New York State Regents Examination in English Language Arts

2019 Technical Report

Prepared for the New York State Education Department
by Pearson

March 2020

Copyright

Developed and published under contract with the New York State Education Department by Pearson.

Copyright © 2020 by the New York State Education Department.

Contents

1.1 INTRODUCTION 1
1.2 PURPOSES OF THE EXAM (STANDARD 12.1) 1
1.3 TARGET POPULATION (STANDARD 7.2) 2

2.1

List of Tables

TABLE 1 TOTAL EXAMINED POPULATION: REGENTS EXAMINATION IN ENGLISH LANGUAGE ARTS 2
TABLE 2 MULTIPLE-CHOICE ITEM ANALYSIS SUMMARY: R

Chapter 2: Classical Item Statistics (Standard 4.10)

This chapter provides an overview of the two most familiar item-level statistics obtained from classical item analysis: item difficulty and item discrimination. The following results pertain to the operational Regents Examination in English Language Arts items.

2.1 ITEM DIFFICULTY

At the most general level, an item's difficulty is indicated by its mean score in some specified group (e.g., grade level).

$$\bar{x} \quad \frac{1}{n} \sum_{i=1}^{n} x_i$$

In the mea

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's productpr4 (epar)-6prod5(-)Tj 0.004 Tc -0.0120.45833 0 Td [(ov)14 ore i iy orrect.aycr

Table 4 Descriptive Statistics in p- value and Point -Biserial Correlation: Regents Examination in English Language Arts

p-value	26	0.70	0.43	0.64	0.71	0.78	0.88
Point-Biserial	26	0.44	0.26	0.38	0.43	0.47	0.86

2.4 OBSERVATIONS AND INTERPRETATIONS

The *p*-values for the MC items ranged from about 0.43 to 0.88, and the *p*-values for the two constructed-response items (Table 3) were 0.64 and 0.71. The difficulty distribution illustrated in Figure 1 shows an acceptable range of item difficulties on the exam. This is consistent with general test development practice, which seeks to measure student ability along a full range of difficulty.

Chapter 3: IRT Calibrations, Equating, and Scaling (Standards 2 and 4.10)

The item response theory (IRT) model used for the Regents Examination in English Language Arts 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 it has become the standard procedure for analyzing item response data in large-scale assessments. According to van der Linden and Hambleton (1997), “The central feature of IRT is the specification of a mathematical function relating the probability of an examinee’s response on a test item to an underlying ability.” Ability in this sense can be thought of as performance on the test and is defined as “the expected value of observed performance on the test of interest” (Hambleton, Swaminathan, and Rogers, 1991). This performance value is often referred to as *T* Performance and *T* 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 English Language Arts items. Generally, item calibration is the process oftnAAa4 (r) This ch1aptiTuty hXe t89s/a7te tXe5(n)94

$$P_{ni}\ X \quad 1 = \frac{\exp\ T_n \quad D_{ij}}{1 \quad \exp\ T_n \quad D_{ij}}.$$

The Rasch model places both performance and item difficulty (estimated in terms of log-odds or logits) on the same continuum. When the model assumptions are met, the Rasch model provides estimates of examinee performance and item difficulty that are theoretically invariant across random samples of the same examinee population.

3.2 SOFTWARE AND ESTIMATION ALGORITHM

Item calibration was implemented via the WINSTEPS 3.60 computer program (Linacre, 2005), which employs unconditional (UCON) joint maximum likelihood estimation (JMLE).

3.3 ITEM DIFFICULTY–STUDENT PERFORMANCE MAP

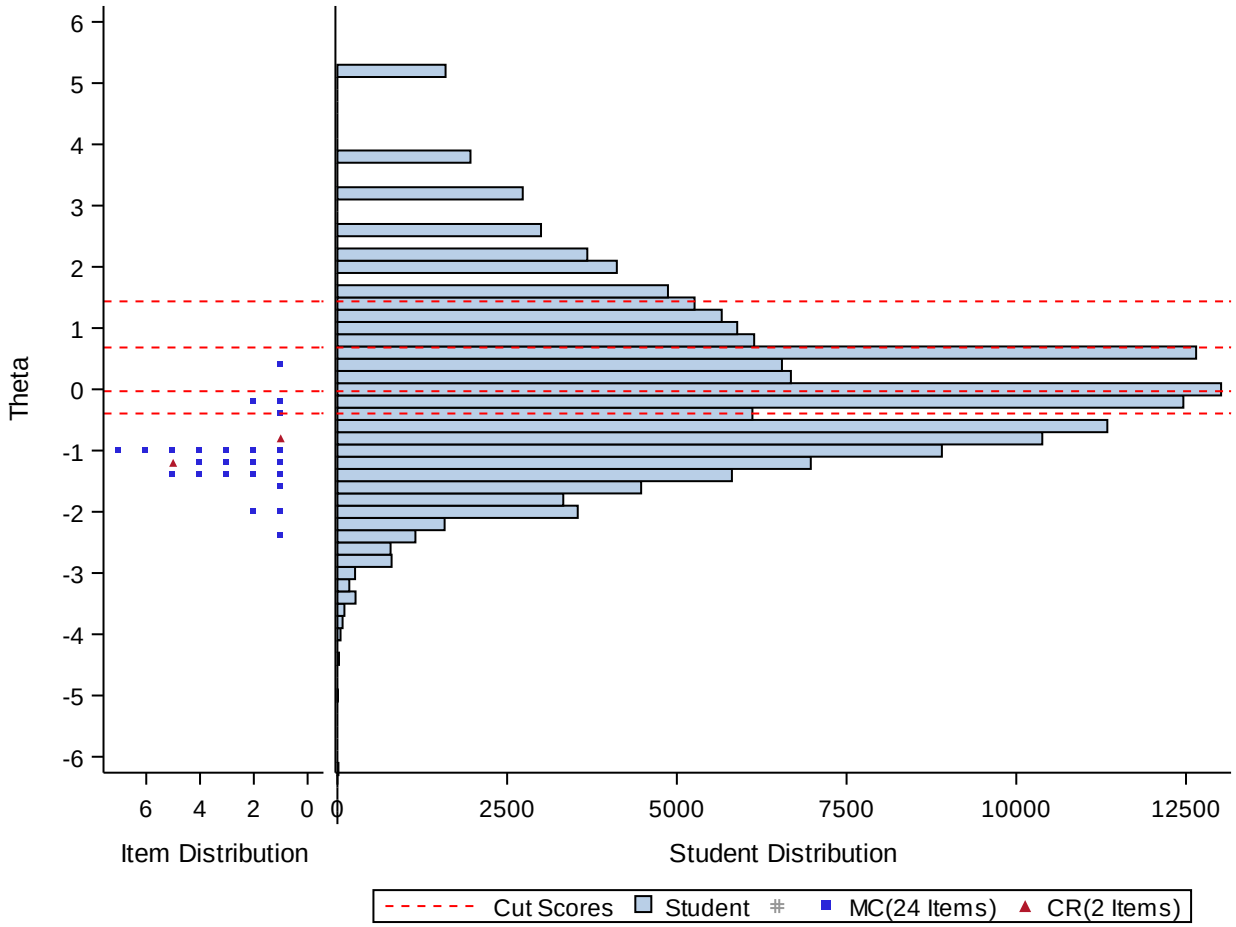

Figure 2 Student Performance Map: Regents Examination in English Language Arts

3.4 CHECKING RASCH ASSUMPTIONS

Since the Rasch model was the basis of all calibration, scoring, and scaling analyses associated with the Regents Examination in English Language Arts, 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

Rasch models assume thh Langu EMC dimensionality, a2

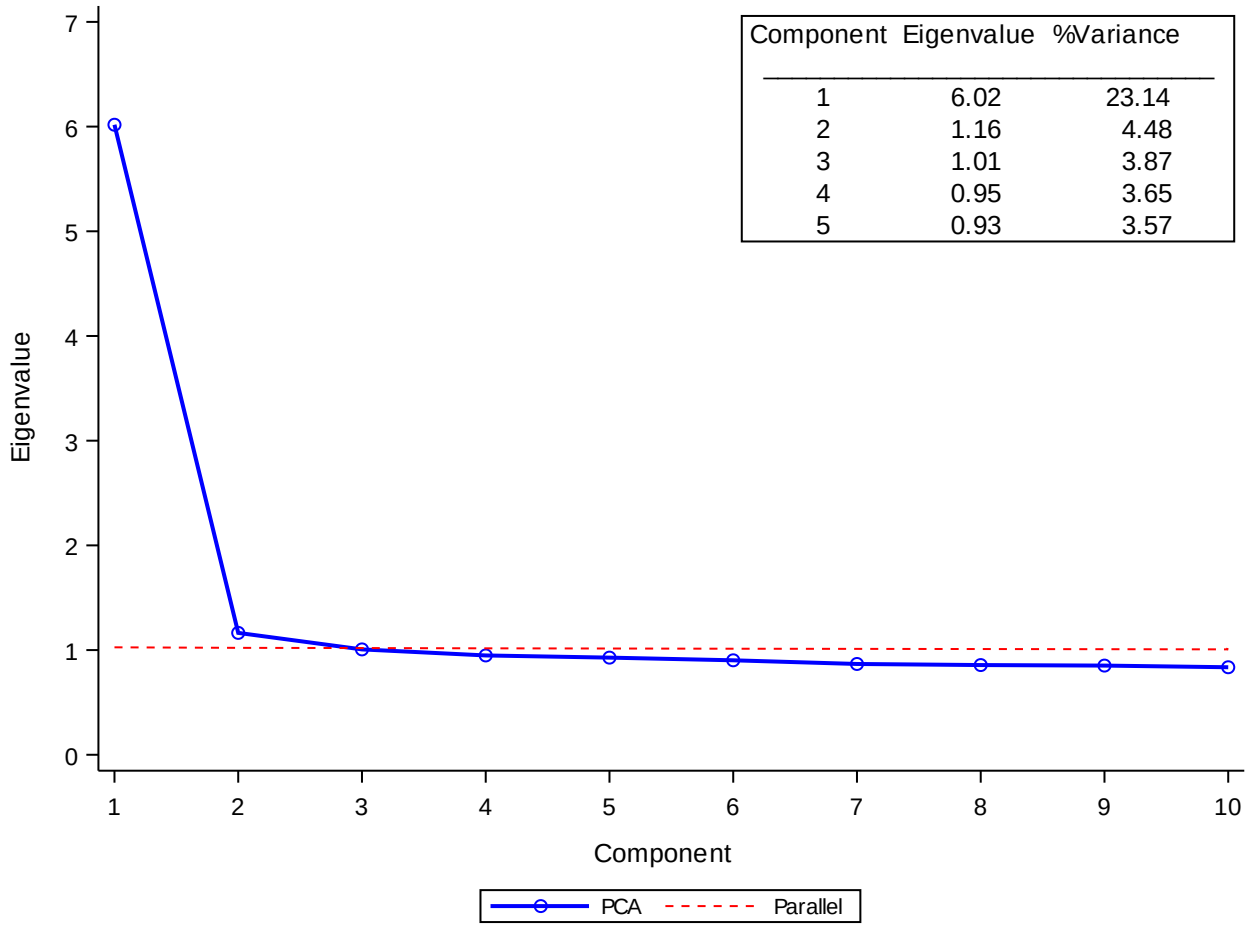

Figure 3 Scree Plot: Regents Examination in English Language Arts

Local Independence

Local independence (LI) is a fundamental assumption of IRT. This means that, for statistical purposes, an examinee's

distinction is important because many indicators of local dependency are actually framed by WLI. For WLI, the conditional covariances of all pairs of item responses, conditioned on the abilities, are assumed to

Table 5 Summary of Item Residual Correlations: Regents Examination in English Language Arts

N	325
Mean	-0.03

be directly compared to scale scores on all previous administrations back to the June 2014 administration.

When the base administration was concluded, the initial raw score-to-scale score relationship was established. Three raw scores were fixed at specific scale scores. Scale scores of 0 and 100 were fixed to correspond to the minimum and maximum possible raw scores. In addition, a standard setting had been held to determine the passing and passing with distinction cut scores in the raw score metric. The scale score points of 55, 65, and 85 were set to correspond to those raw score cuts. A fourth-degree polynomial is required to fit a line exactly to five arbitrary points (e.g., the raw scores corresponding to the five critical scale scores of 0, 55, 65, 85, and 100). The general form of this best-fitting line is:

$$55 = I4\ \hat{U}45^{8} + I3\ \hat{U}45^{7} + I2\ \hat{U}45^{6} + I1\ \hat{U}45^{5} + I0,$$

where SS is the scaled score, RS is the raw score, and m0 through m4 are the transformation constants that convert the raw score into the scale score (please note that m0 will always be equal to zero in this application, since a raw score of zero corresponds to a scale score of zero). A subscript for a person on both dependent and independent variables is not present for simplicity. The above relationship and the values of m1 to m4 specific to this subject were then used to determine the scale scores corresponding to the remainder of the raw scores on the examination. This initial relationship between the raw and scale scores became the base scale.

The Rasch difficulty parameters for the items on the base form were then usehnto e.e scTJ -0.052 Tv

Chapter 4: Reliability (Standard 2)

standard deviation) into account. Consider that a SEM of 3 on a 10-point test would be very different from a SEM of 3 on a 100-point test.

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 precision of individual test scores. SEMs help place “reasonable limits” (Gulliksen, 1950) around observed scores through the construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, *X*, and adding and 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 abo(or)7 (e w)4Crue Se442 Tw 075g

The relationship between and the scale score is not expressible in a simple mathematical form because it is a blend of the third-degree polynomial relationship between the raw and scale scores and the nonlinear relationship between the expected raw and scores. In addition, as the exam is equated from year to year, the relationship between the raw and scale scores moves away from the original third-degree polynomial relationship to one that is also no longer expressible in simple mathematical form. In the absence of a simple mathematical relationship between and the scale scores, the CSEMs that are available for each score via Rasch IRT cannot be converted directly to the scale score metric.

The use of Rasch IRT to scale and equate the Regents Examinations does, however, make it possible to calculate CSEMs by using the procedures described by Kolen, Zeng, and Hanson (1996) for dichotomously scored items and extended by Wang, Kolen, and Harris (2000) to polytomously scored items. For tests such as the Regents Examination in English Language Arts that do not have a one-to-one relationship between raw ( ) and scale scores, the CSEM for each achievable scale score can be calculated by using the compound multinomial distribution to represent the conditional distribution of raw scores for each level of .

Consider an examinee with a certain performance level. If it were possible to measure this examinee's performance perfectly, without any error, this measure could be called the examinee's "true score," as discussed earlier. This score is equal to the expected raw score. However, whenever an examinee takes a test, the observed test score alwaysesould be cal3uonould be

Conditional Standard Error of Measurement Confidence Intervals

CSEMs allow statements regarding the precision of individual test scores. Like SEMs, they help place reasonable limits around observed scaled scores through the construction of an approximate score band. The confidence intervals are constructed by adding and subtracting a multiplicative factor of the CSEM.

Conditional Standard Error of Measurement Characteristics

The relationship between the scale score CSEM and depends both on the nature of the raw-to-scale score transformation (Kolen & Brennan, 2005; Kolen & Lee, 2011) and on whether the CSEM is derived from the raw scores or from (Lord, 1980). The pattern of CSEMs for raw scores and linear transformations of the raw score tend to have a characteristic "inverted-U" shape, with smaller CSEMs at the ends of the score continuum and larger CSEMs toward the middle of the distribution.

Achievable raw score points for these distributions are spaced equally across the score range. Kolen and Brennan (2005, p. 357) state, "When, relative to raw scores, the transformation compresses the scale in the middle and stretches it at the ends, the pattern of the conditional standard errors of measurement will be concave up (U-shaped), even though the pattern for the raw scores was concave down (inverted-U shape)."

Results and Observations

The relationship between raw and scale scores for the Regents Examination tends to be roughly linear from scale scores of 0 to 65 and then concave down from about 65 to 100. In other words, the scale scores track linearly with the raw scores for about the lower 80 percent of the scale score range and then are compressed relative to the raw scores for about the remaining 20 percent of the range, though there are variations. The CSEMs for the Regents Examinations can be expected to have inverted-U shaped patterns, with some variations.

Figure 4 shows this type of CSEM variation for the Regents Examination in English Language Arts where the compression of raw score to scale scores around the cut score 85 changes the shape of the curve slightly. This type of expansion and compression can be seen in Figure 4 by looking at the changing density of raw score points along the scale score range on the horizontal axis. Specifically, at the lower end of the scale, scale scores 0 through 14 span raw scores 0 through 17. Over the range from scale scores 21 to 60, the raw score range is 19 to 30 (40 scale score points to 12 raw score points). Finally, scale scores over the range of 63 to 100 span raw scores of 31 to 56 (38 scale score points for 26 raw score points).

Several factors might affect decision consistency and accuracy. One important factor is the
ceJ 0.017 (esly)4 a (y)4(e)10(s)4 (t)2 2 (he(t)2 7 (esat (he(t)2 6 (s7 0.a (s)4eJ 0.01ob(s7 04 (f)vs7 04

Table 9 Group Means: Regents Examination in English Language Arts

	162,398	77.44	18.01

Table 10 State Percentile Ranking for Scale Score:

Chapter 5: Validity (Standard 1)

Restating the purposes and uses of the Regents Examination in English Language Arts, this exam measures examinee achievement against the New York State Learning Standards. The exam is prepared by teacher examination committees and New York State Education Department subject matter and testing specialists. Additionally,

The Regents Examination in English Language Arts measures student achievement on the NYS P–12 Learning Standards for English Language Arts, consistent with the Model Content Framewo

Overall, the test requires that students closely read eight texts of up to approximately 6,200 words total and that they answer 24 multiple-choice questions, write one source-based argument, and one text-

Figure 7 New York State E ducation Department Test Development Process

Item Review Process

The item review process helps to ensure 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. This process allows high-quality items to be continually developed in a manner that is consistent with the test blueprint. Item review guidelines for multiple-choice items are included in Appendix C.

All reviewers participate in rigorous training designed to assist in a consistent interpretation of the standards throughout the item review process. This is a critical step in item development because consistency between the standards and what the items are asking examinees is a fundamental form of evidence of the validity of the intended score interpretahe ids61b0a4 (k)4 (id)10

- The rubric should be aligned to what is asked for in the item and correspond to the knowledge or skill being assessed.
- Whenever possible, the rubric should be written to allow for alternate approaches and other legitimate methods.

In support of the goal of valid score interpretations for each examinee, such scoring training procedures are implemented for the Regents Examination in English Language Arts. Operational raters are selected based on expertise in the exam subject and are assigned a specific set of items to score. No more than approximately one-half of the items on the test are assigned to any one rater. This increases the consistency of scoring across examinee responses by allowing each rater to focus on a subset of items. It also ensures that no one rater is allowed to score the entire test for any one student. This practice reduces the effect of any potential bias of a single rater on individual examinees. Additionally, raters are not allowed to score the responses of their own students.

Statistical Analysis

One statistic that is useful for evaluating the response processes for multiple-choice items is an item's point-biserial correlation on the distractors. A high point-biserial on a distractor may indicate that students are not able to identify the correct response for a reason other than the

June 2019 Regents Examination in English Language Arts show a range of *p*-values consistent with the targeted exam difficulty. Item *p*-values range from 0.43 to 0.88, with a mean of 0.70. The difficulty distribution illustrated in Figure 1 shows a wide range of item difficulties on the exam. This is consistent with general test development practice, which seeks to measure student ability along a full range of difficulty. Refer to Chapter 2 of this report for additional details.

Item Discrimination

How well the items on a test discriminate between high- and low-performing examinees is an important measure of the structure of a test. Items that do not discriminate well generally provide less reliable information about student performance. Table 2 and Table 3 provide point-biserial values on the correct responses, and Table 2 also provides point-biserial values on the three distractors. The values for correct answers are 0.26 or higher, indicating th1.15 Td 0 Td e28.29 0 Td

5.4 EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

institutions using the exam for placement purposes are advised to examine the scoring rules for the Regents Examination in English Language Arts and to assess their appropriateness for the inferences being made about course placement.

As stated, the nature of validity arguments is not absolute, rather 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 English Language Arts scores for the purposes described.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Barkaoui, Khaled. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Item response theory*

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing, 12*, 54–71.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719-748.

Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*(3), 200-15.

McDonald, R. P. (1979). The structural analysis of multivariate data: A sketch of a general theory. *Multivariate Behavioral Research, 14*, 21–38.

McQueen, J., & Congdon, P. J. (1997, March). *Rater severity in large-scale assessment: Is it invariant?* Paper presented at the Annual Meeting of the Ame

Rasch, G. (1960).

Table A.3 Test Map for June 201 9 Administration

1	MC	1	1	RL.4	0.54	0.35	-	

Appendix B: Raw-to-Theta-to-Scale Score Conversion Tables

Table B.1 Score Table for August 2018 Administration

Table B. 2 Score Table for January 201 9 Administration

0	-6.2135	0.000
1	-4.9781	0.051
2	-4.2405	0.674
3	-3.7912	1.096
4	-3.4604	1.595
5	-3.1948	1.996
6	-2.9707	2.584
7	-2.7754	3.142
8	-2.6014	3.798
9	-2.4438	4.511
10	-2.2991	5.269
11	-2.1648	6.085
12	-2.0391	6.957
13	-1.9206	7.892
14	-1.8080	8.861

41	1.2326	83.520
42	1.4327	84.937
43	1.6489	86.245
44	1.8820	87.517
45	2.1329	88.800
46	2.4014	90.130
47	2.6873	91.503
48	2.9904	92.896
49	3.3115	94.251
50	3.6537	95.494
51	4.0231	96.586
52	4.4307	97.581
53	4.8982	98.474
54	5.4763	99.439
55	6.3325	99.999

3 Score Table for June 201 9 Administration

012.24 68 W n BT

-5.6600	0.000
-4.4308	0.513
-3.7018	1.231
-3.2610	1.896
-2.9385	2.669
-2.6813	3.497
-2.4660	4.402
-2.2799	5.384
-2.1155	6.427
-1.9682	7.514
-1.8346	8.629
-1.7123	10.064
-1.5994	12.682

41	1.2369	83.552
42	1.4406	84.991
43	1.6630	86.326
44	1.9045	87.634
45	2.1646	88.960
46	2.4424	90.329
47	2.7371	91.738
48	3.0484	93.159
49	3.3772	94.524
50	3.7258	95.746
51	4.0988	96.771
52	4.5059	97.765
53	4.9681	98.599

Appendix C: Item Writing Guidelines

pre5.4 ie5.4m(D)-8.1.( )]T /TT1 1 Tc 0 Tc 0

CHECKLIST OF TEST CONSTRUCTION PRINCIPLES
(Multiple-Choice Items)

		YES	NO
1.	Is the item significant?		
2.	Does the item have curricular validity?		
3.	Is the item presented in clear and simple language, with vocabulary kept as simple as possible?		
4.	Does the item have one and only one correct answer?		
5.	Does the item state one single central problem completely in the stem? (See Helpful Hint below.)		
6.	Does the stem include any extraneous material ("window dressing")?		
7.	Are all responses grammatically consistent with the stem and parallel with one another in form?		
8.	Are all responses plausible (attractive to students who lack the information tested by the item)?		
9.	Are all responses independent and mutually exclusive?		
10.	Are there any extraneous clues due to grammatical inconsistencies, verbal associations, length of response, etc.?		
11.	Were the principles of Universal Design used in constructing the item?		

HELPFUL HINT

To determine if the stem is complete (meaningful all by itself):

1. Cover up the responses and read just the stem.

2. Try to turn the stem into a short-answer question by drawing a line after the last word. (If it would not be a good-short answer item you may have a problem with the stem.)

3. The stem must consist of a statement that contains a verb.

Appendix D: Standards Interpretations for

Analyze the impact of the author's choices regarding how to develop and relate elements of a story or drama (e.g., where a story is set, how the action is ordered, how the characters are introduced and developed). *Analyze how and why individuals, events, or ideas develop and interact over the course of a text.* *Central aspect: Items measure students' ability to analyze how different elements of a story interact with one another. Elements may include how a plot is developed in episodes, how the setting shapes the plot, how characters respond to changes in the story, or how specific dialogue and action propel the plot.*

Measures central aspect: The item requires identification or analysis of the interaction between developing story elements (character(s), setting, events, and ideas)

Possible stems may include:

Determine the meaning of words and phrases as they are used in the text, including figurative and connotative meanings; analyze the impact of specific word choices on meaning and tone, including words with multiple meanings or language that is particularly fresh, engaging, or beautiful. *Interpret words and phrases as they are used in a text, including determining technical, connotative, and figurative meanings, and analyze how specific word choices shape meaning or tone.* *Central aspects: Items measure the students' ability to determine the meaning of multiple-meaning words as they are used in text OR analyze the impact of specific word choices.*		

[(P)24 T(r)-.4 T(epar)-.4 T(e)-12.3 (d f)-1.1 (or)-.4 T( )-12.1 (N)-2.9 (Y)-9.

Analyze a case in which grasping point of view requires distinguishing what is directly stated in a text from what is really meant. *Assess how point of view or purpose shapes the content and style of a text.* *Central aspects: Items measure students' ability to assess how point of view or purpose shapes the content and style of a text* OR*to analyze how an author builds a point of view through the narrator, different characters, and/or the audience.*		
Measures central aspect: The item requires students to identify how an author uses specific techniques to distinguish the narrator's point of view from those of the other characters and the audience. Possible stems may include: Stem: How does the narrator's point of view contrast with the point of view of another character in the story? Stem: How does the author help the reader better understand the points of view of X and Y? Stem: What is an effect of telling the story from X's point of view? Stem: How does the author develop X's character in lines #–#?		
Measures supporting aspect: The item requires students to identify author, narrator, or characters' point of view Possible stems may include: Stem: What do lines #-# suggest about author's point of view? Stem: In which sentence does the narrator reveal the irony/suspense/etc. of his situation?		
The item stem reveal: - all necessary points of view - how the point of view is developed using a specific strategy - clues that would limit reliance on text		
The analysis in the item is supported by the text		
THE ITEM MEASURES THIS STANDARD.		

Determine two or more central ideas of a text and analyze their development over the course of the text, including how they interact and build on one another to provide a complex analysis; provide an objective summary of the text. *Determine central ideas or themes of a text and analyze their development; summarize the key supporting details and ideas.* *Central aspect: Items measure students' ability to analyze the development of a central idea across the text, and to summarize the text's ideas.*		
Measures central aspect: The item requires students to identify or summarize how a story's central ideas . Possible stems may include: Stem: The author develops a central idea/claim in the text by Stem: Lines # to # develop a central idea/claim by Stem: Which statement best describes how lines # to # develop the author's claim/central idea?		
Measures supporting aspect: The item requires students to identify or summarize a text's central themes or an author's central claim. Possible stems may include: Stem: Which detail would be most important to include in a summary of the text? Stem: Which direct quote from passage best summarizes a central idea of the text/central claim of the author? Stem: Which statement best states the central idea of the text/ central claim of the author?		

The item stem reveal:

- the central ideas/claims
- how the central ideas/claims develop

*unless the idea/claim is presented in the stem in order to ask the

Determine the meaning of words and phrases as they are used in a text, including figurative, connotative, and technical meanings; analyze how an author uses and refines the meaning of a key term or terms over the course of a text (e.g., how Madison defines faction in Federalist No. 10). *Interpret words and phrases as they are used in a text, including determining technical, connotative, and figurative meanings, and analyze how specific word choices shape meaning or tone.* *Central aspects: Items measure the students' ability to determine the meaning of multiple-meaning words as they are used in text OR analyze the impact of specific word choices.*		

Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on grades 11–12 reading and content. *Determine or clarify the meaning of unknown and multiple-meaning words and phrases by using context clues, analyzing meaningful word parts, and consulting general and specialized reference materials, as appropriate.* *Central aspect: Items measuring this standard test students' ability to determine the meaning of words and phrases in context.*		
Aligned:Item requires that the student glean meaning of an above grade-level or otherwise unfamiliar vocabulary word from context Possible stems may include: Stem: Which word or phrase from sentence X helps the reader understand the meaning of word Y in the passage? Stem: Which line best helps define the meaning of word X? Stem: What is the definition of X as used in lines #–# (figurative/connotative/analogous word use)		
The item stem reveal: - the meaning of the word or phrase		
The text provides the required context		
THE ITEM MEASURES THIS STANDARD.		

Appendix E: Item Review Criteria

New York State Regents Examination in English Language Arts

Multiple-Choice Item Review Checklist

Text _____ Passage ____ Set _____ Item ____
Reviewer _____ Date _____

	The item is free of content that might offend, typecast, or lead to offensive or stereotypic inferences regarding individuals or groups of different genders, ethnicities, locations, religions, socioeconomic status, political views, family situations, ability, or physical or mental condi44.16 1.44 6.1 (d)6S96 30.31.4 (s)8.24 T.6.1 (i)8.1 (s)4.5.2 ET]TJ 0 Tc 0 w	

Appendix F: Tables and Figures for August 2018 Administration

Table F.1 Multiple-Choice Item Analysis Summary: Regents Examination in English Language Arts

		p	

Table F.2 Constructed- Response Item Analysis Summary: Regents Examination in English Language Arts

Table F.3 Descriptive Statistics in p- value and Point -Biserial Correlation: Regents Examination in English Language Arts

Table F.5 Summary of INFIT Mean Square Statistics: Regents Examination in English Language Arts

ELA	26	1.00	0.05	0.92	1.07	[26/26]

Table F.6 Reliabilities and Standard Errors of Measurement: Regents Examination in English Language

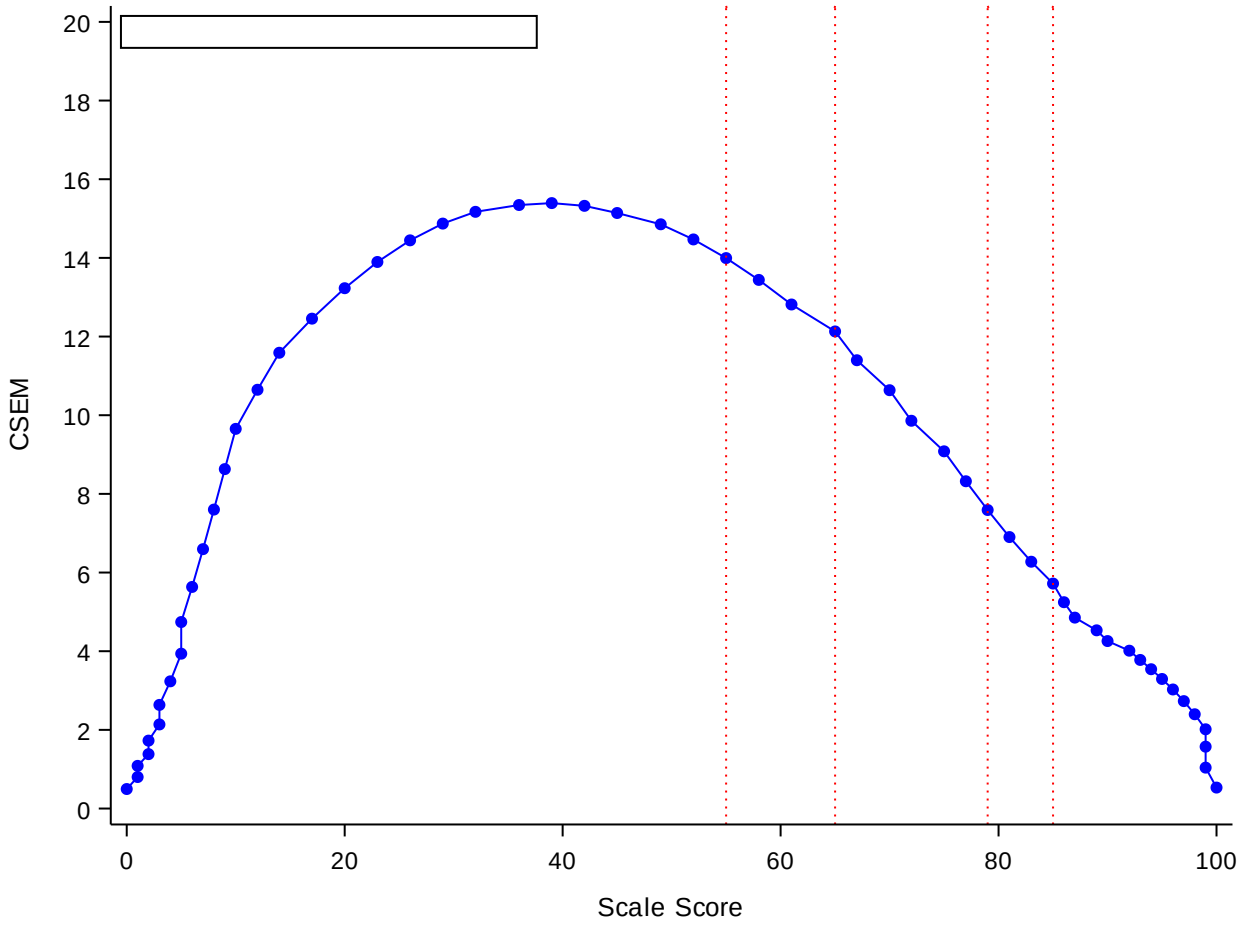
CSEM
Scale Score
0
2
4
6
8
10
12
14
16
18
20
0
20
40
60
80
100

Table F.8 Group Means: Regents Examination in English Language Arts

Appendix G: Tables and Figures for January 2019 Administration

Table G.1 Multiple-Choice Item Analysis Summary: Regents Examination in English Language Arts

		p					
1	108,236	0.76	0.43	0.30	-0.19	-0.16	-0.13
2	108,236	0.67	0.47	0.20	-0.14	-0.09	-0.05
3	108,236	0.78	0.42	0.32	-0.17	-0.16	-0.24
4	108,236	0.82	0.39	0.41	-0.27	-0.18	-0.20
5	108,236	0.78	0.41	0.43	-0.27	-0.18	-0.23
6	108,236	0.72	0.45	0.42	-0.25	-0.18	-0.22
7	108,236	0.82	0.38	0.41	-0.21	-0.21	-0.23
8	108,236	0.47	0.50	0.29	-0.06	-0.20	-0.15
9	108,236	0.54	0.50	0.32	-0.11	-0.16	-0.19
10	108,236	0.55	0.50	0.46	-0.28	-0.27	-0.08
11	108,236	0.68	0.47	0.49	-0.22	-0.33	-0.18

Table G.2 Constructed- Response Item Analysis Summary: Regents Examination in English Language Arts

						p	
25	0	6	108,236				

Table G.8 Group Means: Regents Examination in English Language Arts