New York State Regents Examination in Algebra I (Common Core)

2016 Technical Report

Prepared for the New York State Education Department
by Pearson

March 2017

5.4 EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES 37
5.5 EVIDENCE BASED ON TESTING CONSEQUENCES 38
REFERENCES 40
APPENDIX A: OPERATIONAL TEST MAPS 44
APPENDIX B: RAW-TO-THETA-TO-SCALE SCORE CONVERSION TABLES 47
APPENDIX C: ITEM WRITING GUIDELINES 50
APPENDIX D: TABLES AND FIGURES FOR AUGUST 2015 ADMINISTRATION 55
APPENDIX E: TABLES AND FIGURES FOR JANUARY 2016 ADMINISTRATION 62

1.1 INTRODUCTION

This technical report for the Regents Examination in Algebra I (Common Core) will provide New York State with documentation on the purpose of the Regents Examination, scoring information, evidence of both reliability and validity of the exams, scaling information, and guidelines and reporting information for the August 2015, January 2016, and June 2016 administrations. Chapters 1 ±5 detail results for the June 2016 administration. Results for the August 2015 and January 2016 administrations are provided in Appendix D and E, respectively. As the Standards for Education and Psychological Testing G L V F X V V H V L Q 6 W D Q G D U objective of the documentation is to provide test users with the information needed to help them assess the nature and quality of the test, the resulting scores, and the interpretations based on W K H W H V W V E H L Q J (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p.123).[1] Please note that a technical report, by design, addresses technical documentation of a testing program; other aspects of a testing program (content standards, scoring guides, guide to test interpretation, equating, etc.) are thoroughly addressed and referenced in supporting documents.

1.2 HISTORY

The Board of Regents adopted the Common Core State Standards (CCSS) for English Language Arts & Literacy and Mathematics at its July 2010 meeting and incorporated New York State-specific additions, creating the Common Core Learning Standards (CCLS), at its January 2011 meeting. Based on feedback from the field and to ensure adequate notice and time for students to be prepared to take the new Regents Exams measuring the CCLS, the Department provided an overlap in the administration of the Regents Exams measuring the 2005 Learning Standards with the Regents Exams measuring the CCLS and a phased-in sequence for mathematics.

Students who took the old Regents Exam in addition to the new Regents Exam were allowed to use the higher of the two scores for local transcript purposes, and, similarly, the higher of the two scores was used for institutional accountability for the 2013 ±14 and 2014 ±15 school year results. Such students were able to meet the mathematics exam requirement for graduation by passing either of these exams. The complete memo detailing transition to the Common Core examinations can be located at http://www.p12.nysed.gov/assessment/commoncore/archive/transitionccregents1113rev-arc2.pdf.

1.3 PURPOSES OF THE EXAM

The Regents Examination in Algebra I (Common Core) measures examinee achievement against the New York State (NYS) learning standards. The exam is prepared by teacher examination committees and New York State Education Department (NYSED) subject matter

[1] References to specific Standards will be placed in parentheses throughout the technical report, to provide further context for each section.

and testing specialists and provides teachers and students with important information about student learning and performance against the established curriculum standards. Results of this exam may be used to identify student strengths and needs, in order to guide classroom teaching and learning. The exams also provide students, parents, counselors, administrators, and college admissions officers with objective and easily understood achievement information that may be used to inform empirically based educational and vocational decisions about students. As a state-provided objective benchmark, the Regents Examination in Algebra I (Common Core) is intended for use in satisfying state testing requirements for students who have finished a course in Algebra I. A passing score on the exam counts toward requirements for a high school diploma, as described in the New York State diploma requirements: http://www.nysed.gov/common/nysed/files/programs/curriculum-instruction/currentdiplomarequirements2.pdf. Results of the Regents Examination in Algebra I (Common Core) may also be used to satisfy various locally established requirements throughout the state.

1.4 TARGET POPULATION (STANDARD 7.2)

The examinee population for the Regents Examination in Algebra I (Common Core) is composed of students who have completed a course in Algebra I. Any student, regardless of grade level or cohort, who began their first commencement-level Algebra course in fall 2013 or later was provided with instruction aligned with the NYS P-12 Common Core Learning Standards for Algebra and therefore took or will take the Regents Examination in Algebra I (Common Core). More information about testing requirements can be found at 6 (c (an)Tj 0 Tc e f74 0 T

	August Admin*		January Admin**		June Admin***	
Demographics						

2.2 ITEM DISCRIMINATION

At the most general level, estimates of L W H P G L V F U L P L Q D W L R Q L Q G L F D W differentiate between high and low performance on an item. It is expected that high-performing students (i.e., those who perform well on the Regents Examination in Algebra I [Common Core] overall) would be more likely to answer any given item correctly, while low-performing students (i.e., those who perform poorly on the exam overall) would be more likely to answer the same item incorrectly. Pearson ¶ product-moment correlation coefficient (also commonly referred to as a point-biserial correlation) between item scores and test scores is used to indicate discrimination (Pearson, 1896). The correlation coefficient can range from í1.0 to +1.0. If high

Item	Number	-Value	SD	Point-Biserial	Point-Biserial Distractor 1	Point-Biserial Distractor 2	Point-Biserial Distractor 3
14	238,654	0.58	0.49	0.41	> .23	> .23	> .13
15	238,654	0.57	0.49	0.54	> .24	> .30	> .21
16	238,654	0.66	0.47	0.47	> .22	> .30	> .18
17	238,654	0.59	0.49	0.46	> .19	> .23	> .26
18	238,654	0.55	0.50	0.51	> .16	> .23	> .31
19	238,654	0.55	0.50	0.45	> .23	> .24	> .16
20	238,654	0.49	0.50	0.42	> .19	> .14	> .24
21	238,654	0.55	0.50	0.35	> .15	> .21	> .14
22	238,654	0.46	0.50	0.43	> .11	> .25	> .22
23	238,654	0.33	0.47	0.41	> .28	> .35	0.13
24	238,654	0.27	0.45	0.30	0.01	> .11	> .25

Table 3 Constructed -Response Item Analysis Summary: Regents Examination in Algebra I (Common Core)

The item response theory (IRT) model used for the Regents Examination in Algebra I (Common Core) is based on the work of Georg Rasch (Rasch, 1960). The Rasch model has a long-standing presence in applied testing programs. IRT has several advantages over classical test theory, and has become the standard procedure for analyzing item response data in large-scale assessments. According to van der L L Q G H Q D Q G + D P E O H W R Q ³7 K H of IRT is the specification of a mathematical function relating t K H S U R E D E L O L W \ R I D Q I U H V S R Q V H R Q D W H V W L W H P W R D Q X Q G H U O \ L Q J D E L O L W \ ´ S H U I R U P D Q F H R Q W K H W H V W D Q G L V G H I L Q H G D V ³W K H H [S H F W H V W R I L Q W H U H V W´ than, and Rogers, 1991). This performance value is often referred to as Ɵ Performance and Ɵ will be used interchangeably throughout the remainder of this report.

A fundamental advantage of IRT is that it links examinee performance and item difficulty estimates and places them on the same scale, allowing for an evaluation of examinee performance that considers the difficulty of the test. This is particularly valuable for final test construction and test form equating, as it facilitates a fundamental attention to fairness for all examinees across items and test forms.

This chapter outlines the procedures used for calibrating the operational Regents Examination in Algebra I (Common Core) items. Generally, item calibration is the process of assigning a difficulty, R U L W H P ,´ R U E D W L R Q to each item on an assessment so that all items are placed onto a common scale. This chapter briefly introduces the Rasch model, reports the results from evaluations of the adequacy of the Rasch assumptions, and summarizes the Rasch item statistics.

3.1 DESCRIPTION OF THE RASCH MODEL

The Rasch model (Rasch, 1960) was used to calibrate multiple-choice items, and the partial credit model, or PCM (Wright and Masters, 1982), was used to calibrate constructed-response items. The PCM extends the Rasch model for dichotomous (0, 1) items so that it accommodates the polytomous CR item data. Under the PCM model, for a given item i with m_i score categories, the probability of person n scoring x (x = 0, 1, 2,... m_i) is given by

$$P_{ni}(X = x) = \frac{\exp \sum_{j=0}^{x} (\theta_n - D_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (\theta_n - D_{ij})},$$

where θ_n represents examinee ability, and D_{ij} is the step difficulty of the jth step on item i. D_{ij} can be expressed as , where is the difficulty for item i and is a step deviation value for the jth step. For dichotomous MC items, the RPCM reduces to the standard Rasch P R G H O D Q G W K H V L Q J O H V W H S G L I I L F X O W \ L V U H I H U U H G W predicts the probability of person n getting item i correct as follows:

Figure 2 Student Performance Map: Regents Examination in Algebra I (Common Core)

3.5 CHECKING RASCH ASSUMPTIONS

Since the Rasch model was the basis of all calibration, scoring, and scaling analyses associated with the Regents Examination in Algebra I (Common Core), the validity of the inferences from these results depends on the degree to which the assumptions of the model were met and how well the model fits the test data. Therefore, it is important to check these assumptions. This section evaluates the dimensionality of the data, local item independence, and item fit. It should be noted that only operational items were analyzed, since they are the basis of student scores.

Unidimensionality

Prepared for NYSED by

Table 5 Summary of Item Residual Correlations: Algebra I (Common Core)

Statistic Type	Value
N	666
Mean	-0.02
SD	0.03
Minimum	-0.11
P_{10}	-0.06
P_{25}	-0.04
P_{50}	-0.03
P_{75}	0.00
P_{90}	0.01
Maximum	0.08
>|0.20|	0

Item Fit

An important assumption of the Rasch model is that the data for each item fit the model. WINSTEPS provides two item fit statistics (INFIT and OUTFIT) for evaluating the degree to which the Rasch model predicts the observed item responses for a given set of test items. Each fit statistic can be expressed as a mean square (MnSq) statistic or on a standardized metric (Zstd with mean = 0 and variance = 1). MnSq values are more oriented toward practical significance, while Zstd values are more oriented toward statistical significance. INFIT MnSq values are the average of standardized residual variance (the difference between the observed score and the Rasch-estimated score divided by the square root of the Rasch-model variance). The INFIT statistic is weighted by the ( ) relative to item difficulty, and tends to be affected

would be pure random noise (i.e., all measurement error). If the index achieved a value of 1.0, scores would be perfectly consistent (i.e., contain no measurement error). Although values of 1.0 are never achieved in practice, it is clear that larger coefficients are more desirable because they indicate that the test scores are less influenced by random error.

Coefficient Alpha

Reliability is most often estimated using the formula for Coefficient Alpha, which provides a practical internal consistency index. Coefficient Alpha can be conceptualized as the extent to which an exchangeable set of items from the same domain would result in a similar rank ordering of students. Note that relative error is reflected in this index. Excessive variation in student performance from one sample of items to the next should be of particula

Traditional Standard Error of Measurement Confidence Intervals

The SEM is an index of the random variability in test scores reported in actual score units, which is why it has such great utility for test score users. SEMs allow statements regarding the SUHFLVLRQ RI LQGLYLGXDO WHVW VFRUHV 6(0V KHOS SOD around observed scores, through construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, X, and adding or subtracting a multiplicative factor of the SEM. As an example, students with a given true score will have observed scores that fall between ±1 SEM about two-thirds of the time.4 For ±2 SEM confidence intervals, this increases to about 95 percent.

The Coefficient Alpha and associated SEM for the Regents Examination in Algebra I (Common Core) are provided in Table 7.

Table 7 Reliabilities and Standard Errors of Measurement: Regents Examination in Algebra I (Common Core)

Subject

Figure 4 Conditional Standard Error Plot: Regents Examination in Algebra I (Common Core)

4.3 DECISION CONSISTENCY AND ACCURACY (STAND ARD 2.16)

In a standards-based testing program there is interest in knowing how accurately students are classified into performance categories. In contrast to the Coefficient Alpha, which is concerned with the relative rank-ordering of students, it is the absolute values of student scores that are important in decision consistency and accuracy.

Classification consistency refers to the degree to which the achievement level for each student can be replicated upon retesting using an equivalent form (Huynh, 1976). Decision consistency answers the following question: What is the agreement in classifications between the two non-

		TEST ONE		
		LEVEL I	LEVEL II	MARGINAL
TEST TWO	LEVEL I	M_{11}	M_{12}	$M_{1\bullet}$
	LEVEL II	M_{21}	M_{22}	$M_{2\bullet}$
	MARGINAL	$M_{\bullet 1}$	$M_{\bullet 2}$	1

Figure 5 Pseudo-Decision Table for Two Hypothetical Categories

	TEST ONE				
	LEVEL I	LEVEL II	LEVEL III	LEVEL IV	

Table 10

Restating the purpose and uses of the Regents Examination in Algebra I (Common Core), this exam measures examinee achievement against the New York State learning standards. The exam is p

Content Frameworks for Mathematics provided by the Partnership for the Assessment of Readiness for College and Career (PARCC, 2014). The model content frameworks are located at http://www.parcconline.org/resources/educator-resources/model-content-frameworks/mathematics-model-content-framework. The standards for mathematics are located at http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards-for-mathematics. Clarifications for Algebra I (Common Core) standards are located at http://www.engageny.org/resource/regents-exams-mathematics-algebra-i-standards-clarifications.

Content Validity

Content validity is necessarily concerned with the proper definition of the construct and evidence that the test provides an accurate measure of examinee performance within the defined construct. The test blueprint for the Regents Examination in Algebra I (Common Core) is essentially the design document for constructing the exam. It provides explicit definition of the construct domain that is to be represented on the exam. The test development process (discussed in the next section), is in place to ensure, to the extent possible, that the blueprint is met in all operational forms of the exam.

Table 11 displays domain titles along with their cluster, standard, and targeted proportions of conceptual categories on the exam.

Table 11 Test Blueprint, Regents Examination in Algebra I (Common Core)

Conceptual Category	Percent of Test by Credits	Domains in Algebra I
Number & Quantity	2 t8%	The Real Number System (N-RN) Quantities (N-Q)
Algebra	50 t56%	Seeing Structure in Expressions (A-SSE) Arithmetic with Polynomials and Rational Expressions (A-APR) Creating Equations (A-CED) Reasoning with Equations and Inequalities (A-REI)
Functions	32 t38%	Interpreting Functions (F-IF) Building Functions (F-BF) Linear, Quadratic, and Exponential Models (F-LE)

Only New York State-

Figure 7 New York State Education Department Test Development Process

Item Review Criteria

Item Review Criteria assist in the consistent application of rigorous item reviews intended to assess the quality of the items developed and identify items that require edits or removal from the pool of items to be field tested.

C

characteristics, one unambiguously correct answer and several plausible, but incorrect, answer choices.

Refer to the following link for more detail on the item review criteria: https://www.engageny.org/resource/regents-exams-mathematics-item-criteria-checklist.

Following these reviews, only items that are approved by an assigned educator panel move forward for field testing.

Ongoing attention is also given to the relevance of the standards used to guide curriculum and assessment. Consistent with a desire to assess this relevance, the New York State Education Department (NYSED) is committed to ongoing standards review over time, and periodically solicits thoughtful, specific responses from stakeholders about individual standards within the NYS P-12 Standards.

5.2 Evidence Based on Response Processes

The second source of validity evidence is based on examinee response processes. This standard requires evidence that examinees are responding in the manner intended by the test items and rubrics and that raters are scoring those responses in a manner that is consistent with the rubrics. Accordingly, it is important to control and monitor whether or not construct-irrelevant variance in response patterns has been introduced at any point in the test development, administration, or scoring processes.

The controls and monitoring in place for the Regents Examination in Algebra I (Common Core) include the item development process, with attention paid

Multiple-choice items are scored via local scanning at testing centers, and trained educators score constructed-response items. There are many studies that focus on various elements of producing valid and reliable scores for constructed-response items, but generally, attention to the following all contribute to valid and reliable scores for constructed-response items:

1. Quality training (Hoyt & Kerns, 1999; Lumley & McNamara, 1995; Wang, Wong, and Kwong, 2010; Gorman & Rentsch, 2009; Schleicher, Day, Bronston, Mayes, and Riggo, 2002; Woehr & Huffcutt, 1994; Johnson , Penny, and Gordon, 2008; Weigle, 1998)
2. Detection and correction of rating bias (McQueen & Congdon, 1997; Congdon & McQueen, 2000; Myford, & Wolfe, 2009; Barkaoui, 2011; Patz, Junker, Johnson, and Mariano, 2002)
3. Consistency or reliability of ratings (Congdon & McQueen, 2000; Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar, 2009; McQueen & Congdon, 1997; Myford & Wolfe, 2009; Mero & Motowidlo, 1995; Weinrott & Jones, 1984)
4. Rubric designs that facilitate consistency of ratings (Pecheone & Chung, 2006; Wolfe & Gitomer, 2000; Cronbach, Linn, Brennan, & Haertel, 1995; Cook & Beckman, 2009; Penny, Johnson, & Gordon, 2000; Smith, 1993; Leacock, Gonzalez, and Conarroe, 2014)

The distinct steps for operational test scoring include close attention to each of these elements and begin before the operational test is even selected. After the field test process, during which many more items than appear on the operational test are administered to a representative sample of students, a VHW RI ³DQFKRU´ SDSHUV UHSUHVHQW across the range of possible responses for constructed-response items is selected. The objective of these ³UDQJLQJ´ is to create a training set for scorer training and execution, the scores from which are used to generate important statistical information about the itemtive sample thatQ ae 0 0 612 792 re W* n BT 0 g /TT0 12 Tf 272.09 (m)35 Td [(th)-5 o(m)9 ( )-41 t

halo effect, which occurs when good or poor perform

Differential Item Functioning

Differential item functioning (DIF) for gender was conducted following field testing of the items in 2012–2015. Sample sizes for subgroups based on ethnicity and English language learner status were, unfortunately,

to which the classification decisions based on the first set of test scores matched the decisions based on the second set of test scores. Decision accuracy is an index to determine the extent to which measurement error causes a classification different from that expected from the true score. High decision consistency and accuracy provide strong evidence that the internal structure of a test is sound.

For the Regents Examination in Algebra I (Common Core), both decision consistency and accuracy values are high, indicating very good consistency and accuracy of examinee classifications. The results for the overall consistency across all five performance levels, as well as for the dichotomies created by the four corresponding cut scores, are presented in Table 8. The tabled values are derived with the program BB-Class (Brennan, 2004), using the Livingston and Lewis method. The decision consistency ranged from 0.86 to 0.95, and the decision accuracy ranged from 0.90 to 0.97.

Dimensionality

In addition to model fit, a strong assumption of the Rasch model is that the construct measured by a test is unidimensional. Violation of this assumption might suggest that the test is measuring something other that 9A1sET Q q 0 0 6 Tf 57.6 471.07 Td [(4d [(s)30 (m)-6 (e)-3 (a)-3 (s)1

In terms of predictive validity, time is a fundamental constraint on gathering evidence. The gold standard for supporting the validity of predictive statements about test scores requires empirical evidence of the relationship between test scores and future performance on a defined characteristic. To the extent that the objective of the CCLS is to prepare students for college and career, it will be important to gather evidence of this empirical relationship over time.

Currently, the predictive validity is supported by expert judgments gathered during the standard-setting process for Regents Examination in Algebra I (Common Core). During this process, subject matter experts described

who have completed a course in Algebra I, the exam is most commonly used to inform admissions and course placement decisions. Such uses can be considered reasonable, assuming that the competencies demonstrated in the Regents Examination in Algebra I (Common Core) are consistent with those required in the courses for which a student is seeking enrollment or placement. Educational institutions using the exam for placement purposes are advised to examine the scoring rules for the Regents Examination in Algebra I (Common Core) and to assess their appropriateness for the inferences being made about course placement.

As stated, the nature of validity arguments is not absolute, but it is supported through ongoing processes and studies designed to accumulate support for validity claims. The evidence provided in this report documents the evidence to date that supports the use of the Regents Examination in Algebra I (Common Core) scores for the purposes described.

Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. Journal of Applied Psychology, 80(4), 517 ±524.

Myford, C. M., & Wolfe, E. W. (2009). Monitoring rater performance over time: A framework for detecting differential accuracy and differential scale use. Journal of Educational Measurement, 46(4), 371 ±389.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards. Washington, DC: Authors.

Patz, R. J., Junker, B. W., Johnson, M. S., and Mariano, L. T. (2002). The Hierarchical Rater Model for Rated Test Items and its Application to Large-Scale Educational Assessment Data. Journal of Educational and Behavioral Statistics,

f201.89 48

Weinrott, L., & Jones, B. (1984). Overt verses covert assessment of observer reliability. Child Development, 55, 1125 –1137.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. Journal of Occupational and Organizational Psychology, 67, 189 –205.

Wolfe, E. W., & Gitomer, D. H. (2000). The influence of changes in assessment design on the psychometric quality of scores. Princeton, NJ: Educational Testing Service.

Table A. 2 Test Map for January 2016 Administration

Table A. 3 Test Map for June 2016 Administration

Table B. 2 Score Table for January 2016 Administration

29. Any aspect of the item that provides an unintended clue that can be used to select or eliminate an alternative should be avoided. For example, any term that appears in the stem should not appear in only one of the alternatives.

30. Notation and symbols as presented on Common Core examinations should be used consistently.

31. For example, AB means the length of line segment AB, means line segment AB, and means the number of degrees in the measure of angle A, etc.

Guidelines for Writing Constructed -Response Math Items

1. The item should measure the knowledge, skills, and proficiencies characterized by the standards within the identified cluster.

2. The focus of the problem or topic should be stated clearly and concisely .
   The item should be meaningful, address important knowledge and skills, and focus on key concepts.

3. Include problems that come from a real -world context or problems that make use of multiple representations.
   When using real-world problems, use real-

10. The item should be written to require a specific form of answer.
    3 K U D V H V O L N H ³ L´ ³ W H U P V ´ D Q G ³ L Q V L P S O H V W U D G L
    simplify the writing of the rubric for these types of items.

11. Items that require students to explain in words are encouraged.
    One of the emphases of the Common Core standards is to foster student ability to communicate mathematical thinking. An example is to have students construct viable arguments to make conjectures, analyze situations,

Table D.1 Multiple -Choice Item Analysis Summary: Regents Examination in Algebra I

Table D.3 Descriptive Statistics in ***p***-value and Point-Biserial Correlation: Regents Examination in Algebra I (Common Core)

Statistics	N	Mean	Min	Q1	Median	Q3	Max
-value	37	0.36	0.04	0.20	0.35	0.51	0.71
Point-Biserial	37	0.46	0.15	0.38	0.42	0.55	0.73

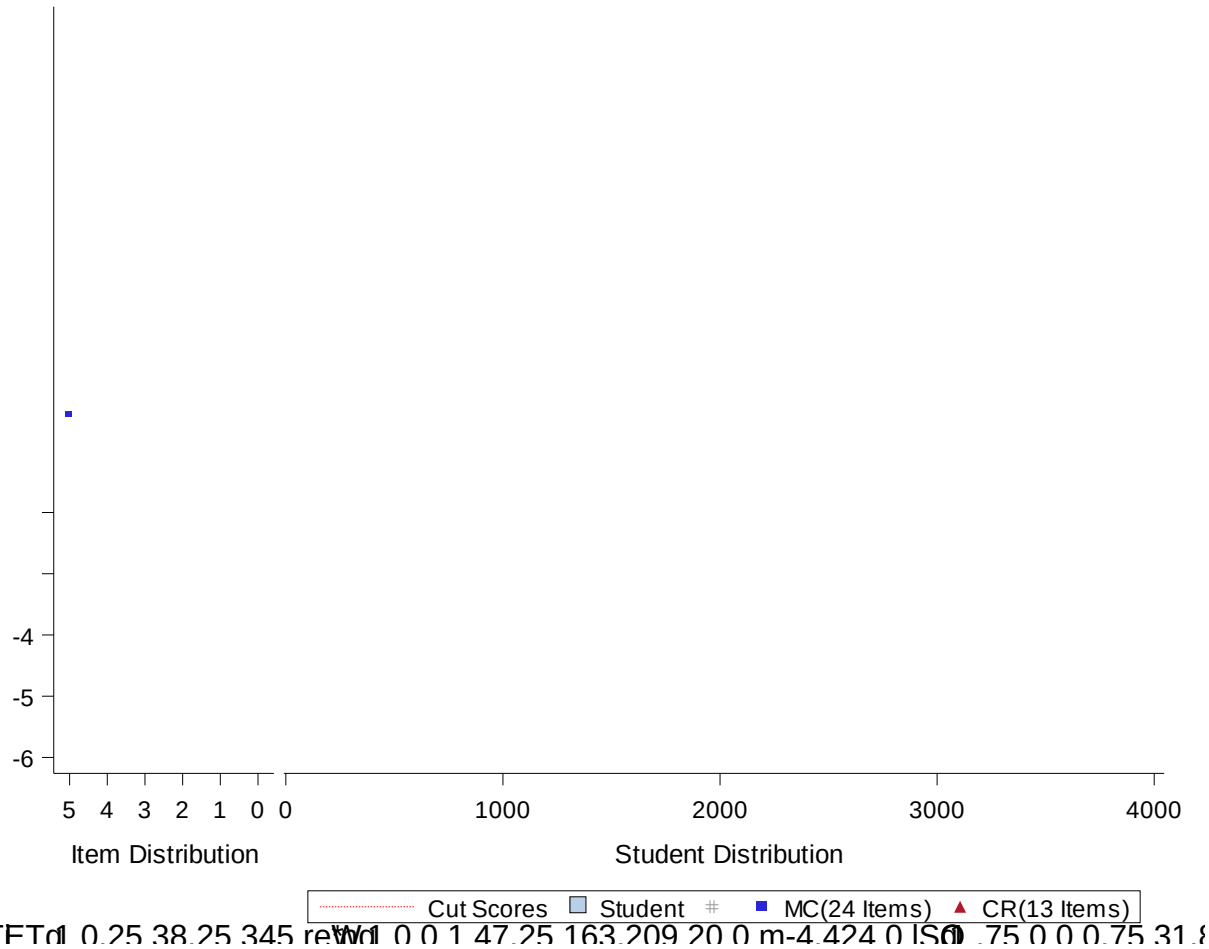

Figure D.2 Student Performance Map: Regents Examination in Algebra I (Common Core)

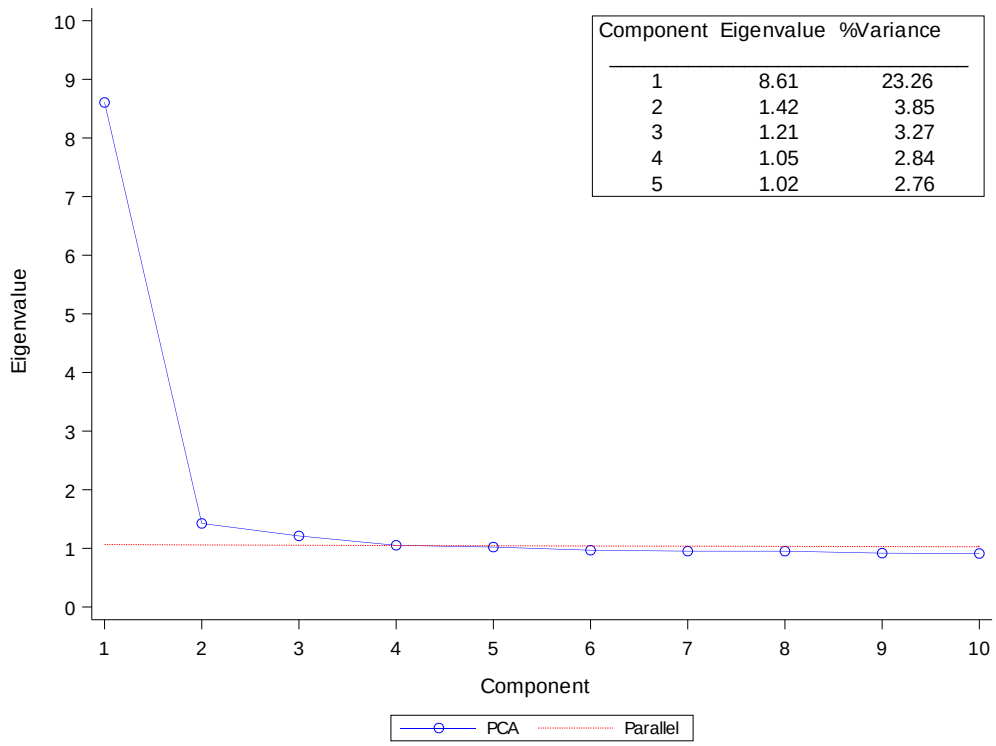

Figure D.3 Scree Plot: Regents Examination in Algebra I (Common Core)

Table D.4 Summary of Item Residual Correlations: Regents Examination in Algebra I (Common Core)

Statistic Type	Value
N	666
Mean	> .003
SD	0.03
Minimum	> .009
P_{10}	> .006
P_{25}	> .004
P_{50}	> .003
P_{75}	> .001
P_{90}	0.01
Maximum	0.10
>|0.20|	0

Table D.5 Summary of INFIT Mean Square Statistics: Regents Examination in Algebra I (Common Core)

	INFITMean Square					
	N	Mean	SD	Min	Max	[0.7, 1.3]
Algebra I (Common Core)	37	1	0.11	0.79	1.28	[37/37]

Table D.6 Reliabilities and Standard Errors of Measurement: Regents Examination in Algebra I (Common Core)

Subject	Coefficient Alpha	SEM
Algebra I (Common Core)	0.90	4.78

Table D.7 Decision Consistency and Accuracy Results: Regents Examination in Algebra I (Common Core)

Statistic	1/2	2/3	3/4	4/5
Consistency	0.88	0.86	0.92	0.99
Accuracy	0.91	0.90	0.95	0.99

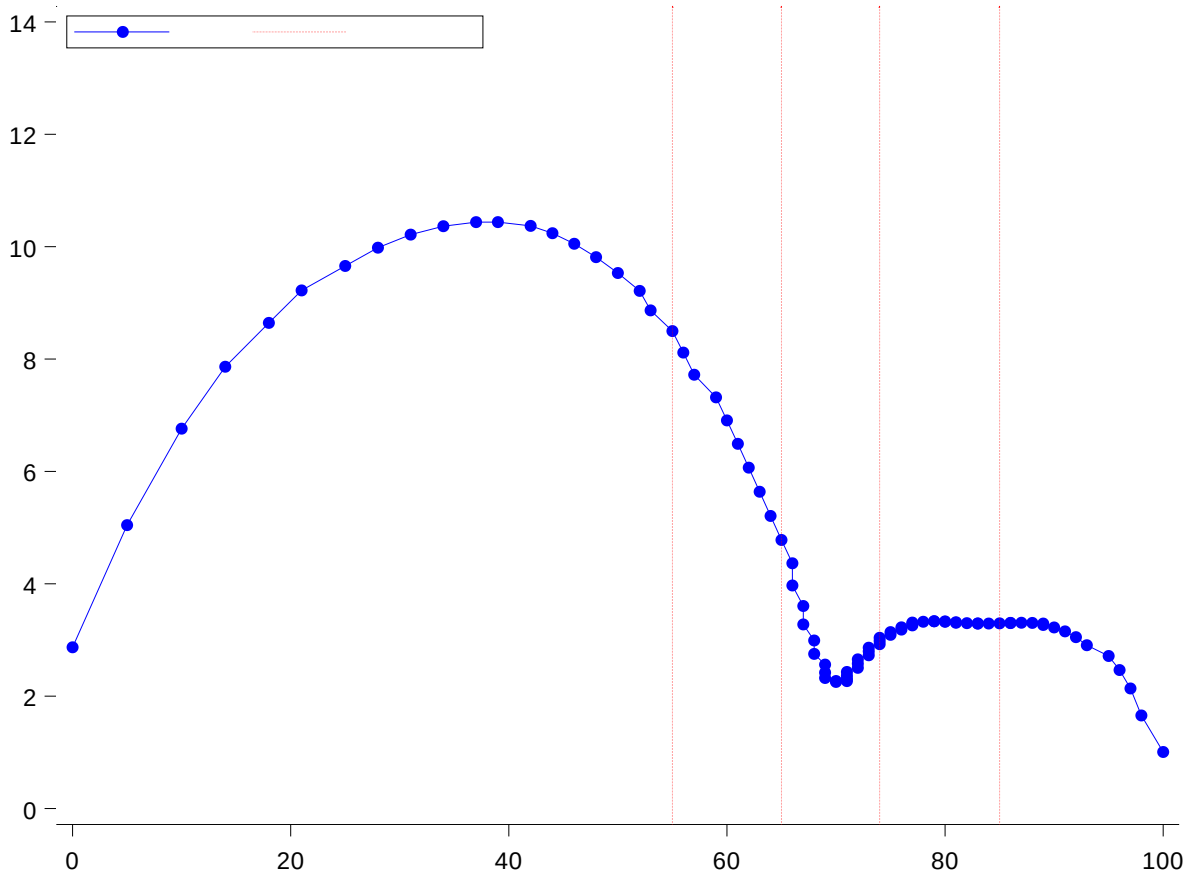

Figure D.4 Conditional Standard Error Plot: Regents Examination in Algebra I (Common Core)

Table D.8 Group Means: Regents Examination in Algebra I (Common Core)

Demographics	Number	Mean Scale Score	SD Scale Score
All Students*	35,846	58.71	

Table E.1 Multiple-Choice Item Analysis Summary: Regents Examination in Algebra I (Common Core)

Item	Number	-Value	SD	Point-Biserial	Point-Biserial Distractor 1	Point-Biserial Distractor 2	Point-Biserial Distractor 3
1	64,843	0.52	0.50	0.38	> .17	> .19	> .16
2	64,843	0.48	0.50	0.36	> .16	> .24	> .10
3	64,843	0.56	0.50	0.45	> .21	> .26	> .15
4	64,843	0.43	0.49	0.40	> .16	> .18	> .17
5	64,843	0.46	0.50	0.35	> .11	> .21	> .16
6	64,843	0.44	0.50	0.35	> .18	> .19	> .07
7	64,843	0.34	0.47	0.35	> .13	> .13	> .15
8	64,843	0.45	0.50	0.29	> .15	> .12	> .15
9	64						

Table E.3 Descriptive Statistics in ***p***-value and Point-Biserial Correlation: Regents Examination in Algebra I (Common Core)

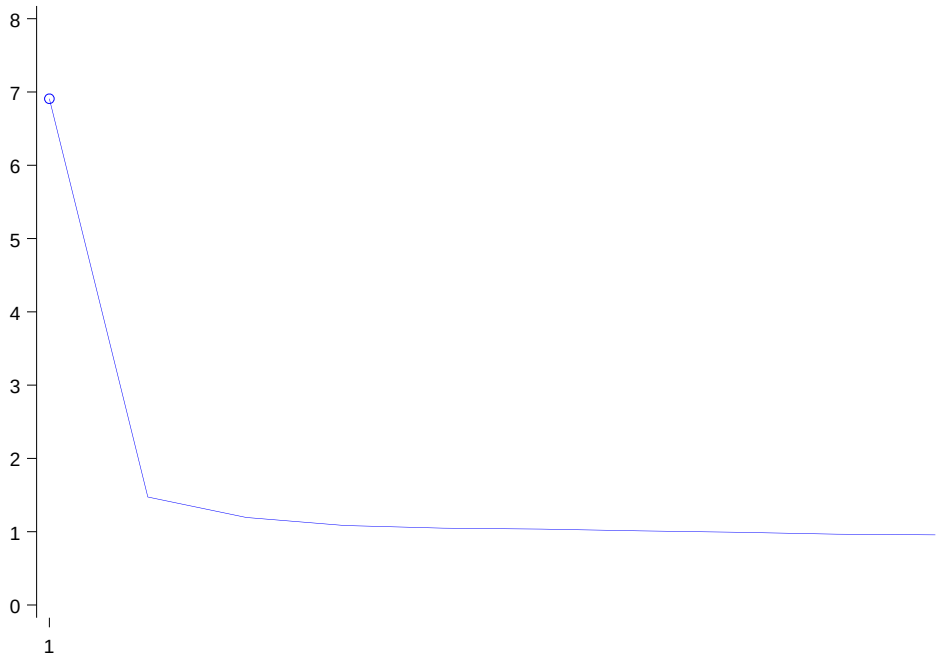
8
7
6
5
4
3
2
1
0
1

Table E.5 Summary of INFIT Mean Square Statistics: Regents Examination in Algebra I (Common Core)

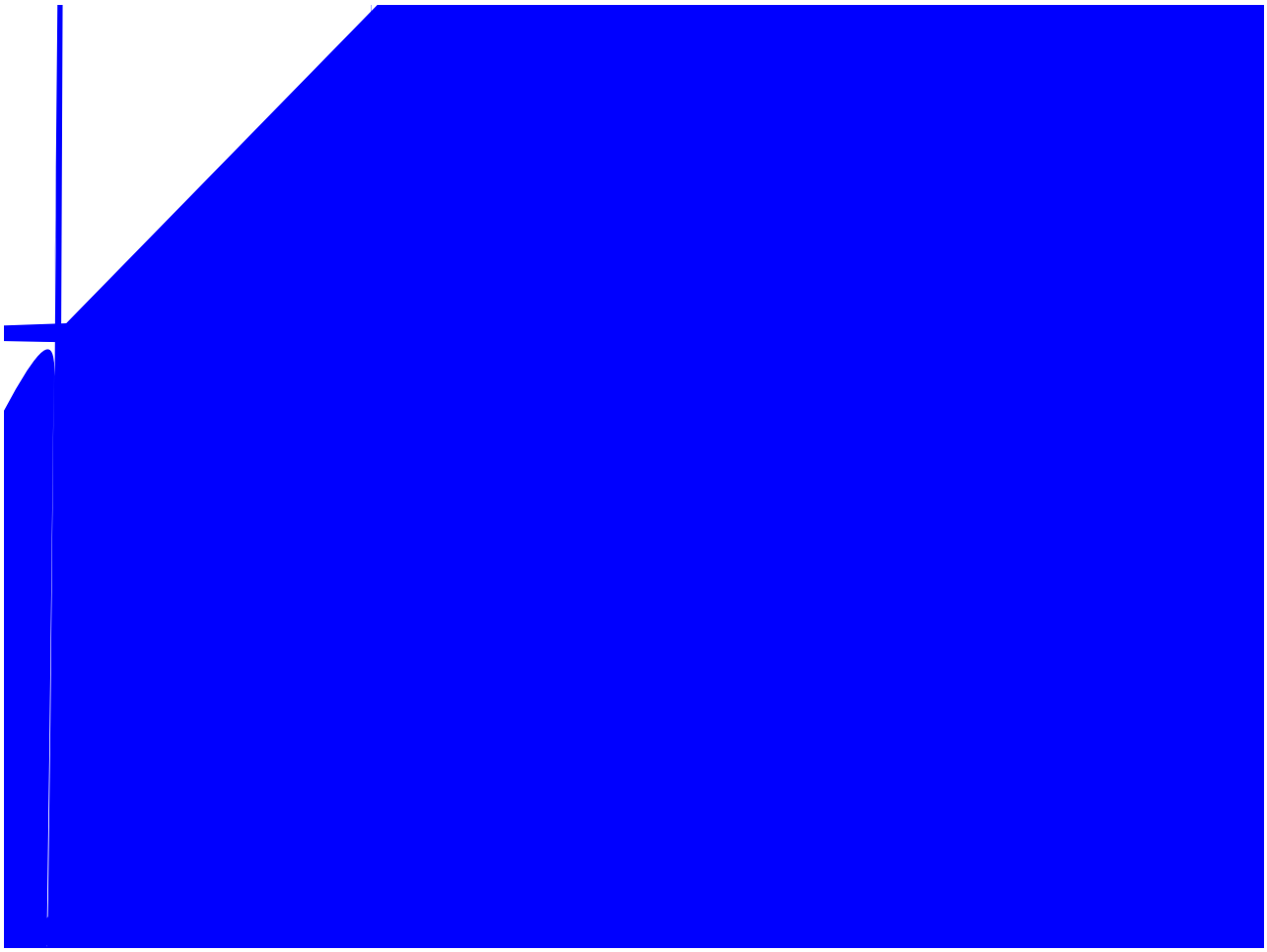

Figure E.4 Conditional Standard Error Plot: Regents Examination in Algebra I (Common Core)

Table E.8 Group Means: Regents Examination in Algebra I (Common Core)

Demographics	Number	Mean Scale Score	SD Scale Score
All Students*	64,843	57.99	11.58
Ethnicity			
American Indian/Alaska Native	526	57.89	12.10
Asian/Native Hawaiian/Other Pacific Islander	5,129	64.46	12.89
Black/African American			