

**TECHNICAL REPORT
NEW JERSEY
BIOLOGY COMPETENCY TEST**

May 2016
Measurement Incorporated™

**January 2017
PTM 1510.85**

**Copyright © 2017 by New Jersey Department of Education
All rights reserved**

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

KIMBERLY HARRINGTON, Acting Commissioner of Education

Table of Contents

Table of Contents	iii
1.1 Description of Assessment	5
1.2 Purpose of Assessment	6
1.3 Organizational Support	7
Part 2: Test Development.....	8
2.1 Test Specifications	8
2.2 Development of Test Items	9
2.3 Item Review Process	13
2.4 Item Use	14
2.5 Quality Control for Test Construction	15
Part 3: Test Administration	16
3.1 Participation	16
3.2 Test Security Procedures	17
3.3 Test Administration Procedures	19
3.4 Test Accommodations	22
Part 4: Scoring.....	28
4.1 MC items	28
4.2 OE and PAT items	28
4.3 Quality Control Procedures in Data Preparation	32
Part 5: Item-Level Statistics	33
5.1 Classical Item Statistics	33
5.2 Speededness	34
5.3 Intercorrelations	35
Part 6: Standard Setting	36
Part 7: Calibration, Scaling and Equating	37
7.1 Calibration and Scaling	37
7.2 Equating	37
Part 8: Test Statistics.....	44
8.1 Summary Statistics	44
8.2 Classical Reliability Estimates of the Test Scores	44
Part 9: Validity	47
Part 10: Reporting	48
10.1 Cycle I Reports	48
10.2 Cycle II Reports.....	48
10.3 State Summary Report.....	49
10.4 Interpreting Reports	49
10.5 Quality Control in Reporting.....	49
Reference	52
Appendix A: Test Specifications	53
Appendix B: Raw Score Frequency Distribution.....	55
Appendix C: Standard Error Plot.....	59
Appendix D: Standard Setting Executive Summary.....	61

Tables and Figures

Table 1.1.1: 2015 NJBCT State Level Raw Score Means	6
Table 1.3.1: Contributors and Their Roles/Responsibilities	7
Table 2.1.1: 2015 NJBCT Total Point	8
Table 2.2.1: Field Test Form Distribution Plan	12
Table 5.1.1: Item Difficulty and Discrimination Summary Statistics by Cluster	33
Table 5.1.2: Frequency Distribution, P-Values and Discrimination Indices by Cluster	34
Table 5.2.1: 2015 NJBCT Percentage of Students Omitting the Last Item of Each Part	34
Table 5.3.1: NJBCT Intercorrelations Among Strands	35
Table 7.2.1: NJBCT Course Titles and Academic Levels	38
Table 7.2.2: 2015 NJBCT Students by Schedule	39
Table 7.2.3: 2015 NJBCT Students by Academic Level	39
Table 7.2.4: 2015 NJBCT Raw Score Mean and Standard Deviation by Schedule and Level	40
Table 7.2.5: Results of Analysis of Variance	40
Table 7.2.6: Summary of Effect Sizes	41
Figure 7.2.1 Biology Item Map	42
Figure 7.2.2 Biology Test Information Function Figure	43
Table 8.1.1: NJBCT Raw Score Means and Standard Deviations	44
Table 8.1.2: 2015 Raw Score and Percent Correct by Cluster	44
Table 8.2.1: Reliability Estimates and SEM, Total Test and by Cluster	45
Table 8.2.2: 2015 Reliability Estimates and SEM, Content Areas/Clusters by Subgroup	46
Table B.1: NJBCT Raw Score Frequency Distribution	56
Table C.1: NJBCT Standard Error Plot	60
Table ES-1: Final Cut Score Recommendations	64
Table ES-2: Percentages of Students Classified at Each Level	65

Part 1: Introduction

1.1 Description of Assessment

In 1975, the New Jersey Legislature passed the Public School Education Act “to provide to all children in New Jersey, regardless of socioeconomic status or geographic location, the educational opportunity which will prepare them to function politically, economically and socially in a democratic society.” An amendment to that act was signed in 1976 which established uniform strands of minimum achievement in basic communication and computation skills. This amendment is the legal basis for the use of a test as a graduation requirement in the state of New Jersey.

Beginning in 1981–1982, ninth-grade students were required to pass the Minimum Basic Skills Test in Reading and Mathematics as one of the requirements for a high school diploma. Students who did not pass both sections of the test had to be retested on those sections not passed.

In 1983, a more difficult test in Reading, Mathematics, and Writing was adopted, the Grade 9 High School Proficiency Test (HSPT9), to measure the basic skills achievements of ninth-grade students. The first due-notice administration of the HSPT9 occurred in the 1983–1984 school year. The first year the test was administered as a graduation requirement was 1985–1986.

In 1988, the New Jersey Legislature passed a law which moved the High School Proficiency Test from the ninth grade to the eleventh grade. The Grade 11 High School Proficiency Test (HSPT11), a rigorous test of essential skills in Reading, Mathematics, and Writing, served as a graduation requirement for all public school students in New Jersey who entered the ninth grade on or after September 1, 1991. Three years of due-notice testing were conducted to allow school districts time to modify curricula and prepare students for the graduation test.

In May 1996, the New Jersey State Board of Education adopted Core Curriculum Content Strands to describe what all students should know and be able to do at the end of fourth grade, eighth grade, and upon completion of a New Jersey public school education. The Core Curriculum Content Strands delineate New Jersey’s expectations for student learning. All New Jersey school districts are required to organize instruction and design curricula so that virtually all students achieve the new content strands. The Core Curriculum Content Strands define the state’s high school graduation requirements and its testing program to measure benchmark achievements toward those requirements in grades 4, 8, and 11.

The Elementary School Proficiency Assessment (ESPA), which was administered to fourth-graders, was designed from its inception in 1997 to align with the content strands, as is the New Jersey Assessment of Skills and Knowledge (ASK), which replaced the ESPA and was expanded to include third-graders. As a result of the No Child Left Behind Act, the New Jersey Assessment of Skills and Knowledge was expanded to the fifth, sixth, and seventh grades in 2006. The Grade Eight Proficiency Assessment (GEPA), which replaced the Grade 8 Early Warning Test (EWT), was aligned with the content standards. The GEPA was used for placement purposes and program planning so that students were provided with the appropriate instruction to enable them to go on and pass the state’s graduation test. The GEPA was replaced by NJ ASK, which served the same purpose as the GEPA. The High School Proficiency Assessment (HSPA) was also

aligned with the content standards and replaced the HSPT11 as the state’s graduation test. In 2004, a general science field test, including physical and biological sciences, occurred preparatory to adding it as part of the HSPA in 2005.

In 2007, the New Jersey Department of Education decided to begin transitioning to the Biology Competency Test (NJBCT) assessments. Currently, Biology is not a graduation requirement but all assessment results are used to satisfy federal requirements under the No Child Left Behind Act. Biology was the first course to get an end-of-course examination in New Jersey. The first operational administration of the NJBCT exam was May 2008 and a standard setting to establish cut scores was conducted in August 2008. The NJDOE decided not to present the cut scores to the State Board and thus only raw scores were reported.

In 2009, the New Jersey Department of Education decided to drop the open-ended items from the NJBCT exam and replace them with an equivalent number of multiple-choice points. NJBCT A total of three 3-point open-ended items were dropped and replaced by 9 multiple-choice items. Again, only raw scores were reported.

In 2010, the New Jersey Department of Education added two Performance Assessment Tasks (PATs) to the test. In 2011 this was reduced to one PAT, because of the test timing. The department also decided to adopt the 2008 standards as a temporary reporting mechanism in order to be in compliance for its high school science assessment. The test was not used as a high school graduation requirement due to the multiple changes to the test.

Table 1.1.1: 2016 NJBCT State Level Raw Score Means

Number of Students	Mean Raw Score	STD Raw Score
102514	29.08	11.10

1.2 Purpose of Assessment

The initial purpose of NJBCT was to provide annual feedback to students, teachers, parents, schools, districts, the State, and the general public in regard to the achievement of individual students and the success of educational programs in relation to the state curriculum. The NJBCT also serves the secondary purpose of providing diagnostic information to teachers regarding strengths and weaknesses of students within the five broad areas assessed. The dual purpose of the tests is reflected in the score reports, which contain not only total scores and designations of achievement level (Exceeded New Jersey Strands through Apprentice) but also cluster level scores for students and programs. No scores for the individual skills that make up the content strands (see the separate content area sections of this document, Appendix A) are reported at any level, which include student/parent, classroom, building, district, and state.

In addition to assessing student knowledge of subject-specific content, a primary focus of NJBCT is to assess students’ critical thinking skills. These skills are assessed through both the multiple-choice (MC) items and the PAT. The test matrix for Biology (Appendix A) reflects the crossing of content with process.

The purpose of the NJBCT test is to accurately measure students’ achievement in Biology. To achieve this purpose, school personnel administering the test play a crucial role. Test administrators can minimize problems that interfere with testing students by addressing the following guidelines:

- Maintaining the security of all testing materials before, during, and after testing.
- Administering the tests in a manner consistent with established testing procedures.
- Establishing testing conditions that prevent opportunities for students to engage in irregular behaviors (intentional or unintentional).
- Communicating with the district NJ HSPA/NJBCT coordinator if questions arise.

1.3 Organizational Support

Measurement Incorporated, the contractor selected by New Jersey to develop, administer, score, and report on the NJBCT, understands the need for proactive communication among all parties involved in a successful assessment program. Given the nature of a project as large and diverse as the NJBCT program, numerous individuals, departments, and the contractor fill specific roles and responsibilities. Table 1.3.1 shows the major contributors to the NJBCT in the past and their general responsibilities.

Table 1.3.1: Contributors and Their Roles/Responsibilities

Contributor	Role/Responsibility
NJ DOE	Oversight of the entire NJBCT program and content strands
NJBCT Program Coordinator	Daily management of the NJBCT program
NJBCT Content Specialist	Content-specific item development and testing issues
NJBCT Item Review Committees*	Review and approval of all test items
Technical Advisory Committee	Advises and assists on technical guidelines of the NJBCT/EOC program.
Educators	Provide feedback on all aspects of the NJBCT/EOC program. Administer the tests to students. Serve on content item review committees.
Measurement Incorporated (MI)	Develop and provide all products and materials relating to the testing, standard setting, scoring, and reporting for the NJBCT/EOC program.
MI NJBCT Project Director	Daily management of the NJBCT/EOC program at MI

It is this need for coordination among various entities within the NJ DOE and Measurement Incorporated that necessitates the use of project management and production schedules, progress reports, conference calls, management and planning meetings, and meeting minutes. Additionally, for each test administration a District Test Coordinator Manual and Test Examiner Manual is provided to the appropriate administrators and teachers involved in the testing process.

Part 2: Test Development

2.1 Test Specifications

Development of NJBCT Test Forms

The Biology Competency Test (NJBCT) is given only in May. The test includes embedded field test items distributed over multiple forms. The position of these items remains unchanged from year to year. Definitions of specific terminology (CPI, Skill, and Cluster) in the test matrix is found in Appendix A of this document.

The NJBCT consists of all independent items. Each test has 45 multiple-choice questions and one Performance Assessment Tasks (PATs). The test measures content from 5 Biology content clusters:

- O - Organization and Development
- M - Matter and Energy Transfer
- I - Interdependence
- H - Heredity and Reproduction
- E - Evolution and Diversity

Within these strands are a number of specific skills and CPIs (see Appendix A).

The NJBCT consists of 57 points:

- 1 point for each of 45 multiple-choice items = 45
- 12 points for the performance assessment tasks=12

Tests have traditionally been assembled using item difficulty parameters derived from field-testing and then equated using the Rasch model. Due to a lack of item development limiting the ability to select new items, the 2016 operational BCT assessment was simply the 2013 operational test, with a different PAT. Table 2.1.1 summarizes the total points possible for Biology on the operational NJBCT administered May 2016. All items are either application or knowledge questions.

Table 2.1.1: 2016 NJBCT Total Point

Biology	Points	Application Points
Organization and Development (O)	11	10
Matter and Energy Transfer (M)	8	6
Interdependence (I)	6	5
Heredity and Reproduction (H)	13	10
Evolution and Diversity (E)	7	6
Performance Assessment Tasks	12	12
Total	57	49

2.2 Development of Test Items

The decision to develop the NJBCT test was made in 2007. Most of the Life Science items in the bank used to construct the HSPA Science test were realigned to the NJBCT strands and those items, along with the items field tested in 2008, were used to construct the NJBCT test. This was problematic in that the HSPA Science items were field tested on the HSPA Science test, which was given to 11th grade students. The NJBCT is given to students, regardless of grade, who have completed a Biology course. The process for developing NJBCT items required a two-year cycle in which the items were:

- Written to NJBCT strands;
- Reviewed by MI content experts;
- Reviewed by state content experts;
- Reviewed by New Jersey teachers and a sensitivity committee to see if the item can be field-tested;
- Reviewed before scoring by a range-finding committee involving state educators;
- Field-tested; and
- Reviewed again by state content experts, New Jersey teachers, and a sensitivity committee after field-testing.

Only an item that was found acceptable at every stage of the two-year cycle was entered into the bank for possible use on an operational test. Although statistical data on test items played an essential role, this cycle of development employed a due process model of validity. This model relied on the expertise of educators participating in the test development process. The strength of this process was dependent on the degree to which the following critical components were integrated into the test development process:

- Recruitment of expert educators familiar with the state's content specifications and population for the assessment;
- Training of item writers and expert reviewers on item writing specifications, content specifications, and the goals and functions of the assessment;
- Careful consideration of individual items by experts to assess the degree to which the items measure the knowledge, skills, and abilities that the assessment was intended to measure with opportunities to reject or revise items per committee recommendation; and
- Careful consideration of sensitivity issues by experts to guarantee that performance on the item was related to classroom achievement and not cultural or social experiences outside the classroom with opportunities to reject or revise items per committee recommendation.

At Measurement Incorporated, item writers, under the supervision of content experts, were instructed on the state specifications and item types necessary for the tests. They were trained on the NJBCT content specifications and directed to write original items tailored to NJ content strands. Content expert reviewers at MI validated (or not) the initial coding of items by item writers to meet NJBCT content strands. At this point in the process, some items were rejected from further consideration on the grounds that the items were not tied closely enough to NJBCT strands or are not at an appropriate level of difficulty.

In reviewing items, NJ educators reviewed the following information beyond the item wording and scoring rubric. First, reviewers confirmed the assignment of each item to a NJ content specification Cluster and Skill. For all content review meetings, MI furnished reviewers with copies of the NJBCT Blueprint to allow committee members to validate assignment of items to NJ content strands. Reviewers accepted or revised an item coding assignment, or rejected an item as not fulfilling any specific part of the content specifications. For each item, both committees also rated each item for a level of difficulty.

Although the content committees were trained to recognize possible bias or lack of cultural sensitivity in test items, a separate sensitivity committee met to review all multiple-choice items that were flagged during the field test as Mantel-Haenszel “C” items (probable differential item functioning, DIF) in NJBCT using student data disaggregated by demographic group for both tests. Like the content committee, the sensitivity committee had the power to reject an item. If either the sensitivity committee or content committee rejected an item, it was considered rejected. If one committee required that the item be revised, that decision outweighed an acceptance by the other committee.

The May 2016 operational test consisted of items that had been previously utilized on the May 2013 operational test. All the items had been previously field tested and cleared for use on the operational assessment by all committees. Thus, the May 2016 BCT consisted entirely of items that had been deemed appropriate for measuring the NJ Biology curriculum and that had been confirmed as being bias-free and culturally sensitive.

At item review sessions, items were presented one-per-page, each with a footer as shown below:

Sensitivity		Content	
*Comments:		*Comments:	
Sensitivity Issue	Yes No	Meets Specifications	Yes No
If yes, identify category and explain*		Appropriate Difficulty	Yes No
		Accurate Coding	Yes No
Definitely Use		Definitely Use	
Revise and Use With Approval		Revise and Use With Approval	
Revise and Resubmit		Revise and Resubmit	
Do Not Use*		Do Not Use*	

At the bottom of each footer is a place for committee members to sign off on their decision:

_____	_____	_____	_____
Sensitivity Sign-off	Date	Content Chairperson's Signature	Date

This is a critical step in the item review process as it records, item by item, the appropriateness of each item for the assessment. Only an item approved by both committees could be field-tested.

The Field Test Form Distribution

Beginning with the first operational administration in May 2010, the NJ DOE began embedding field test items for Biology. Previously, six forms had been distributed to New Jersey schools for the purpose of field testing. In 2016 there was no field testing of new items. However, for the sake of continuity six equivalent test forms were still distributed to New Jersey schools. Each of the six test forms included identical base test items as well as a section of field test items. Note that students earned scores only on the identical common items. No statistical analysis was conducted on the six field test forms. The six 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, which received two forms. Furthermore, the field test forms were distributed across DFG classifications, such that each DFG was represented in the equating blocks. Finally, approximately equal numbers of students (20,250-21,180) were given each test form. Table 2.2.1 shows the field test form distribution plan, by test form and DFG classification.

Table 2.2.1: Field Test Form Distribution Plan

All Testers	DFG													Grand Total
Form	A	B	CD	DE	FG	GH	I	J	N	O	R	S	V	
A	3,210	1,950	1,640	3,210	2,040	3,060	2,460	930		350	240	80	1,150	20,320
B	2,940	1,810	1,500	2,530	2,190	3,750	3,320	750		280	420	90	1,600	21,180
C	2,460	2,000	1,760	3,290	1,970	3,380	3,070	780	330	400	250	80	1,320	21,090
D	2,670	1,800	1,520	2,740	2,670	3,590	2,670	660		280	400	70	1,210	20,280
E	2,320	2,000	1,660	2,520	2,570	3,560	2,800	1,050		280	300	200	1,230	20,490
F	2,330	2,700	1,840	2,810	1,910	3,130	3,430	460		250	300	70	1,020	20,250
Grand Total	15,930	12,260	9,920	17,100	13,350	20,470	17,750	4,630	330	1,840	1,910	590	7,530	123,610

2.3 Item Review Process

Each summer following a field test, the NJ DOE conducted a statistical analysis review session with New Jersey teachers. The teachers on the content and sensitivity committees reviewed the items and evaluated the performance of the items based on field test data. The following variables were included:

- Item ID
- N-count
- p -value
- Biserial
- % answering each option (A-D) and omits
- p -value for bottom 20%
- p -value for top 20%
- % of Whites answering each option (A-D) and omits; N-count for Whites
- % of Blacks answering each option (A-D) and omits; N-count for Blacks
- % of Hispanics answering each option (A-D) and omits; N-count for Hispanics
- % of Males answering each option (A-D) and omits; N-count for Males
- % of Females answering each option (A-D) and omits; N-count for Females
- Total Biology Score for students taking that form

For the meeting, teachers were provided with a training session on how to interpret these statistics. To draw their attention to items that may be problematic, several flags were used. The flags included:

- *Difficulty flag* to indicate that an item has a p -value less than .25 or greater than .95
- *Correlation flag* to indicate an item that has an item-total correlation of less than .25
- *Mantel-Haenszel flags* to indicate any group comparison flagged as “C” for large Differential item functioning (DIF), “B” for moderate DIF, and “A” or blank for no or negligible DIF

Differential Item Functioning Analysis

Differential Item Functioning (DIF) analysis was conducted using the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959). DIF refers to statistically significant differences in item performance by groups whose performance is otherwise comparable on the basis of some criterion, e.g., total score. Consequently, DIF is an indication of possible item bias; the items that were flagged for DIF, then, were reviewed carefully to determine whether or not actual ethnic or gender bias was present. In addition, it is important to consider not only the significance of DIF statistics, but also the extent (effect size) of the difference. The extent of DIF takes into account the effect size and significance of the DIF statistics (see Dorans & Holland, 1993; Zieky, 1993) and is categorized as follows:

- A = No or negligible DIF
- B = Moderate DIF
- C = Large DIF

This coding of Mantel-Haenszel results, while statistically sound, greatly enhances the ability of non-statisticians on bias committees to make appropriate decisions regarding these results by helping to distinguish random chance variation from systematic differences.

2.4 Item Use

At the meetings, teachers were presented with forms similar to those used at initial item development meetings; a sample is shown below:

ITEM CODE AND KEY		Admin: March 2009	Form:	Position:
Sensitivity		Content		
*Comments		*Comments		
Sensitivity Issue	<input type="checkbox"/> Yes <input type="checkbox"/> No	Appropriate Difficulty	<input type="checkbox"/> Yes <input type="checkbox"/> No	
If a sensitivity issue, explain*		P-value = Biserial =		
Mantel-Haenszel Category C <input type="checkbox"/> W-AA <input type="checkbox"/> W-H <input type="checkbox"/> M-F				
<input type="checkbox"/>	Definitely Use		<input type="checkbox"/>	
<input type="checkbox"/>	Revise and Use With Approval **		<input type="checkbox"/>	
<input type="checkbox"/>	Revise and Re-Field Test		<input type="checkbox"/>	
<input type="checkbox"/>	Do Not Use *		<input type="checkbox"/>	

Sensitivity Sign-
Off

Date

Content Chairperson's
Signature

Date

**** Requires director's approval**

- **Accept (Definitely Use):** All content related issues (importance, thematic, grammar, clarity, accuracy, validity, sound measurement, grade-appropriate), all statistical criteria, and all sensitivity issues have been met or exceeded and the item appears suitable for operational use.
- **Revise (Revise and Re-Field Test):** One or more of the content related issues have not been met or the item needs minor changes to make it acceptable. Reviewers provide recommendations on changes to be made to the item that will make the item suitable for re-field testing.
- **Reject (Do Not Use):** Several content related issues, statistical criteria, or sensitivity issues have not been met, or are suspect, or need radical changes to make the item acceptable. In such cases, the item may be vague or ambiguous, inappropriate, or not clearly related to the text or to the standard. Without severe modifications, it is unlikely to be salvaged. Reviewers provide comments as to why the item should be rejected.
- **Revise and Use With Approval:** A very minor content related issue needs to be resolved and the NJ DOE content representative feels it is minor enough to use operationally without re-field testing.

Only items designated as revise and use with approval or accepted by both committees were added to the item bank for possible use on future operational tests.

2.5 Quality Control for Test Construction

Measurement Incorporated and the NJ DOE work very closely together to ensure that the content and editorial quality of the test booklets meet or exceed the state expectations. This requires consistent vigilance and quality control checks during the test booklet assembly process.

The operational test was assembled from the approved NJBCT test designed using field-tested items that are proven to effectively measure the BCT content standards and that are fair to all students. Test booklets were assembled using approved general and style guidelines. The typeset test booklets were then proofread by two editorial staff members for typographical and format errors, and also to determine if the version of the item used was consistent with the version field-tested. Upon completion of this internal review, the test booklet was sent to NJ DOE for a typeset review. MI then responded to any revisions required by NJ DOE, and this process was repeated until NJ DOE approved the test booklet. Once NJ DOE approved the operational test booklet, it was then sent to a proofreading agency external to MI for an independent review. The final approved version of the test booklet was then converted to a pdf electronic file for printing. The pdf version of the test booklet was proofread by editorial staff before submitting to the printing manager. The first copies of the production run of the test booklet were reviewed for possible problems by the MI project management staff.

Ancillary test materials were subject to the same quality control. All typeset copies were first proofread by at least two editorial staff members before being submitted to NJ DOE for a typeset review. Materials were approved by NJ DOE before printing. Approved versions of the ancillary materials were then converted to pdf files for printing.

All accommodated materials were also subject to consistent vigilance and quality control at all stages. The large print test and supporting materials were subject to the same assembly quality control discussed previously. The Braille translation of the test and supporting materials was performed by an independent, certified translation agency. The large print and Braille versions of the test materials were then submitted to NJ DOE for review by specialists from the state commission for the blind. Revisions to the materials were made based on recommendations from these state specialists, and then the accommodated materials were sent to production.

Part 3: Test Administration

3.1 Participation

- **All** New Jersey public high school students, **regardless of grade level**, (grades 9-R12 and AH) who are enrolled in a biology course or content equivalent, any time during the current school year, must take the NJBCT, **regardless of prior testing exposure and experience**. The following course titles are included as exemplars:
 - Resource Center Biology
 - ESL Biology
 - General Biology
 - Life Science
 - Foundations in Biology
 - College Prep Biology
 - Honors Biology
 - Accelerated Biology
 - Advanced Placement Biology (see below)
 - Integrated Science (see below)
 - Other (i.e., Cosmetology, Agricultural Science)

The additional criteria for participation in the NJBCT must be followed:

- Students taking *AP Biology* **without having completed** a Biology prerequisite **must** take the NJBCT.
- Students taking *AP Biology* **after completing** a Biology prerequisite must **not** take the NJBCT.
- Students taking *Integrated Science* as an **alternative** to the Biology offerings listed above **must** take the NJBCT.
- Students taking *Integrated Science* as a **prerequisite** course for a Biology elective must **not** take the NJBCT. These students will take the NJBCT during the school year that they are enrolled in a *Biology* elective.
- Adult high schools, the Department of Corrections, the Juvenile Justice Commission, and the Department of Children & Families facilities must take the NJBCT if they have any eligible students.

Districts are encouraged to align their biology course offerings with the New Jersey High School Biology/Life Science Syllabus. The syllabus can be downloaded from the NJDOE web site at <http://www.state.nj.us/education/aps/cccs/science>.

3.2 Test Security Procedures

RECEIVING MATERIALS

Biology materials were packaged for each school based on information provided by the district on the Biology Material Survey Form. The chief school administrator or designee received all Biology test materials in a single shipment delivered by NCD Package Express Ground Service.

The chief school administrator must have completed the District Authorization to Receive Secure Test Materials Form prior to receiving the shipment of test materials. This form provided space for designating the person(s) who received the materials for the district, along with a space for verifying receipt of the materials. The authorized designee(s) for receiving secure test materials had to be available on the scheduled delivery dates. He/she must have had the authorization form and presented it to the carrier at the time of delivery. If the authorization form was not made available within 15 minutes of the courier's arrival, the courier left to continue with making deliveries, and attempted re-delivery at the end of his or her route.

Upon delivery, the shipment was checked for damage and all boxes were accounted for prior to signing for the shipment. Multiple boxes were marked "1 of 3," "2 of 3," etc. If a problem was discovered—for example, if a box was missing, damaged, or opened—the shipment was to be accepted and the designated authorized personnel were advised to contact Measurement Incorporated staff immediately to resolve the issue. The number of boxes received were recorded on the Delivery Ticket presented by the carrier, and the authorized personnel and the courier were instructed to both keep the copies of the form.

A separate box contains the district overage. One of the boxes for each school and the district overage box will contain a packing list. For each of the materials listed on the packing list, verify that the quantities received match the quantities indicated. Also, verify that a sufficient supply of the test materials has been provided.

SECURE MATERIALS

The following materials were considered secure during the test administration. All materials were shrink-wrapped in packages of ten.

- Test Booklets – The test booklets **must** have been kept secure at all times.
- Answer Folders – Used answer folders **must** have been kept secure at all times. However, the answer folder shrink-wrap may have been opened prior to the test administration date to facilitate the application of Biology ID Labels to the answer folders and any necessary pre-gridding.

NONSECURE MATERIALS

The following materials, although not secure, were stored in a manner to ensure their availability on each testing day.

- Periodic Table of the Elements: Each student had a Periodic Table of the Elements on each day of the test. A supply was provided so students were given a new table each day.
- Examiner Manual: One copy was provided for every 15 students tested. School test coordinators wrote the county, district, and school codes on the front cover of each *Examiner Manual*.
- District/School Test Coordinator Manual: Two copies of the manual were provided for each school that administered the NJBCT.
- Return Shipping Materials: The district overage box contained envelopes, labels, and Federal Express airbills for returning answer folders after the regular and make-up test administrations. All test booklets were returned to Measurement Incorporated in the boxes that were used to ship the materials to the district. Administrators were instructed to store the boxes, envelopes, Federal Express airbills, and labels in a safe place until it was time to use them.
- Biology Header Sheets: The Biology Header Sheets were precoded with the district and school codes and names. A supply of blank orange Biology Header Sheets were also included in the district overage.

MATERIAL SHORTAGES

As soon as the shipment was received, the test materials were transported to a secure location and it was verified that there was a sufficient quantity of all materials needed for testing. Next, the District Receipt Form was completed, and the original was sent to Measurement Incorporated. Designated personnel were instructed to retain a photocopy. The form had to be mailed within ONE day of receipt of test materials.

The district overage was used to make up school shortages of secure and nonsecure test materials. The school test coordinators were instructed to sign the School Security Checklist – District Overage form when they received the overage test booklets from the district test coordinator.

If the district overage was not sufficient to make up school shortages, school test coordinators were instructed to fax the Additional Materials Request Form to Measurement Incorporated. In order to ensure receipt of the additional materials in time for testing, materials were ordered from Measurement Incorporated before the date specified on the manual.

TEST BOOKLETS

- Test coordinators were instructed to open the shrink-wrapped packages of test booklets the day before testing under secure conditions so that sets of materials could be assembled in advance for each examiner.
- Each test booklet section was sealed and not opened until students were directed during the test administration.

ANSWER FOLDERS

- The shrink-wrapped packages of answer folders may have been opened prior to the test administration date to facilitate the application of Biology ID Labels to the answer folders and any necessary pre-gridding.
- Responses to all questions were supposed to be recorded in the answer folder.
- One version of the Biology answer folder was used with the Biology test booklet, which contained both Day 1 and Day 2 test questions.
- Answer folders could not be duplicated or hand-scored.
- Each page of the answer folder was marked with a machine-scannable lithocode number (also referred to as the answer folder number).
- The test booklet **FORM** letter and **TEST BOOKLET NUMBER** were supposed to be recorded on each student's answer folder.
- All applicable "School Use Only" fields were supposed to be gridded before the answer folders were returned to Measurement Incorporated. Detailed instructions were provided in the "School Use Only Gridding Instructions" section of the *Examiner Manual*.

Examiners were instructed to only submit one answer folder for each student to Measurement Incorporated. (Unused answer folders were not to be destroyed.) Examiners were also warned as to how reports could be affected when more than one answer folder was submitted for a student. For instance, multiple answer folders submitted for the same student would have inflated enrollment if it is not possible to merge the two answer folders. Subsequently, this could have a negative impact on the district and school performance data. Examiners were advised to submit an Irregularity Report to document all instances of multiple answer folder usage. This was necessary to facilitate the merging of multiple answer folders used by a student.

3.3 Test Administration Procedures

The New Jersey Biology Competency Test was administered over two days (May 25-26, 2016). Test administrators were advised to not schedule testing after an athletic event or an assembly. All test schedules were to be checked with the appropriate school officials to ensure that other school activities would not interfere with the test administration.

- It was advised that all testing be scheduled in the morning except for adult high school, homebound and bedside students, and for 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 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 *District/School Test Coordinator Manual* and in the *Examiner Manual*.

- Students who were required to test but were absent for the regular test administration were tested on the make-up date.
- Students whose answer folders were voided during testing were considered to have attempted the test section. They could not 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 could not complete the test section during the make-up period or any other time unless additional time was specified in the student's IEP or 504 plan.

Student Rosters

Under the direction of the district test coordinator, the school coordinator prepared a student roster for each examiner. (See Appendix I in the District/School Test Coordinator Manual.) Each roster listed the names of the students each examiner supervised during testing, the names of proctors that assisted the examiner, and the room number. It was ensured that each roster contained any information that must be gridded on a student's answer folder because some students may not have had a Biology Pre-ID Label or may have had information missing from their Biology Pre-ID Labels. Refer to the "School Use Only" section of the *Examiner Manual* for detailed information regarding the gridding of the answer folder. In addition, all test modifications and accommodations for SE and 504 students were indicated when applicable.

Test coordinators distributed the student rosters to examiners in advance of testing to allow the examiners to prepare for the test administration. Examiners returned the student rosters to the school coordinator immediately after testing was completed. The school test coordinator was instructed to keep a copy of the rosters and return the originals to the district test coordinator.

Advance Announcements

Teachers were given advance notice of the testing schedule so they could adjust lesson plans and personal schedules.

Examiners were instructed to advise their students about the time and location of the test administration and given a brief explanation about the test. Students were instructed to bring two #2 pencils if the school did not provide them. No pens of any kind were allowed. All sections of the answer folder were to be marked in pencil.

Students were notified that they could not use dictionaries or other reference materials during testing, and that they were not permitted to have cell phones, MP3 players, or other unauthorized electronics in any testing room, and that any student found to have unauthorized electronics in his or her possession while in a testing room would have his or her answer folder voided (V2) for that test section.

TEST ADMINISTRATION TIMES

- The approximate administration time for the Biology test was as follows:

Biology – Day 1	2 hours, 17 minutes
Biology – Day 2	2 hours, 7 minutes

This schedule included time for gridding student information on the answer folder the day of testing, distributing the test materials, reading directions, test taking, providing breaks for the students, and collecting materials after testing.

- The number of test questions was as follows:

Biology – Day 1	30 multiple-choice questions and 1 performance assessment task
Biology – Day 2	30 multiple-choice questions and 1 performance assessment task

CALCULATORS

- Calculators were **not** permitted for the NJBCT.

GRIDDING STUDENT INFORMATION

The gridding of student, school, and district information on the answer folder was not necessary for students who had a Biology Pre-ID Label with the exception of GRADE, test booklet FORM letter, and TEST BOOKLET NUMBER. These fields had to be gridded for all students. Authorized staff under the supervision of the school test coordinator was instructed to review the Biology Pre-ID Labels to verify the accuracy of the information on the labels.

- Biology Pre-ID Labels were provided for each student whose name was included in the Biology Pre-ID file that the district submitted to Measurement Incorporated in the fall.
- All damaged or unused Biology Pre-ID Labels were to be destroyed.
- All student demographic information has to be hand-gridded on the answer folder for any student who did not have a Biology Pre-ID Label.

BIOLOGY PRE-ID LABELS

- If available, the Biology Pre-ID Label (white with orange bar) was affixed to the upper right-hand corner of each student's answer folder. Required information not contained on the label was gridded on the answer folder for a given student.
- If a student had a Biology Pre-ID Label and transfers to another school in the district, schools were instructed to destroy the Biology Pre-ID Label and hand-grid the student's answer folder.
- Districts could continue to code their own student identification numbers on the answer

folders. This “District/School Student Identification Number” was reported on the Student Sticker and the Individual Student Report.

- GRADE, test booklet FORM letter, and the TEST BOOKLET NUMBER were gridded on the Biology answer folder for all students.
- Districts had the opportunity to ADD or CHANGE some of the student information contained on the labels by gridding the answer folder. This information would override the printed information on the Biology Pre-ID Label.
- Districts had the opportunity to REMOVE certain student information contained on the Biology Pre-ID Labels by gridding the override bubble (*) on the answer folder.
- If an answer folder had a label, the following fields were left blank: student name, date of birth, SID number, and testing site county, district, and school codes. If these fields are gridded in error, erase the gridding as thoroughly as possible.
- The following information **cannot** be corrected by gridding the answer folder: student name, date of birth, SID number, and testing site county, district and school codes. This information could only be corrected through the record change process.
- Labels affixed to an answer folder in error could not be reused. All information had to be hand-gridded on another answer folder.

3.4 Test Accommodations

Accommodations and Modifications of Test Administration Procedures for Special Education Students and Students Eligible under Section 504 of the Rehabilitation Act of 1973

In accordance with the Individuals with Disabilities Education Act (IDEA) of 1997 and the No Child Left Behind Act, all students with disabilities must participate in state assessments. Students with disabilities who are in high school must participate in the New Jersey Biology Competency Test (NJBCT) or the Alternate Proficiency Assessment (APA). The Individual Educational Plan (IEP) team for each student determines which assessment a student will take in accordance with the NJAC 6A:14. The regulations state that each student with disabilities must take the general state assessment, which is the NJBCT, unless the student has not been instructed in any of the knowledge and skills tested and cannot complete any of the types of tasks on the NJBCT. The IEP team determined which assessment the student would take for each content area assessed.

Districts may have used modifications of test administration procedures when administering the NJBCT to special education students or to students eligible under Section 504 of the Rehabilitation Act of 1973. Modifications in the areas listed below may have been used separately or in combination.

Decisions about participation and accommodations and modifications were made by the Individualized Education Program (IEP) or 504 team. Information about test content and item types from the test specifications booklets might have been used to make this determination.

Any accommodations or modifications of test administration procedures for students eligible for special education under the Individuals with Disabilities Education Act (IDEA) or eligible under Section 504 of the Rehabilitation Act of 1973 must have been specified in the student's IEP or 504 accommodation plan. Accommodations or modifications must have been consistent with the instruction and assessment procedures used in the student's classroom. Students eligible for modifications under Section 504 may not have been classified but do have a permanent or temporary impairment in a major life function (for example: performing manual tasks, walking, seeing, hearing, speaking, etc.). Accommodations must have been recorded on the student's answer folder by their unique codes. Advanced planning was integral to implementing accommodations and modifications effectively and ensuring that the security of test materials was maintained.

Acceptable Accommodations or Modifications

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. at home or in a hospital (this will depend 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 administrator 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 the test
2. Administering the Braille version of the test

D. Test Procedures Modifications

1. Administration modifications
 - a. reading directions aloud
 - b. reading test questions aloud; ONLY the teacher who must read aloud or sign the test items is permitted to have a test booklet assigned to him/her for the task
 - 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 to sign or cue the directions or test items
 - 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 ONLY
 - 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. using graph paper for HSPA Mathematics (all students are permitted graph paper for NJASK 3–8)
 - k. using a Braille ruler and talking calculator or large-face calculator
 - l. 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
- m. using calculators for NJASK 3–8 Mathematics (all students are permitted calculators for HSPA)
- 2. Response modifications
 - a. having an examiner record the student’s identifying information on the test booklet and/or answer folder
 - b. dictating oral responses to a scribe (examiner or proctor who writes from dictation)
 - 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 (all editorial functions MUST be disabled)
 - f. providing an augmentative communication device
 - g. using a larger diameter or modified special grip #2 pencil
 - h. circling answers in the test booklet (the examiner subsequently transfers the answers to the answer folder; for NJASK 3–4, the examiner bubbles the student’s answer choice in the scannable test booklet)
 - i. allowing separate additional continuation pages for writing tasks

Other Considerations

Ensure that:

- a. any medication has been appropriately adjusted so it will not interfere with the student’s functioning.
- b. eyeglasses are used, if needed.
- c. hearing aids, FM systems, augmentative communication devices, word processors, or other types of 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.

For students responding on separate sheets of paper:

- a. Responses to open-ended questions and writing tasks, which are written or typed on separate sheets of paper by students eligible for this accommodation, must be placed in the **YELLOW SE/504 Accommodations & Braille/Large-Print Envelope**.
 - 1. The answer folder and separate sheets must be clipped together and placed beneath a separate grade-level Header Sheet.
 - 2. Write on the top of each separate sheet the student’s name, answer folder number, birth date, district name and code, and school name and code. **If these procedures are not followed, these responses cannot be linked to the student’s responses in the answer folder and the student will receive incomplete scores.**
- b. Copies of these pages must be made and retained on file by the school district until scores are received and verified.

For large-print test administration:

- a. Students taking the large-print test:
 1. Will mark their answers in the large-print test booklet and the examiner will transcribe the responses onto the regular answer folder provided.
 2. May be instructed to skip some items identified in the large-print supplemental instructions. The spaces for these items must be left blank on the student's answer folder included in the large-print kit.
- b. Answer folders used for the large-print test:
 1. Must be sent to Measurement Incorporated in the **YELLOW SE/504 Accommodations & Braille/Large-Print Envelope**.
 2. Grid SE/504 ACCOMM = A, B, C, and D. A separate grade-level Header Sheet must be completed for answer folders containing the multiple-choice responses.
- c. For dictations and responses recorded on separate sheets of paper:
 1. Students who dictate responses to open-ended items and writing tasks must indicate all punctuation and must spell all key words.
 2. Responses to open-ended questions and writing tasks recorded on separate sheets of paper must be clipped to the answer folder. The student's name, answer folder number, and birth date as well as the district and school names and codes must be recorded on each of the separate sheets attached.
 3. The district test coordinator must retain a duplicate of all open-ended item responses and writing task responses until scores are reported to and reviewed by district staff.

For Braille test administration:

- a. Students taking the Braille test:
 1. Will dictate their answers to the examiner and/or use a device that produces Braille.
 2. May be instructed to skip some items identified in the Braille supplemental instructions. The spaces for these items must be left blank on the student's answer folder included in the Braille kit.
 3. Should be instructed to bring a Braille ruler and a talking calculator to the test session.
- b. Answer folders used for the Braille test:
 1. Must be sent to Measurement Incorporated in the **YELLOW SE/504 Accommodations & Braille/Large-Print Envelope**.
 2. Grid SE/504 ACCOMM = A, B, C, and D. A separate grade-level Header Sheet must be completed for answer folders containing the multiple-choice responses.
- c. For dictations and responses recorded in Braille:
 1. Students who dictate responses for the open-ended items and writing tasks must indicate all punctuation and must spell all key words.
 2. Responses to the writing task and open-ended questions recorded in Braille must be transcribed. The transcriptions of the Brailled responses, along with the student's own Brailled work, must be clipped to the answer folder. The student's name, answer folder number, and birth date as well as the district and school names and codes must be recorded on each of the

separate sheets attached.

3. The district test coordinator must retain a duplicate of all open-ended item responses and writing task responses until scores are reported to and reviewed by district staff.

For students who communicate using sign language:

- a. An interpreter will be needed to interpret oral directions and test items (do NOT interpret the reading passages in Language Arts Literacy). The interpreter must be able to communicate in the mode used by the student, American Sign Language or a form of Manually Coded English, depending upon the student's communication system. The interpreter must be instructed to interpret so as not to give the answer to the student through the use of a particular sign or finger spelling.
- b. Students using American Sign Language for open-ended item responses 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.
- c. Students using Signed English or cued speech will sign or cue the responses to the interpreter who will transliterate (word for word) into spoken English, and a scribe will record the responses in the answer folder.

For students needing a scribe:

The student will dictate the response at first without indicating punctuation and spelling. This allows the student to get his/her ideas on paper. The scribe should write what is said, without capitals or punctuation and without the student seeing, but with correct spelling. Once the student has finished the response, the scribe must identify key words and ask the student to spell them. The scribe can underline the words and write the student's spelling above the word or write the student's spelling at the bottom of the page. The scribe should then show the piece to the student and ask him/her to indicate what words should be capitalized and where punctuation should occur. The student also has the opportunity to edit the piece at this point. The scribe must not advise or lead the student in any way.

Part 4: Scoring

4.1 MC items

The answer keys approved by NJ DOE were used to score the multiple-choice items after the responses were scanned. Each item had a key associated with the item (A, B, C, or D), which had been supplied and verified by the NJBCT content specialist. All correct answers were assigned the value of “1” while incorrect answers were assigned the value of “0.” At no time in this process was the original scanned answer overwritten, in case the key was determined to be incorrect during the post-scoring quality assurance check. After scoring was completed, simple item statistics were provided to the appropriate NJBCT content specialist to ensure that the correct keys were applied. If a key changed, then the process was repeated until the scoring file was correct. The key-check data that were provided to the Office of Student Learning Assessment for review are provided in Part 5.1. Each file contained the following information:

- proportion 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 Rpb (CFLAG).

4.2 OE and PAT items

The open-ended (OE) questions were dropped from the 2009 exam so there was no OE scoring. However, two performance assessment tasks were added to the 2010 test. This was reduced to one performance assessment task in 2011 and onward. The following sections detail the selection of scorers, range finding activities, field test range finding, the development of scoring guides, team leader training and qualifying, scorer training/qualifying, and the monitoring of scorer performance.

Scorer Selection

Because MI has been conducting the handscoring of writing and open-ended items for many years, a large pool of qualified, experienced readers was available for scoring NJBCT. MI routinely maintains supervisors’ evaluations and performance data for each person who works on each scoring project in order to determine employment eligibility for future projects. Many experienced readers were utilized for scoring NJBCT.

New readers are often needed to complement MI’s existing reader pool. MI’s procedures for selecting new readers are very thorough. After advertising in local newspapers, with the job service, and elsewhere, and receiving applications, staff in the human resources department review the applications and schedule interviews for qualified applicants. Qualified applicants are those with a four-year college degree in English, language arts, education, mathematics, science, or a related field. Each qualified applicant must pass an interview by experienced MI staff, write an acceptable essay, and receive good recommendations from references. The information about each applicant is then reviewed before offering employment.

In selecting team leaders, MI's management staff and scoring directors review the files of all the returning staff. They look for people who are experienced team leaders with a record of good performance on previous projects and also consider readers who have been recommended for promotion to the team leader position.

MI is an equal opportunity employer that actively recruits minority staff. Historically, temporary staff on major projects averages about 70% female, 30% male, 76% Caucasian and 24% minority. MI strongly opposes illegal discrimination against any employee or applicant for employment with respect to hire, tenure, terms, conditions, or privileges of employment, or any matter directly or indirectly related to employment, because of race, color, religion, sex, age, handicap, national origin, ancestry, veteran status, or sexual orientation.

Range Finding

Range finding meetings were conducted to establish "true" scores for a representative sample of student responses to the PATs. Between 100 and 220 sample responses per task were chosen by MI leadership personnel either from the available field test responses or from the current test administration. For items using specific rubrics, the rubrics were discussed and refined. The sample responses brought to the NJBCT range finding meetings were selected from a broad range of New Jersey school districts in order to ensure that the sample was representative of overall student performance. The range finding committees consisted of NJ DOE content specialists, NJ teacher representatives, MI management personnel, as well as the scoring director responsible for Biology.

Developing Scoring Guides

After the range finding meeting, MI management and the scoring director developed training materials consisting of an anchor set (examples of student responses for each score point) and training/qualifying sets for each task using the responses scored at range finding. Anchor sets usually consisted of three, or more, annotated examples of each score point in score point order. Training/qualifying sets consisted of clearly anchored student responses in random score point order.

Team Leader Training and Qualifying

After the anchor responses, training, and qualifying responses were identified and finalized, team leader training was conducted by the scoring director for each task. The procedures were similar to those for training scorers, but more comprehensive, dealing with resolution of discrepant scores, identification of non-scorable responses, unusual prompt treatment, alert situation responses (e.g., child-in-danger), and other duties performed only by team leaders. Team leaders took careful notes on the training responses in preparation for discussion with the scorers, and the scoring directors counseled them on the application of the rubric and training techniques. During scoring, team leaders responded to questions, spot-checked scorers, and counseled scorers having difficulty with the criteria. Furthermore, they were charged with monitoring the scoring patterns of each scorer throughout the project, conducting retraining as necessary,

performing resolution readings, and maintaining a professional working environment. Team leaders worked 7.75 hours per day, excluding breaks.

Scorer Training/Qualifying

All scorers were trained using the rubrics, as well as the anchor, training, and qualifying responses selected during the range finding meetings and approved by the NJ DOE. Scorers were assigned to a scoring group consisting of one team leader and 10-12 scorers. Each scorer's work was easily identifiable throughout the scoring session.

After contracts and nondisclosure forms were signed, training began. Scorer training followed the same format as team leader training. The scoring director introduced the set of anchor responses and thoroughly discussed each score point. This presentation was followed by practice scoring on the training sets. Scorers broke into teams to discuss the responses in the training sets. This arrangement gave scorers an opportunity to discuss any possible points of confusion or problems in understanding the criteria in a small group setting.

Team leaders analyzed the results of each training set, and identified which responses were giving scorers difficulty. The scoring director monitored each team leader during this process, listening to the team leaders' explanations and adding additional information when necessary. If a particular response or type of response seemed to be causing difficulty across teams, the problem was discussed with the room at large to ensure that everyone heard the same explanation.

Scorers were required to demonstrate their ability to score accurately by attaining 90% adjacent agreement (within one point) on the qualifying sets before they read packets of actual responses. Any reader unable to meet the standards set by the NJ DOE was dismissed. All scorers were made aware of this stipulation when they were hired.

Training was carefully orchestrated so that scorers understood how to apply the rubric in scoring the responses, learned how to reference the scoring guide, developed the flexibility needed to deal with a variety of responses, and retained the consistency needed to score all responses accurately. In addition to completing all of the initial training and qualifying, a significant amount of time was allotted for demonstrations of using the "Virtual Scoring Center" (MI's image scoring application). This included instructions on how to send notes with questions about scoring to team leaders and how to identify possible "content concern" alerts. Scorers generally worked 7.0 hours per day, excluding breaks.

Monitoring Scorer Performance

Each student response was be scored holistically by two independent readers using the Registered Holistic Scoring Method. The two independent scores, if identical or adjacent, were combined to produce the student’s final score on each task. If the two scores differed by more than one score point, the response was scored by a third reader. The final score was determined by an algorithm supplied by the NJ DOE.

MI project management constantly monitored the quality of each scorer’s work throughout the project. The scoring director and team leaders used a variety of methods to ensure that the student responses were being scored in accordance with the NJDOE rubrics and directions.

The first line of monitoring was the team leader, who worked with each reader on a daily basis. During training, readers were reminded that they should refer any questions not addressed in their guides and training sets to their team leader. This allowed the team leader to see what responses caused questions and to make sure that these responses received the correct score. Also, the team leader used those questions to focus their read-behinds for that reader. Also, if the question was one that the team leader could not answer, it went to the scoring director, who could then cover that issue with the entire room, if needed.

Team leaders also read behind each reader on a regular basis. Using the VSC system the team leader could set a percentage of each reader’s work to be sent automatically to the team leader for review. The percentage can be anything from 0 to 100. The team leader then identified any response, or type of response, that needed to be discussed with the reader. All such responses were explained to the reader, using guide and training responses as references, and the reader was given the opportunity to correct the score(s). This allowed for immediate feedback to readers.

Each day the data application continuously scanned the readers’ work and generated daily status reports. These reports showed the total number of responses read, the number of third readings required, and the percentage agreement of each reader, both perfect and adjacent. The reports also showed score point distributions. Scoring directors are experienced in examining the reports and used the information to determine the need for retraining of individual readers or the group as a whole. For individual readers, the scoring director and team leaders could see if a particular reader was scoring “high” or “low” according to the room average and, because of the specificity of the reports, they could pinpoint the score point(s) with which the reader was having difficulty. In addition, the VSC system produced a “post-scoring read behind report” which showed exactly how each reader had scored an individual response when compared to the second reader on the same response. This helped focus the team leader’s work with that reader. With daily reports, it was possible to see immediately the results of retraining.

In addition to the statistical information from the daily and cumulative reports, there was the validity set report. Each day members of each team took a validity set—a pre-scored and selected set of responses representing the range of scores in each item. The data application recorded the results and a validity report was printed out daily. Team leaders checked on individual readers and see which score points may have problematic for them. Also, the scoring director could see the cumulative responses to each set and see if there was any room-wide issue.

Retraining was an ongoing process once scoring began. If a whole team or a whole group was having difficulty with a particular type of response, large group training sessions were conducted. Standard retraining procedures included room-wide discussions led by the scoring director, team discussions conducted by team leaders, spot-checking of individual scorers by team leaders, and discussions between team leaders and individual scorers.

Scorers were dismissed when, in the opinion of the scoring director and the project director, they had been counseled, retrained, and given every reasonable opportunity to improve and were still performing below the acceptable standard.

4.3 Quality Control Procedures in Data Preparation

Machine Scoring/Linking of Answer Documents

In order to ensure the quality of the testing materials, Measurement Incorporated, the Office of Student Learning Assessment, and all sub-contractors worked together to rigorously proof all materials prior to printing/production.

As mentioned in Part 2, all items underwent multiple reviews to ensure that the operational and field test items used were valid and fair for all students. All assessment materials were submitted to rigorous editing and proofreading procedures. All materials to be developed were first checked by the editorial staff of Measurement Incorporated prior to being typeset to assure continuity existed across all documents.

Prior to typesetting of any documents, sample layouts based on the approved Style Guide were provided to NJBCT staff for review and approval. Typeset page proofs were then prepared and thoroughly proofread by well-trained staff members who read the documents in their entirety for typographical errors and for potential problems in context. Copies of the page proofs were provided to the NJDOE for review and approval. Upon approval of the page proofs, printer's proofs were produced. Printer's proofs of all documents were proofread by two staff members and two independent editors and then provided to the NJDOE for final approval prior to printing. All forms necessary for test administration were approved by NJDOE prior to final production.

Additionally, all accommodated materials were also reviewed for accuracy and quality at multiple stages. The first stage of review involved content specialists ensuring that the items being used on the tests were still valid in the accommodated format. Once this was completed, the large print test followed the quality control procedures discussed previously, while the other formats underwent additional quality control procedures. The Braille test was reviewed by an independent contractor that double-checked the integrity of the translation from print to Braille.

Part 5: Item-Level Statistics

5.1 Classical Item Statistics

Tables 5.1.1 and 5.1.2 present the item analysis summary for NJBCT. The item analysis provides statistical information about the items that helps identify items that may be mis-keyed. The following information was presented in each item analysis:

Classical difficulties: The proportion of students who answered each item correctly (PC) was presented. The proportion correct or p -value (for multiple-choice items) was used to identify items that were potentially too difficult or too easy. Generally, items are selected so that the p -values are greater than 0.25 and less than 0.95. Any mis-keys are corrected and the test is rescored.

Item Discriminations: The item-total correlation was presented as a measure of item discrimination. The item discrimination measures the relationship between the item score and the total score. The higher the correlation the better the item discriminates. The point biserial for the keyed response was presented for multiple-choice items.

Table 5.1.1: Item Difficulty and Discrimination Summary Statistics for Multiple-Choice Items by Cluster

Test Section Cluster	Item Difficulty			Item Discrimination
	Number of Items	Mean	STD	Mean
Biology	45	0.51	0.11	0.36
Organization and Development	11	0.49	0.09	0.38
Matter and Energy Transfer	8	0.53	0.11	0.39
Interdependence	6	0.47	0.12	0.32
Heredity and Reproduction	13	0.51	0.14	0.35
Evolution and Diversity	7	0.56	0.07	0.34

Table 5.1.2: Frequency Distribution for Multiple-Choice P-Values and Discrimination Indices by Cluster

	Number of Items	Difficulty							Discrimination					
		Median	P < 0.25	0.25 ≤ P < 0.40	0.40 ≤ P < 0.55	0.55 ≤ P < 0.70	0.70 ≤ P < 0.85	P ≥ 0.85	Median	RPB < 0.25	0.25 ≤ RPB < 0.35	0.35 ≤ RPB < 0.45	0.45 ≤ RPB < 0.55	RPB ≥ 0.55
Biology	45	0.53	0	9	17	18	1	0	0.38	5	13	19	8	0
Organization and Development	11	0.51	0	2	5	4	0	0	0.38	1	2	5	3	0
Matter and Energy Transfer	8	0.52	0	1	4	3	0	0	0.39	0	2	5	1	0
Interdependence	6	0.48	0	2	2	2	0	0	0.32	1	3	1	1	0
Heredity and Reproduction	13	0.54	0	4	3	5	1	0	0.37	1	5	7	0	0
Evolution and Diversity	7	0.59	0	0	3	4	0	0	0.38	2	1	1	3	0

5.2 Speededness

The NJBCT was intended to provide sufficient time for almost all students to respond to all of the items. The percentage of students omitting an item provided information about speededness. If the percentage was low, it implied that speededness was probably not an issue. If the percentage of omits was high, speededness may have been a problem, although other factors may be contributing to students omitting items besides not having enough time.

Table 5.2.1 presents data regarding omit rates. Table 5.2.1 shows the percentage of students omitting the last two multiple-choice items. The omit rates for the multiple-choice items were extremely low. Only between 1.04% and 1.97% of students omitted one of the last two multiple-choice items.

Table 5.2.1: 2016 NJBCT Percentage of Students Omitting the Last Item of Each Part

Biology Multiple Choice	
Order	Percent Omitting
14	1.04
15	1.05
29	1.14
30	1.14
44	1.95
45	1.97

5.3 Intercorrelations

Table 5.3.1 shows the correlations between students' scores and content clusters. The correlations between strand scores and Biology total score tended to be relatively high. They ranged from 0.49 to 0.85.

Table 5.3.1: NJBCT Intercorrelations Among Strands

	Major Content Clusters and Item Types					
Major Content Clusters and Item Types	BIO	O	M	I	H	E
BIO Biology	1.00					
O Organization and Development	0.85	1.00				
M Matter and Energy Transfer	0.81	0.65	1.00			
I Interdependence	0.70	0.55	0.53	1.00		
H Heredity and Reproduction	0.85	0.67	0.63	0.53	1.00	
E Evolution and Diversity	0.76	0.59	0.57	0.49	0.58	1.00

Part 6: Standard Setting

In August of 2008, standard setting for the Biology test took place. After completion, the NJDOE decided against submitting the cut scores to the State Board of Education. Initially, NJDOE decided to report raw scores only. However, in 2011 the results of the standard setting were provisionally approved as a temporary reporting mechanism in order to meet NCLB compliance. The executive summary of the Standard Setting Report for NJBCT is in Appendix D.

Part 7: Calibration, Scaling and Equating

7.1 Calibration and Scaling

The NJBCT test was calibrated with WINSTEPS using the Item Response Theory (IRT) Rasch model to estimate the difficulty of each item on the test and the IRT-based proficiency score (theta) for each examinee. WINSTEPS is a program written by Linacre (Mesa Press, 2005) to conduct Rasch analyses. The program is a Windows-based program that is widely used for high stakes tests.

Based on the results of the NJBCT standard settings in 2008 and the reporting policy adopted by NJ DOE, the total score in the NJBCT test was reported as a scale score ranging from 100 to 300. The scale scores of 200 and 250 were provisionally designated as the cut points for proficient and advanced students respective. The scale score range for each provisional proficiency standard follows:

<u>Proficiency Standard</u>	<u>Scale Score Range</u>
Partially Proficient	100 – 199
Proficient	200 – 249
Advanced Proficient	250 – 300

The scaling for NJBCT involved two steps. First, the raw score to IRT-based proficiency score (theta) scale table for the current year was obtained from the WINSTEPS anchored calibration. Then, the raw score to scale score conversion table (scoring table) was derived via a simple linear transformation of theta, where the linear transformation constants (slope and intercept) were determined from the standard setting. Per NJ DOE policy, the linearly transformed scores were rounded to whole numbers and the lower and upper bounds were set to 100 and 300 respectively. An additional pseudo rounding step was applied to always ensure that the cut scores of 200 and 250 actually appeared in the scoring table. Table B.1 in Appendix B contains the 2016 BCT raw score frequency distributions.

7.2 Equating

Equating refers to the statistical process of adjusting the current-year IRT calibration to the base scale. Equating or linking a calibration is tantamount to maintaining the continuity of the base scale and proficiency standards over time. The equating process for the NJ EOC Biology examination was carried out in three steps. Step 1 involved calibrating all of the operational (i.e., scored) items taken on a particular, current test form using WINSTEPS. All examinees in the sample were used in this first calibration step. In Step 2, an iterative stability analysis was carried out using common anchor items that had been previously calibrated and equated to the base scale. This group of items is collectively referred to as the anchor-item link set. The 2016 BCT operational assessment consisted of all the items from the 2013 operational assessment, excepting the PAT. Thus, the 2016 anchor set consisted of all 45 multiple-choice questions. The stability analysis determined which items remained in the anchor-item link set by iteratively deleting from the link set any item demonstrating a logit displacement (item-level equating discrepancy)—after applying the mean-shift equating constant—of 0.3 or smaller. The stability analysis was terminated when all of the remaining linking items had discrepancies of less than

0.3. In Step 3, WINSTEPS was rerun, now fixing the item parameter estimates for the anchor items remaining in the link set at their previously calibrated/equated values. This final step effectively equated all of the items on the current test form to the base scale and generated a scoring table that was used to compute IRT-based scores on the base scale for any examinees taking the current test form. The IRT-based scores, in turn, were used in the scaling process as described earlier.

Only 4 of the 45 items were flagged during the stability analysis. That indicates that the difficulty of the items was stable from 2013 to 2016. Thus, 41 multiple-choice items were utilized as anchor items. A more extensive review of the equating procedures may be found in the 2016 BCT Equating Report.

Course and Scheduling Results

The NJBCT is for all New Jersey public high school students, regardless of grade level, who are enrolled in a first-time Biology course at any time (full year or block schedule) during the 2015–16 school year. Table 7.2.1 provides examples of course titles and academic levels.

Table 7.2.1: NJBCT Course Titles and Academic Levels

Course Title	Academic Level
Resource Center Biology	Low
ESL Biology	Low
Integrated Science¹	Low
General Biology	General
Life Science	General
Foundations in Biology	General
College Prep Biology	Advanced
Honors Biology	Advanced
Accelerated Biology	Advanced
Advanced Placement Biology²	Advanced

¹ If alternative to biology offerings listed

² If no prerequisite biology course is required

Some students take their Biology course in a full year schedule while others are on block scheduling. Table 7.2.2 shows the breakdown of the 102,514 General Education students who

took the test by schedule, but were not voided. Table 7.2.3 shows the breakdown of the same 102,514 General Education students by academic level. Table 7.2.4 shows the mean raw scores by schedule and level.

Table 7.2.2: 2016 NJBCT Students by Schedule

Schedule	Number of Students	Percent
Full year	92,046	89.79
Block (1st Semester)	4,364	4.26
Block (2nd Semester)	4,833	4.71
Other	1,271	1.24
Total	102,514	100.00

Table 7.2.3: 2016 NJBCT Students by Academic Level

Academic Level	Number of Students	Percent
Low	5,287	5.16
General	32,846	32.04
Advanced	63,346	61.79
Other	1,035	1.01
Total	102,514	100.00

Table 7.2.4: 2016 NJBCT Raw Score Mean and Standard Deviation by Schedule and Level

Schedule	Level	Number of Students	Mean	STD
Full year	Low	4,853	18.01	6.89
	General	29,227	24.27	9.35
	Advanced	57,767	33.17	10.51
	Other	199	15.71	7.84
	Overall	92,046	29.51	11.14
Block (1st Semester)	Low	169	15.78	5.56
	General	1,523	24.22	9.44
	Advanced	2,619	27.47	9.97
	Other	53	32.12	9.23
	Overall	4,364	25.94	9.99
Block (2nd Semester)	Low	243	17.01	5.70
	General	1,923	24.29	9.69
	Advanced	2,630	27.43	9.42
	Other	37	31.39	8.78
	Overall	4,833	25.69	9.71
Other	Low	22	15.00	3.71
	General	173	16.92	7.81
	Advanced	330	22.99	9.29
	Other	746	22.94	10.50
	Overall	1,271	22.00	10.04
Any Schedule	Low	5,287	17.88	6.80
	General	32,846	24.23	9.38
	Advanced	63,346	32.64	10.58
	Other	1,035	22.33	10.71
	Overall	102,514	29.08	11.10

To determine whether the group differences were significant, an analysis of variance was conducted. Results of that analysis are shown in Table 7.2.5.

Table 7.2.5: Results of Analysis of Variance

Source	SS	DF	MS	F	p
Between	114406.01	2	57203.00	470.13	<.0001
Within	12318260.67	101240	121.67	–	–
Total	12432666.68	101242	–	–	–

As Table 7.2.5 shows, the differences among groups are significant beyond the .0001 chance level (i.e., less than one chance in 10,000 that these results would have occurred randomly had no real differences existed). Owing to the size of the groups, however, nearly any difference will be considered statistically significant. Therefore, it was reasonable to examine effect size as well. Cohen (1988) has defined effect size as the difference between groups, relative to the standard deviation of scores. Table 7.2.6 summarizes effect sizes. In this table, the z-score for each group is the mean score for that group expressed as a standardized difference between that group's total

score mean and the population's total score mean. In this instance, the population is the 102,514 students who indicated a schedule.

Table 7.2.6: Summary of Effect Sizes

Schedule	Z	Effect
Full Year vs All	0.0383	Negligible
1st Semester vs All	0.2830	Small
2nd Semester vs All	0.3055	Small
Full Year vs 1st Semester	0.3213	Small
Full Year vs 2nd Semester	0.3438	Small
1st Semester vs 2nd Semester	0.0226	Negligible

Because the effect for Full Year was effectively based on a comparison of the Full-Year group with itself (i.e., most of the total population was in the Full-Year group), it was also possible to examine differences between groups, using the standardized difference (z-score). The z-score between Full-Year and 2nd Semester, for example was 0.3438, while that between Full-Year and 1st Semester was 0.3213. According to Cohen (1988), effects smaller than .2 are negligible. Those between .2 and .49 are small, those between .5 and .79 are moderate, and those .8 or above are large. Thus, the differences between Full-Year and either 1st or 2nd Semester can be considered small.

Item Map and Test Information Function

Item map for Biology is presented in Figure 7.2.1. It is essentially organized as two vertical plots with the left side indicating where students' abilities are clustered and the right side indicating where the item difficulties are clustered. The items were generally centered around the proficient cut score of 0.211. This indicates that there were an ample number of items around the cut score point to adequately differentiate between students who should be categorized as proficient or partially proficient. The interpretation of the test information function in Figure 7.2.2 is similar. Ideally, the peak of the test information function should be where the most important decisions are taking place. In the case of BCT the categorization of students into proficiency standards is of paramount importance. Both the proficient and advanced proficient cut scores were close to the peak of the test information function. Thus, the 2016 BCT assessment effectively maximized test information exactly where it was most needed.

Figure 7.2.1 Biology Item Map

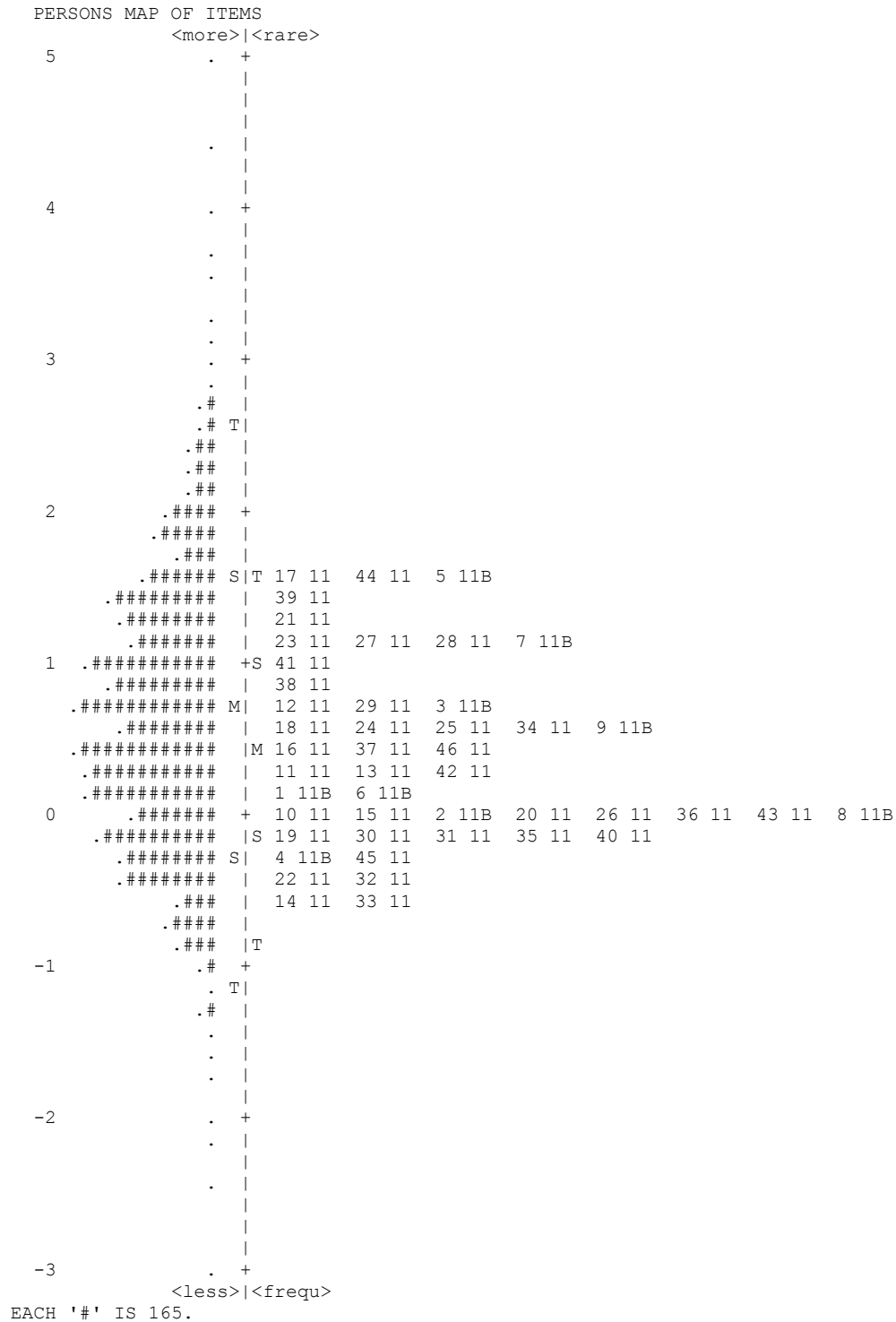
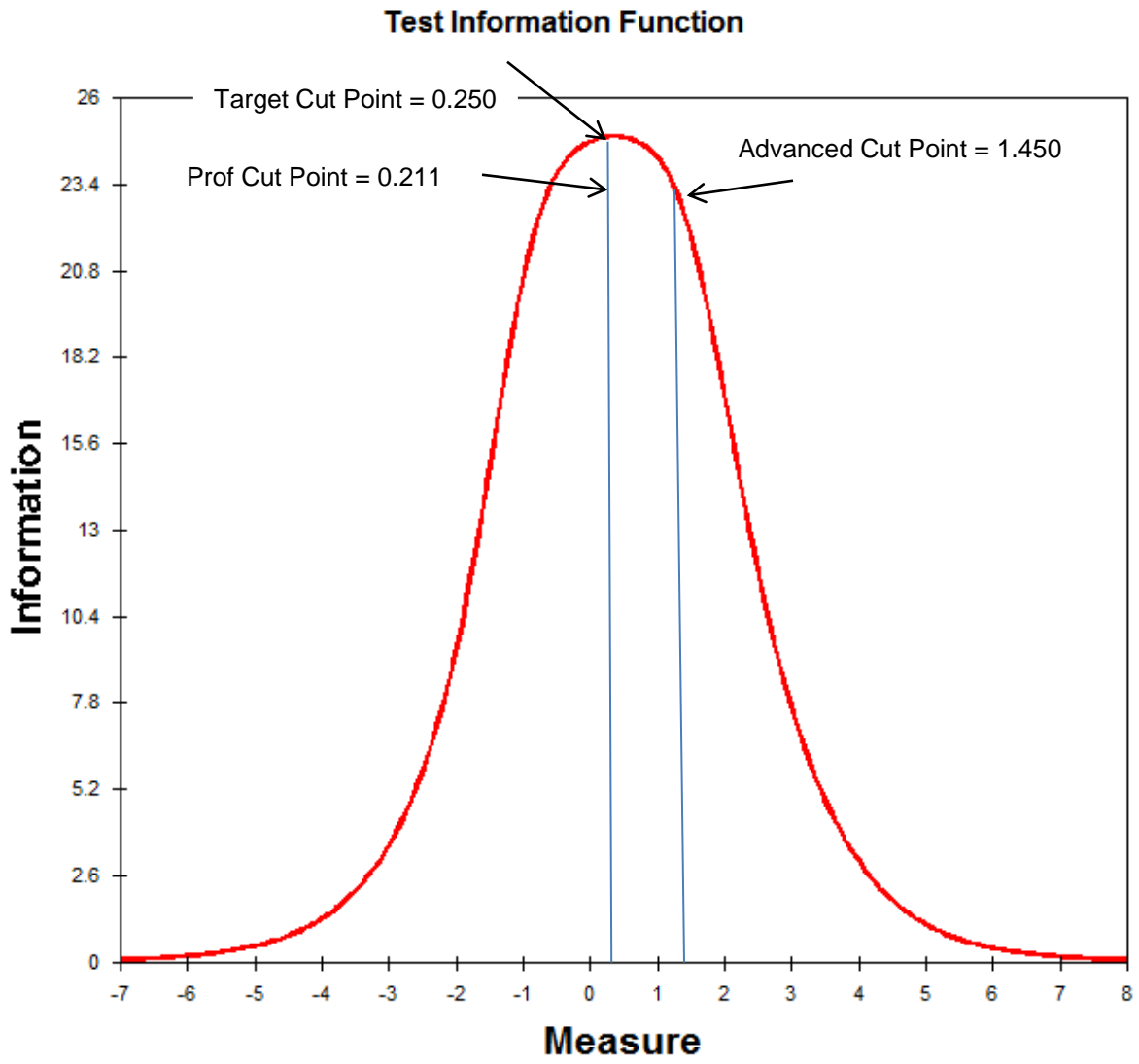


Figure 7.2.2 Biology Test Information Function Figure



Part 8: Test Statistics

8.1 Summary Statistics

Mean and standard deviation of students' raw scores on each content area are given in Table 8.1.1 for the May 2016 test. These data are based on all the general education students who took Biology in 2016, except those students who were voided.

Table 8.1.1: NJBCT Raw Score Means and Standard Deviations

Test	Number of Points	Raw Score Mean	Standard Deviation	Number Tested
Biology	57	29.08	11.10	102,514

Table 8.1.2 reports the mean and standard deviation for students' raw scores by cluster for the May 2016 test. The number of raw score points for both multiple-choice are presented. Additionally, the mean percent correct score for each standard is also presented.

Table 8.1.2: 2016 NJBCT Means and Standard Deviations of Students' Raw Score and Percent Correct by Cluster

Test Section / Cluster	Number of Possible Points	Raw Score		Percent Correct	
		Mean	Standard Deviation	Mean	Standard Deviation
Biology*	57	29.08	11.10	51.02	19.47
Organization and Development	11	5.35	2.67	48.64	24.24
Matter and Energy Transfer	8	4.21	2.08	52.57	26.02
Interdependence	6	2.81	1.51	46.86	25.13
Heredity and Reproduction	13	6.62	2.85	50.94	21.92
Evolution and Diversity	7	3.90	1.75	55.76	24.95

* Raw Score calculations of Mean and Standard Deviation include the PAT items.

8.2 Classical Reliability Estimates of the Test Scores

Table 8.2.1 summarizes reliability estimates and standard error of measurements (SEMs) for NJBCT total test and by clusters. The reliability coefficients are based on Cronbach's alpha measure of internal consistency. The SEMs are expressed in terms of the raw score metric. When evaluating these results it is important to remember that reliability is partially a function of test length and thus reliability is likely to be greater for clusters that have more items. Table C.1 in Appendix C shows the standard error plot of the 2016 BCT assessment.

Table 8.2.1: 2016 NJBCT Reliability Estimates and Standard Errors of Measurement for the Total Test and by Cluster

Test Section / Clusters	Number of Possible Points	Reliability	Raw Score SEM
Biology	57	0.86	4.20
Organization and Development	11	0.69	1.49
Matter and Energy Transfer	8	0.64	1.25
Interdependence	6	0.45	1.12
Heredity and Reproduction	13	0.68	1.61
Evolution and Diversity	7	0.52	1.21

Subgroups of interest, as requested by NJ DOE were analyzed for their reliability as well. Table 8.2.2 contains the reliability estimates and SEMs in NJBCT for subgroups requested at the whole test level as well as at the cluster level.

Subgroups

Table 8.2.2: 2016 NJBCT Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters for Subgroups

Subgroup	N	Mean	Test Level			Organization and Development			Matter and Energy Transfer			Interdependence			Heredity and Reproduction			Evolution and Diversity		
			Rel	SD	SEM	Rel	SD	SEM	Rel	SD	SEM	Rel	SD	SEM	Rel	SD	SEM	Rel	SD	SEM
General Education	87,701	30.5	0.85	10.88	4.18	0.68	2.65	1.49	0.63	2.04	1.25	0.45	1.50	1.12	0.68	2.83	1.61	0.50	1.70	1.21
Special Education	14,813	20.7	0.79	8.38	3.87	0.50	2.11	1.48	0.52	1.82	1.26	0.31	1.34	1.11	0.49	2.26	1.62	0.43	1.61	1.22
Limited English Proficient	4,790	19.2	0.77	8.27	3.98	0.48	2.05	1.47	0.52	1.79	1.24	0.23	1.27	1.11	0.46	2.20	1.62	0.35	1.51	1.22
Male	51,913	28.6	0.87	11.42	4.18	0.70	2.72	1.49	0.66	2.12	1.24	0.47	1.53	1.11	0.70	2.91	1.61	0.55	1.80	1.20
Female	50,435	29.6	0.85	10.73	4.17	0.67	2.61	1.50	0.62	2.04	1.26	0.43	1.48	1.12	0.66	2.78	1.61	0.47	1.69	1.22
White	50,293	31.8	0.84	10.18	4.07	0.65	2.56	1.50	0.59	1.97	1.25	0.42	1.46	1.11	0.65	2.73	1.61	0.47	1.66	1.21
Black or African American	15,293	22.8	0.81	9.60	4.15	0.59	2.32	1.50	0.58	1.94	1.26	0.34	1.38	1.12	0.56	2.45	1.63	0.47	1.69	1.23
Asian	9,946	37.5	0.87	10.74	3.93	0.73	2.69	1.39	0.65	1.95	1.16	0.48	1.50	1.09	0.73	2.90	1.52	0.49	1.59	1.14
Native Hawaiian or Pacific Isl.	176	31.7	0.84	10.60	4.18	0.68	2.64	1.50	0.54	1.88	1.27	0.39	1.45	1.14	0.66	2.78	1.62	0.48	1.66	1.19
Hispanic or Latino	24,802	24.2	0.81	9.59	4.14	0.58	2.33	1.51	0.57	1.95	1.27	0.35	1.40	1.13	0.57	2.49	1.63	0.43	1.64	1.24
American Indian or Alaskan	118	27.2	0.84	10.73	4.24	0.71	2.73	1.48	0.55	1.91	1.29	0.43	1.50	1.13	0.62	2.63	1.62	0.41	1.60	1.23
Economically Disadvantaged	33,505	23.7	0.81	9.50	4.14	0.57	2.31	1.51	0.57	1.94	1.27	0.34	1.39	1.13	0.56	2.46	1.63	0.43	1.65	1.24
Non-Econ. Disadvantaged	69,009	31.7	0.86	10.87	4.13	0.69	2.67	1.49	0.63	2.04	1.24	0.46	1.50	1.11	0.69	2.85	1.60	0.51	1.71	1.20

Part 9: Validity

The New Jersey Department of Education developed a comprehensive set of assessments that measure student achievement of the Core Curriculum Content Strands. The validity of the NJBCT scores was based on the alignment of the NJBCT assessment to the Core Curriculum Content Strands and the knowledge and skills expected of high school students.

The Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p.13-16) note the following possible sources of validity evidence:

- Evidence based on test content presented in part 2.1. Construct Validity describes how the test content relates to the construct. Part 2.1 describes the test content that describes the construct.
- Evidence based on internal structure of the test presented in parts 2.2 and 2.3. Content Validity describes the item review and development process. The internal structure of the test is described.
- Evidence based on relations to other variables presented in part 5.3. Concurrent Predictive validity describes the relationship between the test scores and an external test.
- Evidence based on consequences of testing is presented in part 1.2.


For an assessment like NJBCT, one intended to measure student's performance in relation to the Core Curriculum Content Strands, content validity was the most relevant and important source of validity evidence. The section of this technical report on "Test Development," presents validity evidence based on the test content. A description of the test specification development is followed by the procedures for test item development. Details about item writing as well as task, prompt, and passage selection are included. Part 2 delineates the review work of the New Jersey Assessment Content Committees.

Part 10: Reporting

10.1 Cycle I Reports

Individual Student Report (ISR)

The Individual Student Report (ISR) is a report of how each individual student did on the NJBCT. Two copies of the report were produced for every student tested, one for the student's permanent folder after the results were analyzed, and the other for the student's parent/guardian to be shared in a manner determined by the local district. A sample ISR is shown below.

TEST DATE: MAY 2013 REPORT PRINTED: 1/14/2013		New Jersey Biology Competency Test Individual Student Report						
COUNTY: 88 ANY COUNTY DISTRICT: 7777 ANY DISTRICT SCHOOL: 666 ANOTHER SCHOOL								
Student Information				Student Scores				
Student Name: AVELLA, LAURA				Your Raw Score	Your Scale Score	State Scale Score Mean	Proficiency Level	Pass
NJBCT ID No.: 0098900126		SID (NJ SMART) No.:		30.5	204	208.8	PROFICIENT	YES
District/School ID No.:		Grade: 9						
Answer Folder No.: 45388		Date of Birth: 03/27/98						
LEP: 1		Sex: F						
SE:		Title I:						
Special Form:								
Partially Proficient / Not Pass: Scale Score BELOW 200 Proficient / Pass: Scale Score AT OR ABOVE 200 but BELOW 250 Advanced Proficient / Pass: Scale Score AT OR ABOVE 250								
Cluster and Performance Assessment Points								
Biology								
Multiple Choice				Performance Assessment				
Cluster	Your Points	Just Proficient Mean		Performance Assessment Task	Your Points	Just Proficient Mean		
Organization and Development	8.0 out of 11	4.8		Performance Assessment Task	6.0 out of 12	5.0		
Matter and Energy Transfer	1.0 out of 8	4.6						
Interdependence	3.0 out of 6	3.2						
Heredity and Reproduction	6.0 out of 13	6.8						
Evolution and Diversity	6.5 out of 7	3.9						
TOTAL MULTIPLE CHOICE POINTS	24.5 out of 45			TOTAL PERFORMANCE ASSESSMENT POINTS	6.0 out of 12			
TOTAL RAW SCORE POINTS 30.5 out of 57								
88-7777-666								

10.2 Cycle II Reports

Performance by Demographic Group Report

The Performance by Demographic Group Report is a one-page report that presents test results for NJBCT. The test results are broken down within the report by student subgroup (GE, SE, LEP, and Total), gender, migrant status, ethnicity, and economic status. Individual reports were produced for the school, district, statewide population, and the DFG of the district. For Special

Needs districts, reports were also produced for the statewide Special Needs population, and the statewide non-Special Needs population.

10.3 State Summary Report

With only raw scores being reported, there was no State Summary Report.

10.4 Interpreting Reports

Raw Score Data for Special Needs and Non-Special Needs: Data included the mean number of points obtained by all the listed student groups (Total Students, GE, SE, and LEP) for each cluster in the 31 Special Needs or Abbott districts and the mean number of points for these same listed student groups for each cluster in all other non-Special Needs districts.

Raw Score Data for DFG: Data included the mean number of points obtained by all the listed student groups (Total Students, GE, SE, and LEP) for each cluster in the same DFG as the district listed on this report.

10.5 Quality Control in Reporting

Quality control procedures for score reporting

MI employed stringent quality control procedures to ensure that reporting on all levels was completely accurate. MI thoroughly tested, reviewed, and proofread all reporting deliverables prior to this year's deliveries to NJ DOE or other parties such as other state offices and districts.

MI used structured testing methodologies to test all software programs that were used in the production of score reports. All scanning and scoring programs were fully tested and reviewed prior to the start of live material processing and were continually monitored throughout the process. MI's QA staff developed and utilized independent queries to validate all software programs and programmatically produced deliverables for reporting. Before reporting any results, the first step was to ensure that the data were accurate. Each program was tested to ensure that all data were included or excluded as appropriate (with particular attention to Braille and Large Print students as well as any other special equating situations) and to ensure that all programmatic calculations were performed accurately and according to the reporting rules provided by the State Contract Manager. Next, it was verified that all formatting rules were followed and that reports were presented as they were designed by MI and NJ DOE with all appropriate fonts, footnotes, line separations, sections, and headings. All aspects of the reports were tested to ensure that valid values were verified, all valid codes were included on all student records, correct scores were attributed to the correct student, cluster scores were accurately aggregated and totaled, and appropriate student totals were reported in all aggregate reports. In addition, during phase II of Cycle I Quality Control (QC) review, it was ensured that all appropriate record changes made by districts had been applied. The testing process described above is inclusive of data files, electronic reports and printed reports.

MI worked cooperatively with NJ DOE to arrange for NJ DOE QC review of selected districts, schools, and students prior to release of any score reporting. MI assisted the NJ DOE staff with

this QC review in various ways. The guidelines that were developed for QC sample selection were a joint collaboration between MI and NJ DOE staff. The QC sample was based on these guidelines. After the initial proposed selection was made, the list was sent to NJ DOE for approval and any additions that they wished to make. The QC sample consisted of students enrolled in 8-9 QC districts plus miscellaneous individual student records. The QC districts included the following:

- A Charter school,
- At least one school district with multiple schools, and
- Districts taking the alternate form and any other districts with students receiving any special equating due to testing irregularities.

Altogether, the QC districts were distributed over the full range of DFGs, but did not necessarily represent each DFG.

In addition to the QC districts, miscellaneous individual student records also included:

- A minimum of five Braille and five large-print students testing because these students may have had certain items omitted, leading to different maximum cluster scores. If there were not enough Braille and large print students in the QC districts, more students from other districts were added.
- Any miscellaneous problem answer folders that NJ DOE wanted to monitor. These might have been added before or after the initial QC sample had been selected.

The QC process was reviewed after each test administration, and improvements were developed in order to streamline the process and make it much more efficient and effective. The current methodology for selecting the sample for NJBCT has historically provided the best coverage of test scenarios for review and allowed for the effective identification of students who represent all scenarios. Furthermore, it allowed the NJ DOE staff to review fewer students to determine the correctness of the processing and programs. Through improvements to the QC process, a significant decrease in the amount of time needed to complete the review has been achieved.

During the QC process, live answer folders were hand-scored for multiple-choice (MC) items using scoring masks, and reader score sheets were reviewed and compared to student records to ensure that the hand scores were appropriately applied. A selection of students was presented to ensure coverage of each type of demographic coding scenario as well as any overrides that were done by MI according to coding rules as developed by NJ DOE.

Live reports were reviewed for the selection of students, schools, and districts. Reports were available to the State Contract Manager immediately following review and approval by MI QA staff. MI provided printed and electronic copies of a sample of all reports to the State Contract Manager for each phase of NJ DOE QC review. These reports were made available via download from the secure FTP site. All reports were provided in a timely manner for each phase and in advance whenever possible.

Preliminary reporting quality control sample

MI worked with NJ DOE to select certain districts and other special educational entities to include in the quality control sample for score reporting. As discussed in the previous section, this sample represents regular schools as well as Figureer schools, educational programs administered by the state, and private schools for the disabled. The sample at each grade level had up to 10 schools in eight regular districts and up to three special districts or state facilities.

MI made arrangements for the early processing of answer folders after the first spring week of testing so that they could be scanned and scored first. This allowed for additional time to prepare the reports for examination by MI's QA staff prior to NJ DOE review. This also allowed for the introduction of a limited amount of live data into the systems, so that it was possible to troubleshoot any problematic situations that became apparent prior to operating at higher capacity. Once these documents were scanned, a copy of the data was saved, and subsequently used to create the sample reports. Using a copy allowed other processes such as record change requests to be applied to the original data without causing changes to the quality control reports.

After MI's QA staff checked the reports for accuracy, a complete set of Cycle I reports for these students was generated for comparison to the actual answer documents. MI provided original answer folders and photocopies of the answer documents and scannable reader score sheets so that NJ DOE staff could compare the gridding of demographic and multiple choice bubbles to the score reports.

Overall, the QC review for the May 2016 NJBCT administration went smoothly. All data and reports were approved for release at the conclusion of the visit allowing for reports to be produced in a timely and accurate manner.

Reference

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N.J., & Holland, P.W. (1993). DIF detection and Description: Mantel-Haenszel and Standardization. In P.W. Holland, and H. Wainer (Eds.), *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Zieky, M. (1993). Practical Questions in the Use of DIF Statistics in Test Development, IN P.W. Holland, and H. Wainer (Eds.), *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Appendix A: Test Specifications

BIOLOGY COMPETENCY TEST MATRIX

	Comprehension/Recall Understand Core Concepts and Principles (10%)	Application Make connections among scientific explanations (60%)	Data Analysis Evaluate mathematical, and computational skills (30%)
Organization and Development (O)			
Biochemistry	A1	B1	C1
Cellular transport	A2	B2	C2
Cell Growth and Division	A3	B3	C3
Homeostasis and Gene Regulation	A4	B4	C4
Organization and Development	A5	B5	C5
Matter and Energy Transfer (M)			
Biosynthesis	A6	B6	D6
Energy Flow	A7	B7	C7
Photosynthesis and Cellular Respiration	A8	B8	C8
Interdependence (I)			
Ecology	A9	B9	C9
Heredity and Reproduction (H)			
Genetics	A10	B10	C10
Heredity	A11	B11	C11
Evolution and Diversity (E)			
Natural Selection/Organic Evolution	A12	B12	C12
Diversity of Life	A13	B13	C13

Test Layout

Operational Base Test:

57 points

Parts 1,2, & 5 - 15 MC Part 4 PAT

Embedded Field Test:

Part 6 - 15 MC Part 3 PAT

Timing: 30 - 45 Minutes per Part

Reporting

By Strand:

- 1. Organization and Development**
- 2. Matter and Energy Transfer**
- 3. Interdependence**
- 4. Heredity and Reproduction**
- 5. Evolution and Diversity**

Appendix B: Raw Score Frequency Distribution

Table B.1: NJBCT Raw Score Frequency Distribution

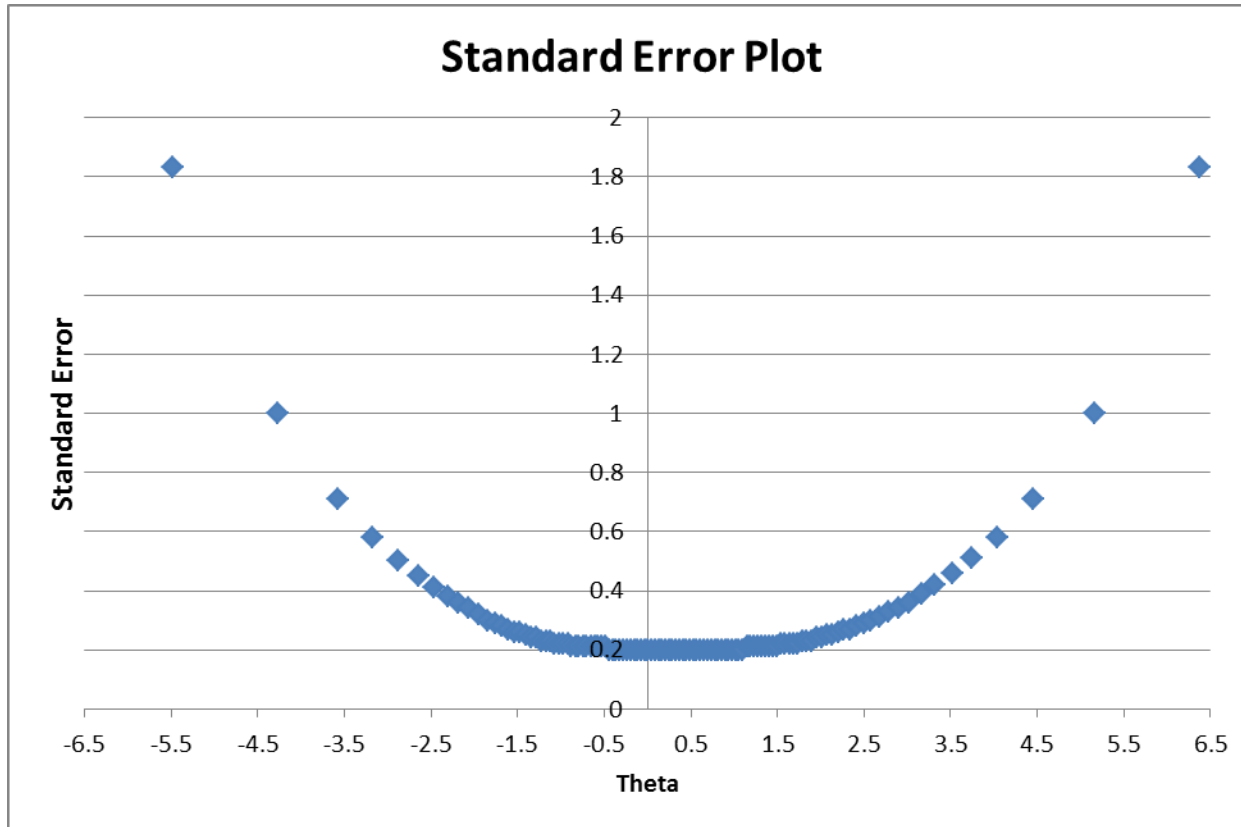
Raw Score	Theta	SE	Freq	%	Cum. Freq	Cum. %
0.0	-5.48	1.83	5	0.00	5	0.00
0.5	-4.27	1.00	0	0.00	5	0.00
1.0	-3.58	0.71	10	0.01	15	0.01
1.5	-3.17	0.58	0	0.00	15	0.01
2.0	-2.88	0.50	26	0.03	41	0.04
2.5	-2.65	0.45	0	0.00	41	0.04
3.0	-2.47	0.41	65	0.06	106	0.10
3.5	-2.31	0.38	0	0.00	106	0.10
4.0	-2.18	0.36	119	0.12	225	0.22
4.5	-2.06	0.34	1	0.00	226	0.22
5.0	-1.95	0.32	188	0.18	414	0.40
5.5	-1.85	0.30	2	0.00	416	0.41
6.0	-1.76	0.29	310	0.30	726	0.71
6.5	-1.68	0.28	7	0.01	733	0.72
7.0	-1.61	0.27	408	0.40	1141	1.11
7.5	-1.53	0.26	9	0.01	1150	1.12
8.0	-1.47	0.26	549	0.54	1699	1.66
8.5	-1.40	0.25	24	0.02	1723	1.68
9.0	-1.34	0.24	636	0.62	2359	2.30
9.5	-1.28	0.24	34	0.03	2393	2.33
10.0	-1.23	0.23	894	0.87	3287	3.21
10.5	-1.17	0.23	39	0.04	3326	3.24
11.0	-1.12	0.23	1111	1.08	4437	4.33
11.5	-1.07	0.22	65	0.06	4502	4.39
12.0	-1.02	0.22	1339	1.31	5841	5.70
12.5	-0.97	0.22	101	0.10	5942	5.80
13.0	-0.92	0.22	1522	1.48	7464	7.28
13.5	-0.88	0.21	155	0.15	7619	7.43
14.0	-0.83	0.21	1755	1.71	9374	9.14
14.5	-0.79	0.21	229	0.22	9603	9.37
15.0	-0.74	0.21	1926	1.88	11529	11.25
15.5	-0.70	0.21	344	0.34	11873	11.58
16.0	-0.65	0.21	2084	2.03	13957	13.61
16.5	-0.61	0.21	423	0.41	14380	14.03
17.0	-0.57	0.21	2147	2.09	16527	16.12
17.5	-0.53	0.21	526	0.51	17053	16.63
18.0	-0.48	0.21	2235	2.18	19288	18.81
18.5	-0.44	0.20	637	0.62	19925	19.44
19.0	-0.40	0.20	2310	2.25	22235	21.69
19.5	-0.36	0.20	681	0.66	22916	22.35
20.0	-0.32	0.20	2295	2.24	25211	24.59
20.5	-0.28	0.20	749	0.73	25960	25.32
21.0	-0.23	0.20	2446	2.39	28406	27.71
21.5	-0.19	0.20	806	0.79	29212	28.50

Raw Score	Theta	SE	Freq	%	Cum. Freq	Cum. %
22.0	-0.15	0.20	2398	2.34	31610	30.83
22.5	-0.11	0.20	852	0.83	32462	31.67
23.0	-0.07	0.20	2376	2.32	34838	33.98
23.5	-0.03	0.20	815	0.80	35653	34.78
24.0	0.01	0.20	2422	2.36	38075	37.14
24.5	0.05	0.20	906	0.88	38981	38.03
25.0	0.09	0.20	2306	2.25	41287	40.27
25.5	0.13	0.20	897	0.88	42184	41.15
26.0	0.17	0.20	2285	2.23	44469	43.38
26.5	0.21	0.20	881	0.86	45350	44.24
27.0	0.25	0.20	2331	2.27	47681	46.51
27.5	0.29	0.20	931	0.91	48612	47.42
28.0	0.33	0.20	2331	2.27	50943	49.69
28.5	0.37	0.20	979	0.95	51922	50.65
29.0	0.41	0.20	2191	2.14	54113	52.79
29.5	0.45	0.20	890	0.87	55003	53.65
30.0	0.49	0.20	2039	1.99	57042	55.64
30.5	0.53	0.20	912	0.89	57954	56.53
31.0	0.57	0.20	2167	2.11	60121	58.65
31.5	0.61	0.20	937	0.91	61058	59.56
32.0	0.65	0.20	2091	2.04	63149	61.60
32.5	0.70	0.20	980	0.96	64129	62.56
33.0	0.74	0.20	2024	1.97	66153	64.53
33.5	0.78	0.20	957	0.93	67110	65.46
34.0	0.82	0.20	1843	1.80	68953	67.26
34.5	0.86	0.20	891	0.87	69844	68.13
35.0	0.90	0.20	1825	1.78	71669	69.91
35.5	0.94	0.20	903	0.88	72572	70.79
36.0	0.98	0.20	1788	1.74	74360	72.54
36.5	1.02	0.20	867	0.85	75227	73.38
37.0	1.06	0.20	1655	1.61	76882	75.00
37.5	1.10	0.20	885	0.86	77767	75.86
38.0	1.15	0.21	1564	1.53	79331	77.39
38.5	1.19	0.21	819	0.80	80150	78.18
39.0	1.23	0.21	1571	1.53	81721	79.72
39.5	1.27	0.21	814	0.79	82535	80.51
40.0	1.32	0.21	1451	1.42	83986	81.93
40.5	1.36	0.21	762	0.74	84748	82.67
41.0	1.41	0.21	1390	1.36	86138	84.03
41.5	1.45	0.21	736	0.72	86874	84.74
42.0	1.50	0.21	1378	1.34	88252	86.09
42.5	1.54	0.22	723	0.71	88975	86.79
43.0	1.59	0.22	1221	1.19	90196	87.98
43.5	1.64	0.22	668	0.65	90864	88.64
44.0	1.69	0.22	1105	1.08	91969	89.71
44.5	1.74	0.22	606	0.59	92575	90.30

Raw Score	Theta	SE	Freq	%	Cum. Freq	Cum. %
45.0	1.79	0.23	1028	1.00	93603	91.31
45.5	1.84	0.23	535	0.52	94138	91.83
46.0	1.89	0.23	942	0.92	95080	92.75
46.5	1.95	0.24	508	0.50	95588	93.24
47.0	2.01	0.24	841	0.82	96429	94.06
47.5	2.07	0.25	458	0.45	96887	94.51
48.0	2.13	0.25	801	0.78	97688	95.29
48.5	2.20	0.26	365	0.36	98053	95.65
49.0	2.26	0.27	694	0.68	98747	96.33
49.5	2.34	0.27	344	0.34	99091	96.66
50.0	2.41	0.28	581	0.57	99672	97.23
50.5	2.50	0.29	247	0.24	99919	97.47
51.0	2.58	0.30	533	0.52	100452	97.99
51.5	2.68	0.31	196	0.19	100648	98.18
52.0	2.78	0.33	459	0.45	101107	98.63
52.5	2.90	0.34	151	0.15	101258	98.77
53.0	3.02	0.36	363	0.35	101621	99.13
53.5	3.16	0.39	111	0.11	101732	99.24
54.0	3.32	0.42	271	0.26	102003	99.50
54.5	3.52	0.46	65	0.06	102068	99.56
55.0	3.75	0.51	222	0.22	102290	99.78
55.5	4.04	0.58	27	0.03	102317	99.81
56.0	4.46	0.71	146	0.14	102463	99.95
56.5	5.16	1.00	0	0.00	102463	99.95
57.0	6.37	1.83	51	0.05	102514	100.00

Appendix C: Standard Error Plot

Table C.1: NJBCT Standard Error Plot



Appendix D: Standard Setting Executive Summary

New Jersey End-of-Course Biology Standard Setting Executive Summary

**Michael B. Bunch
Measurement Incorporated**

The New Jersey End of Course Biology Test (Biology Test) is a recent addition to the assessment program for the state. It is the first of a planned series of end-of-course tests for high school subjects and was administered for the first time in the spring of 2008 on a no-fault basis; i.e., there were no sanctions for students who did not perform at a satisfactory level on the test in 2008.

The New Jersey Department of Education (NJDOE) conducted standard setting for the Biology Test week of August 5-7, 2008. Educators from throughout the state of New Jersey participated in this three-day meeting. Staff of Measurement Incorporated (MI), the contractor for the New Jersey High School Proficiency Assessment (NJ HSPA), facilitated the meeting, along with NJDOE staff.

The main goals of the meeting were to (1) describe and delineate the thresholds of performance, for each subject area, that are indicative of Partially Proficient, Proficient, and Advanced Proficient performance, and (2) establish recommended cut scores for the test that differentiate Partially Proficient from Proficient and Proficient from Advanced Proficient performance (i.e., two cut scores to yield three performance levels). These recommendations are designed to help inform the New Jersey State Board of Education (Board) as it completes its task of establishing performance strands for all statewide assessments.

From August 5 to 7, 2008, MI staff met with representatives of the NJDOE and 27 educator-panelists from around the state to recommend performance strands. The panelists, nominated by district superintendents, were chosen specifically to represent the demographics of educators throughout the state. A profile of the 27 panelists is provided in the report (Table 2). Following an introduction to the process, an overview of the test development process by NJDOE staff, and specific standard-setting techniques by Dr. Michael Bunch (Senior Vice President of MI), the panelists divided into two independent groups (14 in one group, 13 in the other), each led by a facilitator from MI. The purpose of the two-group approach was to test for and neutralize any facilitator effects; i.e., influence on the final outcomes by the person facilitating the group. Panelists in both groups received a thorough orientation and practice exercises to prepare them for their task. MI staff provided additional information to panelists as they proceeded through three rounds of setting strands, discussing decisions, and settling on final recommendations.

In accordance with a plan approved by the New Jersey Technical Advisory Committee (TAC), a group of nationally recognized experts in testing, MI staff employed the bookmark standard-setting procedure. This procedure is the most widely used standard-setting procedure for statewide assessments and is thoroughly documented in the plan submitted to the NJ TAC, as well as in a recent book by Gregory Cizek and Michael Bunch. Dr. Bunch provided the orientation to the bookmark procedure. In this procedure, panelists review all test items in a specially formatted test booklet that places the easiest item on page one, the most difficult item

on the final page, and all items in between ordered by difficulty. Panelists place a bookmark at the point in the test booklet where they believe the probability of a minimally *Proficient* or minimally *Advanced Proficient* student would begin to have less than a two-thirds chance of answering correctly. These page numbers are then mathematically translated into ability estimates, which are translated into raw scores. The average of the panelists' ability estimates is then translated into a raw score which becomes the cut score. The procedure is more fully described in the body of the report.

NJDOE convened biology educators from across the state on May 30 to review and revise performance level descriptors (PLDs) for *Partially Proficient*, *Proficient*, and *Advanced Proficient*. These PLDs were to provide the guidance standard-setting panelists would need in order to complete their tasks the week of August 5-7. These PLDs are included in an appendix to the report.

Throughout the three days of the August standard-setting meeting, an external evaluator observed the process and took notes. The evaluator has prepared a separate report and will submit it for review by the NJDOE and the TAC in response to a TAC recommendation. The evaluator, Dr. Robert Lissitz, is well known in the educational assessment field, having served for many years as the head of the educational measurement department at a major university. He also participated in the review of standard setting for the NJ HSPA Science Test in 2007.

Results and Discussion

Panelists considered each test in three rounds. During Round 1, each panelist placed two bookmarks, one for *Proficient* and one for *Advanced Proficient*. MI staff analyzed the data for Round 1 and led discussions of the results. Panelists reviewed the results of Round 1 and then repeated the process of placing bookmarks in Round 2. After Round 2, MI staff again analyzed the data and presented results to the panelists, along with score distributions showing percentages of students who would be classified as *Partially Proficient*, *Proficient*, and *Advanced Proficient* on the basis of the Round 2 cut scores. After discussion of these results, panelists once again placed two bookmarks in Round 3. These bookmarks defined the final cut scores (averaged over all panelists) to be forwarded to the NJDOE.

From Round 1 to Round 2, neither the *Proficient* nor the *Advanced Proficient* cut score changed. In both instances, the two cut scores were 30.0 and 41.5 for *Proficient* and *Advanced Proficient*, respectively. What did change from Round 1 to Round 2 was the range of cut scores. Predictably, after having an opportunity to discuss their Round 1 bookmarks, many of the outliers moved toward the group mean. From Round 2 to Round 3, the *Proficient* cut score dropped by one point (to 29.0), while the *Advanced Proficient* cut score went up half a point (to 42.0). There were group differences on the final round. Group 1 set a *Proficient* cut score of 26, while Group 2 set a cut score of 31. Similarly, Group 1 set an *Advanced Proficient* cut score of 41, while Group 2 set a cut score of 43. In analyzing the differences between the two groups, it was clear that they were not general. Rather, in Group 1, there was a single table (three panelists) that consistently held out for lower cut scores. They moderated toward the group mean in Round 2 but reverted in Round 3 after viewing impact data.

Final recommended performance strands are reported in Table ES-1. The table includes the total number of points possible on the test. The Proficient and Advanced Proficient Cut Score column includes both the raw score mean and the mean expressed in terms of a percentage of the Total Points Possible. The final column in Table ES-1 shows the total number of points possible for the test.

**Table ES-1
Final Cut Score Recommendations**

Proficient Cut Score		Advanced Proficient Cut Score		Total Points Possible
Raw Score	% Correct	Raw Score	% Correct	
29	54	42	78	54

The recommended *Proficient* cut score is 29 out of 54 points, or 54 percent of the total score possible. This outcome is consistent with the goal of more rigorous strands, reflected in performance level descriptors developed by NJDOE with the assistance of New Jersey educators.

The impact of these cut scores on New Jersey students is summarized in Table ES-2. The final column of the table shows the total percentage of students whose scores would place them in the *Proficient* or *Advanced Proficient* category. The impact data reflect the performances of all 100,069 regular students tested.

**Table ES-2
Percentages of Students Classified at Each Level**

Group	Number Tested	% Partially Proficient	% Proficient	% Advanced Proficient	% Proficient or Above
All Students	100,069	49.1	33.4	17.5	50.9
By Race/Sex					
Asian	7,774	27.3	36.0	36.8	72.7
African American	15,608	79.0	17.7	3.3	21.0
Hispanic	17,234	73.5	21.9	4.7	26.5
White	57,376	36.1	41.0	22.9	63.9
Female	49,395	50.9	33.8	15.4	49.1
Male	50,578	47.3	33.0	19.7	52.7
By Status					
LEP	3,432	90.6	8.3	1.1	9.4
Special Education	13,748	82.0	14.7	3.4	18.1
Economically Disadvantaged	20,424	77.3	19.2	3.6	22.7

Measurement Incorporated believes and can document that the standard setting process for the Biology End of Course Test was sound, both in conception and execution, representative of the highest strands in contemporary educational measurement, and representative of strands operating among state assessment systems nationwide; that the participants, New Jersey teachers, found it to be so; and that, as New Jersey's assessment vendor, with wide experience implementing assessment programs in other states, MI stands behind the validity of the Biology test standard setting results and the process which produced them, and is prepared to assist the NJDOE in communicating this validity to stakeholders and federal peer reviewers.