245.91818**	**250**	**13**	**237**	**263**
27.0	0.70790	0.18505	73.98639	198.50548	250.88045	**251**	**14**	237	265
28.0	0.77801	0.18952	73.98639	198.50548	256.06763	**256**	**14**	242	270
29.0	0.85182	0.19485	73.98639	198.50548	261.52857	**262**	**14**	248	276
30.0	0.93019	0.20125	73.98639	198.50548	267.32688	**267**	**15**	252	282
31.0	1.01426	0.20906	73.98639	198.50548	273.54692	**274**	**15**	259	289
32.0	1.10566	0.21882	73.98639	198.50548	280.30927	**280**	**16**	264	296
33.0	1.20678	0.23140	73.98639	198.50548	287.79078	**288**	**17**	271	300
34.0	1.32145	0.24826	73.98639	198.50548	296.27480	**296**	**18**	278	300
35.0	1.45612	0.27203	73.98639	198.50548	306.23854	**300**	**20**	280	300
36.0	1.62303	0.30814	73.98639	198.50548	318.58761	**300**	**20**	280	300
37.0	1.84945	0.37036	73.98639	198.50548	335.33961	**300**	**20**	280	300
38.0	2.22145	0.51322	73.98639	198.50548	362.86255	**300**	**20**	280	300
39.0	3.19419	1.29685	73.98639	198.50548	434.83207	**300**	**20**	280	300

Table H.38: Science Grade 4: Special Equating

SCI Grade 4 Special Equating Theta To Scale Score Table									
Raw Score	**Theta**	**Standard Error**	**Slope**	**Intercept**	**Unrounded**	**Scale Score**	**Scale SE**	**Lower SS**	**Upper SS**
0.0	-2.72654	1.29436	73.98639	198.50548	-3.22137	**100**	**27**	100	127
1.0	-1.76309	0.50874	73.98639	198.50548	68.06082	**100**	**27**	100	127
2.0	-1.39874	0.36601	73.98639	198.50548	95.01776	**100**	**27**	100	127
3.0	-1.17792	0.30407	73.98639	198.50548	111.35543	**111**	**22**	100	133
4.0	-1.01566	0.26796	73.98639	198.50548	123.36046	**123**	**20**	103	143
5.0	-0.88529	0.24393	73.98639	198.50548	133.00607	**133**	**18**	115	151
6.0	-0.77491	0.22668	73.98639	198.50548	141.17269	**141**	**17**	124	158
7.0	-0.67817	0.21370	73.98639	198.50548	148.33013	**148**	**16**	132	164
8.0	-0.59124	0.20362	73.98639	198.50548	154.76177	**155**	**15**	140	170
9.0	-0.51164	0.19563	73.98639	198.50548	160.65108	**161**	**14**	147	175
10.0	-0.43768	0.18921	73.98639	198.50548	166.12312	**166**	**14**	152	180
11.0	-0.36810	0.18400	73.98639	198.50548	171.27109	**171**	**14**	157	185
12.0	-0.30197	0.17979	73.98639	198.50548	176.16381	**176**	**13**	163	189
13.0	-0.23858	0.17639	73.98639	198.50548	180.85381	**181**	**13**	168	194
14.0	-0.17734	0.17369	73.98639	198.50548	185.38473	**185**	**13**	172	198
15.0	-0.11776	0.17160	73.98639	198.50548	189.79284	**190**	**13**	177	203
16.0	-0.05942	0.17005	73.98639	198.50548	194.10921	**194**	**13**	181	207
17.0	**-0.00197**	**0.16900**	**73.98639**	**198.50548**	**198.35973**	**200**	**13**	**187**	**213**
18.0	0.05493	0.16841	73.98639	198.50548	202.56955	**203**	**12**	191	215
19.0	0.11158	0.16827	73.98639	198.50548	206.76088	**207**	**12**	195	219
20.0	0.16829	0.16858	73.98639	198.50548	210.95665	**211**	**12**	199	223
21.0	0.22535	0.16933	73.98639	198.50548	215.17831	**215**	**13**	202	228
22.0	0.28308	0.17053	73.98639	198.50548	219.44955	**219**	**13**	206	232
23.0	0.34179	0.17223	73.98639	198.50548	223.79329	**224**	**13**	211	237
24.0	0.40185	0.17445	73.98639	198.50548	228.23691	**228**	**13**	215	241
25.0	0.46366	0.17725	73.98639	198.50548	232.81001	**233**	**13**	220	246
26.0	0.52769	0.18071	73.98639	198.50548	237.54736	**238**	**13**	225	251
27.0	0.59450	0.18494	73.98639	198.50548	242.49039	**242**	**14**	228	256
28.0	**0.66476**	**0.19009**	**73.98639**	**198.50548**	**247.68867**	**250**	**14**	**236**	**264**
29.0	0.73936	0.19635	73.98639	198.50548	253.20806	**253**	**15**	238	268
30.0	0.81941	0.20402	73.98639	198.50548	259.13067	**259**	**15**	244	274
31.0	0.90644	0.21350	73.98639	198.50548	265.56970	**266**	**16**	250	282
32.0	1.00258	0.22541	73.98639	198.50548	272.68275	**273**	**17**	256	290
33.0	1.11096	0.24074	73.98639	198.50548	280.70140	**281**	**18**	263	299
34.0	1.23643	0.26110	73.98639	198.50548	289.98447	**290**	**19**	271	300
35.0	1.38717	0.28940	73.98639	198.50548	301.13718	**300**	**21**	279	300
36.0	1.57829	0.33144	73.98639	198.50548	315.27746	**300**	**21**	279	300
37.0	1.84250	0.40103	73.98639	198.50548	334.82540	**300**	**21**	279	300
38.0	2.27505	0.54878	73.98639	198.50548	366.82822	**300**	**21**	279	300
39.0	3.32336	1.31615	73.98639	198.50548	444.38889	**300**	**21**	279	300

Table H.39: Science Grade 8

SCI Grade 8 OP Theta To Scale Score Table									
Raw Score	Theta	Standard Error	Slope	Intercept	Unrounded	Scale Score	Scale SE	Lower SS	Upper SS
0.0	-5.34059	1.83251	33.51206	205.16086	26.18669	**100**	**24**	100	124
1.0	-4.11917	1.01245	33.51206	205.16086	67.11899	**100**	**24**	100	124
2.0	-3.40076	0.72490	33.51206	205.16086	91.19439	**100**	**24**	100	124
3.0	-2.96960	0.59938	33.51206	205.16086	105.64345	**106**	**20**	100	126
4.0	-2.65577	0.52571	33.51206	205.16086	116.16054	**116**	**18**	100	134
5.0	-2.40602	0.47629	33.51206	205.16086	124.53017	**125**	**16**	109	141
6.0	-2.19661	0.44046	33.51206	205.16086	131.54793	**132**	**15**	117	147
7.0	-2.01486	0.41318	33.51206	205.16086	137.63875	**138**	**14**	124	152
8.0	-1.85320	0.39166	33.51206	205.16086	143.05631	**143**	**13**	130	156
9.0	-1.70674	0.37426	33.51206	205.16086	147.96449	**148**	**13**	135	161
10.0	-1.57213	0.35993	33.51206	205.16086	152.47555	**152**	**12**	140	164
11.0	-1.44696	0.34796	33.51206	205.16086	156.67025	**157**	**12**	145	169
12.0	-1.32947	0.33785	33.51206	205.16086	160.60758	**161**	**11**	150	172
13.0	-1.21828	0.32926	33.51206	205.16086	164.33379	**164**	**11**	153	175
14.0	-1.11233	0.32191	33.51206	205.16086	167.88439	**168**	**11**	157	179
15.0	-1.01077	0.31561	33.51206	205.16086	171.28788	**171**	**11**	160	182
16.0	-0.91290	0.31021	33.51206	205.16086	174.56770	**175**	**10**	165	185
17.0	-0.81814	0.30558	33.51206	205.16086	177.74330	**178**	**10**	168	188
18.0	-0.72599	0.30164	33.51206	205.16086	180.83144	**181**	**10**	171	191
19.0	-0.63603	0.29830	33.51206	205.16086	183.84618	**184**	**10**	174	194
20.0	-0.54791	0.29551	33.51206	205.16086	186.79927	**187**	**10**	177	197
21.0	-0.46128	0.29321	33.51206	205.16086	189.70242	**190**	**10**	180	200
22.0	-0.37587	0.29137	33.51206	205.16086	192.56468	**193**	**10**	183	203
23.0	-0.29140	0.28996	33.51206	205.16086	195.39545	**195**	**10**	185	205
24.0	**-0.20763**	**0.28895**	**33.51206**	**205.16086**	**198.20275**	**200**	**10**	**190**	**210**
25.0	-0.12434	0.28833	33.51206	205.16086	200.99397	**201**	**10**	191	211
26.0	-0.04130	0.28807	33.51206	205.16086	203.77681	**204**	**10**	194	214
27.0	0.04169	0.28816	33.51206	205.16086	206.55798	**207**	**10**	197	217
28.0	0.12484	0.28860	33.51206	205.16086	209.34451	**209**	**10**	199	219
29.0	0.20834	0.28939	33.51206	205.16086	212.14276	**212**	**10**	202	222
30.0	0.29240	0.29052	33.51206	205.16086	214.95979	**215**	**10**	205	225
31.0	0.37722	0.29200	33.51206	205.16086	217.80228	**218**	**10**	208	228
32.0	0.46300	0.29383	33.51206	205.16086	220.67694	**221**	**10**	211	231
33.0	0.54997	0.29603	33.51206	205.16086	223.59149	**224**	**10**	214	234
34.0	0.63835	0.29861	33.51206	205.16086	226.55328	**227**	**10**	217	237
35.0	0.72840	0.30160	33.51206	205.16086	229.57104	**230**	**10**	220	240
36.0	0.82037	0.30503	33.51206	205.16086	232.65315	**233**	**10**	223	243
37.0	0.91459	0.30895	33.51206	205.16086	235.81065	**236**	**10**	226	246
38.0	1.01140	0.31343	33.51206	205.16086	239.05496	**239**	**11**	228	250
39.0	1.11121	0.31857	33.51206	205.16086	242.39980	**242**	**11**	231	253
40.0	1.21455	0.32449	33.51206	205.16086	245.86293	**246**	**11**	235	257
41.0	**1.32203**	**0.33136**	**33.51206**	**205.16086**	**249.46481**	**250**	**11**	**239**	**261**
42.0	1.43445	0.33943	33.51206	205.16086	253.23223	**253**	**11**	242	264
43.0	1.55285	0.34901	33.51206	205.16086	257.20006	**257**	**12**	245	269
44.0	1.67858	0.36051	33.51206	205.16086	261.41353	**261**	**12**	249	273
45.0	1.81349	0.37451	33.51206	205.16086	265.93465	**266**	**13**	253	279
46.0	1.96008	0.39180	33.51206	205.16086	270.84718	**271**	**13**	258	284
47.0	2.12191	0.41352	33.51206	205.16086	276.27044	**276**	**14**	262	290
48.0	2.30417	0.44137	33.51206	205.16086	282.37834	**282**	**15**	267	297
49.0	2.51483	0.47815	33.51206	205.16086	289.43799	**289**	**16**	273	300
50.0	2.76708	0.52886	33.51206	205.16086	297.89141	**298**	**18**	280	300
51.0	3.08526	0.60395	33.51206	205.16086	308.55428	**300**	**20**	280	300
52.0	3.52330	0.73066	33.51206	205.16086	323.23390	**300**	**20**	280	300
53.0	4.25180	1.01831	33.51206	205.16086	347.64744	**300**	**20**	280	300
54.0	5.48253	1.83653	33.51206	205.16086	388.89173	**300**	**20**	280	300

Table H.40: Science Grade 8: Special Equating

SCI Grade 8 Special Equating Theta To Scale Score Table									
Raw Score	Theta	Standard Error	Slope	Intercept	Unrounded	Scale Score	Scale SE	Lower SS	Upper SS
0.0	-5.57084	1.83445	33.51206	205.16086	18.47054	**100**	**20**	100	120
1.0	-4.34459	1.01577	33.51206	205.16086	59.56470	**100**	**20**	100	120
2.0	-3.61968	0.72921	33.51206	205.16086	83.85793	**100**	**20**	100	120
3.0	-3.18247	0.60424	33.51206	205.16086	98.50973	**100**	**20**	100	120
4.0	-2.86298	0.53090	33.51206	205.16086	109.21650	**109**	**18**	100	127
5.0	-2.60791	0.48168	33.51206	205.16086	117.76442	**118**	**16**	102	134
6.0	-2.39347	0.44597	33.51206	205.16086	124.95075	**125**	**15**	110	140
7.0	-2.20698	0.41873	33.51206	205.16086	131.20041	**131**	**14**	117	145
8.0	-2.04082	0.39722	33.51206	205.16086	136.76878	**137**	**13**	124	150
9.0	-1.89008	0.37980	33.51206	205.16086	141.82039	**142**	**13**	129	155
10.0	-1.75138	0.36543	33.51206	205.16086	146.46851	**146**	**12**	134	158
11.0	-1.62232	0.35340	33.51206	205.16086	150.79357	**151**	**12**	139	163
12.0	-1.50109	0.34322	33.51206	205.16086	154.85624	**155**	**12**	143	167
13.0	-1.38632	0.33454	33.51206	205.16086	158.70242	**159**	**11**	148	170
14.0	-1.27694	0.32710	33.51206	205.16086	162.36797	**162**	**11**	151	173
15.0	-1.17208	0.32070	33.51206	205.16086	165.88204	**166**	**11**	155	177
16.0	-1.07103	0.31519	33.51206	205.16086	169.26844	**169**	**11**	158	180
17.0	-0.97321	0.31044	33.51206	205.16086	172.54659	**173**	**10**	163	183
18.0	-0.87813	0.30637	33.51206	205.16086	175.73291	**176**	**10**	166	186
19.0	-0.78535	0.30291	33.51206	205.16086	178.84216	**179**	**10**	169	189
20.0	-0.69451	0.29998	33.51206	205.16086	181.88640	**182**	**10**	172	192
21.0	-0.60527	0.29754	33.51206	205.16086	184.87702	**185**	**10**	175	195
22.0	-0.51735	0.29556	33.51206	205.16086	187.82340	**188**	**10**	178	198
23.0	-0.43048	0.29399	33.51206	205.16086	190.73459	**191**	**10**	181	201
24.0	-0.34441	0.29283	33.51206	205.16086	193.61897	**194**	**10**	184	204
25.0	-0.25891	0.29205	33.51206	205.16086	196.48425	**196**	**10**	186	206
26.0	**-0.17376**	**0.29163**	**33.51206**	**205.16086**	**199.33780**	**200**	**10**	**190**	**210**
27.0	-0.08874	0.29157	33.51206	205.16086	202.18700	**202**	**10**	192	212
28.0	-0.00365	0.29188	33.51206	205.16086	205.03854	**205**	**10**	195	215
29.0	0.08171	0.29254	33.51206	205.16086	207.89913	**208**	**10**	198	218
30.0	0.16757	0.29356	33.51206	205.16086	210.77648	**211**	**10**	201	221
31.0	0.25414	0.29496	33.51206	205.16086	213.67761	**214**	**10**	204	224
32.0	0.34166	0.29676	33.51206	205.16086	216.61059	**217**	**10**	207	227
33.0	0.43036	0.29897	33.51206	205.16086	219.58311	**220**	**10**	210	230
34.0	0.52052	0.30163	33.51206	205.16086	222.60456	**223**	**10**	213	233
35.0	0.61243	0.30477	33.51206	205.16086	225.68465	**226**	**10**	216	236
36.0	0.70641	0.30845	33.51206	205.16086	228.83411	**229**	**10**	219	239
37.0	0.80284	0.31271	33.51206	205.16086	232.06568	**232**	**10**	222	242
38.0	0.90214	0.31764	33.51206	205.16086	235.39343	**235**	**11**	224	246
39.0	1.00481	0.32333	33.51206	205.16086	238.83411	**239**	**11**	228	250
40.0	1.11143	0.32989	33.51206	205.16086	242.40717	**242**	**11**	231	253
41.0	**1.22273**	**0.33748**	**33.51206**	**205.16086**	**246.13706**	**250**	**11**	**239**	**261**
42.0	1.33954	0.34629	33.51206	205.16086	250.05160	**251**	**12**	239	263
43.0	1.46297	0.35658	33.51206	205.16086	254.18800	**254**	**12**	242	266
44.0	1.59437	0.36870	33.51206	205.16086	258.59148	**259**	**12**	247	271
45.0	1.73553	0.38313	33.51206	205.16086	263.32205	**263**	**13**	250	276
46.0	1.88887	0.40055	33.51206	205.16086	268.46078	**268**	**13**	255	281
47.0	2.05774	0.42202	33.51206	205.16086	274.11997	**274**	**14**	260	288
48.0	2.24704	0.44914	33.51206	205.16086	280.46380	**280**	**15**	265	295
49.0	2.46435	0.48467	33.51206	205.16086	287.74631	**288**	**16**	272	300
50.0	2.72233	0.53363	33.51206	205.16086	296.39175	**296**	**18**	278	300
51.0	3.04470	0.60660	33.51206	205.16086	307.19503	**300**	**20**	280	300
52.0	3.48473	0.73109	33.51206	205.16086	321.94134	**300**	**20**	280	300
53.0	4.21229	1.01702	33.51206	205.16086	346.32338	**300**	**20**	280	300
54.0	5.44022	1.83507	33.51206	205.16086	387.47384	**300**	**20**	280	300

Table H.41: Science Grade 8: Braille

SCI Grade 8 Braille Theta To Scale Score Table									
Raw Score	Theta	Standard Error	Slope	Intercept	Unrounded	Scale Score	Scale SE	Lower SS	Upper SS
0.0	-5.33126	1.83262	33.51206	205.16086	26.49936	**100**	**24**	100	124
1.0	-4.10958	1.01264	33.51206	205.16086	67.44037	**100**	**24**	100	124
2.0	-3.39077	0.72518	33.51206	205.16086	91.52917	**100**	**24**	100	124
3.0	-2.95919	0.59973	33.51206	205.16086	105.99231	**106**	**20**	100	126
4.0	-2.64492	0.52614	33.51206	205.16086	116.52414	**117**	**18**	100	135
5.0	-2.39471	0.47678	33.51206	205.16086	124.90919	**125**	**16**	109	141
6.0	-2.18481	0.44103	33.51206	205.16086	131.94338	**132**	**15**	117	147
7.0	-2.00255	0.41381	33.51206	205.16086	138.05128	**138**	**14**	124	152
8.0	-1.84036	0.39236	33.51206	205.16086	143.48661	**143**	**13**	130	156
9.0	-1.69333	0.37503	33.51206	205.16086	148.41388	**148**	**13**	135	161
10.0	-1.55813	0.36077	33.51206	205.16086	152.94471	**153**	**12**	141	165
11.0	-1.43233	0.34888	33.51206	205.16086	157.16053	**157**	**12**	145	169
12.0	-1.31418	0.33886	33.51206	205.16086	161.11998	**161**	**11**	150	172
13.0	-1.20229	0.33035	33.51206	205.16086	164.86965	**165**	**11**	154	176
14.0	-1.09560	0.32309	33.51206	205.16086	168.44505	**168**	**11**	157	179
15.0	-0.99326	0.31688	33.51206	205.16086	171.87467	**172**	**11**	161	183
16.0	-0.89456	0.31158	33.51206	205.16086	175.18231	**175**	**10**	165	185
17.0	-0.79892	0.30706	33.51206	205.16086	178.38741	**178**	**10**	168	188
18.0	-0.70584	0.30323	33.51206	205.16086	181.50671	**182**	**10**	172	192
19.0	-0.61489	0.30001	33.51206	205.16086	184.55463	**185**	**10**	175	195
20.0	-0.52571	0.29734	33.51206	205.16086	187.54323	**188**	**10**	178	198
21.0	-0.43797	0.29518	33.51206	205.16086	190.48358	**190**	**10**	180	200
22.0	-0.35136	0.29349	33.51206	205.16086	193.38606	**193**	**10**	183	203
23.0	-0.26561	0.29223	33.51206	205.16086	196.25972	**196**	**10**	186	206
24.0	**-0.18048**	**0.29139**	**33.51206**	**205.16086**	**199.11260**	**200**	**10**	**190**	**210**
25.0	-0.09572	0.29094	33.51206	205.16086	201.95309	**202**	**10**	192	212
26.0	-0.01112	0.29087	33.51206	205.16086	204.78821	**205**	**10**	195	215
27.0	0.07356	0.29117	33.51206	205.16086	207.62601	**208**	**10**	198	218
28.0	0.15851	0.29183	33.51206	205.16086	210.47286	**210**	**10**	200	220
29.0	0.24395	0.29285	33.51206	205.16086	213.33613	**213**	**10**	203	223
30.0	0.33010	0.29424	33.51206	205.16086	216.22319	**216**	**10**	206	226
31.0	0.41718	0.29599	33.51206	205.16086	219.14142	**219**	**10**	209	229
32.0	0.50540	0.29812	33.51206	205.16086	222.09786	**222**	**10**	212	232
33.0	0.59501	0.30064	33.51206	205.16086	225.10087	**225**	**10**	215	235
34.0	0.68626	0.30358	33.51206	205.16086	228.15885	**228**	**10**	218	238
35.0	0.77943	0.30696	33.51206	205.16086	231.28116	**231**	**10**	221	241
36.0	0.87482	0.31083	33.51206	205.16086	234.47788	**234**	**10**	224	244
37.0	0.97279	0.31526	33.51206	205.16086	237.76106	**238**	**11**	227	249
38.0	1.07375	0.32034	33.51206	205.16086	241.14443	**241**	**11**	230	252
39.0	1.17821	0.32620	33.51206	205.16086	244.64510	**245**	**11**	234	256
40.0	**1.28679**	**0.33299**	**33.51206**	**205.16086**	**248.28384**	**250**	**11**	**239**	**261**
41.0	1.40027	0.34096	33.51206	205.16086	252.08679	**252**	**11**	241	263
42.0	1.51968	0.35041	33.51206	205.16086	256.08847	**256**	**12**	244	268
43.0	1.64637	0.36178	33.51206	205.16086	260.33411	**260**	**12**	248	272
44.0	1.78216	0.37564	33.51206	205.16086	264.88471	**265**	**13**	252	278
45.0	1.92957	0.39279	33.51206	205.16086	269.82473	**270**	**13**	257	283
46.0	2.09214	0.41437	33.51206	205.16086	275.27278	**275**	**14**	261	289
47.0	2.27507	0.44212	33.51206	205.16086	281.40314	**281**	**15**	266	296
48.0	2.48638	0.47881	33.51206	205.16086	288.48458	**288**	**16**	272	300
49.0	2.73928	0.52949	33.51206	205.16086	296.95978	**297**	**18**	279	300
50.0	3.05816	0.60458	33.51206	205.16086	307.64610	**300**	**20**	280	300
51.0	3.49703	0.73128	33.51206	205.16086	322.35354	**300**	**20**	280	300
52.0	4.22657	1.01887	33.51206	205.16086	346.80193	**300**	**20**	280	300
53.0	5.45813	1.83687	33.51206	205.16086	388.07404	**300**	**20**	280	300

APPENDIX I
ITEM MAPS

Figure I.13: Item Map Science Grade 4

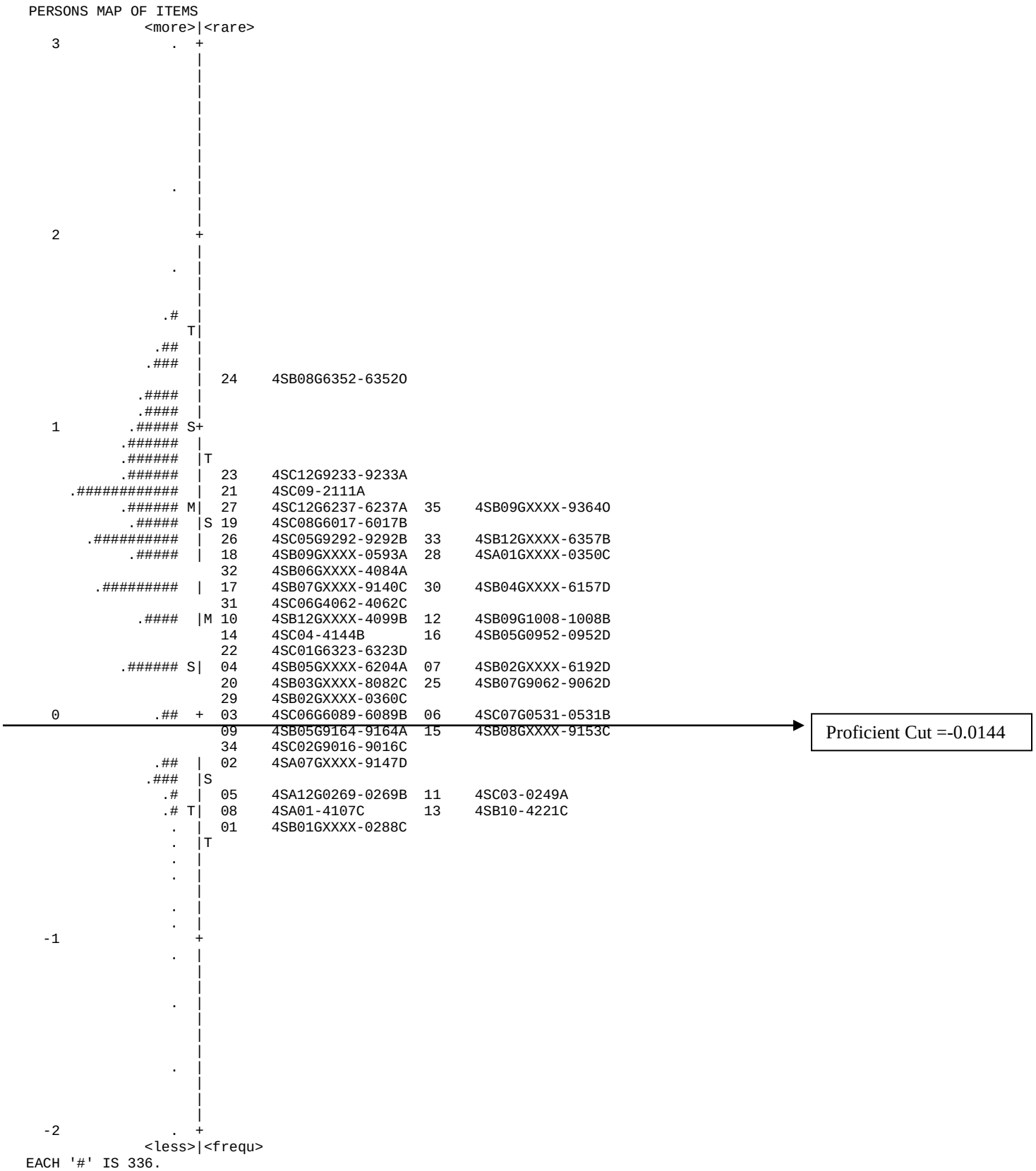

Figure I.14: Item Map Science Grade 8