



Pearson

**Indiana Statewide Testing for
Educational Progress-Plus (ISTEP+)
Technical Report - 2018/2019**

*Indiana Department of Education
Pearson*

Contents

Chapter 1: Introduction.....	1
Chapter 2: Argument Based Validation	3
Test Score Interpretation and Use Validation.....	3
Validity Claims and Methodology Overview	3
Chapter 3: Academic Content Standards	7
Chapter 4: Test Development	8
Blueprint Design	8
Passage Development	9
Passage Review and Selection	9
Passage and Item Specifications.....	10
Item Writing and Review.....	11
Chapter 5: Forms Construction	13
Test Design	13
Item Selection and Form Development.....	14
Test Construction and Pre-Equating.....	14
Chapter 6: Administration.....	15
Participation Requirements.....	15
Special Populations.....	16
Chapter 7: Scoring	17
Scoring Rubrics for Open-Ended Items	17
Anchor Papers.....	17
Recruiting of Handscorers	18
Training and Qualification of Handscorers	18
Handscoring Process and Validity	19
Inter-rater Agreement	20
Valid and Invalid Test Attempts	21
Chapter 8: Students	22
Chapter 9: Equating and Item Analysis	23
Spring 2019 Operational Post-Equating.....	23
Spring 2019 First Time Administration Operational and Field Test Item Analysis	25
Scaling	26

ISTEP+ Spring 2019 Field Test Item Equating	27
Chapter 10: Achievement Standards.....	28
Chapter 11: Reliability	29
Internal Consistency	29
Classical Standard Error of Measurement	30
Conditional Standard Error of Measurement.....	30
Classification Consistency and Accuracy.....	31
Chapter 12: Reporting.....	33
Report Design Process.....	34
Chapter 13: Validity Argument Synthesis	35
Evidence Related to System Coherence	35
Evidence from the Design and Development Phase	36
Evidence from the Scoring and Analysis Phase	37
Evidence from the Reporting Phase.....	37
Synthesis of Evidence Related to System Coherence.....	38
Evidence Related to Comparability and Procedural Quality	38
Evidence from the Design and Development Phase	38
Evidence from the Administration Phase.....	39
Evidence from the Scoring and Analysis Phase	40
Evidence from the Reporting Phase.....	41
Synthesis of Evidence Related to Comparability and Procedural Quality	41
Evidence Related to Accessibility and Fairness.....	42
Evidence from the Design and Development Phase	42
Evidence from the Administration Phase.....	43
Evidence from the Scoring and Analysis Phase	43
Synthesis of Accessibility and Fairness Evidence	43
Summary of Validity Evidence	44
References	46
Appendix A: ISTEP+ Test Blueprints and Item Counts.....	47
Reporting Category Blueprint for Winter 2018 Retest.....	47
Reporting Category Blueprint for Retest Spring 2019.....	48
Reporting Category Blueprint for First Time Administration (FTA) Spring 2019.....	49
Reporting Category Blueprint for Summer 2019.....	50

Mathematics	51
Reporting Category Blueprint for Winter 2018 Retest.....	51
Reporting Category Blueprint for Spring 2019 Retest.....	52
Reporting Category Blueprint for Spring 2019 FTA	53
Reporting Category Blueprint for Summer 2019.....	54
Reporting Category and Strand Priority Blueprint Analysis.....	55
Reporting Category and Strand Priority Blueprint Analysis.....	56
Appendix B: ISTEP+ Testing Times	58
Timing for Winter 2018 and Spring 2019 Retest.....	58
Timing for Spring 2019 First Time Administration.....	61
Appendix C: ISTEP+ Test Item Statistics	64
Appendix D: ISTEP+ DIF Flags.....	65
Appendix E: Stability Check Summary: Spring 2019	66
Appendix F: Equating Constants and LOSS and HOSS.....	67
Appendix G: ISTEP+ Test Curves.....	68
Appendix H: ISTEP+ Classification Consistency and Accuracy Statistics.....	86
Appendix I: ISTEP+ Summary Data for Test and Reporting Categories	89
Appendix J: ISTEP+ Scale Score Descriptive Data and Distribution.....	114
Appendix K: ISTEP+ Performance Level Distribution	126
Appendix L: Quality Criteria Checklist	132

List of Tables

Table 1. History of Content Standards Adoption and Baseline Scale Years	2
Table 2. Validity Evidence by Validity Theme and Operational Assessment Phase.....	6
Table 3. Administration of ISTEP+ Spring 2019 Forms and Modes.....	13
Table 4. Validity Evidence by Validity Theme and Operational Assessment Phase.....	45
Table 5. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Strand: Winter 2018 Retest.....	47
Table 6. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Strand: Retest Spring 2019.....	48
Table 7. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Reporting Category FTA Spring 2019.....	49
Table 8. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Reporting Category Summer	50
Table 9. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Winter 2018 Retest.....	51
Table 10. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Spring 2019 Retest	52
Table 11. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Spring FTA	53
Table 12. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Summer	54
Table 13. ISTEP+ Grade 10 E/LA Operational Blueprint Analysis for 2016, 2017,2018,and 2019	55
Table 14. ISTEP+ Grade 10 Mathematics Operational Blueprint Analysis for 2016, 2017, 2018 and 2019	56
Table 15. Operational Test Times: English/Language Arts ISTEP+ Part 1: Winter 2018 and Spring 2019 Retest.....	58
Table 16. Operational Test Times: Mathematics ISTEP+ Part 1 : Winter 2018 and Spring 2019 Retest.....	58
Table 17. Operational Test Times: English/Language Arts ISTEP+ Part 2: Winter 2018 and Spring 2019 Retest.....	59
Table 18. Operational Test Times: Mathematics ISTEP+ Part 2: Winter 2018 and Spring 2019 Retest.....	60

Table 19. Operational Test Times: English/Language Arts ISTEP+ Part 1: Spring 2019 FTA ...	61
Table 20. Operational Test Times: Mathematics ISTEP+ Part 1: Spring 2019 FTA.....	61
Table 21. Operational Test Times: English/Language Arts ISTEP+ Part 2: Spring 2019 FTA ...	62
Table 22. Operational Test Times: Mathematics ISTEP+ Part 2: Spring 2019 FTA.....	63
Table 23. Summary of Operational Item P-Values and Item-Total Correlations across Testing Modes: Spring 2019.....	64
Table 24. Summary items flagged for DIF – English/Language Arts Spring 2019.....	65
Table 25. Summary items flagged for DIF – Mathematics Spring 2019	65
Table 26. E/LA Anchor Items: Grade 10 Spring 2019.....	66
Table 27. Mathematics Anchor Items: Grade 10 Spring 2019	66
Table 28. Original Equating Constants: Grade 10 E/LA and Mathematics.....	67
Table 29. HOSS, LOSS, and Performance Level Cut Scores: Grade 10 E/LA and Mathematics	67
Table 30. Classification Consistency and Accuracy – E/LA Winter 2018.....	86
Table 31. Classification Consistency and Accuracy – Mathematics Winter 2018.....	86
Table 32. Classification Consistency and Accuracy – E/LA Sprint Retest 2019.....	87
Table 33. Classification Consistency and Accuracy – Mathematics Sprint Retest 2019	87
Table 34. Classification Consistency and Accuracy – E/LA Spring 2019.....	88
Table 35. Classification Consistency and Accuracy – Mathematics Grade 10 Spring 2019.....	88
Table 36. Summary Data E/LA Grade 10 Winter 2018.....	89
Table 37. Summary Data Mathematics Grade 10 Winter 2018.....	93
Table 38. Summary Data E/LA Grade 10 Spring Retest 2019.....	98
Table 39. Summary Data Mathematics Grade 10 Spring Retest 2019	101
Table 40. Summary Data E/LA Grade 10 Spring 2019.....	106
Table 41. Summary Data Mathematics Grade 10 Spring 2019.....	109
Table 42. Scale Score Statistics E/LA Grade 10 Winter 2018	114
Table 43. Scale Score Statistics Mathematics 10 Winter 2018.....	116
Table 44. Scale Score Statistics E/LA Grade 10 Spring Retest 2019	118
Table 45. Scale Score Statistics Mathematics Grade 10 Spring Retest 2019.....	120
Table 46. Scale Score Statistics E/LA Grade 10 Spring 2019	122
Table 47. Scale Score Statistics Mathematics Grade 10 Spring 2019.....	124
Table 48. Performance Level Distribution E/LA Grade 10 Winter 2018	126
Table 49. Performance Level Distribution Mathematics Grade 10 Winter 2018.....	127
Table 50. Performance Level Distribution E/LA Grade 10 Spring Retest 2019	128

Table 51. Performance Level Distribution Mathematics Grade 10 Spring Retest 2019.....	129
Table 52. Performance Level Distribution E/LA Grade 10 Spring 2019	130
Table 53. Performance Level Distribution Mathematics Grade 10 Spring 2019.....	131

List of Figures

Figure 1. Test Development Process	8
Figure 2. Excerpt from the ISTEP+ Individual Student Report Overall Student Performance	33
Figure 3. Excerpt from the ISTEP+ Individual Student Report Performance by Strand.....	34
Figure 4. Excerpt from the ISTEP+ Individual Student Report Performance on Open-Ended Items.....	34
Figure 5. Test Characteristic Curves for ISTEP+ E/LA Grade 10 Winter 2018	68
Figure 6. Test Information Curves for ISTEP+ E/LA Grade 10 Winter 2018	69
Figure 7. Standard Error of Measure Curves for ISTEP+ E/LA Grade 10 Winter 2018	70
Figure 8. Test Characteristic Curves for ISTEP+ Mathematics Grade 10 Winter 2018	71
Figure 9. Test Information Curves for ISTEP+ Mathematics Grade 10 Winter 2018	72
Figure 10. Standard Error of Measure Curves for ISTEP+ Mathematics Grade 10 Winter 2018	73
Figure 11. Test Characteristic Curves for ISTEP+ E/LA Grade 10 Spring Retest 2019	74
Figure 12. Test Information Curves for ISTEP+ E/LA Grade 10 Spring Retest 2019	75
Figure 13. Standard Error of Measure Curves for ISTEP+ E/LA Grade 10 Spring Retest 2019.	76
Figure 14. Test Characteristic Curves for ISTEP+ Mathematics Grade 10 Spring Retest 2019.	77
Figure 15. Test Information Curves for ISTEP+ Mathematics Grade 10 Spring Retest 2019.....	78
Figure 16. Standard Error of Measure Curves for ISTEP+ Mathematics Grade 10 Spring Retest 2019	79
Figure 17. Test Characteristic Curves for ISTEP+ E/LA Grade 10 Spring 2019	80
Figure 18. Test Information Curves for ISTEP+ E/LA Grade 10 Spring 2019	81
Figure 19. Standard Error of Measure Curves for ISTEP+ E/LA Grade 10 Spring 2019	82
Figure 20. Test Characteristic Curves for ISTEP+ Mathematics Grade 10 Spring 2019	83
Figure 21. Test Information Curves for ISTEP+ Mathematics Grade 10 Spring 2019	84
Figure 22. Standard Error of Measure Curves for ISTEP+ Mathematics Grade 10 Spring 2019	85
Figure 23. Scale Score Distribution E/LA Grade 10 Winter 2018.....	115
Figure 24. Scale Score Distribution Mathematics Grade 10 Winter 2018	117
Figure 25. Scale Score Distribution E/LA Grade 10 Spring Retest 2019.....	119
Figure 26. Scale Score Distribution Mathematics Grade 10 Spring Restest 2019	121
Figure 27. Scale Score Distribution E/LA Grade 10 Spring 2019.....	123
Figure 28. Scale Score Distribution Mathematics Grade 10 Spring 2019	125

Acronyms and Abbreviations

3PL 3-parameter logistic model

AERA American Educational Research Association

APA American Psychological Association

CCC Case Conference Committee

CR Constructed-response item

CSEM Conditional standard error of measurement

DOK Depth of knowledge

ECA End of Course Assessment

E/LA English/Language Arts

ER Extended-response item

ESL English as a second language

GPC Generalized partial credit model

GR Gridded response

HOSS Highest obtainable scale score

IDOE Indiana Department of Education

INCC Indiana Common Core

IN SBOE Indiana State Board of Education

IPI Indiana Performance Index

IRT Item response theory

LEP Limited English proficient

LOSS Lowest obtainable scale score

MA Mathematics

MC Multiple-choice item

NCME National Council on Measurement in Education

OE Open-ended item

SC Science

SEM Standard error of measurement

SES Socioeconomic status

SS Social Studies

TEI Technology-enhanced item

WP Writing prompt

Chapter 1: Introduction

The purpose of the Indiana Statewide Testing for Educational Progress-Plus (ISTEP+) program is to measure student achievement beginning in Grade 10. ISTEP+ is the high school accountability assessment for Indiana through School Year 2019-2020 and also serves as a Graduation Qualifying Exam (GQE) for students in cohorts 2019- 2022. ISTEP+ assesses student achievement levels in English/Language Arts and Mathematics according to the Indiana Academic Standards that were adopted by the Indiana State Board of Education in 2014. The ISTEP+ assessment is criterion-referenced and is designed to measure students' mastery of the standards. The Indiana Academic Standards are the foundation for the ISTEP+ assessments.

This technical report provides information about the development and technical characteristics of the ISTEP+ assessments for the winter 2018 retest administration, the spring 2019 retest administration, and spring 2019 first time administration (FTA). The chapters outline general information about the construction of the ISTEP+ assessments, statistical analysis of the results, and the interpretation of the scores on the tests. The information in this report can serve as a resource for educators in explaining assessment information to students, parents, teachers, school corporation boards, and the general public. At the end of each chapter, both elements from the United States Department of Education Peer Review Critical Elements and specific standards from the *Standards for Educational and Psychological Testing* (2014) that are relevant to the content of the chapter appear for easy reference.

ISTEP+ assessments are broken into parts for the purpose of test administration. Part 1, which contains applied skills items that are handscored, was administered for the spring 2019 first time administration during a window in early spring — February 25, 2019, to March 22, 2019. Part 2, which includes multiple-choice, gridded response, and technology-enhanced items that are machine-scored, is administered in a later window, which started April 8, 2019, and extended through May 3, 2019. The spring 2019 FTA ISTEP+ forms contained embedded field test items.

For ISTEP+ E/LA and MA in grade 10, 2016 was the baseline year of administration. Retest administrations were available for the first time for grade 10 E/LA and MA beginning in winter of 2017. The winter 2018 and spring 2019 retest administrations occur within a single window, and do not contain embedded field test items. Winter 2018 retest window opened November 22, 2018 and closed December 11, 2018. The spring 2019 retest window opened February 11, 2019 and closed March 8, 2019.

It should be noted that Human Resources Research Organization (HumRRO) provided independent replication of the analyses and reporting for the ISTEP+ Winter 2018 and the Spring 2019 administrations. HumRRO reviewed the preliminary item analysis (PIA) and key checks, the pattern scoring results, and the state data files. Pearson used a program called IRTPro to calculate the theta estimates, whereas HumRRO created its own program to estimate the thetas.

Statistical analyses are provided independently for the spring 2019 FTA and spring 2019 retest administrations. These administrations are not combined for any statistical analyses.

Table 1. History of Content Standards Adoption and Baseline Scale Years

Content Area - Grade	Content Standards	Baseline Year/Scale
E/LA and MA - Grade 10	2014	2016

Critical Elements: 1.1, 1.2, 1.3, 1.4

Standards: 1.1, 4.1, 5.3, 7.0, 7.1, 7.3, 12.1, 12.2, 12.4, 13.4, 13.5

Chapter 2: Argument Based Validation

A large-scale standardized assessment, such as ISTEP+, requires evidence to support the meaningfulness of the inferences made from the scores (validity) and the consistency with which the scores are derived (reliability, equating accuracy, and freedom from processing errors). Validity refers to the degree to which evidence and theory support the interpretations of test scores for the proposed uses of tests (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Test validation is an ongoing process of gathering evidence from many sources to evaluate the trustworthiness of the desired score interpretation or use. Reliability is also a necessary element for validity: inferences from test scores cannot be valid if they are not also reliable.

This chapter explains the argument-based approach to validation around which this technical report is organized. Following is a brief outline of the components of an argument-based approach to validation. Then the chapter continues to identify the claims of the ISTEP+ assessments and provides a framework, including studies with methodologies and results, for gathering validity evidence and an ongoing program of gathering evidence.

Test Score Interpretation and Use Validation

As alluded to previously, validity is a property of the proposed interpretations and uses of test score, and it refers to the degree to which evidence supports those intended interpretations and uses. The IDOE must make valid and reliable decisions about students, schools, and corporations based on scores from student assessments. These decisions relate to students' ability to graduate as well as decisions by schools and corporations about remediation programs for students. In addition, the IDOE intends for student and school-level scores to be used by local educators to make decisions about curriculum and instruction that, over time, will lead to improvements in student achievement.

To support these intended uses, ISTEP+ scores must provide information that reflects what students know and can do in relation to the academic expectations defined in the academic content it measures and achievement standards applied to scores. This is the primary claim that all other claims depend upon. Through the validity evaluation process, the IDOE gathers evidence related to this claim and to the decisions that rely upon it.

Validity Claims and Methodology Overview

Using Kane's (1990) framework for argument-based validation, an interpretive argument for ISTEP+ is outlined below with the claims for the validity argument following. Those claims are explicated in the chapters that follow.

The purpose of the ISTEP+ program is to measure student achievement in the subject areas of E/LA and Mathematics. In particular, ISTEP+ reports student achievement levels according to

the Indiana Academic Standards that were adopted by the Indiana State Board of Education. The assessments are administered in two parts: Part 1 consists of applied skills (open-ended) items, and Part 2 includes multiple-choice and technology-enhanced items. Both Part 1 and Part 2 are required components of the ISTEP+ program and are used to measure student mastery of the Indiana Academic Standards.

More specifically, ISTEP+ assessments are intended to ensure that students possess the knowledge, skills, and abilities necessary to be successful beyond high school. The performance expectations explicated in the performance levels for ISTEP+ are as follows:

- *Did Not Pass*: Students are on a trajectory to need remediation after leaving high school.
- *Pass*: Students are on track to enter a two-year higher education institution, a technical program, or the workplace without the need for remediation after leaving high school.
- *Pass Plus*: Students are on track to enter a four-year college or university without the need for remediation after leaving high school.

There are some major inferences represented in the proposed uses and interpretations of ISTEP+ scores and they are:

1. Proficiency in each grade and subject is necessary in order to be prepared for a two-year higher education program, technical program, or the workplace, as represented by the performance expectations for the *Pass* performance level that students can enter the workforce or a higher education institution without the need for remediation.
2. Mastery in each grade and subject is necessary in order to be prepared for a four-year college or university, as represented by the performance expectations for the *Pass+* performance level that students can enter a university setting without the need for remediation.
3. Proficiency and/or mastery of the content being taught in each preceding grade of any subject area is necessary in order to reach proficiency and/or mastery in the subsequent grade.
4. The content being taught to students in each subject contains the knowledge and skills necessary to be successful on the ISTEP+.
5. The ISTEP+ assessments for each subject area measure the requisite skills required in subsequent grades, are reliable, and are not influenced substantially by systematic error.
6. The cut scores for the ISTEP+ assessments are appropriate and quantify the level of knowledge and skill a student must have to be successful beyond high school.

These inferences can be used to start outlining an agenda of validity studies that are organized around three critical validity claims described as:

- a. **System Coherence**: The assessment and its operational system have been designed to yield scores that reflect students' knowledge and skills in relation to academic expectations. This theme corresponds to aspects of key validity issues of test content, response processes, internal structure, and relations with other variables as described in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 2014).
- b. **Comparability or Procedural Quality**: The assessment system operates as intended (e.g., administration, scoring, analyses, and reporting). This theme corresponds to

aspects of the administration and procedural analyses and reporting structures to inform validity issues.

- c. **Accessibility and Fairness:** Students take the assessment under conditions that allow them to demonstrate what they know and can do in relation to academic expectations. This theme corresponds to aspects of the administration procedures, response processes, and internal structure to inform validity issues.

The table below shows the program of study. The program of study can also be organized into processes used to ensure standardization and potential studies that could be conducted to further improve the validity of score interpretations.

Table 2. Validity Evidence by Validity Theme and Operational Assessment Phase

Validity Theme	Assessment Phases and Forms of Validity Evidence			
	Test Design and Construction	Administration Process	Scoring and Psychometric Analyses	Reporting
System Coherence	Process: <ul style="list-style-type: none"> • Development of content standards • ISTEP+ blueprints • Item specifications • Item writing/reviews • Pilot tests 		Process: <ul style="list-style-type: none"> • Scoring accuracy • Standard setting 	Process: <ul style="list-style-type: none"> • Types of scores reported • Report design
Comparability/ Procedural Quality	Process: <ul style="list-style-type: none"> • Attention to alignment • Forms construction and mode comparability • Anchor item selection 	Process: <ul style="list-style-type: none"> • Test calendars and administration structure • Production of test materials • Distribution of test materials • Administration training • Monitoring administration • Return and processing of materials 	Process: <ul style="list-style-type: none"> • Reliability • Scaling • Equating • Test level mode comparability 	Process: <ul style="list-style-type: none"> • Report production • Report distribution
Accessibility and Fairness	Process: <ul style="list-style-type: none"> • Universal design principles • Bias reviews 	Process: <ul style="list-style-type: none"> • Eligibility criteria • Accommodations • Monitoring 	Process: <ul style="list-style-type: none"> • Differential item functioning analyses 	

In the chapters that follow, evidence related to the major validity inferences is organized into chapters that align with the operational phases of the testing program: adoption of academic content standards, test development, forms construction, administration, scoring, psychometric analyses, and reporting. Each of these chapters is structured to address each of the validity claims presented above. A final chapter is then presented as a review of the entire validity argument; it does this by integrating into a coherent validity narrative the specific claims discussed in the preceding chapters.

Critical Elements: 3.1, 3.2, 3.3, 3.4

Standards: 1.1, 1.2, 1.5, 1.6, 1.7, 1.8, 1.9, 1.10, 1.11, 1.12, 1.13, 1.14, 1.15, 1.16, 1.17, 1.25, 7.1, 12.1, 12.2, 12.3, 12.4

Chapter 3: Academic Content Standards

The Indiana Academic Standards in E/LA and Mathematics are the content standards to which all ISTEP+ assessments are aligned. The Indiana Academic Standards were designed using a process “to identify, evaluate, synthesize, and create high-quality, rigorous standards for Indiana students” (Indiana Academic Standards, 2016). They have been validated by several different groups as indicating college and career readiness, including “the Indiana Education Roundtable, the Indiana Commission for Higher Education, the Indiana Department of Education, the Indiana State Board of Education, and the Indiana Center for Education and Career Innovation” (Indiana Academic Standards, 2016). The college- and career-readiness indication means that “students who successfully master these objectives for what they should know and be able to do in Mathematics and English/Language Arts disciplines by the time they graduate from high school will be ready to go directly into the workplace or a postsecondary educational opportunity without the need of remediation” (Indiana Academic Standards, 2016).



The Indiana Academic Standards for E/LA and Mathematics were developed using a collaborative process which involved teams of K-12 educators, parents representing school corporations from throughout Indiana, professors from post-secondary public and private institutions, and representatives from Indiana businesses and industries. A public review process was used to revise and approve the final Academic Standards. All Academic Standards can be found on the IDOE website: <http://www.doe.in.gov/standards>. 2017-2018 was the last year the ISTEP+ MA grades 3-8, E/LA grades 3-8, SC and SS studies test were administered. Information regarding these ISTEP+ tests can be found in the 2017-2018 technical report. The 2018-2019 ISTEP+ was only administered as a high school exam for E/LA and Mathematics.

Originally, the Indiana Academic Standards for E/LA and Mathematics were approved and adopted in 2000. Starting in the 2011-2012 academic year and through the 2013-2014 academic year, schools in Indiana began a gradual implementation of the E/LA and MA Common Core State Standards (CCSS), or Indiana’s Common Core (INCC) standards. In November 2013, Public Law 286 was passed by the Indiana General Assembly, which required the review of the INCC standards and adoption of college- and career-ready educational standards. After a comprehensive development and review process, the IN SBOE adopted the college- and career-ready Indiana Academic Standards in April 2014. The Standards were used to assess Mathematics and E/LA for grade 10 starting in spring 2017.

More information regarding the measured standards for each subject is presented in blueprints in Appendix A and discussed in Chapter 4.

Critical Elements: 1.1-1.4, 3.1

Standards: 1.11, 4.1, 4.2, 7.2, 12.4, 12.8, 12.9

Chapter 4: Test Development

The test development process is shown below. It is a complex, multi-stage process that begins with blueprint development. All stages of the process include involvement of IDOE assessment specialists, Pearson assessment specialists, and K-12 educators. The graphic below shows a high-level overview of the steps in the test development process.

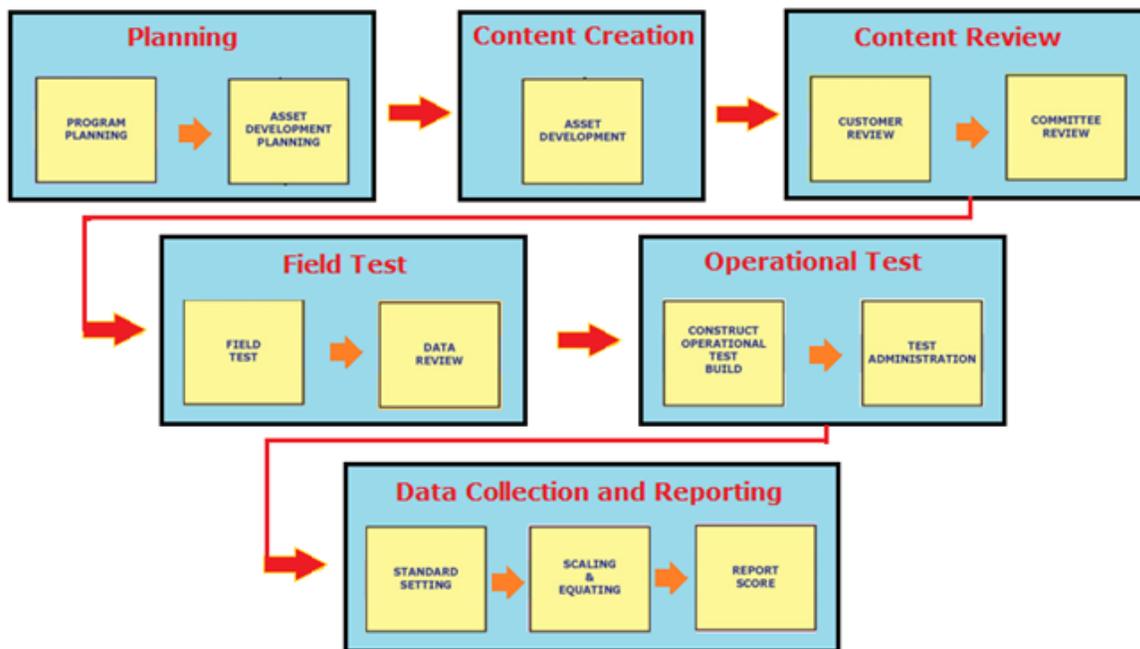


Figure 1. Test Development Process

The remainder of this chapter is focused on the top row of activities, including planning, content creation, and content review.

Blueprint Design

Indiana's Academic Standards serve as the foundation for test development and item writing activities for ISTEP+. In the first step of the test development process, a blueprint for each assessment must be identified. Because the spring 2019 first time administration and retest administrations (winter 2018 and spring 2019) were a continuation of an existing assessment program, blueprints for each subject were available from the baseline year each test was first administered. The final blueprints at the reporting category level are shown in Appendix A.

Detailed blueprints containing target point ranges for each standard can be found here: <https://www.doe.in.gov/sites/default/files/assessment/grade-10-math-public-facing-blueprint-final.pdf> <https://www.doe.in.gov/sites/default/files/assessment/grade-10-ela-public-facing-blueprint-final.pdf>.

Passage Development

Indiana's testing contractor, Pearson, used a subcontractor to search for and write the needed passages.

Measurement Incorporated content experts were trained by Pearson assessment specialists. This training included an overview of the test-development cycle and the learning outcomes and measurement specifications for each reporting category to which the passages were intended to align. Additionally, the subcontractor was presented with information regarding the selection and documentation of reliable sources. They were also presented with the scope of the testing program, security issues, use of the measurement specifications, and issues related to bias and sensitivity. Experienced Pearson assessment specialists, copy editors, and research librarians led the training and provided specific and evaluative feedback to participants throughout the meetings and process of passage writing.

Passage writers and searchers were asked to submit both literary and nonfiction passages that were between 650 and 1100 words as defined by the specifications. These passages needed to support the development of at least 12 items from a variety of standards in the four reporting categories assessed on ISTEP+ E/LA. In addition, the passages needed to be free of grammar errors, allow for inferences, and be accessible and interesting for students.

Passage assignments included a range of commissioned, permissioned, and public domain texts, with no more than 60 percent of the passages requiring permissions. Commissioned texts are those where writers are hired to provide a specific type of text. Permissioned texts are those that are found by content experts that meet the passage specifications and whose authors are requested to allow the text to appear on an assessment. Public domain texts are those that are freely available in the public domain that do not require permission for use. Measurement Incorporated assessment specialists performed the initial review and revision of the passages selected and developed for ISTEP+ E/LA. Additional Measurement Incorporated staff, including editors, universal design specialists, and research librarians, reviewed the passages to ensure the quality of the passages being proposed for use. Pearson assessment specialists then reviewed the passages. Approved passages were provided to the IDOE assessment specialists for their review. The IDOE staff determined which passages would receive further review by Indiana educators at the passage content and fairness review meetings. The criteria for passage selection is outlined in the ISTEP+ Passage Specifications, which can be found on the IDOE website. <https://www.doe.in.gov/sites/default/files/assessment/istep-ela-passage-specifications-2017-final.pdf>

Passage Review and Selection

A group of educators, some who were invited for their content expertise and others for their fairness and sensitivity expertise, were convened to review the ISTEP+ E/LA passages approved by the IDOE. Members were selected from a pool of applicants provided by the IDOE staff to represent the diversity of students taking the assessment.

Seven Indiana educators participated in the grade 10 passage content review June 12-13, 2018. The meetings started with IDOE staff and Pearson facilitators providing a brief overview of the test development process and specific training regarding the feedback being sought from the educators. Educators were asked to evaluate the content of the passages to determine which ones should be eligible for use and to offer guidance to the IDOE and Pearson for item development based on specific passage content. Educators evaluated the passages on the following criteria:

- logical and consistent organization;
- appropriate vocabulary;
- consistent story line with a clear resolution;
- correct grammar and mechanics;
- content that allows students to make inferences and supports assessment of multiple standards;
- content that supports the development of at least 10 items; and
- text that appeals to a broad audience.

Educators were asked to identify any potential bias and sensitivity concerns in the passages, including, but not limited to, the following sources:

- controversial, inflammatory, or insensitive content;
- regional and geographic differences in language usage and topic familiarity;
- gender, culture, and race stereotypes;
- socio-economic differences and stereotypes; and
- accessibility concerns.

Pearson and Measurement Incorporated assessment specialists worked with the IDOE leadership to facilitate the committee member discussions and to ensure a fair review process. The members of the review committees made one of two decisions about each passage: accept or reject from further consideration. Of the 17 passages presented to the committees, 15 passages were accepted, and two passages were rejected.

After the committee reviews, Pearson assessment specialists received feedback from the IDOE as to which passages to move forward for item development. Items were developed for the recommended fifteen accepted passages.

Passage and Item Specifications

Documentation of item specifications include information about the assessed standards and some limited clarifications that assist with understanding the content or nuances of the content. Item specifications also provide guidelines for the level of cognitive complexity, expressed as Webb's Depth of Knowledge categories, for which items can be written and the types of items that can be used to assess the content (e.g., multiple-choice, gridded, etc.). Item specifications also prioritize certain content standards, indicating those where more emphasis is given on the test. The item specifications provide restrictions and requirements for item context and

vocabulary available for use in items. Item specifications were strictly followed by item writers to ensure accurate measurement of the intended knowledge and skills. Passage and item specifications were provided to the content and fairness teacher panels to verify.

The item specifications documentation that was carried over from the previous vendor was reviewed and approved by IDOE in collaboration with Pearson assessment specialists. These specifications were used by Pearson to evaluate the development of individual items that addressed a specific content or skill in the standards at a specific level of cognitive complexity. , It was determined that additional detail and clarity needed to be added to the E/LA passage specifications, and grade 10 E/LA and mathematics item specifications. The item specifications were updated to provide some sample stems and items, to include depth of knowledge (DOK)limits, and to add content limits. Both the passage and item specifications were released on the IDOE website and were used up to and including the 2019 development.

Item Writing and Review

Indiana’s testing contractor, Pearson, used subcontractors for EL/A and mathematics development who have extensive experience developing items for standardized achievement tests. The subcontractors went through rigorous training led by Pearson assessment specialists. The training included an overview of the test-development process and the learning outcomes and measurement specifications for each reporting category that the items were intended to assess. The assessment specialists articulated the best practices in the creation of high-quality multiple-choice items that are free of bias and sensitivity concerns. Additionally, prospective item writers were presented with the scope of the testing program, cautioned about security issues, and trained to use Pearson’s item development platform.

Item writing assignments included passage-based items, stand-alone items, selected-response items, constructed-response items, and technology-enhanced items. Passage-based items are referring to E/LA items where a reading passage is required in order to provide a response to a question as opposed to a stand-alone item that does not require or include a passage. ISTEP+ items are written to various levels of complexity, including depth of knowledge (DOK) levels 1-4, depending on the content area. Measurement Incorporated performed the initial review and revision of the items developed for ISTEP+. Pearson then reviewed a portion of each batch of items prior to submission to IDOE. Additional Measurement Incorporated staff, including editors and universal design specialists, reviewed the items to ensure the quality of the items. Approved items were provided to IDOE assessment specialists for their review. IDOE staff then reviewed and provided item-level feedback. Feedback was incorporated prior to taking items to committee reviews.

Groups of Indiana educators, some who were invited for their content expertise and others for their fairness and accessibility expertise, were convened to review the items. Members were selected from a pool of applicants by IDOE staff to represent the diversity of students taking the assessment. IDOE staff and Pearson facilitators provided a brief overview of the test development process and specific training regarding the feedback being sought from the

educators. Educators were asked to evaluate the content of the items to determine which ones should be eligible for use on ISTEP+. Educators evaluated the items for the following criteria:

- appropriately measures the intended standard;
- contains precise language;
- assesses unique ideas from the passage;
- is appropriately difficult for the grade level of the examinees; and
- contains plausible distractors but only one correct answer.

In addition to content reviews, the IDOE engaged educators to review items for potential fairness and accessibility issues prior to placing items on operational test forms. Pearson facilitators and members of the IDOE staff provided a brief overview of the test development process and provided specific training regarding the feedback being sought from the educators. Educators were asked to identify any potential concerns in the items including, but not limited to, the following sources:

- controversial, inflammatory, or insensitive content;
- regional and geographic differences in language usage and topic familiarity;
- gender, culture, and race stereotypes; and
- socioeconomic differences and stereotypes.

The Pearson assessment specialists and universal design experts worked with IDOE leadership to facilitate the committee member discussions according to quality criteria. The members of the review committees made one of three decisions about each item: accept, accept with revisions, or recommend removal from further consideration. The complete quality criteria and universal design checklists can be found in Appendix M.

Meetings to review these items were held in July and August of 2018. Seven educators participated in the mathematics review, and 10 educators in the E/LA. In E/LA, 148 items were reviewed 2 rejected from the item bank. In MA, 117 items were reviewed with 0 rejected.

Critical Elements: 1.1, 2.1, 2.2, 3.1, 4.2, 4.4, 4.5, 4.6, 5.3

Standards: 1.1, 1.11, 3.1, 3.2, 3.5, 3.11, 4.0, 4.1, 4.2, 4.6, 4.7, 4.8, 4.10, 4.12, 4.16, 4.18, 7.4, 7.5, 12.4, 12.8

Chapter 5: Forms Construction

The forms construction phase of developing the ISTEP+ assessments included several activities designed to ensure the production of high-quality assessment instruments that accurately measure the achievement of students with respect to the knowledge, skills, and abilities contained within the Indiana Academic Standards.

Test Design

ISTEP+ assessments for E/LA and MA winter 2018 retest, spring 2019 retest, spring 2019 first time assessment (FTA), and summer 2019 retest grade 10 were established assessments pulled from the existing item bank. To prepare the item bank for future administrations, additional items needed to be field tested.

As mentioned previously, the ISTEP+ assessments are structured into two parts. In 2016, each part had its own booklet or online test session composed of the operational items and another booklet or online test session composed of the field test items. Starting with the spring 2017 administration, each part had its own book, which included both operational items and embedded field test items. Since the field test items were embedded within the test book, whether the items were operational or field test, was not explicitly known by test administrators or students. The embedding of field test items within the operational administration was done to counter the perception that the awareness of operational versus field test items may impact student motivation.

For field testing, multiple forms of grade- and subject-specific assessments were created for the spring 2019 FTA. The number of forms varied across tests from one to sixteen forms, displayed in table 3.

Table 3. Administration of ISTEP+ Spring 2019 Forms and Modes

Administration	E/LA				Mathematics			
	Part 1		Part 2		Part 1		Part 2	
	O	P	O	P	O	P	O	P
Spring FTA	9	1	10	1	16	1	16	1

Field test items from the spring 2018 administration, with acceptable statistics, were added to the appropriate item bank following IDOE data review, for use in spring 2019 test construction. Breach forms were built for mathematics and E/LA, also without field test items. Retest forms were built without field test items, resulting in shorter forms.

Item Selection and Form Development

Various strategies used for item selection and form development for the 2018-2019 forms. Those strategies are outlined below.

English/Language Arts and Mathematics Grade 10

Winter 2018 had a 100% refresh for Part 1 items. The goal for part 2 was a 10-25% refresh. The winter retest did not include any field test items. A new breach form was created. The breach forms were based on the Summer 2018 forms, with a goal of being as different from the operational forms that would be administered in 2018-19 as possible based on the item bank available. The breach forms were designed to be used as a full form, and Part 1 of the breach could not be matched with Part 2 of the winter retest, for example. The breach forms were available for the first time on paper in winter of 2018. The breach form did not include any field test items.

Spring 2019 retest was a reuse of the winter 2017 form. The spring 2019 FTA used the winter 2018 form as the base. Part 1 was refreshed at 100% and the Part 2 was refreshed at 55% of the items for E/LA and 45% for mathematics.

The summer 2019 retest used the spring 2019 FTA as a base. Part 1 was refreshed at 100% and Part 2 at 35% for E/LA and 61% for mathematics.

Test Construction and Pre-Equating

Pre-administration equating occurs as part of the test development process and supports the construction of new forms that are similar in difficulty and psychometric characteristics to those previously administered. All retest forms were pre-equated, and the FTA was post-equated.

The pre-equating process links each item on a newly constructed test form to one or more previously used test forms. In this way, the difficulty level of the newly developed test forms can be placed onto the same base (operational) scale as previously administered forms and compared to those previously administered forms. In the pre-equating stage, the IRT parameter estimates associated with a proposed set of operational items are used to estimate the overall difficulty and precision of the newly developed form. For items that have not yet been administered operationally, the field-test item parameters are used for the pre-equating process. These values represent the best guess as to how the items will perform when administered to the total-testing population.

Critical Elements: 1.1, 2.1, 2.2, 3.1, 4.2, 4.4, 4.5, 4.6, 5.3

Standards: 1.1, 1.11, 3.1, 3.2, 3.5, 4.0, 4.1, 4.2, 4.4, 4.6-4.10, 4.12, 4.16, 4.18, 7.2, 7.4, 7.5, 12.3, 12.4, 12.8

Chapter 6: Administration

Indiana students in grade 10 were administered ISTEP+ assessments in E/LA and MA. The first time administration assessments were given in two separate windows. The Part 1 and Part 2 assessments were administered during the same online or paper-and-pencil window for the retest administrations.

Administration	Online	Paper-and-Pencil
Winter 2018 Retest (Parts 1 and 2)	11/12 – 12/11/18	11/19 – 12/7/18
Spring 2019 Retest (Parts 1 and 2)	2/11 – 3/8/19	2/18 – 3/1/19
Spring 2019 FTA Part 1	2/25 – 3/22/19	2/25 – 3/13/19
Spring 2019 FTA Part 2	4/8 – 5/3/19	4/15 – 5/1/19

It should be noted that there was one operational form for each grade and content area. While Part 1 contains applied skills items that are hand-scored, Part 2 includes multiple-choice, gridded response, and technology-enhanced items that are machine-scored. Both parts for each assessment were offered in paper or online.

The ISTEP+ assessments are not intended to be timed or speeded tests. However, students were given logical time limits for the administration to occur within a one-day period. The time limits, approved by the IDOE, were derived from previous administrations as a baseline. The final time allotment took into account the characteristics of the population, estimated time needed per item, and past experience from previous administrations. Information regarding the suggested testing times and structure of the parts and sections are summarized by subject in Appendix B. The tables in Appendix B were taken from the spring 2018 Test Coordinator's Manual and show time allotted to practice items and instructions as well as actual test working time.

Participation Requirements

According to federal regulations, all students must be assessed on grade-level content standards. Consequently, all students are expected to participate in the ISTEP+ assessment program. Prior to the 2018-2019 administrations, school corporations and schools were provided information about the tests. This information is updated annually and available at <http://www.doe.in.gov/assessment>. Test administrators were also provided a test administration manual for the subject-area tests that included logistical guidelines, sample items, scoring rubrics, guidance for student participation, and accommodations guidelines for special populations.

Special Populations

Assessment accommodations are changes in testing materials or procedures that enable students to participate in assessments in a way that assess abilities rather than disabilities. Assessment accommodations are generally grouped into the following categories.

- Presentation (e.g., repeat directions, read aloud, large print braille)
- Equipment and materials (e.g., calculator, amplification equipment, manipulatives)
- Response (e.g., scribe records response, point)
- Setting (e.g., separate room, study carrel, student's home)
- Timing/Scheduling (e.g., extended time, frequent breaks)

A Case Conference Committee (CCC) is used to make assessment and accommodation decisions for students with disabilities based on individual student need in accordance with state and federal guidelines. Students with disabilities must be provided accommodations based on individual need. The IDOE makes determinations as to whether accommodations threaten the validity of test scores.

Information about the accommodations provided to students administered the ISTEP+ assessments may be found in various locations including:

- the Accessibility and Accommodations for Instruction and Assessment Resource Guide and Toolkit at <https://www.doe.in.gov/sites/default/files/specialed/accessibility-and-accommodations-instruction-and-assessment-resource.pdf>;
- Appendix C of the Indiana Assessment Program Manual at <https://www.doe.in.gov/sites/default/files/assessment/2018-19-program-manual-7-18-18-final.pdf>; and
- guidance regarding appropriate testing format for students – either computer-based or paper-and-pencil – at <https://www.doe.in.gov/sites/default/files/specialed/patins-computer-pencil-letter-checklists.pdf>.

Special forms were provided for paper for students in need of large print, Braille, or read-aloud accommodations. For online administrations, special forms were provided for audio administrations, calculator usage, and audio/calculator. Online test takers could also use color contrast, or the browser zoom feature to enlarge text.

Critical Elements: 2.3, 4.2, 5.1, 5.3

Standards: 3.1, 3.4, 3.5, 3.9, 3.10, 4.2, 4.15, 4.16, 6.0-6.7, 7.7-7.9, 7.13, 12.9

Chapter 7: Scoring

The 2018-2019 ISTEP+ operational assessments included items that were machine-scored and items that were scored by trained human scorers (called handscorers). Multiple-choice (MC), gridded (GR), and technology-enhanced (TE) items were machine scored. Open-ended (OE) items, including constructed-response (CR), extended-response (ER), and writing prompts (WP), were handscored.

ISTEP+ assessments are administered online and on paper. For items that are machine-scored, regardless of mode, the scoring mechanism is the PearsonAccess^{next} online platform. Paper and pencil items are scanned and are then transferred into the same system that is used to capture online responses (i.e. PearsonAccess^{next}). Items are scored dichotomously and polytomously. Items on which students responded with multiple marks or that were missing, or left blank were treated as incorrect.

Scoring Rubrics for Open-Ended Items

The three types of OE, or handscored, items were administered during Part 1 of the ISTEP+ assessments. The E/LA assessments are the only assessments with WPs. Each item was scored using a holistic rubric. The rubrics were developed by the IDOE. For some items, a student's single response was scored using two different rubrics in which case the two scores are combined for the total item score. Specifically, E/LA WPs are scored using two rubrics -- once using a 6-point rubric related to writing applications and once using a 4-point rubric related to language conventions. Similarly, MA CRs are scored using two rubrics -- once for content and once for process.

Although rubrics for similar item types share some characteristics, handscoring materials that guide the training and scoring of every item are specific to the items. Anchor papers, training papers, qualification sets, and validity papers are developed and used to ensure specificity, reliability, and validity in scores. Scoring rubrics for OE items can be found by grade and subject, along with sample responses and scoring notes, on the IDOE's website at:

<https://www.doe.in.gov/assessment/istep-sample-items-and-scoring>

Anchor Papers

Anchor papers are actual students' responses that exemplify the most common responses for each score point in an OE item. Anchor papers are selected during a process called range-finding. IDOE engaged with educators beginning in 2018 to review students' responses as part of a range-finding process. The range-finding process included training and collaborative review of field test items with IDOE, Indiana educators, Pearson assessment specialists, and Pearson scoring directors. These staff met to review the proposed anchor papers along with their applicable scoring rubrics to select the best set of responses to use in training handscorers. The anchor papers were selected from student responses at each score point across subject areas,

reviewed by IDOE staff, Indiana educators, the Pearson handscoring team and approved accordingly.

Responses that clearly represented a score point as well as those that were on the line between score points were discussed. Based on these discussions, any clarifications or annotations that could be made to rubrics or training materials were documented to make explicit why a response earns a specific score point. The IDOE releases materials related to OE items, including anchor papers, on an annual basis. These can be found by grade and subject at <https://www.doe.in.gov/assessment/istep-sample-items-and-scoring> under the heading ISTEP+ Sample Items and Scoring.

Recruiting of Handscorers

Pearson worked with other agencies to recruit, interview, and select highly qualified handscorers. IDOE requires that a percentage of its contracts be expended on minority-owned and women-owned businesses. Businesses fitting those criteria were contracted assist with recruiting of handscorers. Pearson requires that all handscorers and supervisors possess at least a bachelor's degree and they must complete a screening interview. Pearson will initially recruit individuals with previous experience scoring similar assessments. Each potential handscorer completes a pre-interview activity where he or she is introduced to the process of scoring with examples. The applicant's trainability and ability to understand and implement the standards set forth in the sample scoring guide are key determinants in being approved as a handscorer.

Pearson has ready access to well-qualified scoring staff. Scorer trainees who fail to meet our training and qualifying requirements are dismissed from the project. After being hired, scorers may also be dismissed if their scoring performance does not continuously meet Pearson's validity and/ or inter-rater agreement standards.

Training and Qualification of Handscorers

Prior to scoring, Pearson developed and the IDOE approved various scoring materials to train and qualify handscorers. These scoring materials included anchor, practice and qualification papers. In addition to the previously described anchor papers, Pearson selected and annotated two sets of 10 papers for each OE item referred to as practice sets. Each practice paper, its applicable score point, and annotation were approved by the IDOE. Additionally, Pearson selected and the IDOE approved two sets of 10 qualification sets for each OE item. Handscorers were trained by studying/reviewing the anchor papers for their assigned item, reviewing both sets of 10 practice papers, and meeting or exceeding the minimum percentage of exact or adjacent agreement required for the two qualification sets.

Scorers were required to meet the qualification criteria for their content area on at least 1 of 2 qualification sets. Requirements listed in the following chart.

Content Area	Exact Agreement	Adjacent Agreement
Writing (for each trait)	70%	90%
Mathematics (for each trait)	70%	90%
E/LA	80%	90%

Materials used for checking the reliability of handscorers during live scoring (i.e., “validity papers”) were also identified by Pearson and approved by IDOE prior to and throughout live scoring as needed.

The spring 2019 FTA contained one mathematics item serving as an operational field test and therefore, that item had no field test responses nor field test training materials. Consequently, scorers for this Operational Field Test item were trained and qualified on a similar item (a baseline item) and then moved to the Operational Field Test item. Scorers on this item began training approximately a week after the other scorers on this administration. During that week, a Pearson scoring director reviewed “live” student responses and built a draft anchor set, which went through the rangefinding process with Indiana educators and IDOE content staff. The scoring director also built an annotated 10 paper bridge set that was approved by IDOE content staff. Scorers who had qualified on the baseline item then reviewed the Operational Field test item and its anchor and also took the bridge set before scoring student responses.

Handscoring Process and Validity

Handscorers were rigorously trained and had to meet qualifying requirements before being permitted to score. Even after qualification, handscorers are monitored daily to ensure integrity and consistency in scoring by making use of validity papers.

Validity papers are pre-scored papers not previously seen by scorers, which are distributed on a regular basis throughout a project to monitor consistency in scoring over time. Validity responses are interspersed with and are indistinguishable from other student responses. True scores for these papers are loaded into the system and a report is run that shows what percentage of accuracy a scorer has achieved in scoring against the true score on the validity papers. Validity papers are used as a check to ensure that scorers, as well as scoring supervisors, are not drifting from the training materials and are continuing to score in a way that is valid based on the rubrics and training materials.

Validity Standards

- Writing - 65% perfect agreement; 96% perfect plus adjacent agreement
- Reading - 80% perfect agreement; 96% perfect plus adjacent agreement
- Mathematics - 65% perfect agreement; 96% perfect plus adjacent agreement

In the event that a handscorer began to “drift” away from scoring papers accurately, that handscorer went through a recalibration process whereby they were required to review and

pass a specified set of papers to correct their scoring before being permitted to continue scoring.

To address IDOE's concerns regarding response mode, Pearson scoring directors created a bridge set consisting of 10 annotated responses for all items. For those items with anchor sets of handwritten responses, the bridge set consisted of 10 typed responses. For items with anchors sets of typed responses, the bridge set consisted of 10 handwritten responses.

For those items with training materials consisting of handwritten responses, scorers took the bridge set consisting of typed responses at the start of the scoring window—the online responses are the first responses scored, followed by hand written responses. For those items with training materials of typed (online) responses, the bridge set of handwritten responses was distributed later in the scoring window when handwritten responses entered the scoring pool.

Inter-rater Agreement

Inter-rater agreement describes how consistent or reliable handscorers are at providing the same ("perfect") score or adjacent scores across first and second readings of an OE item. To capture and ensure inter-rater agreement, approximately five percent of all papers were read twice by two different scorers. When scores between the first and second reads did not agree (or if they differed by more than one point), papers were read a third time and, if necessary, a fourth time. Handscorers provided most of the scores for the first and second reads.

Supervisors and Scoring Directors performed the third and fourth reads. Thus, for the ISTEP+, any item that required a second read was read repeatedly until the score was resolved by more experienced handscorers. The items were not given the mean of scores or the most frequent score, as is sometimes the case in other score resolution approaches. Instead, if the first and second score are adjacent, the higher of the two scores is the final score. If not, and the response goes to scoring resolution, the Supervisory staff provides the final score. If it ultimately ends up in Adjudication, the score resulting from that process is the final score.

Inter-rater Agreement Targets

- Writing - 65% perfect agreement; 95% perfect plus adjacent agreement
- Reading - 80% perfect agreement; 96% perfect plus adjacent agreement
- Mathematics - 65% perfect agreement; 95% perfect plus adjacent agreement

Valid and Invalid Test Attempts

Validation rules for the ISTEP+ assessments were applied to the 2018-2019 administrations. A test session could be invalidated if a student did any of the following:

- Worked in a section other than the one being administered,
- Cheated,
- Marked most or all answers randomly,
- Left the section completely blank, or
- Lost a significant amount of time during the test session.

Invalid test attempts are determined by individual test examiners and reported to the principal or test coordinator. If a student had an invalid test attempt, it was not used in item or test-level analyses. Alternatively, a valid test attempt for either part would be defined by a single response to an item in a section, excluding E/LA Part 1 Section 2, which contained all field test items and was excluded from attemptedness determination.

Attemptedness rules apply to both parts for a determination of attempted or not. It should be noted that HumRRO replicated Pearson's attemptedness rules for the winter 2018 and spring 2019 administrations, and all attemptedness issues were resolved with the close guidance of IDOE.

Chapter 8: Students

The operational items on the ISTEP+ assessments were administered to all eligible students in the appropriate grade level during the spring 2019 first time administration window.

Spring 2019 FTA operational score results were reviewed based on student characteristics, such as gender, ethnicity, disability status, socioeconomic status, and English language learner status. A student's disability status is defined by whether he/she is receiving special education services (SPED). A student's socioeconomic status was classified into two groups, low and high. Appendix I shows the proportion of students in each subgroup who took the ISTEP+ assessments for the first time in spring 2019. The student represented in the tables in Appendix I are only a sample of the total public-school students that took the ISTEP+ assessments during the spring 2019 first time administration and received scores. The students in these samples were used in the analysis described in Chapter 9: Equating and Item Analysis, Chapter 10: Reliability, and the appendices.

It should be noted that both winter 2018 retest and spring 2019 retest results can be found in the same appendix.

Critical Elements: 3.1, 4.1-4.7

Standards: 1.8, 1.10, 2.19, 4.9, 4.10

Chapter 9: Equating and Item Analysis

In order to maintain the same performance standards across different administrations of a particular test and different forms within the same administration, a statistical procedure called equating is employed. Equating is used to transform the scores of one administration or forms of a test to the same scale as the scores of a second administration or form of the test. For the ISTEP+ spring 2019 first time administration, a post-equating process was used to ensure the IRT parameters for all items were on the same scale.

The statistical analyses of spring 2019 first time administration were conducted only with the first-time testers. However, it should be noted that students' scale scores for both winter 2018 and spring 2019 retest administrations were generated by pre-equating method. Therefore, the statistical results of both administrations (i.e., winter 2018 and spring 2019 retest) can be found in the appendices.

Spring 2019 Operational Post-Equating

Spring 2019 first time administration was a continuation of previous administrations of the assessments. Therefore, a post-equating process was used to maintain the scale and performance standards from the previous administrations. The operational post-equating process used for grade 10 E/LA and Mathematics assessments was completed using the following steps.

- Free calibration of operational items
- Stability check and creation of equating constants using anchor items
- Creation of final item parameters for scoring purposes.

Grade 10 E/LA and Mathematics Assessments

A subset of the operational items was identified to serve as anchors for post-equating the spring 2019 first time administration such that items in spring 2019 would be on the same scale as the previous administration. The following guidelines were used to identify the anchor sets:

- Only items from Part 2 of the tests were used (i.e., items that were machine-scored)
- Items were treated the same in scoring between administrations (e.g., if the item was a multi-part item and scored polytomously in both administrations)
- Blueprint representation

Because the ISTEP+ assessments have a large number of items reused annually, the number of anchor items for E/LA and Mathematics is 15 and 18 respectively, and detailed information can be found in Appendix E.

Calibration

The commercial software IRTPro (Scientific Software International, Inc., 2011) was used for item calibration at each stage in the process. All multiple-choice items were calibrated using the

three-parameter logistic model (3PL), defined by three parameters: item difficulty or location (b_j), item discrimination (a_j), and the pseudo-guessing parameter (c_j). The 3PL model estimates the probability that a student with an ability estimate θ responds correctly to item j is:

$$P_j(\theta) = c_j + (1 - c_j) \times \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}}$$

For any polytomous items, the generalized partial credit (GPC) model was used, defining the following parameters: item difficulty or location (b_j), item discrimination (a_j), and category step difficulty parameters (d_{iv}). The GPC model estimates the probability that a student with an ability estimate θ earns a score of k to item j is:

$$P_{im}(\theta_j) = \frac{\exp \left[\sum_{k=0}^m Da_i(\theta_j - b_i + d_{ik}) \right]}{\sum_{v=0}^{M_i-1} \exp \left[\sum_{k=0}^v Da_i(\theta_j - b_i + d_{iv}) \right]},$$

For the free calibration, an incomplete data matrix was created that contained student responses for all items regardless of testing mode and were calibrated simultaneously using a concurrent calibration (Kolen & Brennan, 2004; Lord & Wingersky, 1984). Items were calibrated such that a single set of item parameters was obtained for common items. As described previously in Chapter 7, common items which demonstrated mode DIF were calibrated separately, so there were separate item parameters for each mode. For the anchored calibrations of all subjects, operational and field test items were calibrated simultaneously, with the item parameters for the operational items used as the anchors.

Once the item calibration was completed, additional analyses were completed to evaluate the items against additional criteria, including item-model fit and item level analyses including examination of differential item functioning (DIF). For each item, a graph showing the model-based item characteristic curve along with the empirical performance of the students on the item was created to evaluate the fit of the data to the IRT model.

Spring 2019 First Time Administration Operational and Field Test Item Analysis

Additional item-level analyses were completed for operational and field test items. For selected-response items a key check analysis is performed. For selected-response items, the key check analysis flagged items where:

- N-count < 200
- P-value ≤ 0.20
- Item-total correlation < 0.20
- Distractor selected by 40% or more examinees
- Distractor item-total correlation ≥ 0.05

An adjudication process is employed for gridded and technology-enhanced items. This process involves a review of every student response provided to these items and its scoring resolution (i.e., correct or incorrect) to ensure that all possible correct responses are being scored as such. This prevents errors in scoring based on unexpected or creative response formats provided by students. Adjudication reports with all possible response and their score are provided to the Pearson assessment specialists for review. If there are uncertainties about the scoring rule associated with a given response, the Pearson assessment specialist consulted with the IDOE staff for a final determination. It should be noted that HumRRO replicated Pearson's operational analyses for spring 2019 first time administration, and all scoring issues were resolved with guidance of IDOE.

DIF analyses were performed for each item to examine if some trait unrelated to student proficiency accounted for students' item performance. It is historically presented as performance of a focal group who may be a minority or under-served, relative to a reference group who is presumed to be a member of the majority group or are not under-served. The analyses described below are statistical methods for flagging DIF. However, a statistical flag does not dictate that an item exhibits bias in favor of one group or another. A statistical DIF flag requires further investigation by content experts.

Where sufficient data were available, three DIF comparisons were made for each item:

Reference Group	Focal Group
Male	Female
White	African American
White	Hispanic

The Mantel-Haenszel Delta DIF (MHD; Dorans, & Holland, 1992) statistic was used for flagging possible DIF. It is based on the traditional Mantel-Haenszel (Mantel, 1963) statistic, but transformed to the ETS delta scale. The three categorical labels applied to possible DIF judgments correspond to negligible (A), moderate (B), or severe (C). The number of operational

items administered with B or C DIF flags across ISTEP+ assessments can be found in Appendix D: ISTEP+ DIF Flags.

When items displayed extreme C DIF favoring either the reference or focal group, those items were forwarded to Pearson content specialists for further review to determine if issues of sensitivity or bias should prevent the item from contributing to a student's score. None of the flagged items were considered to have issues of sensitivity or bias and were deemed to be construct-relevant to measurement of the content domain for the specific tests.

The following software packages were used during the equating process.

- IRT calibration and ability estimates for fit: IRTPro
- Creation of scale scores using pattern scoring: ISEV

Scaling

Scaling is the process where raw scores are converted to scale scores. For the ISTEP+ assessments, a common method called pattern scoring is used to transform student raw scores into scale scores.

The method of scaling referred to as pattern scoring takes the pattern of correct and incorrect responses into account in derivation of a students' scale score. In fact, pattern scoring takes into account the pattern of student responses, as well as characteristics of the items themselves.

Pattern scoring can be contrasted with scaling that relies solely on the number of items answered correctly. In a method of number correct scoring, any student receiving a particular raw score would obtain the same scale score regardless of which items they answered correctly. So, a student obtaining a raw score of 40 by answering the easiest 40 questions would obtain the same scale as a student answering the 40 most difficult questions correctly. In contrast, pattern scoring would result in these two students obtaining different scale scores because item parameters (e.g., discrimination parameter of an item) of the items a student answers correctly are taken into account for the purpose of scoring. Pattern scoring is thought to provide a more precise estimate of student ability than the method of number correct scoring.

The software package Operational Scoring: IRT Score Estimation (ISE V1.3.f; Chien & Shin, 2012) is used to perform the pattern scoring process and provide student scores on the scale score metric, using the student scored responses and the item response theory (IRT) item parameters for the operational items on the scale score metric.

It is noteworthy to mention that ISE is a computer program written in C++ language that can estimate IRT pattern scores for the 1-, 2-, and 3-parameter logistic IRT models for dichotomous items, and the generalized partial credit model for polytomous items. The item parameters resulting from the equating process, described in Chapter 9, are transformed from the theta/logit metric, which are provided by IRTPro, to the scale score metric. This transformation is useful for two primary reasons – (1) to put the scores onto a metric that is useful/useable to stakeholders

and (2) to maintain a score scale that is comparable year over year. A linear transformation is performed on each item parameter using the following formulas:

$$a\text{-parameter:} \quad a_{scale} = \frac{a_{base}}{r}$$

$$b\text{-parameter:} \quad b_{scale} = r \times b_{base} + l$$

$$c\text{-parameter:} \quad c_{scale} = c_{base}$$

$$d\text{-parameter:} \quad d_{i-scale} = r \times d_{i-base}$$

The variables a_{base} , b_{base} , c_{base} , and d_{i-base} are the IRT item parameters on the theta metric. The variables a_{scale} , b_{scale} , c_{scale} , and $d_{i-scale}$ are the IRT item parameters on the scale score metric. The variables r and l are the scaling slope and intercept values presented in Appendix F.

The IDOE also uses a common practice of setting upper and lower boundaries for the scale score metric. Without these boundaries, the transformation of student responses that are aberrant might result in scale scores that are drastically different (e.g., extremely high or extremely low) from the majority of the other scale scores. Therefore, many testing programs assign a reasonable lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS). The LOSS and HOSS were determined for grades 10 ISTEP+ mathematics and E/LA assessments in the first year the scale was established, and that rationale can be found in the 2016 ISTEP+ Technical Report. The ISTEP+ LOSS and HOSS values for the ISTEP+ grade 10 assessments are shown in Appendix F. It should be noted that differential item functioning (DIF) analysis on test mode comparability was not conducted for 2019 assessment because students who took a paper test were less than 10% of the whole IN population.

Critical Elements: 3.2, 3.4, 4.1, 4.4

Standards: 1.9, 1.10, 1.12, 2.2, 2.3, 2.4, 2.5, 2.7, 2.13, 2.19, 3.4, 3.8, 4.8, 4.10, 4.18, 4.20, 4.21, 5.2, 5.6, 5.16, 6.4, 6.5, 6.8, 6.9, 7.7, 7.8, 7.10, 7.11, 12.16

ISTEP+ Spring 2019 Field Test Item Equating

The process used to complete the field test item equating is an anchored item equating process. In this process the item parameters from the operational items from the 2019 administration were used to calculate equating constants (i.e., to calculate Stocking-Lord equating constant) and the item parameters for the field test items were freely calibrated, placing the item parameters for the field test items on the same scale as the operational items.

As mentioned previously, field test items are reviewed for all the same criteria as outlined previously. The result of such reviews is to determine if items are eligible to be placed in the item bank for future test construction or if items need to be updated and field tested again.

Critical Elements: 3.1, 4.3-4.6

Standards: 4.10, 5.1, 5.2, 5.6, 5.12, 5.13, 5.14, 5.15, 5.16, 6.9

Chapter 10: Achievement Standards

Test scores in and of themselves do not imply student competence. Rather, the interpretation of test scores permits inferences about student competence. In order to make valid interpretations, a process of evaluating expected and actual student performance on assessments must be completed. This evaluation of expected versus actual performance is typically referred to as standard setting. Academic achievement standards are set to identify the level of performance students need to demonstrate on the ISTEP+ assessments to be classified into defined achievement levels.

It should be noted that in spring 2016, IDOE worked with Pearson to conduct a standard-setting meeting to establish achievement standards for grade 10 ISTEP+ assessments, which were also maintained through equating in spring 2019. The ISTEP+ cut scores are presented in Appendix F.

Critical Elements: 6.1-6.3

Standards: 2.14, 5.21, 5.22, 5.23

Chapter 11: Reliability

Reliability refers to the expectation that repeated administrations of the same test should generate consistent results. Reliability is a critical technical characteristic of any measurement instrument because unreliable scores cannot be interpreted as valid indicators of students' knowledge and skills. For the spring 2018 administration, reliability for ISTEP+ was estimated using statistical measures such as internal consistency, classical standard error of measurement, conditional standard error of measurement, and classification accuracy.

Internal Consistency

Internal consistency is a measure of the consistency with which students respond to items within a test. ISTEP+ contains items that are dichotomously and polytomously scored; therefore, Cronbach's alpha was used to estimate reliability. The formula for calculating coefficient alpha is:

$$\alpha = \left(\frac{N}{N-1} \right) \times \left(1 - \frac{\sum_{i=1}^N S_{Y_i}^2}{S_X^2} \right)$$

Where N is the number of items on the test, $S_{Y_i}^2$ is the sample variance of the i^{th} item and S_X^2 is the observed score sample variance for the test. As a general rule, reliability coefficients ranging from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are considered excellent (Nunnally & Bernstein, 1994).

Because internal consistency estimates typically decrease as the number of test items decrease, internal consistency estimates for individual reporting categories can be noticeably lower than those for the full assessment.

In spring 2019, the internal consistency estimates for total score ranged between 0.91 for E/LA, and between 0.94 for Mathematics. As expected, however, the estimates for each strand score were noticeably lower. It should be noted that those of winter 2018 retest and spring 2019 retest are slight lower than those of spring 2018. Coefficient alpha for the overall test and by reporting category and subgroup can be found in Appendix I.

Classical Standard Error of Measurement

The classical standard error of measurement (SEM) represents the amount of variance in a score that results from random factors other than what the assessment is intended to measure. Because underlying traits such as academic achievement cannot be measured with perfect precision, the SEM is used to quantify the margin of uncertainty in test scores. For example, factors such as chance error and differential testing conditions can cause a student's observed score (the score achieved on a test) to fluctuate above or below his or her true score (the student's expected score). The SEM is calculated using both the standard deviation and the reliability of test scores, as follows:

$$SEM = \sigma_x \sqrt{(1 - P'_{xx})}$$

Where P'_{xx} is the reliability estimate and σ_x is the standard deviation of raw scores on the test. A standard error provides some sense of the uncertainty or error in the estimate of the true score using the observed score. For example, suppose a student achieves a raw score of 50 on a test with an SEM of 3. Placing a one-SEM band around this student's score would result in a raw score range of 47 to 53. If the student took the test 100 times and 100 similar raw score ranges were computed, about 68 of those score ranges would include the student's true score.

It is important to note that the SEM provides an estimate of the average test score error for all students regardless of their individual proficiency levels. It is generally accepted that the SEM varies across the range of student proficiencies (Peterson, Kolen, & Hoover, 1989). For this reason, it is useful to report test-level SEM but also individual score-level estimates. Individual score-level estimates are commonly referred to as conditional SEMs.

SEM for E/LA and Mathematics were 3.48 and 3.59, respectively. More detailed results, including SEM by subgroup, are provided in Appendix I. It should be also noted that winter 2018 retest and spring 2019 retest results can be found in the same appendix.

Conditional Standard Error of Measurement

It is important to note that the SEM index provides only an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, conditional standard error of measurement (CSEM) provides a reliability estimate at each score point on a test. Like the SEM, the CSEM reflects the amount of variance in a score resulting from random factors other than what the assessment is designed to measure, but it provides an estimate conditional on proficiency. The CSEM is usually smallest, and thus scores are most reliable, near the middle of the score distribution. Typically, achievement tests included relatively large numbers of moderately difficult items. Because these items are usually well-matched to a majority of students' ability, they provide the most reliable estimates of ability. It is desirable for an achievement test where students are classified into pass/fail categories, that the CSEM be lowest at the cut score for passing.

IRT methods are used for estimating CSEM and are presented in graph form in Appendix G. Generally, the lowest point of the curve occurs at the location of the Pass and Pass+ cut score.

This is always the case for the Pass cut; however, the Pass+ cut does not always follow this trend. It should be also noted that winter 2018 and spring 2019 retest results can be found at the same appendix.

Classification Consistency and Accuracy

ISTEP+ scores are used to classify students into performance levels. For the vast majority of students, these classifications are accurate reflections of their performance. However, all test scores contain error, so some students might be misclassified. To better understand the expected degree of misclassification, an analysis of and accuracy of student classifications into performance levels was completed.

Classification consistency is defined as the extent to which two classifications of a single student agree from two independent administrations of the same test (or two parallel forms of the test). Classification accuracy is defined as the agreement between the classifications using observed cut scores and true classifications based on known true cut scores (Livingston & Lewis, 1995). Classification consistency refers to the agreement between two observed classifications results, while classification accuracy refers to the agreement between the observed classification outcome and the true classification result.

To represent classification consistency, a contingency table with the three classifications for ISTEP+ can be created.

	Did Not Pass	Pass	Pass+	Sum
Did Not Pass	P_{11}	P_{21}	P_{31}	P_{-1}
Pass	P_{12}	P_{22}	P_{32}	P_{-2}
Pass+	P_{13}	P_{23}	P_{33}	P_{-3}
Sum	P_{1-}	P_{2-}	P_{3-}	1.0

The procedure for calculating classification consistency was Cohen's kappa (1960), which is recommended by Swaminathan, Hambleton, and Algina (1974). The formula for Cohen's kappa is:

$$kappa = \frac{P - P_c}{1 - P_c},$$

where P is defined as the sum of the diagonal values of the contingency table, representing the proportion of events where both classifications matched, and P_c is the chance probability of a consistent classification under two completely random assignments. The chance probability P_c is the probability obtained by multiplying the marginal probability of the first event and the corresponding marginal of the second administration:

$$P_c = (P_{-1} \times P_{1-}) + (P_{-2} \times P_{2-}).$$

A simulation procedure (Kim, Kim, & Barton, 2007) was used for estimating classification consistency and accuracy, which involves the generation of item responses using item parameters based on IRT models. Using the examinee's ability, selected from the ability distribution from a single administration of the test, two sets of item responses are generated using a set of item parameters. These two sets of item responses are considered as an examinee's responses on two administrations of the same form.

Appendix H presents the classification consistency and accuracy values for the ISTEP+ 2019 Spring 2019 administration as well as the two other administrations (i.e., winter 2018 and spring 2019 retests). The values of the classification consistency and accuracy depends on several different factors, such as the reliability of the actual test form, the distribution of scores, the number of cut scores, and the location of each cut score. The classification consistency calculated using Cohen's kappa (kappa) represents the agreement of the classification between the two parallel forms with the consideration of the probability of a correct classification by chance.

The classification consistency calculated using Cohen's kappa has a range of values across the mode combinations and grade-level. Generally, Cohen's kappa values for Mathematics test are in general higher than those for E/LA test. This is likely due to the nature of E/LA having more items requiring human-scoring as opposed to machine scoring. Classification consistency and accuracy should be considered together. The classification accuracy represents the agreement between the observed classification based on the actual test form and the true classification given the modeled form. In general, accuracy values of Mathematics are higher than those of E/LA. Winter 2018 and spring 2019 retests can be found in Appendix H.

Critical Elements: 4.1

Standards: 2.11, 2.13, 2.14, 2.15, 2.16, 2.19, 4.10, 5.6, 6.9, 7.4

Chapter 12: Reporting

After the administration, several informative and user-friendly reports are created, filtered through various quality checks, and made accessible to various stakeholders. This chapter provides a short overview of the report development process, the types of scores reported, and a brief description of each type of report.

One of the goals of the ISTEP+ assessments is to ensure that all students are proficient before moving on to the next grade. As such, it is important that any reports to parents and/or teachers clearly characterize a student's proficiency status. ISTEP+ student score reports show both the student's individual scale score, the passing scale score, and the student's proficiency status – Did Not Pass, Pass, or Pass+. The excerpts below are from the interpretive guide that goes along with the various reports. The interpretive guide details the types of reports provided and defines the information that is presented within them. The interpretive guide is made available to stakeholders in various ways including listserv emails as well as posting on the Indiana Resource Center website at: <http://indiana.pearsonaccessnext.com/reporting-resources/>.

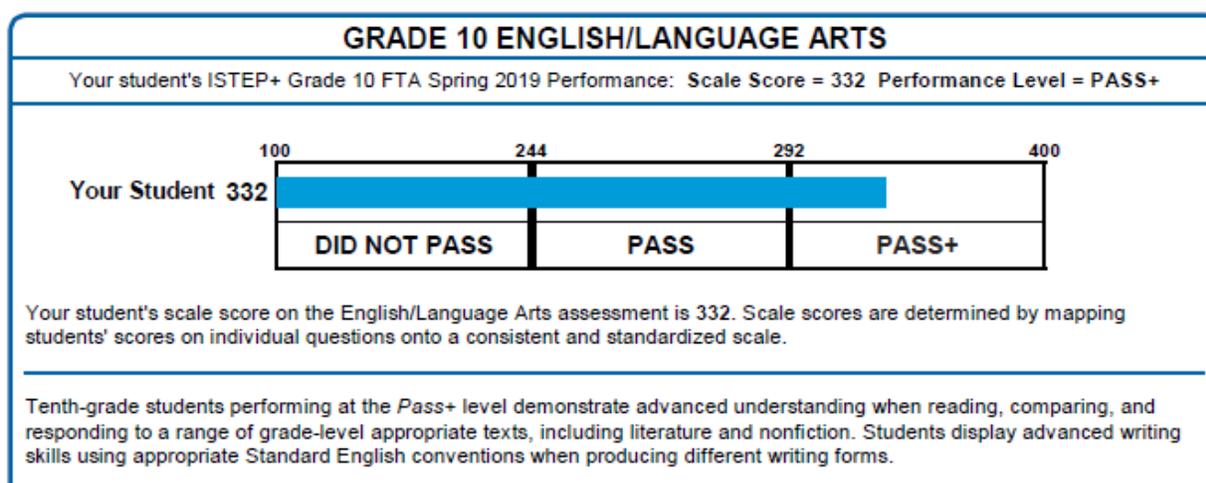


Figure 2. Excerpt from the ISTEP+ Individual Student Report Overall Student Performance

Additionally, it is intended that teachers can use information from the ISTEP+ assessment to guide instructional decisions. As a result, information is also presented on score reports showing more detailed information about how students are performing within a reporting category. These scores are called Indiana Performance Index (IPI) scores. The IPI is a score earned across a subset of items that are associated with a clearly defined skill domain (i.e. reporting category). It is an estimate of the number of items a student would be expected to answer correctly if there had been 100 similar items for that reporting category. It can also be regarded as the percent of items a student would answer correctly out of 100 similar items. The maximum possible IPI value is 100 and the minimum possible value is 0. By focusing on each reporting category, the IPI provides diagnostic information to help identify a student's relative

academic strengths and weaknesses. In most cases, an item is associated with only one reporting category. Occasionally an item is associated with multiple reporting categories.

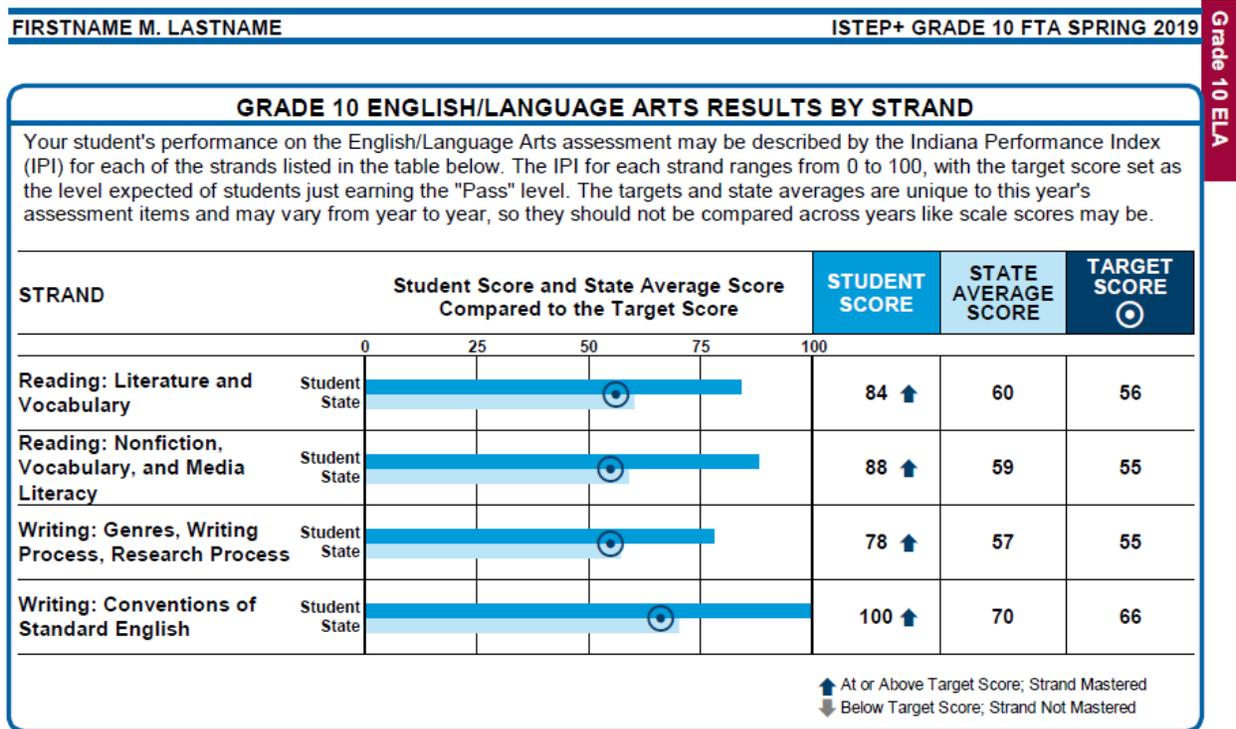


Figure 3. Excerpt from the ISTEP+ Individual Student Report Performance by Strand

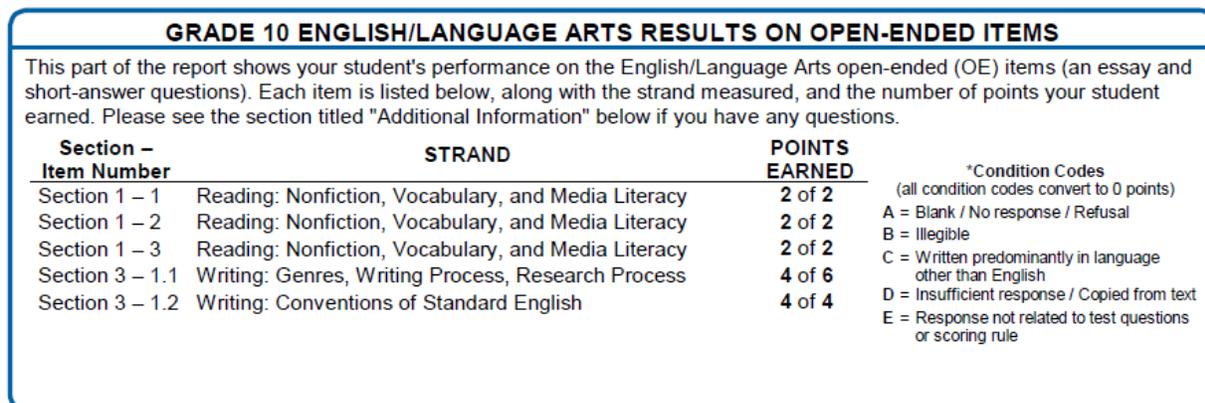


Figure 4. Excerpt from the ISTEP+ Individual Student Report Performance on Open-Ended Items

Report Design Process

Reports were created by Pearson and proposed to IDOE. They were reviewed and approved by IDOE prior to the 2018-2019 administrations.

Critical Elements: 6.4

Standards: 7.11, 8.7, 8.8, 9.8, 9.16

Chapter 13: Validity Argument Synthesis

As described in Chapter 2, the IDOE is employing an argument-based approach to validity (Kane, 1990) to ensure that the combined evidence about its assessment system is comprehensive and addresses critical features of the assessments that relate to score interpretations and uses. Toward this end, this final chapter provides a synthesis of the information presented in Chapters 3 through 12 to address the inferences and assumptions presented in Chapter 2.

The inferences and assumptions presented in Chapter 2 are presented again below:

1. Proficiency in each grade and subject is necessary in order to be prepared for a two-year higher education program, technical program, or the workplace.
2. Mastery in each grade and subject is necessary in order to be prepared for a four-year college or university.
3. Proficiency and/or mastery of the content being taught in each preceding grade of any subject area is necessary in order to reach proficiency and/or mastery in the subsequent grade.
4. The content being taught to students in each grade and subject contains the knowledge and skills necessary to be successful on the ISTEP+.
5. The ISTEP+ assessments for each subject area measure the requisite skills required in subsequent grades, is reliable, and is not influenced substantially by systematic error.
6. The cut scores for the ISTEP+ assessments are appropriate and quantifies the level of knowledge and skill a student must have to be successful.

To support the inferences and assumptions outlined above, one must also assume that the ISTEP+ scores provide a snapshot that reflects what students know and can do in relation to academic expectations. To demonstrate support for this assumption, the IDOE must have evidence related to three additional claims:

1. The assessments and the operational system have been designed to yield scores that reflect students' knowledge and skills in relation to academic expectations.
2. The assessment system operates as intended (e.g., administration, scoring, analyses, and reporting).
3. Students take the assessments under conditions that allow them to demonstrate what they know and can do in relation to academic expectations.

The concepts underlying these claims reflect three critical validity themes that were presented in Chapter 2 – system coherence, comparability and procedural quality, and accessibility and fairness.

Evidence Related to System Coherence

System coherence involves the degree to which an assessment and its operational system have been designed to yield scores that reflect students' knowledge and skills in relation to academic

expectations. Without evidence of system coherence, one cannot interpret test scores as being reflective of any particular construct, body of knowledge, or skill set and, therefore, cannot use test scores to make decisions about any student, group of students, program, or other entity.

Evidence related to system coherence comes from process documentation associated with design and development, scoring and analysis, and reporting phases of testing. The primary threats to system coherence are a lack of confirmed alignment between the target measurement constructs and (a) test items and forms and (b) scoring and reporting methods.

Evidence from the Design and Development Phase

Although many reviewers focus attention on post-hoc evaluation evidence of alignment to support claims related to system coherence, some of the most crucial evidence for this aspect of validity comes from the development process. Alignment must be built into the assessment from the outset. Subsequent evaluations are necessary but are not sufficient to support strong alignment.

The IDOE engaged in several processes both during the initial design phases for the ISTEP+ assessments and as part of ongoing efforts to ensure that ISTEP+ remains aligned with the content and skill expectations it is intended to measure.

- **Development of aligned test specifications:** The IDOE established the 2014 Indiana Academic Standards and then turned to the revision of its assessments to ensure alignment between the two. During this process, the IDOE convened multiple panels of local educators, curriculum specialists, administrators, and university professors to review the new standards, as well as the parameters necessary for aligned item development so as to identify those that were suitable for large-scale assessment. With the guidance of content and measurement experts at their testing vendor, the IDOE then evaluated the appropriateness of existing blueprints for ISTEP+ and made adjustments where necessary. Each test form is developed by matching items to the blueprint.
- **Item development:** Item specifications were created by content and assessment experts using the blueprints as a guide. These specifications were used by item writers to guide the development of individual items that address a specific content or skill element in the standards at a specific level of cognitive complexity. An annual analysis of the test item banks and the blueprints is conducted to identify needs for new item development. To ensure that new items conform to historical item specifications, an annual training of item writers is conducted. All items are subject to internal content reviews, and each item is associated with a code that connects it to a particular element in the standards.
- **Item reviews:** Each year, all new items and item passages are reviewed by the IDOE and by content review committees made up of Indiana educators and/or other stakeholders. Review panelists are selected based on recognized expertise in their target content area and grade level. Items may be accepted, edited, or rejected by these panels based on the predetermined criteria outlined in the passage and item

specifications (found on IDOE's website) and the quality criteria checklist (found in Appendix M).

- **Item tryouts:** Items that survive the review processes are included as field test items on operational test forms. Data related to item difficulty, discrimination, reliability, and possible bias is gathered. These data are reviewed by content and psychometric experts at Pearson and by content experts at the IDOE. Items with out-of-range statistics are reviewed, edited, and re-field tested or removed from the item bank. Only items that are deemed acceptable by these data review panels are maintained in the item banks for inclusion on subsequent test forms as operational items.

Evidence from the Scoring and Analysis Phase

Evidence of system coherence during the scoring and analysis phase of the assessment relates to the design of the scoring process, the standard-setting process, and outcomes. In addition, it relates to studies that evaluate item functioning and the alignment of the performance levels to the content and skill expectations and progressions defined in the content standards and the assessments.

- **Scoring accuracy:** Performance scoring involves rigorous training and qualification for handscorers. There is also a rigorous monitoring process to evaluate whether scorers maintain the level of rigor expected during the entire scoring process.
- **Standard setting/Equipercenile equating:** Achievement standards were set for many of the ISTEP+ assessments in a previous administration. Standard setting meetings for the grade 10 assessments were held in spring 2016. The meetings were planned in collaboration with the IDOE and TAC to ensure a reliable and valid process was being used. Throughout the meetings, IDOE staff, technical advisors, and Pearson facilitators monitored the process and completed the standard setting activities with fidelity.

Evidence from the Reporting Phase

System coherence evidence in the reporting phase of an assessment program involves the decisions and the manifestation of these decisions about which scores are reported and how they are explained. In addition, because the IDOE expects educators to use assessment information to guide decisions related to curriculum and instruction, reporting evidence encompasses information about how educators interpret and use ISTEP+ scores.

- **Types of scores reported:** The IDOE reports total test scores and subtest scores that correspond to the structure of the Indiana Academic Standards (i.e., reporting categories). The items that contribute to the subtest scores are those that conform to the relevant strands in the content standards; these scores are referred to as the Indiana Performance Index scores. In addition, the IDOE reports students' scores in relation to performance levels: *Did Not Pass*, *Pass*, and *Pass+*. As noted in the classification accuracy and classification consistency analyses, the performance level decisions made using ISTEP+ scores are appropriately accurate.

- **Report design:** The ISTEP+ score reports reflect the content and achievement standards scores. These reports were initially developed based on input from a range of Indiana stakeholders and in consideration of professional standards for assessment reporting. As a follow-up to ensure that the reports were interpretable by stakeholders, the IDOE and its vendor reviewed the reports prior to the 2019 administration.

Synthesis of Evidence Related to System Coherence

The IDOE has collected a large body of evidence related to the system coherence of the ISTEP+ assessments in relation to the academic content standards. Evidence from the development and review process indicates that the ISTEP+ test forms adequately reflect the content and skill expectations defined in the academic content standards. Furthermore, the scoring, analysis, and reporting processes all reflect alignment with the structure of the content standards. Taken together, this evidence supports the system coherence of ISTEP+.

Evidence Related to Comparability and Procedural Quality

Comparability is a critical aspect of assessment score interpretation. If scores are not comparable, then they cannot be interpreted as being reflective of students' knowledge and skills and cannot be interpreted as having the same meaning from one site to another or one year to another. Such comparisons are important even when the primary point of the assessment system does not involve such comparisons. If scores cannot be compared from one school to another because of differences in administration conditions, one cannot assume they are meaningful in either school. Likewise, if scores are not comparable over time, it is impossible to warrant interpretation at any single point in time.

Thus, evidence related to comparability involves the degree to which the assessments yield scores that are comparable in meaning across sites and time. The primary threats to comparability are the use of test forms that are not comparable in terms of content and skill coverage, differences in test administration conditions across sites or years, flaws in equating strategies, and variation in reporting formats or dissemination across sites or years.

Evidence from the Design and Development Phase

For test scores to be comparable over sites and time, test forms must be built to support comparability.

- **Alignment:** Throughout the design and development phase of item development, test design, and forms construction, alignment to Indiana academic content standards remains a central goal and focus during training and reviews.
- **Forms construction:** Each ISTEP+ test form is developed to match a blueprint that remains constant across years. Items are selected from the bank to match the blueprint in terms of content and difficulty as well as to support overall test quality in terms of precision. While specific contextual or textual aspects of items differ over time, the match to the blueprint specifications remains the same.

During the 2016-2017 academic year, an analysis of the assessments to an expanded blueprint, including both reporting categories and standards was performed to better align the assessments to the intended content and skill expectations. As a result of the analysis, it was concluded that the assessments met reporting category expectations but did not consistently meet the standards expectations. These expanded blueprints are used to guide future test construction activities and analysis.

- **Design of anchor item set:** In addition to creating forms based on a common blueprint, a pre-administration equating strategy is used to avoid score “drift” that could occur if there were no common items on forms from year to year. The items that are common from one year to another are anchor items; more than 50% of the multiple-choice items are identified from one year to also appear on the test form for the subsequent year. These anchor items are selected to be as broadly representative of the full operational test as possible. The remaining items are selected from the test bank to complete the match to the blueprint. Each anchor and operational form is analyzed to confirm psychometric similarities with the original base form.

In addition to being aligned to a common blueprint, each test form is subject to an extensive content review to ensure that no item provides a contextual clue to another and that the range of contexts, character names, and keyed answer choices vary appropriately across the full form.

If the psychometric or content reviews reveal items that are inadequate in terms of content or performance, these items are replaced until the form meets pre-specified criteria for both statistical and content comparability.

Evidence from the Administration Phase

The test administration process must be as similar as possible across sites and time to support interpretations of score meaning. The term “standardized testing” characterizes these comparable condition requirements.

- **Test calendar and session structure:** The IDOE selects a window for each part of the ISTEP+ assessments that is at approximately the same point in the spring of each year. The structure of the testing sessions is the same across schools. Similarly, the IDOE selects a window for online retesting and paper retesting in the summer of each year. The timing and duration of those windows is as similar as possible year after year.
- **Production of testing materials:** ISTEP+ assessments are administered online and in paper-and-pencil modes. Pearson adheres to a quality control protocol for the design, printing, and online publishing of all testing materials. Design specifications for test booklets and the response areas of the answer documents remain the same across school years. Similar publishing requirements for online presentations exist and are maintained consistently across years.
- **Distribution of testing materials:** Each year, the IDOE submits student enrollments to Pearson, which allows Pearson to load information into the online testing system and to distribute an appropriate number of testing documents to each school, including some degree of overage to allow for damage of documents or late-enrollees for sites

testing paper-and-pencil. All documents are distributed via a trackable system and shrink-wrapped to ensure that exact counts are maintained throughout the distribution process.

All secure documents (i.e., those that contain operational test questions) are assigned unique barcodes, and a unique shipping and receipt report is created for each administration site.

- **Administration training:** Test administration manuals that are identical in substantive content are created each year. These documents are disseminated to all test coordinators. Pearson incorporates diagrams, instructions, and charts to aid in administrators' understanding of the administration process.

Test administration training is provided for corporation and regional personnel, who, in turn, provide school-level training on the test administration procedures within their residing corporations. Online administration training is also provided to online test coordinators and online test examiners.

- **Monitoring of test administration:** The IDOE selects and monitors schools based on previous irregularities and a random sample. IDOE utilizes a checklist and rating scale to follow up with CTCs to ensure fidelity of the assessment program.
- **Return and processing of testing materials:** Pearson organizes and color-codes testing materials and provides return shipment materials to every site to facilitate the return of materials to the scoring facility. All materials must be returned, including testing booklets and answer documents that were not used.

Once materials arrive at the scoring facility, a report is produced to compare the materials received to the materials expected. If secure materials are found to be missing, the standard process is that an Irregularity Report is completed to report the missing barcode range to Pearson. Pearson then takes the steps necessary to account for missing materials. All answer documents are scanned, and the quality of data capture is monitored continuously. Pearson engages in several rounds of reviews to ensure that any potential data-capture errors are identified and corrected.

Evidence from the Scoring and Analysis Phase

Once response data have been captured, the process for generating scores and conducting analyses must have procedural fidelity and also address issues related to score comparability.

- **Reliability of scores and classifications:** Pearson conducts several analyses related to the reliability of the ISTEP+ scores. These analyses address internal consistency (coefficient alpha), the standard error of measurement (estimating likely error associated with multiple retests), the conditional standard error of measurement (estimating the standard error at each score point), and the accuracy and consistency of performance-level classifications.

- **Scaling:** The 3PL model and GPC model, along with pattern scoring, are used to place student responses onto the same scale. Also, item-model fit is examined each year to ensure that responses conform adequately to the standardized score scale.
- **Equating:** A two-stage approach to equating is used to maximize comparability in scale scores across administrations. The first stage is a pre-administration equating. Pre-administration equating ensures that test forms are built to be consistent through the conformance to a blueprint and statistical targets. The post-administration equating process utilizes a common-item nonequivalent groups equating design, which compares form statistics for the just-administered form to those for the form used in the previous year. It should be noted that differential item functioning (DIF) analysis on test mode comparability was not conducted for 2019 assessment because students who took a paper test were less than 10% of the whole IN population

Two of Pearson’s research scientists independently conduct each of the operational analyses. Their results are compared to ensure a complete match.

Evidence from the Reporting Phase

Comparability evidence in the reporting phase relates to the quality and fidelity of the processes for producing and disseminating score reports.

- **Report production:** Pearson prepares for production of score reports by generating mock reports using non-operational data to ensure that the reports print/place properly and to confirm calculations that yield reported scores. A specific group is assigned to verify the quality and accuracy of the mock reports. After the process is verified, a set of reports at each reporting level (student, classroom, school, corporation, and state) is generated using actual operational data. These reports are again verified for accuracy and appearance. During the subsequent production of the full set of reports, random samples are pulled and reviewed for accuracy and appearance. All reports are then sent to the IDOE for review and approval. Once approved, paper score reports, and electronic PDF versions of the reports are assembled.
- **Report Dissemination:** Reports are delivered to Indiana schools, corporations, and the IDOE via a secure shipping process. All reports are also published electronically via a password-protected online site.

Synthesis of Evidence Related to Comparability and Procedural Quality

The IDOE has gathered significant evidence related to comparability and procedural quality. Test forms are constructed to match a blueprint and a pre-equating strategy is used to avoid score “drift.” The test administration is the same across sites and years. Quality control protocol is followed for the design, printing, and publishing of testing materials. All secure materials are assigned unique barcodes, and shipping and receipt reports are created for each administration site. Test administration training is provided, with live question-and-answer sessions for training support purposes. All materials must be returned after the administration. All answer documents are scanned, and the quality of data capture is monitored continuously. Several analyses are

conducted related to the reliability of the ISTEP+ scores. The 3PL and GPC IRT models, along with pattern scoring, are used to place student responses onto the same scale. Two of Pearson's research scientists independently conduct the operational analyses and the results are compared for a complete match. Comparability of online and paper scores is ensured through mode adjustments derived from common mode study methods. Mock-reports are created to confirm the proper print/place and calculations that yield reported scores. Quality and accuracy of mock-reports are verified by a specific group. Then, reports are created with the actual data and verified once more for accuracy and appearance. Following the IDOE approval, paper score reports are sent to Indiana schools, corporations, and the IDOE via a secure shipping process. All reports are also published electronically via a password-protected online site.

Taken together, this evidence from the general test design, administration, scoring and analysis, and reporting procedures strongly supports the comparability and procedural quality of the ISTEP+ assessments.

Evidence Related to Accessibility and Fairness

Accessibility and fairness involve the degree to which students take the assessment under conditions that allow them to demonstrate what they know and can do in relation to academic expectations. If students do not have full access to the assessment (perhaps because they are unable to see well enough to read the questions), do not have the ability to use a response mode that can be captured under the allowed testing conditions, have limited language proficiency, or are easily distracted and require a modified setting in which to take the test, we cannot know whether scores reflect students' knowledge and skills or some other construct-irrelevant factor(s). Evidence of accessibility and fairness comes from the design and development, administration, and scoring and reporting phases of testing.

Evidence from the Design and Development Phase

A test cannot be accessible to all students if it has not been designed to be so. The student population in Indiana, as any student population elsewhere, includes students who have many different needs for accessing items and recording responses. Both research-based design practices and criterion-driven reviews of items are necessary to support accessibility to test items.

- **Design principles:** The principles of Universal Design (UD) are adhered to as specified by the National Center on Educational Outcomes. These principles include guidelines for item presentation and response modes that result in test forms and answer documents that are accessible without accommodation to nearly all students. In addition, UD principles are meant to drive the development of items that, when necessary, can be accommodated during the test administration process.
- **Fairness Reviews:** The IDOE convenes a review of fairness concern annually as part of the Content and Fairness review described in Chapter 4 "Item Writing and Review". Members of these committees represent the full range of the student population, including students with disabilities and students with limited English proficiency as well

as all regions of the state. Committee members review all new items and passages using a protocol that allows them to identify potential sources of bias. Predetermined criteria are used within this review to ensure a thorough examination of items through this lens. The IDOE and Pearson use committee input to decide whether to remove items from the item bank, revise items, or leave items as is.

Evidence from the Administration Phase

An assessment may be designed to be accessible and fair, but the administration conditions may not allow for full and appropriate access to the testing opportunity.

- **Eligibility for Accommodations:** The IDOE has established a list of allowable accommodations based on extant research as well as a set of criteria that CCCs are to use in determining whether and which accommodations a student should use or whether the student is eligible for the ISTEP+. The IDOE publishes guidelines for selection of accommodations and eligibility criteria for the assessment on their website in both the *Indiana Assessment Program Manual* and on the ISTEP+ web page.

Evidence from the Scoring and Analysis Phase

Evidence of accessibility and fairness from the scoring and analysis phase involves the statistical analysis of how test items function for different groups of students.

- **Bias and sensitivity analyses:** The Mantel-Haenszel and Standardized Mean Difference procedures are used to check for the potential presence of differential item functioning (DIF). This analysis indicates when it is more likely for a reference group to answer a particular item correctly or earn a higher score than the focal group at the same proficiency level. Items that are flagged are presented to content experts who review each item and make recommendations to the IDOE for what to do with an item. For example, they can recommend that an item be maintained as is because no obvious source of bias is apparent or that the item be either discarded from the item bank or revised for future use.

Synthesis of Accessibility and Fairness Evidence

The IDOE has gathered evidence related to the accessibility and fairness of the ISTEP+ assessment. The general test design, administration (including accommodations), and scoring procedures support accessibility and fairness. During the test design phase, consideration was given to UD principles and bias and sensitivity were considered during multiple stages of the process. Additionally, extensive documentation is provided to CCCs to guide selection of appropriate accommodations during the administration. Finally, the scoring and reporting phases of the process addressed issues of accessibility and fairness through evaluating statistical aspects of bias via DIF analyses.

Taken together, this evidence supports the accessibility and fairness of the ISTEP+.

Summary of Validity Evidence

The evidence presented above in relation to the three themes of system coherence, comparability and procedural quality, and accessibility and fairness relate to the three primary validity claims presented in the opening of this chapter and show that the IDOE has been successful in its initial attempts to establish a validity argument for ISTEP+ scores.

Table 4. Validity Evidence by Validity Theme and Operational Assessment Phase

Validity Theme	Assessment Phases and Forms of Validity Evidence			
	Test Design and Construction	Administration Process	Scoring and Psychometric Analyses	Reporting
System Coherence	Process: <ul style="list-style-type: none"> Development of content standards ISTEP+ blueprints Item specifications Item writing/reviews Field tests 		Process: <ul style="list-style-type: none"> Scoring accuracy Standard setting 	Process: <ul style="list-style-type: none"> Types of scores reported Report design
Comparability/ Procedural Quality	Process: <ul style="list-style-type: none"> Attention to alignment Forms construction Anchor item selection 	Process: <ul style="list-style-type: none"> Test calendars and administration structure Production of test materials Distribution of test materials Administration training Monitoring administration Return and processing of materials 	Process: <ul style="list-style-type: none"> Reliability Scaling Equating 	Process: <ul style="list-style-type: none"> Report production Report distribution
Accessibility and Fairness	Process: <ul style="list-style-type: none"> Universal design principles Bias reviews 	Process: <ul style="list-style-type: none"> Eligibility criteria Accommodations Monitoring 	Process: <ul style="list-style-type: none"> Differential item functioning analyses 	

Critical Elements: 3.1-3.4

Standards: 1.1-1.25, 3.3

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 27-46.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1-24.
- Indiana Academic Standards. (2016, August 9). Retrieved from <http://www.doe.in.gov/standards>.
- Kane, M. T. (1990). *An Argument-Based Approach to Validation*. ACT Research Report Series 90-13. Retrieved from https://forms.act.org/research/researchers/reports/pdf/ACT_RR90-13.pdf.
- Kim, D., Kim, J., & Barton, K. (2007). *Estimating classification consistency and classification accuracy with pattern scoring*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179-197).
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement*, 8, 453-461.
- Nunnally, J., & Bernstein, L. (1994). *Psychometric theory*. New York: McGraw-Hill Higher, INC.
- Peterson, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational Measurement* (pp. 221-262). New York, NY: Macmillan.
- Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced test: A decision-theoretic formulation. *Journal of Educational Measurement*, 11(4), 263-267.

Appendix A: ISTEP+ Test Blueprints and Item Counts

ENGLISH/LANGUAGE ARTS

Reporting Category Blueprint for Winter 2018 Retest

Table 5. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Strand: Winter 2018 Retest

Standard	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.RLV	Reading: Literature and Vocabulary	19-24	30-40%	15	19	31.1%
10.RNV	Reading: Nonfiction, Vocabulary, and Media Literacy	19-24	30-40%	19	23	37.7%
10.WC	Writing: Conventions of Standard English	7-12	10-20%		7	11.5%
10.WG	Writing: Genres, Writing Process, Research Process	7-12	10-20%	6	12	19.7%
	Total	61	100%	44	61	100%

Reporting Category Blueprint for Retest Spring 2019

Table 6. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Strand: Retest Spring 2019

Standard	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.RLV	Reading: Literature and Vocabulary	19-24	30-40%	17	23	37.7%
10.RNV	Reading: Nonfiction, Vocabulary, and Media Literacy	19-24	30-40%	20	20	32.8%
10.WC	Writing: Conventions of Standard English	7-12	10-20%	4	8	13.1%
10.WG	Writing: Genres, Writing Process, Research Process	7-12	10-20%	5	10	16.4%
	Total	61	100%	46	61	100%

Reporting Category Blueprint for First Time Administration (FTA) Spring 2019

Table 7. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Reporting Category FTA Spring 2019

Reporting Category	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.RLV	Reading: Literature and Vocabulary	19-24	30-40%	17	21	34.4%
10.RNV	Reading: Nonfiction, Vocabulary, and Media Literacy	19-24	30-40%	16	20	32.8%
10.WC	Writing: Conventions of Standard English	7-12	10-20%	5	9	14.8%
10.WG	Writing: Genres, Writing Process, Research Process	7-12	10-20%	6	11	18%
	Total	61	100%	44	61	100%

Reporting Category Blueprint for Summer 2019

Table 8. ISTEP+ Grade 10 E/LA Operational Blueprint and Number of Items and Points by Reporting Category Summer

Reporting Category	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.RLV	Reading: Literature and Vocabulary	19-24	30-40%	17	22	36.1%
10.RNV	Reading: Nonfiction, Vocabulary, and Media Literacy	19-24	30-40%	15	19	31.1%
10.WC	Writing: Conventions of Standard English	7-12	10-20%	5	9	14.8%
10.WG	Writing: Genres, Writing Process, Research Process	7-12	10-20%	6	11	18%
	Total	61	100%	43	61	100%

Mathematics

Reporting Category Blueprint for Winter 2018 Retest

Table 9. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Winter 2018 Retest

Standard	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.DS	Data Analysis, Statistics, and Probability	7-14	9-19%	7	11	15.1%
10.GM	Geometry and Measurement	3-11	4-14%	4	4	5.5%
10.LEI	Linear Equations, Inequalities, and Functions	21-29	28-38%	21	23	31.5%
10.MP	Mathematical Process	3-11	4-14%	4	9	12.3%
10.NSC	Number Sense, Expressions, and Computation	8-16	11-21%	11	11	15.1%
10.QEF	Quadratic & Exponential Equations and Functions	4-11	5-15%	4	9	12.3%
10.SEI	Systems of Equations and Inequalities	3-11	4-14%	6	7	9.6%
	Total	73	100%	53	73	100%

Reporting Category Blueprint for Spring 2019 Retest

Table 10. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Spring 2019 Retest

Standard	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.DS	Data Analysis, Statistics, and Probability	7-14	9-19%	6	11	15.1%
10.GM	Geometry and Measurement	3-11	4-14%	5	8	11%
10.LEI	Linear Equations, Inequalities, and Functions	21-29	28-38%	21	23	31.5%
10.MP	Mathematical Process	3-11	4-14%	4	9	12.3%
10.NSC	Number Sense, Expressions, and Computation	8-16	11-21%	10	11	15.1%
10.QEF	Quadratic & Exponential Equations and Functions	4-11	5-15%	3	4	5.5%
10.SEI	Systems of Equations and Inequalities	3-11	4-14%	7	7	9.6%
	Total	73	100%	52	73	100%

Reporting Category Blueprint for Spring 2019 FTA

Table 11. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Spring FTA

Standard	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.DS	Data Analysis, Statistics, and Probability	7-14	9-19%	8	12	16.4%
10.GM	Geometry and Measurement	3-11	4-14%	5	5	8.9%
10.LEI	Linear Equations, Inequalities, and Functions	21-29	28-38%	19	21	28.8%
10.MP	Mathematical Process	3-11	4-14%	4	9	12.3%
10.NSC	Number Sense, Expressions, and Computation	8-16	11-21%	11	12	16.4%
10.QEF	Quadratic & Exponential Equations and Functions	4-11	5-15%	5	8	11%
10.SEI	Systems of Equations and Inequalities	3-11	4-14%	5	6	8.2%
	Total	73	100%	53	73	100%

Reporting Category Blueprint for Summer 2019

Table 12. ISTEP+ Grade 10 Mathematics Operational Blueprint and Number of Items and Points by Strand: Summer

Standard	Reference	Blueprint		Actual		
		Points	Percent	Items	Points	Percent
10.DS	Data Analysis, Statistics, and Probability	7-14	9-19%	8	11	15.1%
10.GM	Geometry and Measurement	3-11	4-14%	4	5	6.8%
10.LEI	Linear Equations, Inequalities, and Functions	21-29	28-38%	20	24	32.9%
10.MP	Mathematical Process	3-11	4-14%	4	9	12.3%
10.NSC	Number Sense, Expressions, and Computation	8-16	11-21%	11	11	15.1%
10.QEF	Quadratic & Exponential Equations and Functions	4-11	5-15%	6	9	12.3%
10.SEI	Systems of Equations and Inequalities	3-11	4-14%	4	4	5.5%
	Total	73	100%	53	73	100%

Reporting Category and Strand Priority Blueprint Analysis

Table 13. ISTEP+ Grade 10 E/LA Operational Blueprint Analysis for 2016, 2017, 2018, and 2019

Reporting Category	Standard	Min. items based on priority	Actual Items							
			SP 16	SP 17	WT 17*	SP 18	SU 18	WT 18	SP 19 FTA	SU 19
Reading: Literature and Vocabulary	9-10.RL.2.1	2	2	2	3	3	2	3	3	2
	9-10.RL.2.2	2	3	3	2	2	3	2	2	2
	9-10.RL.2.3	2	6	6	5	4	6	4	5	5
	9-10.RL.3.1	2	2	2	2	2	2	2	3	2
	9-10.RL.3.2	1	2	2	1	1	2	1	1	2
	9-10.RL.4.2	0	0	0	0	0	0	0	0	0
	9-10.RV.2.1	2	1	1	1	1	1	1	1	1
	9-10.RV.2.3	0	0	0	0	0	0	0	0	0
	9-10.RV.2.4	0	0	0	0	0	0	0	0	0
	9-10.RV.3.1	2	1	1	2	3	1	3	2	2
	9-10.RV.3.3	0	0	0	0	0	0	0	0	0
Reading: Nonfiction, Vocabulary, and Media Literacy	9-10.ML.2.1	1	1	1	1	1	1	1	1	1
	9-10.RN.2.1	2	3	2	3	2	2	3	3	1
	9-10.RN.2.2	2	0	0	0	1		1	0	2
	9-10.RN.2.3	2	1	3	3	3	3	2	3	2
	9-10.RN.3.2	2	3	4	3	3	4	3	3	3
	9-10.RN.3.3	2	7	5	5	3	5	3	3	3
	9-10.RN.4.1	1	0	0	1	2		1	1	0
	9-10.RN.4.3	1	0	0	0	0		0	0	0
	9-10.RV.2.1	2	2	1	1	1	1	1	1	1
	9-10.RV.2.3	0	0	0	0	0	0	1	1	1
	9-10.RV.2.4	0	0	1	0	0	0	1	1	1
	9-10.RV.3.2	2	0	0	3	3	0	2	3	1
	9-10.RV.3.3	0	1	1	0	0	1	0	0	1
Writing: Genres, Writing Process, Research Process	9-10.W.3.1	1	2	1	1	1	1	1	1	1
	9-10.W.3.2	1	0	1	1	1	1	1	1	1
	9-10.W.3.3	1	0	0	0	1	0	1	0	1
	9-10.W.4	2	2	2	2	2	2	2	2	2
	9-10.W.5	1	1	1	1	1	1	1	1	1
Writing: Genres, Writing Process, Research Process	Prompt		1	1	1	1	1	1	1	1
	9-10.W.6.1b	2	1	1	1	1	1	1	2	2
	9-10.W.6.1e		1	1	1	1	1	1	1	1
	9-10.W.6.2b	2	1	2	2	2	2	2	2	2

*WT17 became the Spring 19RT

Reporting Category and Strand Priority Blueprint Analysis

Table 14. ISTEP+ Grade 10 Mathematics Operational Blueprint Analysis for 2016, 2017, 2018 and 2019

Reporting Category	Standard	Min. items based on priority	Actual Items								
			SP 16	SP 17	*W T 17	SP 18	SU 18	WT 18	SP 19 FTA	SU 19	
Number Sense, Expressions, and Computation	AI.RNE.2	1	0	1	1	1	1	1	2	2	1
	AI.RNE.3	0	1	1	1	1	1	1	1	1	0
	AI.RNE.4	0	2	1	1	1	1	1	1	1	1
	AI.RNE.5	0	0	0	0	0	0	0	0	0	1
	AI.RNE.6	1	1	1	2	2	2	2	1	1	1
	AI.RNE.7	0	1	2	2	0	2	0	0	0	1
	8.NS.1	1	2	2	2	2	2	2	1	1	2
	8.NS.2	1	0	0	0	1	0	1	1	1	1
	8.NS.3	0	1	1	1	1	1	1	2	2	1
	8.NS.4	0	1	1	1	1	1	1	1	0	2
8.C.1	1	0	0	0	0	0	0	1	2	0	
Geometry & Measurement	8.GM.1	1	1	1	1	1	1	1	1	1	1
	8.GM.2	1	1	1	0	0	0	0	0	0	0
	8.GM.4	0	0	1	0	0	0	0	0	1	0
	8.GM.6	0	0	0	0	0	0	0	0	1	1
	8.GM.8	1	2	1	2	2	2	2	2	1	1
	8.GM.9	0	1	0	0	0	0	0	1	1	1
Data Analysis, Statistics & Probability	AI.DS.1	1	0	0	1	0	1	0	0	1	0
	AI.DS.3	1	2	0	1	2	1	2	2	2	1
	AI.DS.4	0	1	0	0	0	0	0	0	0	0
	AI.DS.6	1	1	1	1	1	1	1	1	1	1
	8.DSP.1	1	2	1	0	1	0	1	1	1	2
	8.DSP.3	2	1	3	1	1	1	1	2	1	3
	8.DSP.5	0	0	0	0	0	0	0	1	1	0
8.DSP.6	0	0	0	0	0	0	0	0	1	1	

Reporting Category	Standard	Min. items based on priority	Actual Items							
			SP 16	SP 17	*W T 17	SP 18	SU 18	WT 18	SP 19 FTA	SU 19
Linear Equations, Inequalities, and Functions	Al.L.1	2	0	2	3	2	3	1	1	2
	Al.L.2	2	1	1	0	0	0	0	0	2
	Al.L.3	1	2	2	2	2	2	2	2	3
	Al.L.4	1	2	4	3	2	3	3	2	2
	Al.L.5	2	1	1	2	3	2	3	1	1
	Al.L.6	1	0	1	1	1	1	2	2	2
	Al.L.7	1	1	3	2	2	2	1	0	2
	Al.L.8	0	2	1	1	1	1	1	2	2
	Al.L.11	0	1	1	1	1	1	1	1	0
	Al.F.2	2	0	0	1	2	1	1	2	1
	Al.F.3	0	0	0	0	0	0	0	1	0
	Al.F.4	1	0	0	0	1	0	0	0	0
	8.AF.2	0	1	1	1	1	1	1	1	0
	8.AF.3	1	1	1	1	1	1	1	1	1
	8.AF.5	1	2	2	2	2	2	2	1	1
	8.AF.6	1	3	0	0	1	0	1	1	0
8.AF.7	1	2	1	1	1	1	1	1	1	
Systems of Equations and Inequalities	Al.SEI.1	1	1	0	1	1	1	0	1	0
	Al.SEI.2	1	3	3	3	2	3	2	2	1
	Al.SEI.3	2	1	2	3	3	3	2	1	2
	Al.SEI.4	1	0	1	1	1	1	2	1	1
	8.AF.8	0	0	0	0	1	0	0	0	0
Quadratic & Exponential Equations and Functions	Al.QE.1	1	2	2	1	1	1	1	1	1
	Al.QE.3	0	1	0	0	0	0	0	0	0
	Al.QE.4	1	1	2	1	1	1	1	2	2
	Al.QE.5	1	1	0	0	0	0	0	1	1
	Al.QE.6	1	3	1	2	2	2	1	1	2
	Al.QE.7	0	0	1	0	0	0	1	0	0
Mathematical Process	PS.1-PS.8									

*WT17 became the Spring 19RT

Appendix B: ISTEP+ Testing Times

Timing for Winter 2018 and Spring 2019 Retest

Table 15. Operational Test Times: English/Language Arts ISTEP+ Part 1: Winter 2018 and Spring 2019 Retest

English/Language Arts: ISTEP+ Part 1		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	30 minutes
	Total Time	35 minutes
Section 2	Instructions	5 minutes
	Working Time	65 minutes
	Total Time	70 minutes

Table 16. Operational Test Times: Mathematics ISTEP+ Part 1 : Winter 2018 and Spring 2019 Retest

Mathematics: ISTEP+ Part 1		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	30 minutes
	Total Time	35 minutes
Section 2	Instructions	5 minutes
	Working Time	20 minutes
	Total Time	25 minutes

Table 17. Operational Test Times: English/Language Arts ISTEP+ Part 2: Winter 2018 and Spring 2019 Retest

English/Language Arts: ISTEP+ Part 2		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	28 minutes
	Total Time	33 minutes
Section 2	Instructions	5 minutes
	Working Time	30 minutes
	Total Time	35 minutes
Section 3	Instructions	5 minutes
	Working Time	45 minutes
	Total Time	50 minutes

Table 18. Operational Test Times: Mathematics ISTEP+ Part 2: Winter 2018 and Spring 2019 Retest

Mathematics: ISTEP+ Part 2		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	35 minutes
	Total Time	40 minutes
Section 2	Instructions	5 minutes
	Working Time	35 minutes
	Total Time	40 minutes
Section 3	Instructions	5 minutes
	Working Time	30 minutes
	Total Time	35 minutes

Timing for Spring 2019 First Time Administration

Table 19. Operational Test Times: English/Language Arts ISTEP+ Part 1: Spring 2019 FTA

English/Language Arts: ISTEP+ Part 1		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	30 minutes
	Total Time	35 minutes
Section 2	Instructions	5 minutes
	Working Time	65 minutes
	Total Time	70 minutes
Section 3	Instructions	5 minutes
	Working Time	65 minutes
	Total Time	70 minutes

Table 20. Operational Test Times: Mathematics ISTEP+ Part 1: Spring 2019 FTA

Mathematics: ISTEP+ Part 1		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	40 minutes
	Total Time	45 minutes
Section 2	Instructions	5 minutes
	Working Time	38 minutes
	Total Time	43 minutes

Table 21. Operational Test Times: English/Language Arts ISTEP+ Part 2: Spring 2019 FTA

English/Language Arts: ISTEP+ Part 2		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	47 minutes
	Total Time	52 minutes
Section 2	Instructions	5 minutes
	Working Time	38 minutes
	Total Time	43 minutes
Section 3	Instructions	5 minutes
	Working Time	39 minutes
	Total Time	44 minutes

Table 22. Operational Test Times: Mathematics ISTEP+ Part 2: Spring 2019 FTA

Mathematics: ISTEP+ Part 2		
Test Administration Times Online and Paper-and-Pencil		Grade 10
Section 1	Instructions	5 minutes
	Working Time	45 minutes
	Total Time	50 minutes
Section 2	Instructions	5 minutes
	Working Time	33 minutes
	Total Time	38 minutes
Section 3	Instructions	5 minutes
	Working Time	26 minutes
	Total Time	31 minutes

Appendix C: ISTEP+ Test Item Statistics

Table 23. Summary of Operational Item P-Values and Item-Total Correlations across Testing Modes: Spring 2019

	Grade	Mean Item p-value	Mean Item Total Correlation
E/LA	10	0.54	0.43
Mathematics	10	0.48	0.46

Appendix D: ISTEP+ DIF Flags

Table 24. Summary items flagged for DIF – English/Language Arts Spring 2019

Grade	Base	Focus	B-Flag				C-Flag			
			Reference		Focal		Reference		Focal	
			N	%	N	%	N	%	N	%
10	M	F	2	4%	1	2%	0	0%	0	0%
	W	AA	1	2%	0	0%	0	0%	0	0%
	W	H	0	0%	0	0%	0	0%	0	0%

Table 25. Summary items flagged for DIF – Mathematics Spring 2019

Grade	Base	Focus	B-Flag				C-Flag			
			Reference		Focal		Reference		Focal	
			N	%	N	%	N	%	N	%
10	M	F	3	4%	0	0%	0	0%	0	0%
	W	AA	2	3%	1	1%	0	0%	0	0%
	W	H	0	0%	1	1%	0	0%	1	1%

Appendix E: Stability Check Summary: Spring 2019

Table 26. E/LA Anchor Items: Grade 10 Spring 2019

	Grade						
	3	4	5	6	7	8	10
No. of operational items							45
No. of initial anchor items							14
No. of final anchor items							13
No. of operational points							61
No. of initial anchor points							15
No. of final anchor points							14

Table 27. Mathematics Anchor Items: Grade 10 Spring 2019

	Grade						
	3	4	5	6	7	8	10
No. of operational items							57
No. of initial anchor items							17
No. of final anchor items							16
No. of operational points							73
No. of initial anchor points							18
No. of final anchor points							17

Appendix F: Equating Constants and LOSS and HOSS

Table 28. Original Equating Constants: Grade 10 E/LA and Mathematics

Grade	Original Scaling Year	Equating Constants		Scaling Constants	
		Slope (a)	Intercept (b)	Slope (r)	Intercept (l)
E/LA	2016	1.0209	0.0736	46.0447	249.3767
Mathematics	2016	1.0557	0.0482	46.4555	250.7266

Table 29. HOSS, LOSS, and Performance Level Cut Scores: Grade 10 E/LA and Mathematics

Subject	Year	LOSS	Pass	Pass Plus	HOSS
E/LA	2019	100	244	292	400
Mathematics	2019	100	271	339	400

Appendix G: ISTEP+ Test Curves

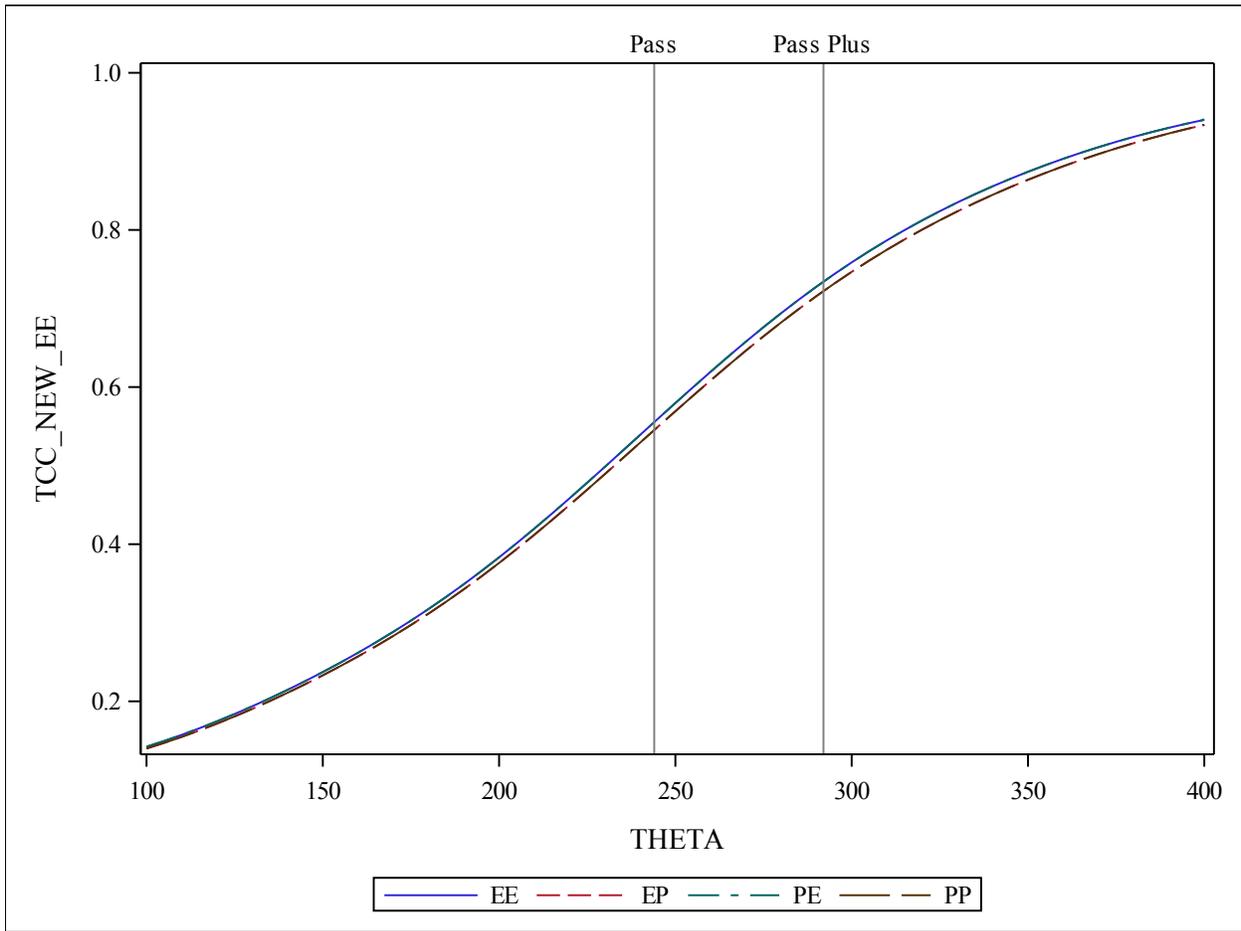


Figure 5. Test Characteristic Curves for ISTEP+ E/LA Grade 10 Winter 2018

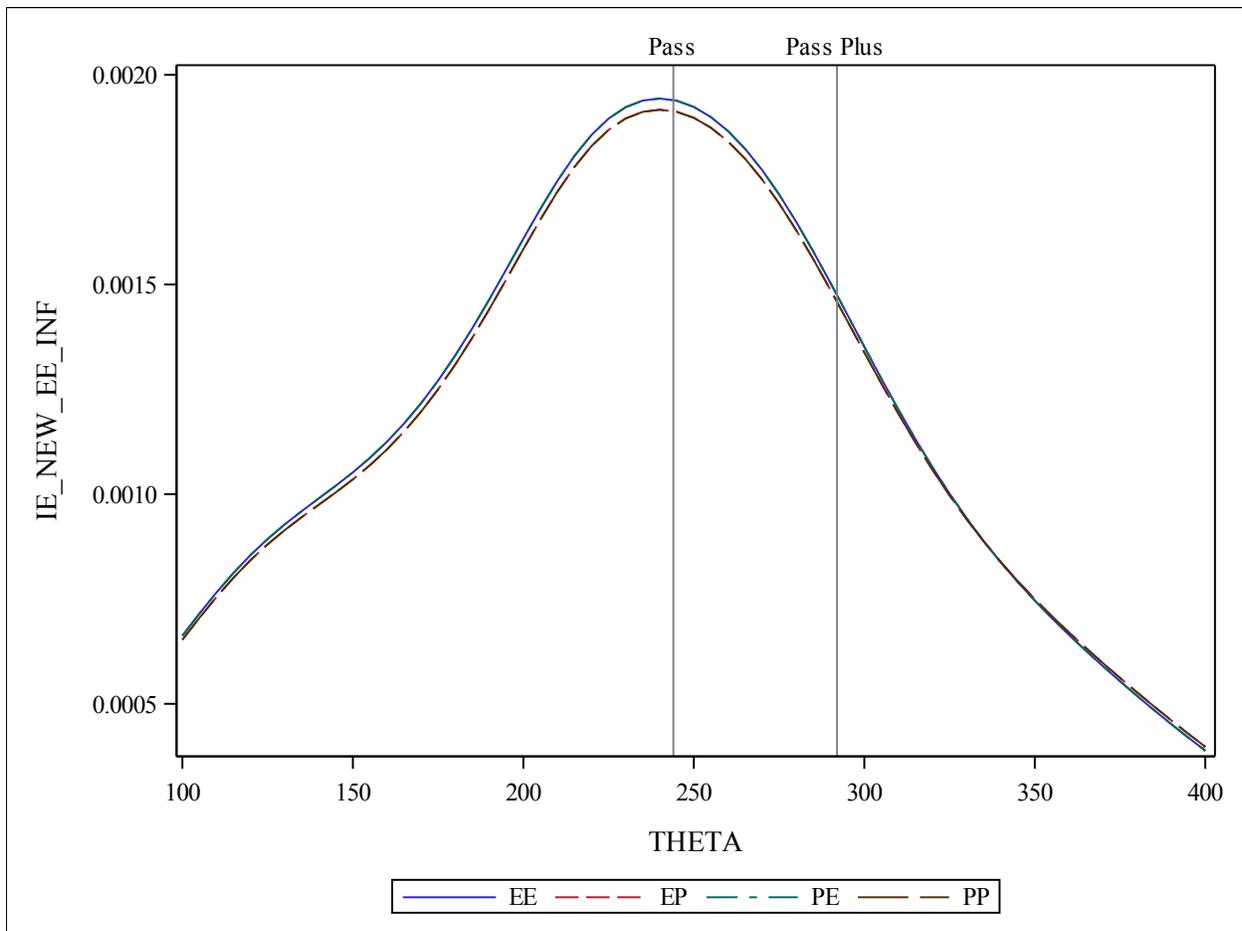


Figure 6. Test Information Curves for ISTEP+ E/LA Grade 10 Winter 2018

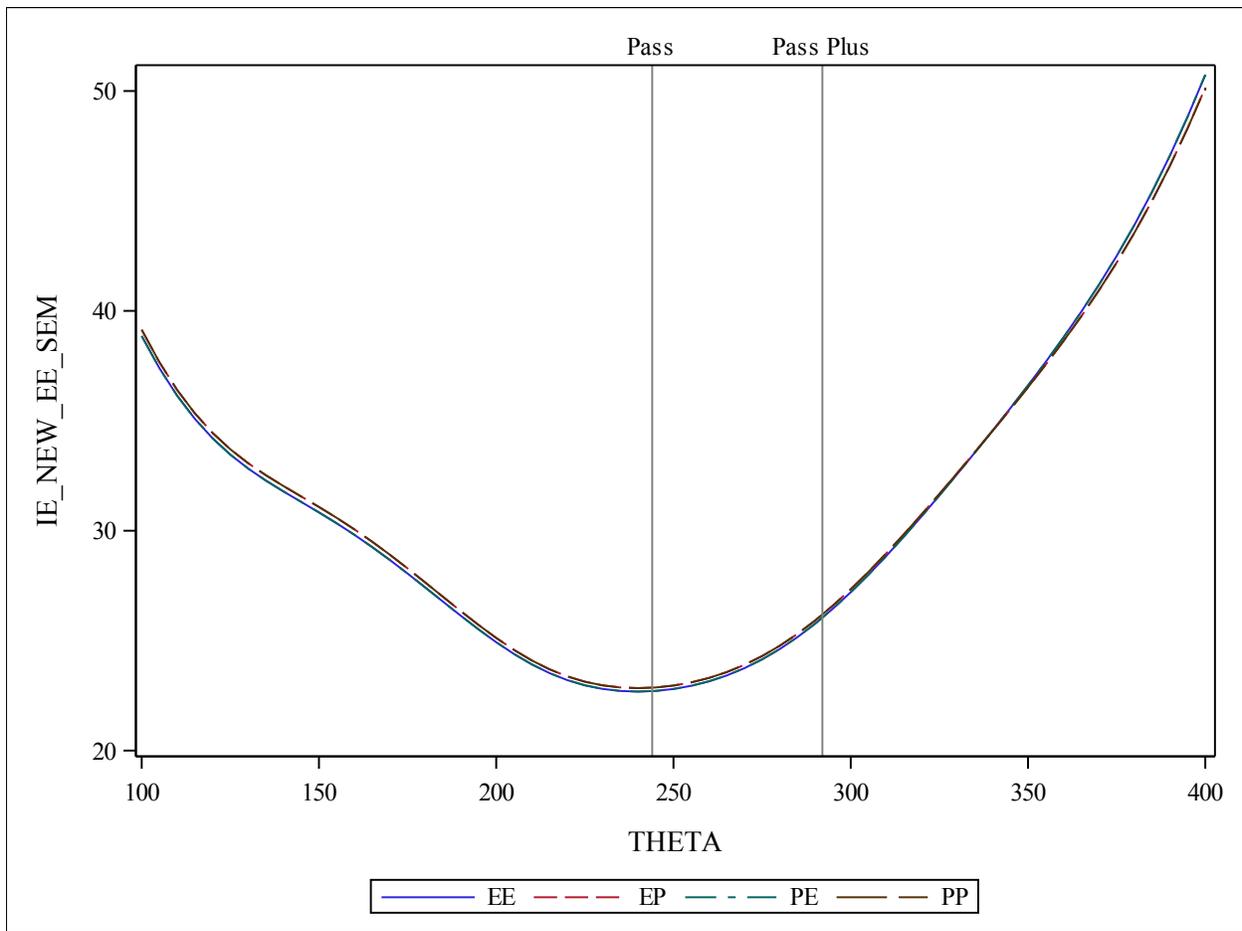


Figure 7. Standard Error of Measure Curves for ISTEP+ E/LA Grade 10 Winter 2018

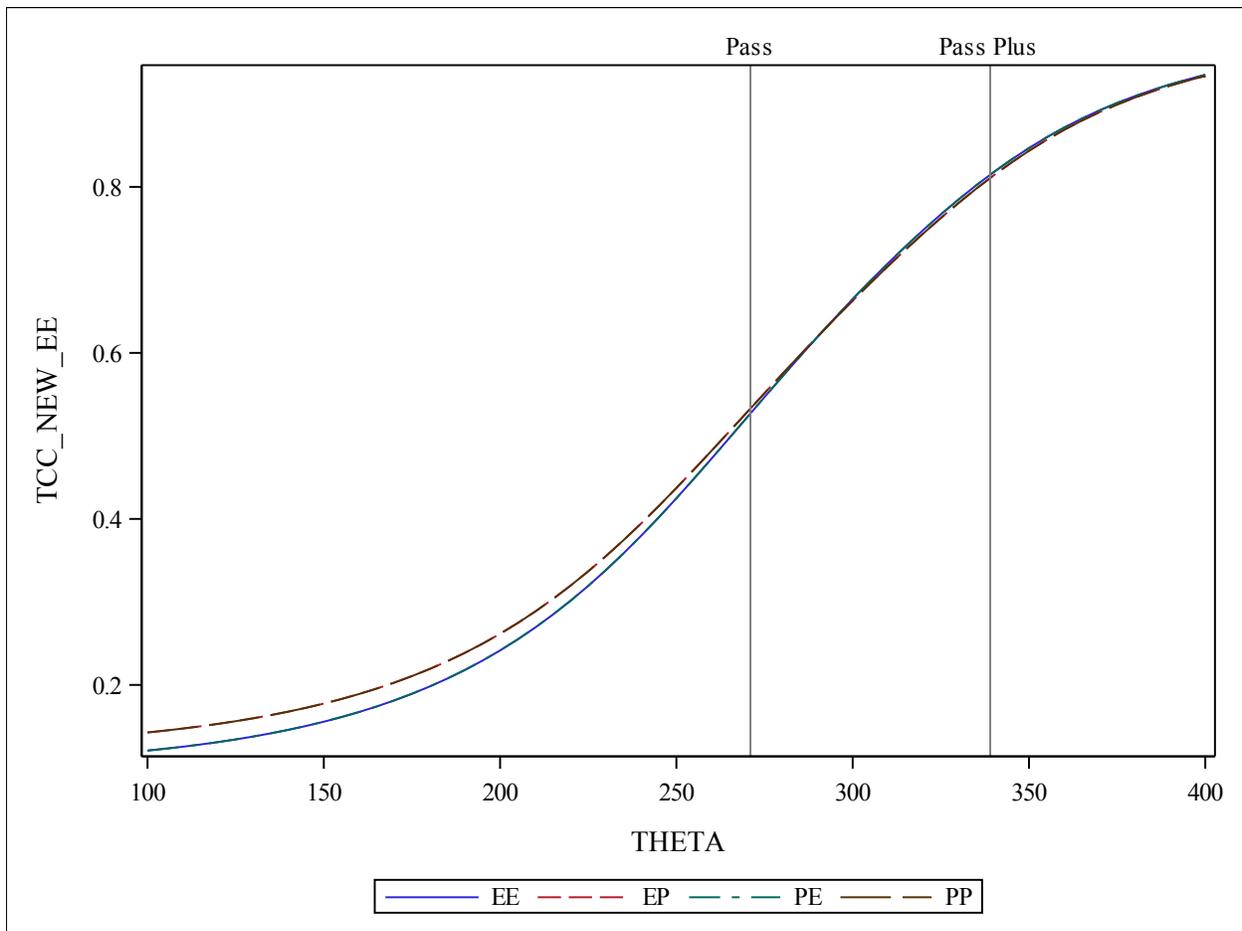


Figure 8. Test Characteristic Curves for ISTEP+ Mathematics Grade 10 Winter 2018

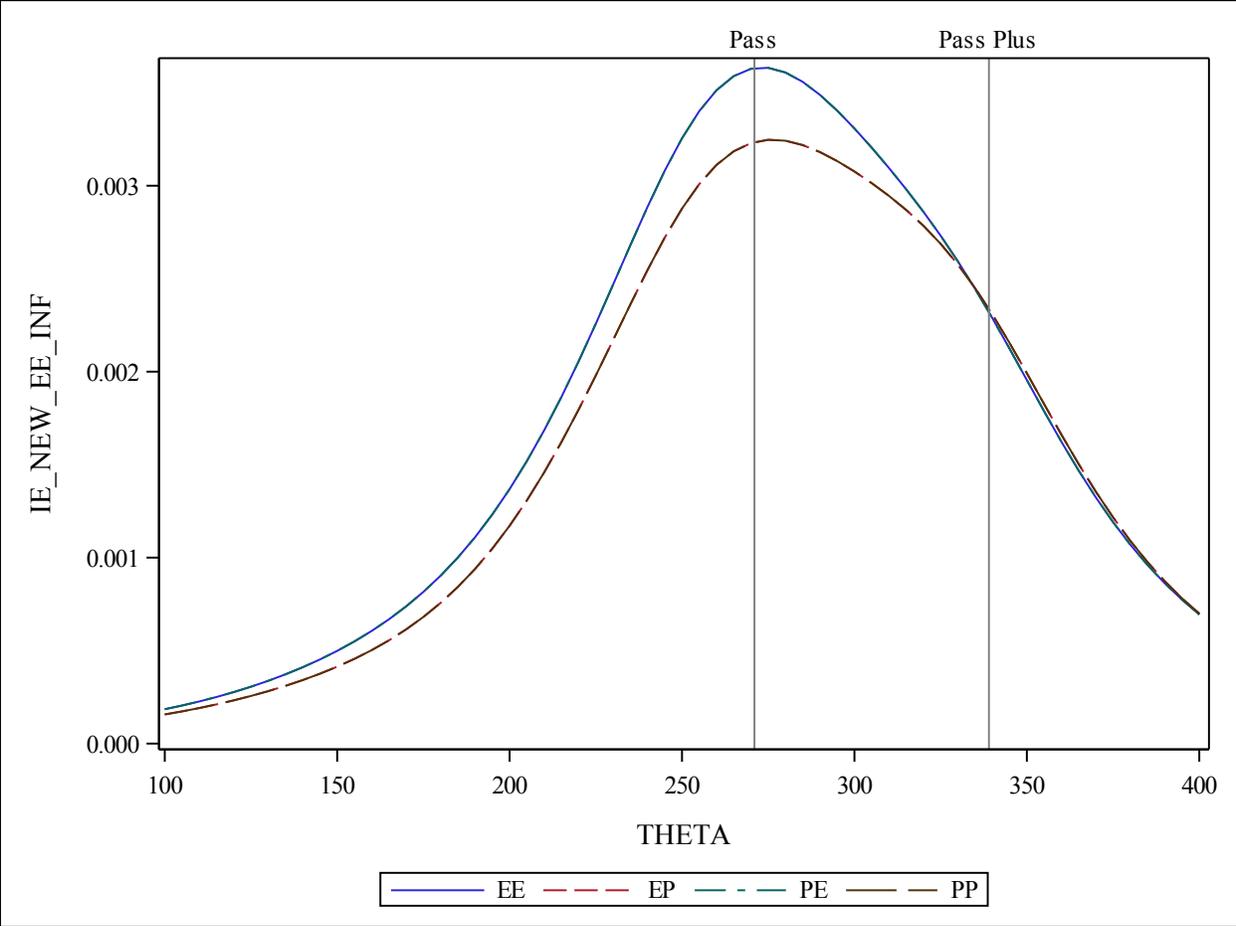


Figure 9. Test Information Curves for ISTEP+ Mathematics Grade 10 Winter 2018

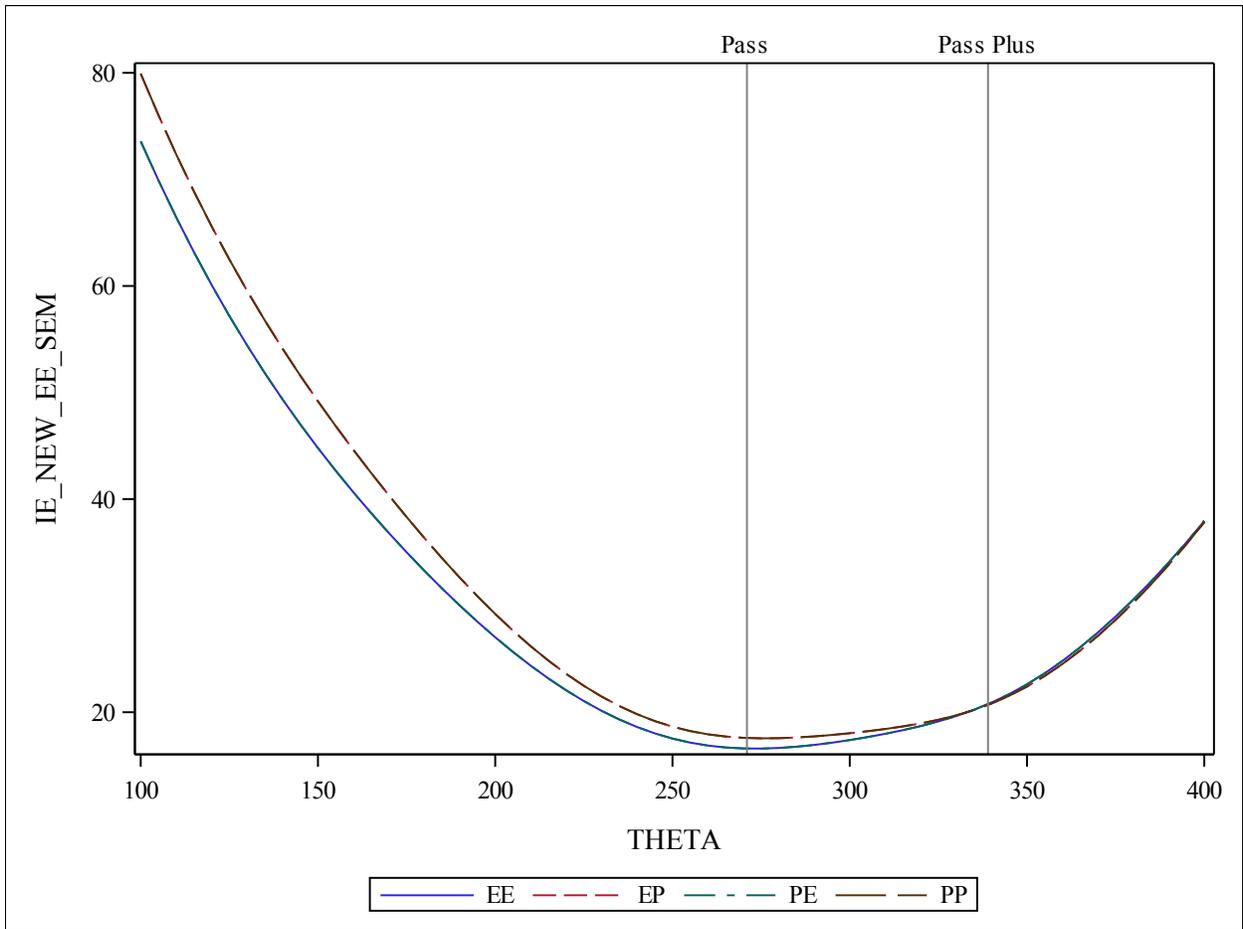


Figure 10. Standard Error of Measure Curves for ISTEP+ Mathematics Grade 10 Winter 2018

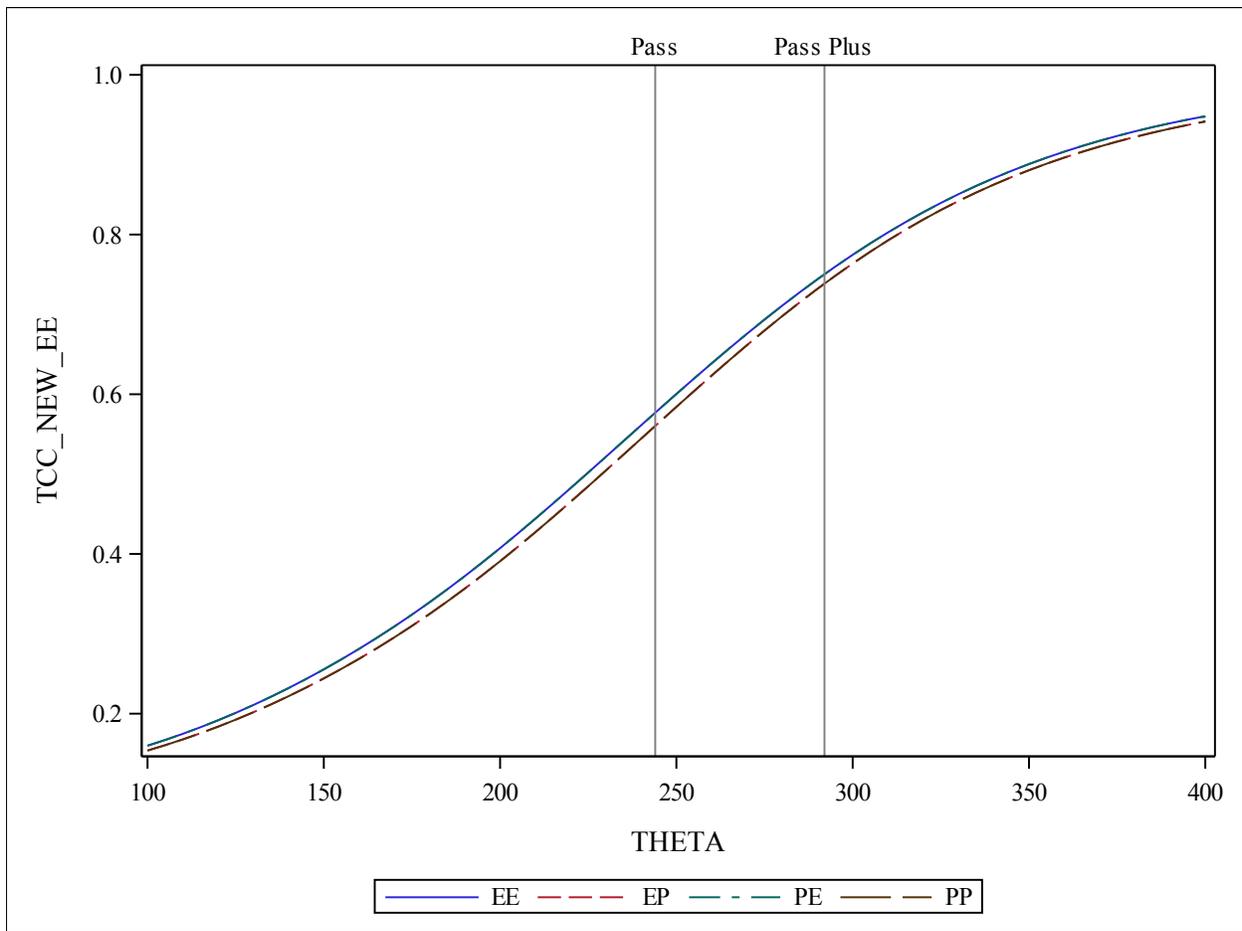


Figure 11. Test Characteristic Curves for ISTEP+ E/LA Grade 10 Spring Retest 2019

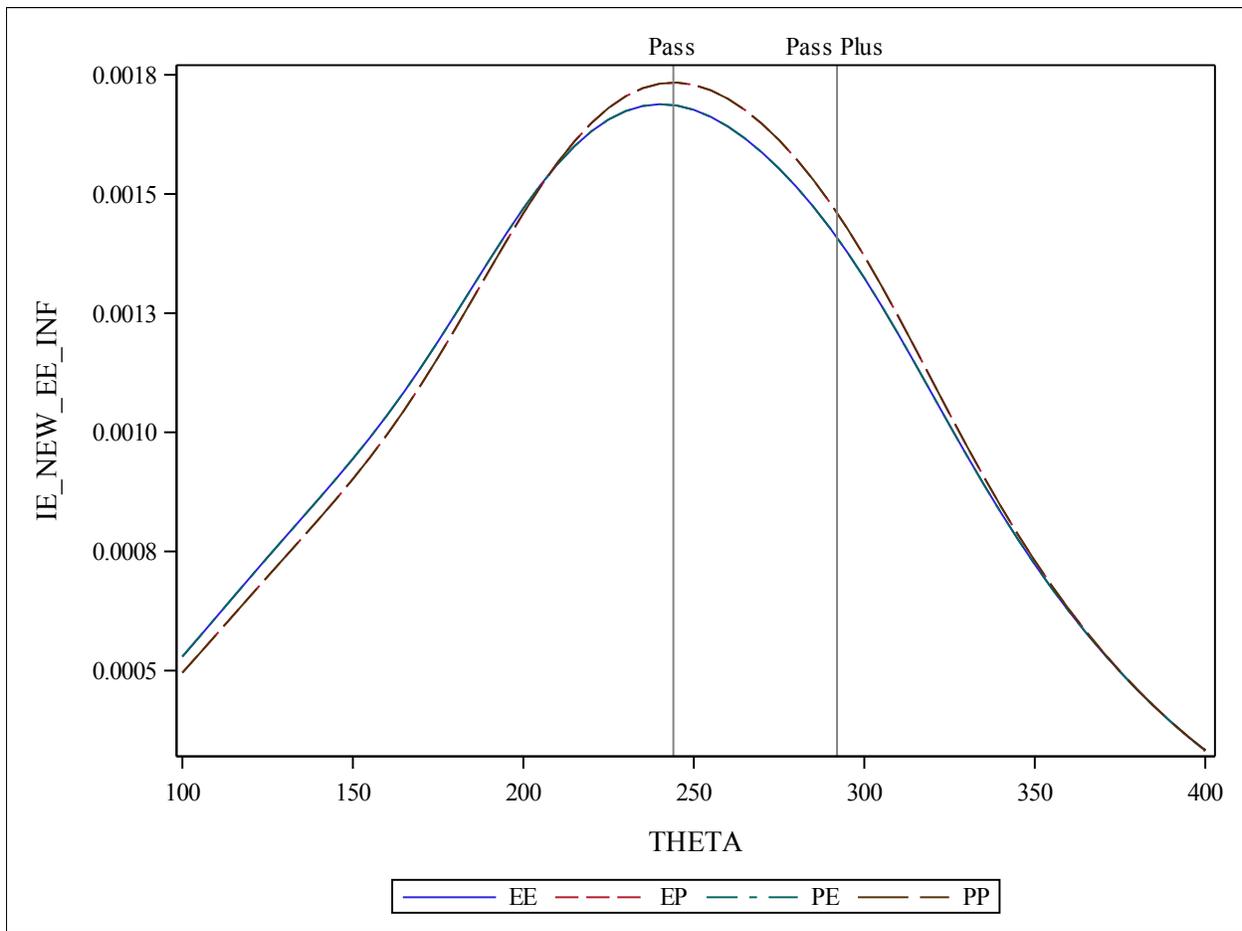


Figure 12. Test Information Curves for ISTEP+ E/LA Grade 10 Spring Retest 2019

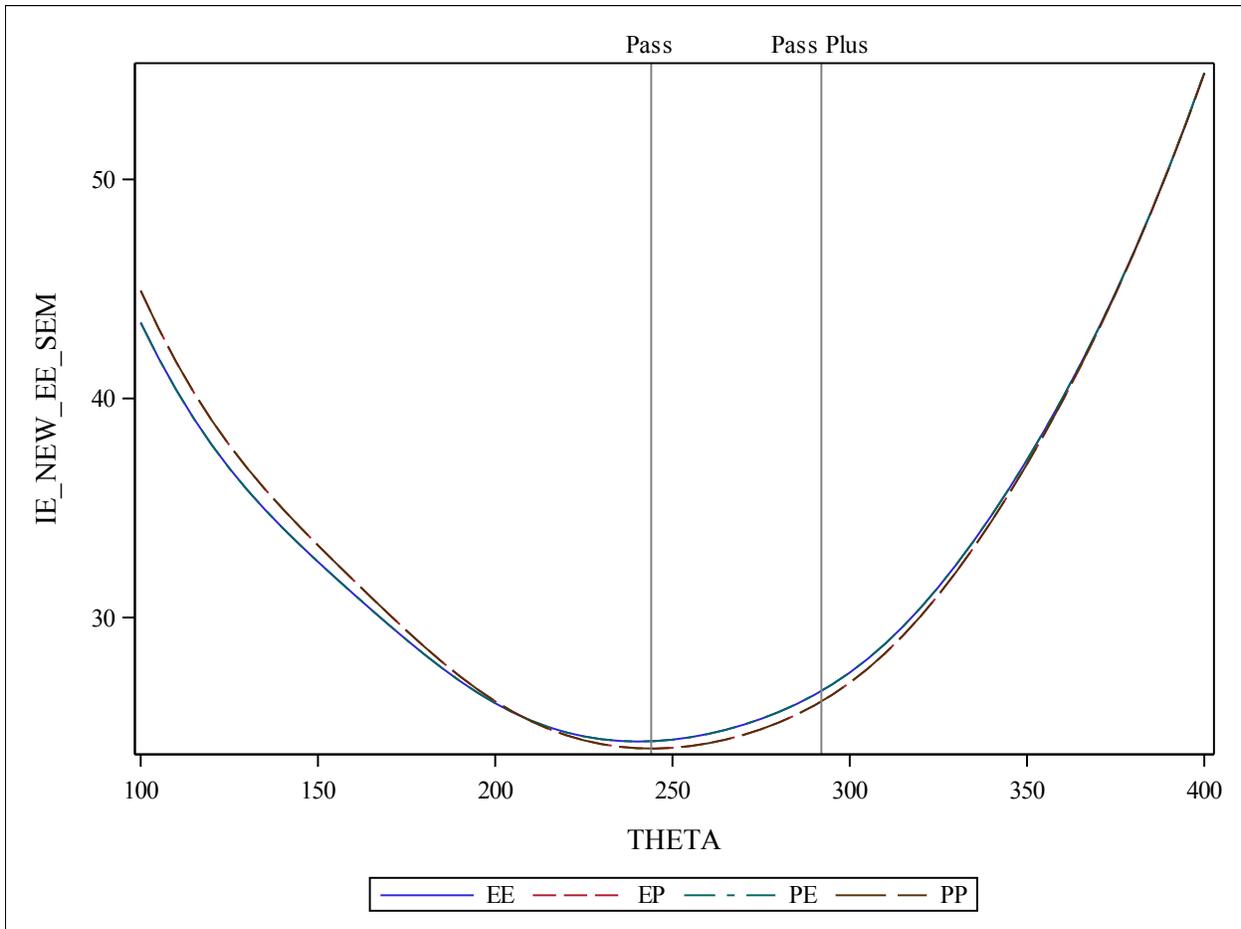


Figure 13. Standard Error of Measure Curves for ISTEP+ E/LA Grade 10 Spring Retest 2019

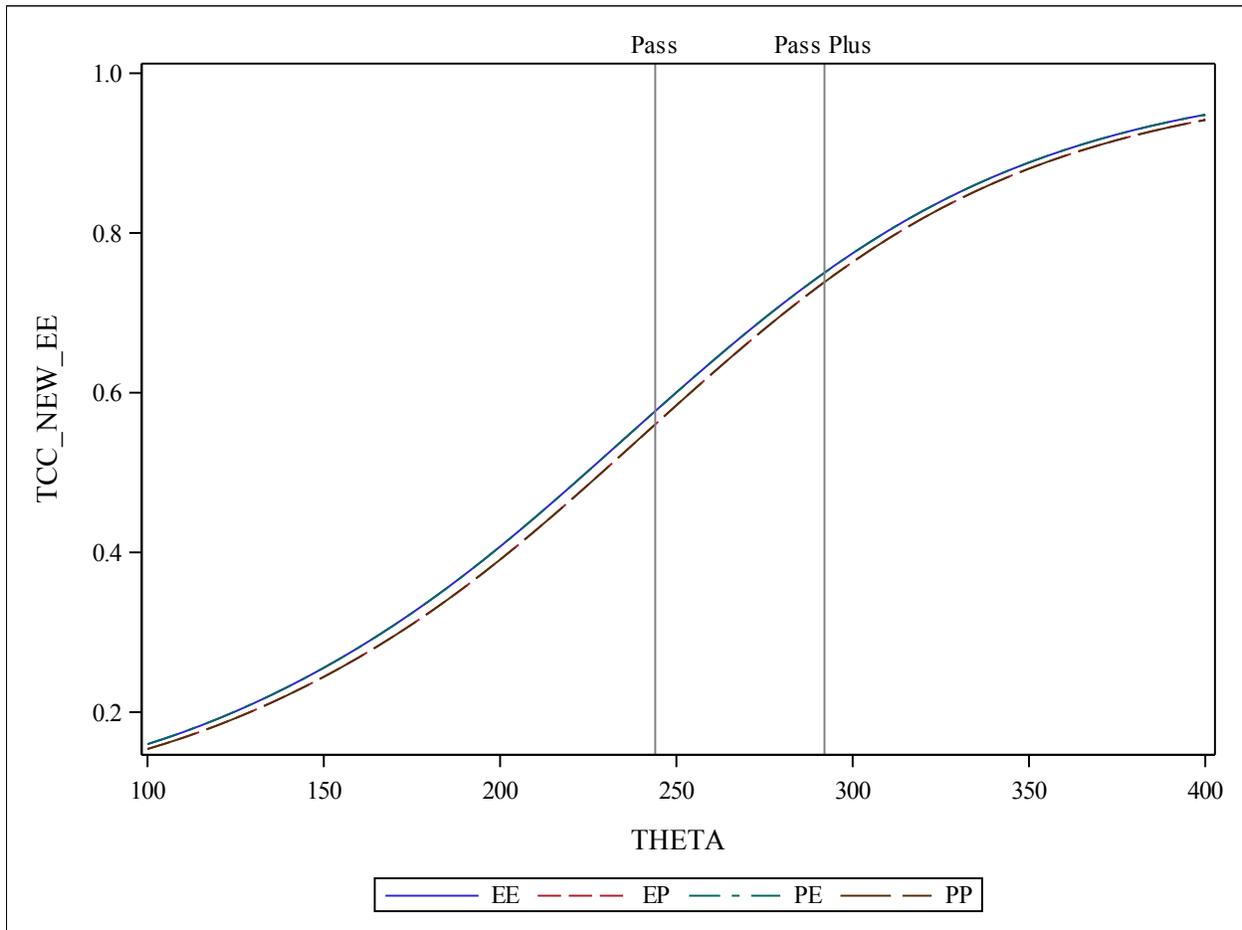


Figure 14. Test Characteristic Curves for ISTEP+ Mathematics Grade 10 Spring Retest 2019

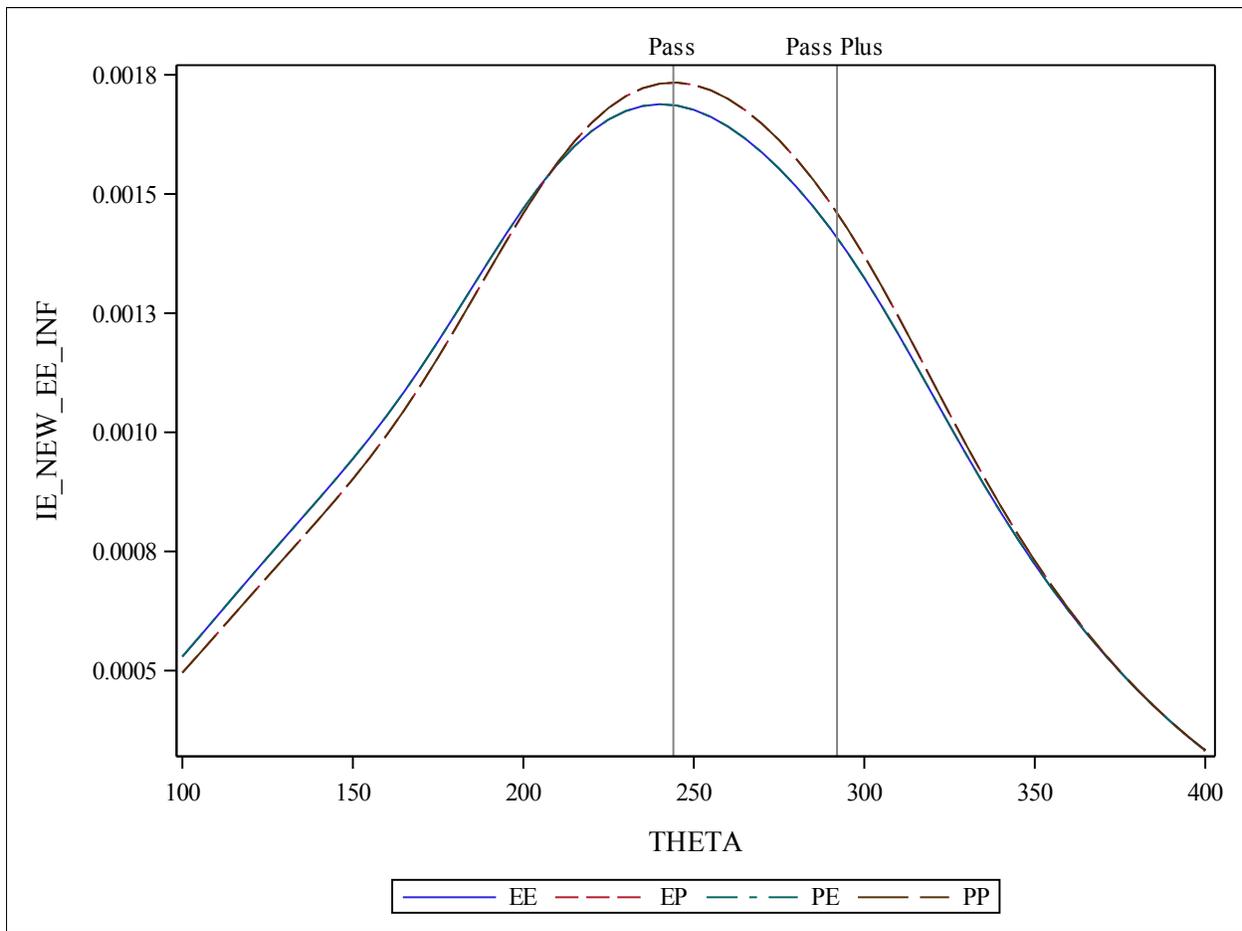


Figure 15. Test Information Curves for ISTEP+ Mathematics Grade 10 Spring Retest 2019

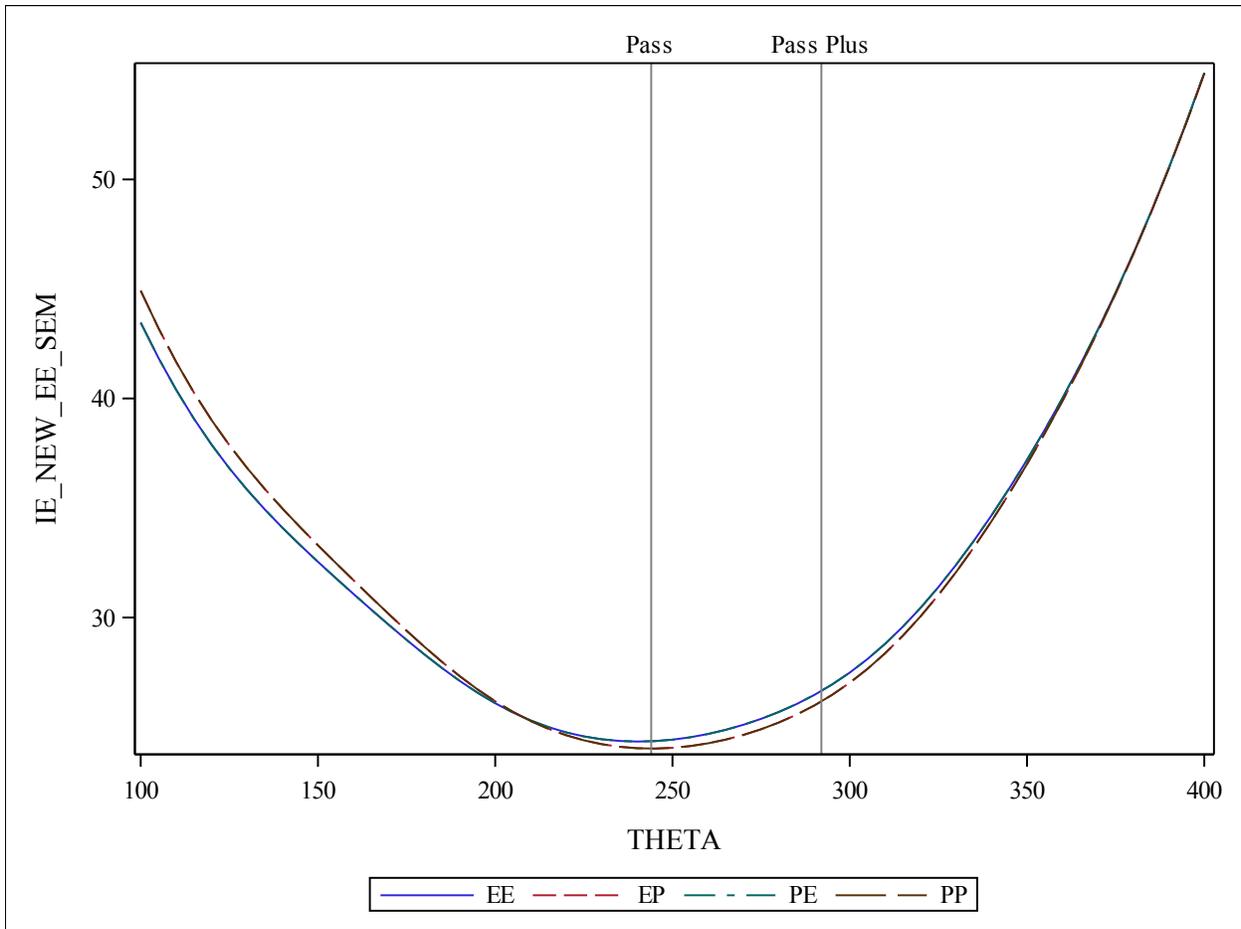


Figure 16. Standard Error of Measure Curves for ISTEP+ Mathematics Grade 10 Spring Retest 2019

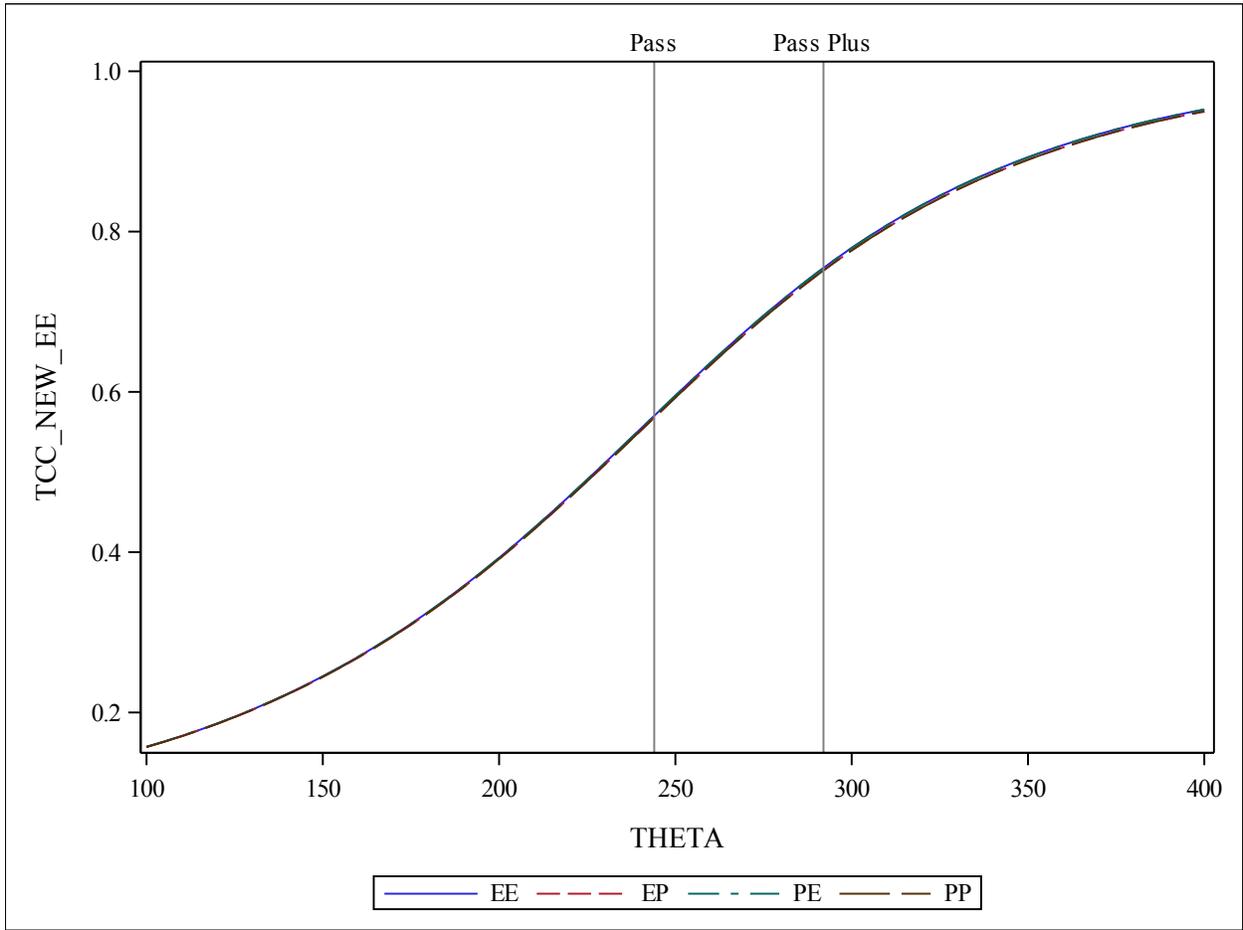


Figure 17. Test Characteristic Curves for ISTEP+ E/LA Grade 10 Spring 2019

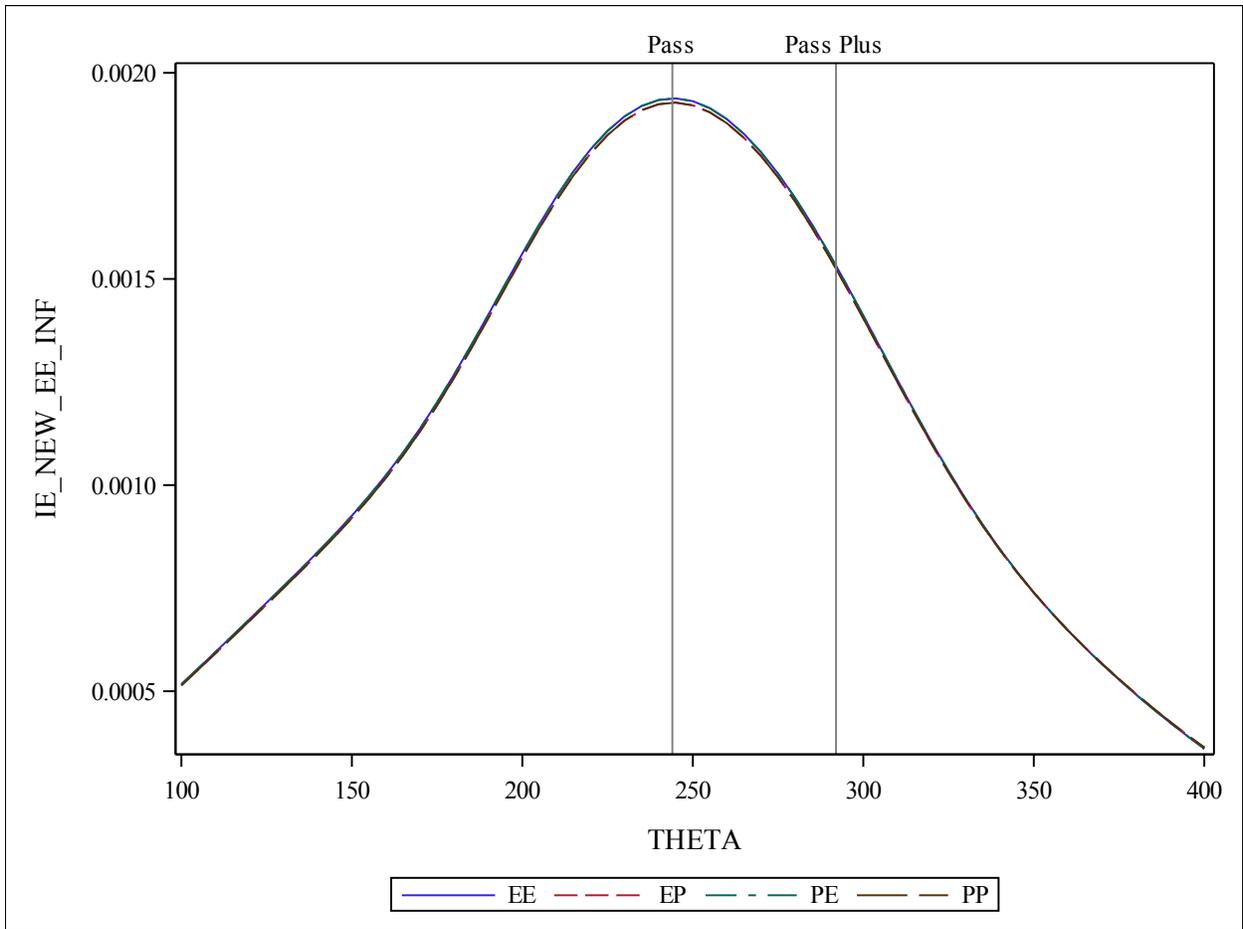


Figure 18. Test Information Curves for ISTEP+ E/LA Grade 10 Spring 2019

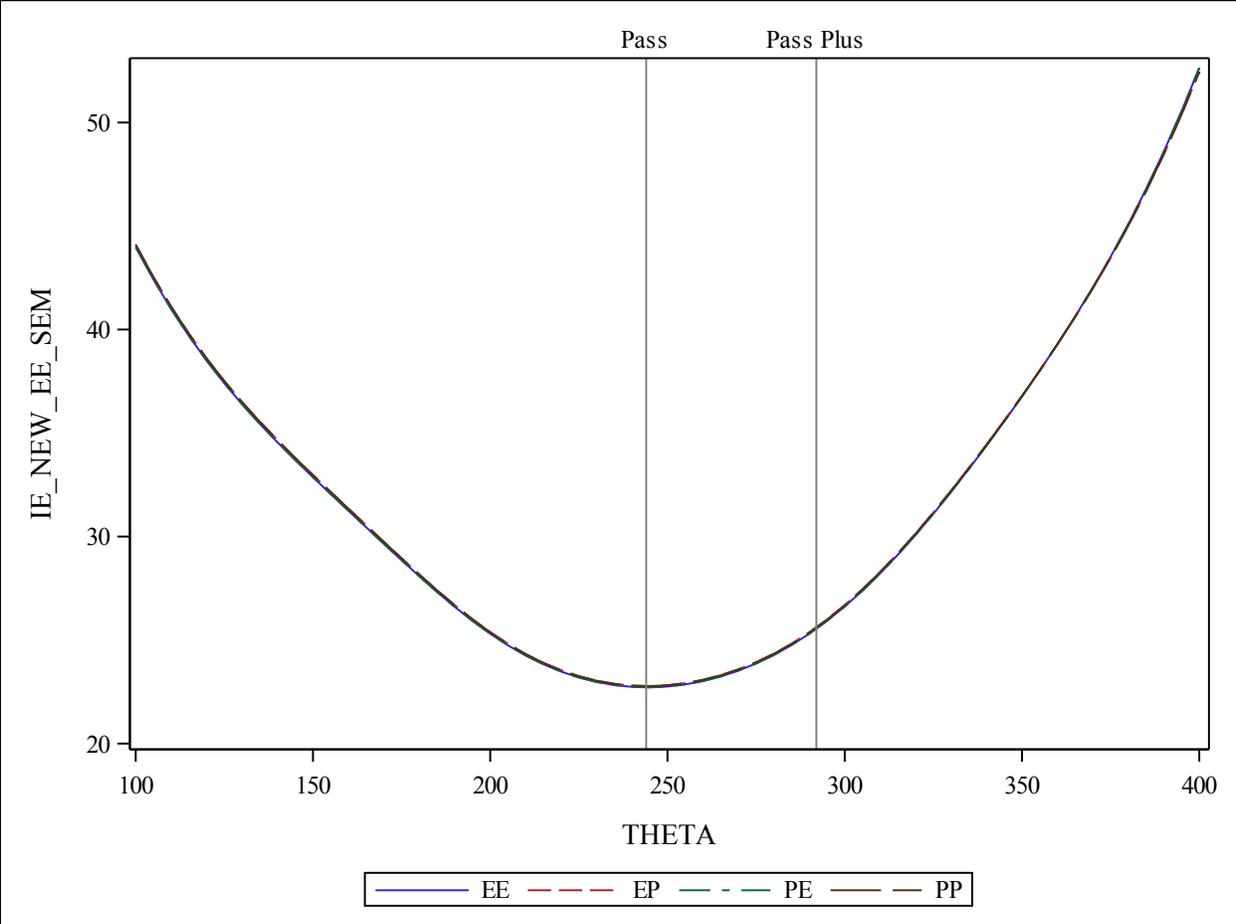


Figure 19. Standard Error of Measure Curves for ISTEP+ E/LA Grade 10 Spring 2019

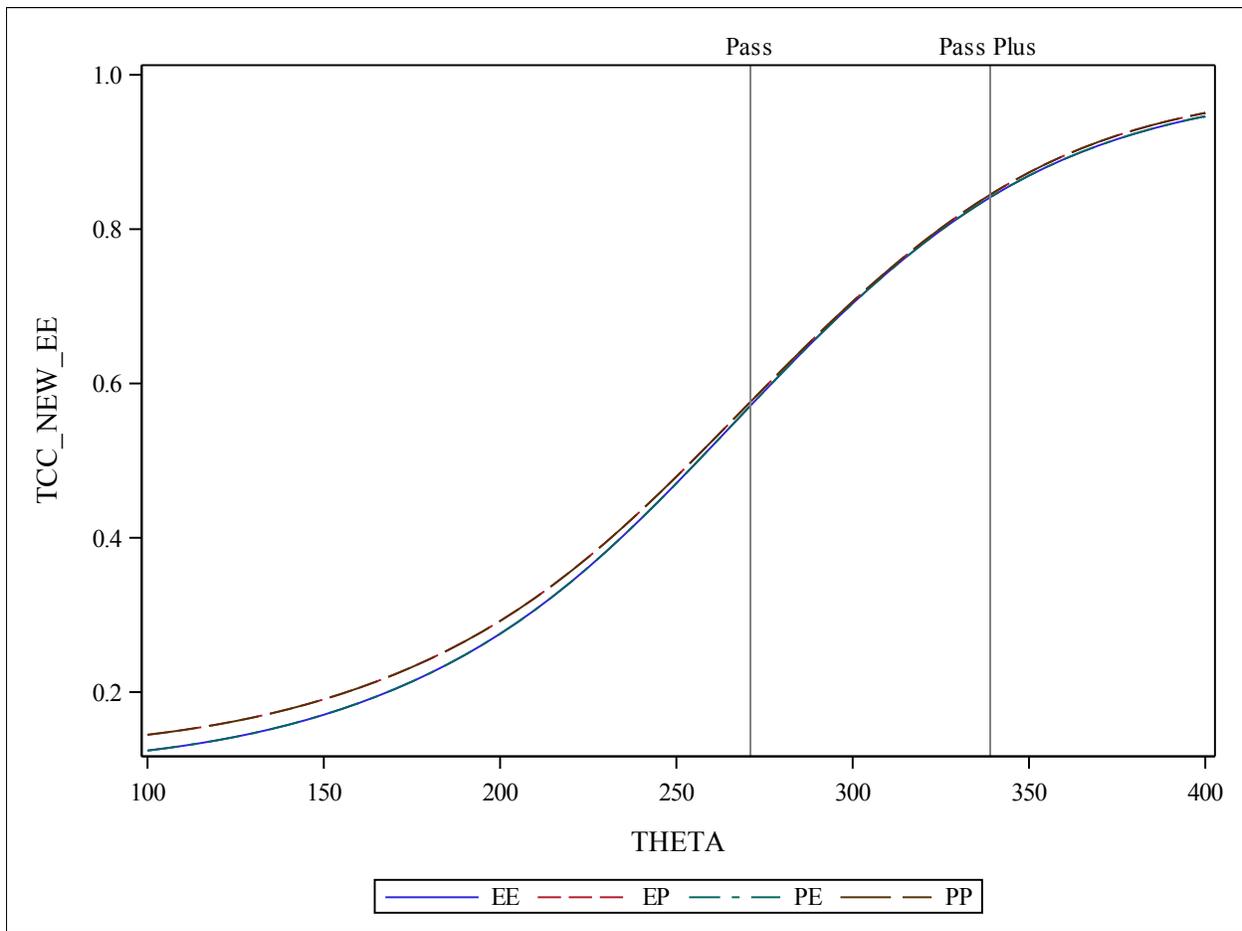


Figure 20. Test Characteristic Curves for ISTEP+ Mathematics Grade 10 Spring 2019

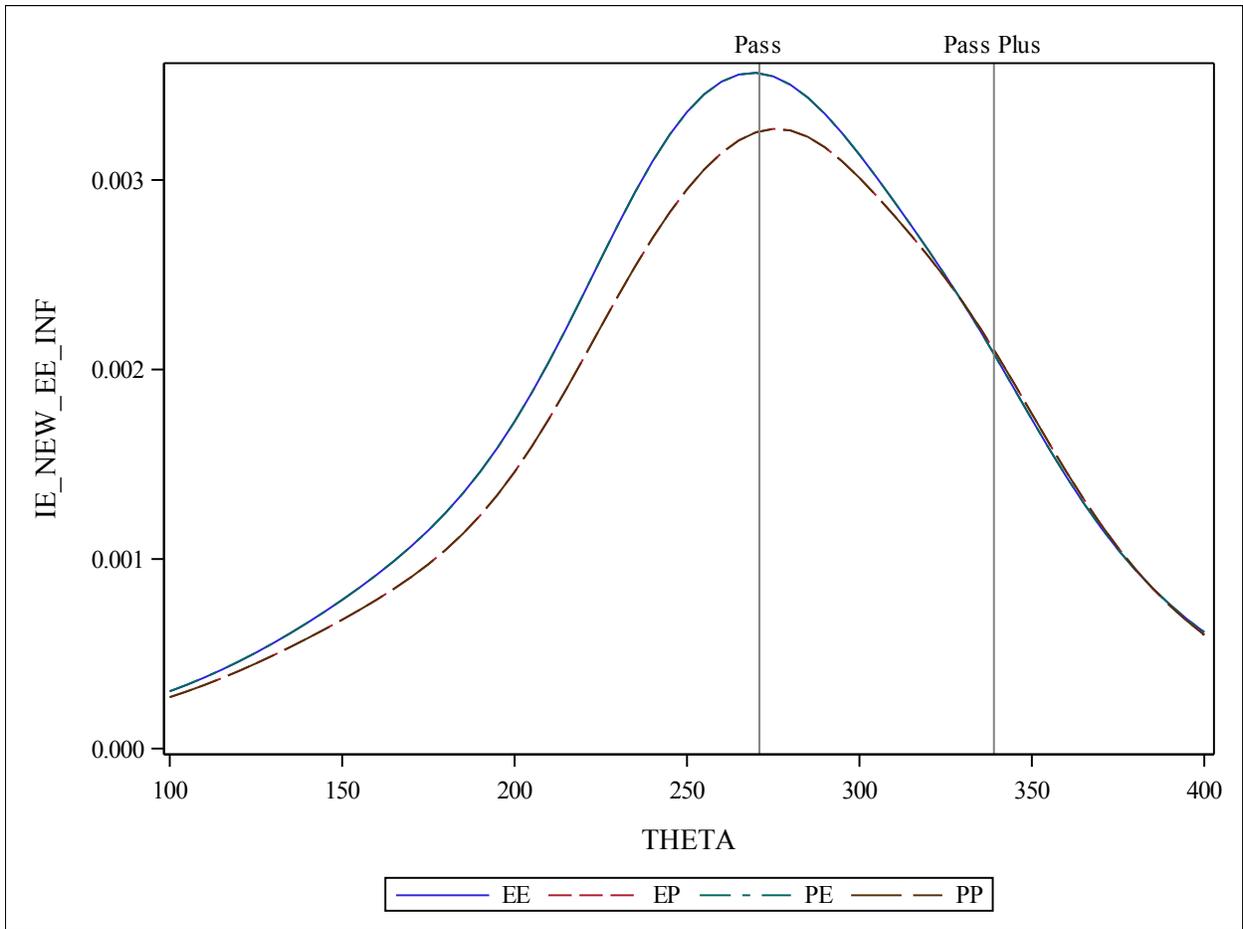


Figure 21. Test Information Curves for ISTEP+ Mathematics Grade 10 Spring 2019

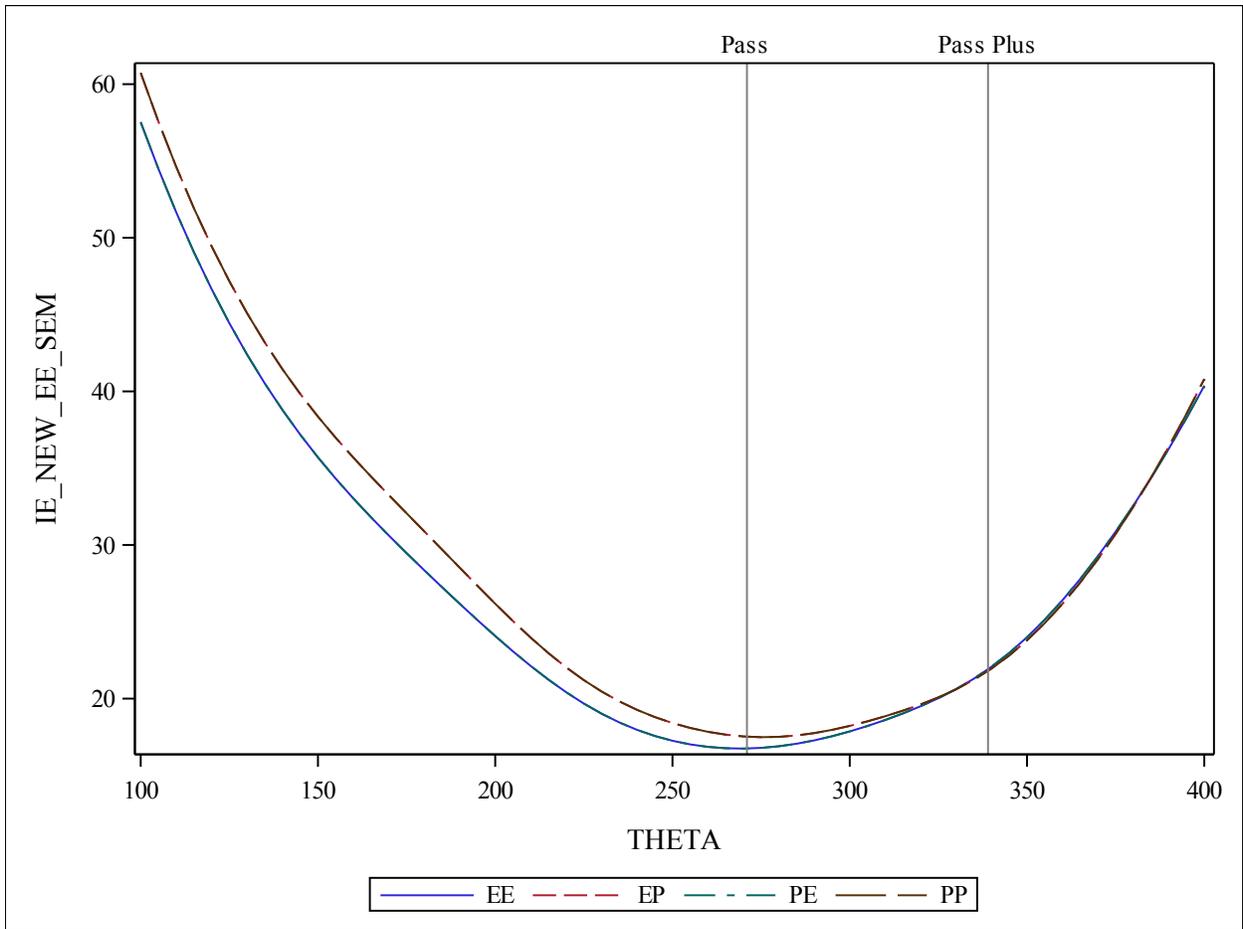


Figure 22. Standard Error of Measure Curves for ISTEP+ Mathematics Grade 10 Spring 2019

Appendix H: ISTEP+ Classification Consistency and Accuracy Statistics

Table 30. Classification Consistency and Accuracy – E/LA Winter 2018

Mode	Level	Consistency	Chance	Kappa	Accuracy	False Positive	False Negative
Online-Online	Pass	0.82020	0.61812	0.52917	0.87058	0.05203	0.07739
	PassP	0.93969	0.89824	0.40737	0.95953	0.00434	0.03612
Paper-Paper	Pass	0.85486	0.67199	0.55752	0.89633	0.02640	0.07726
	PassP	0.95343	0.91028	0.48091	0.96769	0.00272	0.02959

Table 31. Classification Consistency and Accuracy – Mathematics Winter 2018

Mode	Level	Consistency	Chance	Kappa	Accuracy	False Positive	False Negative
Online-Online	Pass	0.89637	0.81302	0.44574	0.90048	0.07470	0.02481
	PassP	0.99691	0.99482	0.40304	0.99782	0.00110	0.00107
Paper-Paper	Pass	0.88745	0.76895	0.51288	0.90760	0.02780	0.03329
	PassP	0.96479	0.92199	0.54860	0.96643	0.00077	0.00149

Table 32. Classification Consistency and Accuracy – E/LA Sprint Retest 2019

Mode	Level	Consistency	Chance	Kappa	Accuracy	False Positive	False Negative
Online-Online	Pass	0.74531	0.62337	0.32379	0.79429	0.13055	0.07474
	PassP	0.92075	0.90447	0.17034	0.95088	0.00436	0.04434
Paper-Online	Pass	0.74676	0.62322	0.32788	0.79315	0.13047	0.07596
	PassP	0.92251	0.90585	0.17694	0.95124	0.00434	0.04400
Online-Paper	Pass	0.74549	0.62454	0.32213	0.79513	0.13079	0.07365
	PassP	0.92444	0.90728	0.18515	0.95313	0.00434	0.04210
Online-Paper	Pass	0.74274	0.62386	0.31603	0.79348	0.13108	0.07502
	PassP	0.92337	0.90626	0.18253	0.95202	0.00440	0.04315

Table 33. Classification Consistency and Accuracy – Mathematics Sprint Retest 2019

Mode	Level	Consistency	Chance	Kappa	Accuracy	False Positive	False Negative
Online-Online	Pass	0.84001	0.73884	0.38739	0.88451	0.01598	0.09909
	PassP	0.99606	0.99488	0.23027	0.99750	0.00026	0.00181
Paper-Online	Pass	0.83644	0.73650	0.37926	0.87939	0.01649	0.10369
	PassP	0.99581	0.99461	0.22314	0.99739	0.00026	0.00193
Online-Paper	Pass	0.89761	0.84431	0.34238	0.92473	0.02904	0.04579
	PassP	0.99718	0.99597	0.29969	0.99811	0.00024	0.00122
Online-Paper	Pass	0.89695	0.84438	0.33781	0.92534	0.02920	0.04503
	PassP	0.99715	0.99592	0.30149	0.99817	0.00023	0.00117

Table 34. Classification Consistency and Accuracy – E/LA Spring 2019

Mode	Level	Consistency	Chance	Kappa	Accuracy	False Positive	False Negative
Online-Online	Pass	0.84574	0.50923	0.68568	0.79292	0.20136	0.00390
	PassP	0.91970	0.85822	0.43364	0.80290	0.19340	0.00189
Paper-Online	Pass	0.84398	0.50887	0.68233	0.79487	0.19907	0.00425
	PassP	0.91999	0.85936	0.43113	0.80323	0.19297	0.00198
Online-Paper	Pass	0.84161	0.51452	0.67374	0.77955	0.21538	0.00326
	PassP	0.92796	0.88176	0.39071	0.79185	0.20532	0.00102
Online-Paper	Pass	0.84325	0.51538	0.67654	0.77778	0.21735	0.00306
	PassP	0.92781	0.88072	0.39481	0.79224	0.20482	0.00113

Table 35. Classification Consistency and Accuracy – Mathematics Grade 10 Spring 2019

Mode	Level	Consistency	Chance	Kappa	Accuracy	False Positive	False Negative
Online-Online	Pass	0.91040	0.68160	0.71861	0.83786	0.15658	0.00167
	PassP	0.98450	0.97645	0.34178	0.95748	0.03820	0.00043
Paper-Online	Pass	0.90934	0.68102	0.71578	0.83844	0.15571	0.00196
	PassP	0.98474	0.97658	0.34840	0.95746	0.03825	0.00040
Online-Paper	Pass	0.90288	0.60059	0.75684	0.90021	0.08712	0.00879
	PassP	0.97350	0.95118	0.45705	0.96611	0.02731	0.00270
Online-Paper	Pass	0.90274	0.60113	0.75616	0.89740	0.08888	0.00984
	PassP	0.97310	0.95081	0.45311	0.96574	0.02746	0.00291

Appendix I: ISTEP+ Summary Data for Test and Reporting Categories

Table 36. Summary Data E/LA Grade 10 Winter 2018

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
All	Overall Test	41,786	83	61	26.56	9.25	0.86	3.47	0.23	0.20	0.34
	10.RLV		27	24	7.67	3.49	0.65	2.07	0.22	0.19	0.33
	10.RNV		37	23	9.65	4.51	0.76	2.21	0.23	0.20	0.36
	10.WC		8	8	4.08	1.54	0.24	1.35	0.29	0.27	0.39
	10.WG		11	12	4.56	1.95	0.37	1.55	0.20	0.18	0.28
Male	Overall Test	24,304	83	61	25.80	9.41	0.86	3.48	0.23	0.20	0.34
	10.RLV		27	24	7.40	3.50	0.65	2.07	0.21	0.19	0.33
	10.RNV		37	23	9.45	4.56	0.77	2.20	0.22	0.19	0.37
	10.WC		8	8	3.93	1.55	0.23	1.36	0.28	0.26	0.38
	10.WG		11	12	4.41	2.00	0.39	1.56	0.19	0.17	0.28
Female	Overall Test	17,476	83	61	27.62	8.92	0.85	3.42	0.24	0.21	0.34
	10.RLV		27	24	8.03	3.45	0.64	2.07	0.23	0.20	0.33
	10.RNV		37	23	9.94	4.42	0.75	2.23	0.24	0.20	0.36
	10.WC		8	8	4.30	1.51	0.26	1.30	0.31	0.28	0.40
	10.WG		11	12	4.77	1.86	0.34	1.52	0.21	0.19	0.28
American Indian / Alaska Native	Overall Test	104	83	47	27.09	8.41	0.83	3.48	0.24	0.21	0.33
	10.RLV		27	16	7.83	3.25	0.58	2.11	0.22	0.20	0.33
	10.RNV		37	20	10.01	4.28	0.73	2.24	0.24	0.21	0.37
	10.WC		8	7	4.03	1.38	0.21	1.23	0.29	0.27	0.37
	10.WG		11	8	4.75	1.61	0.12	1.51	0.21	0.19	0.26

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
African American	Overall Test	8,180	83	56	24.79	8.71	0.85	3.34	0.22	0.20	0.33
	10.RLV		27	19	7.17	3.46	0.65	2.05	0.20	0.19	0.33
	10.RNV		37	23	8.71	4.31	0.75	2.17	0.21	0.19	0.36
	10.WC		8	7	3.80	1.51	0.21	1.35	0.27	0.26	0.38
	10.WG		11	11	4.14	1.85	0.33	1.51	0.18	0.17	0.27
Asian	Overall Test	1,005	83	59	26.20	9.76	0.86	3.61	0.23	0.19	0.34
	10.RLV		27	19	7.72	3.33	0.62	2.05	0.22	0.19	0.33
	10.RNV		37	23	9.25	4.61	0.77	2.22	0.22	0.18	0.36
	10.WC		8	7	4.33	1.69	0.32	1.39	0.31	0.28	0.42
	10.WG		11	11	4.58	2.02	0.36	1.62	0.20	0.17	0.29
Hispanic	Overall Test	6,513	83	56	25.81	8.89	0.85	3.39	0.23	0.20	0.33
	10.RLV		27	19	7.56	3.47	0.65	2.06	0.22	0.20	0.33
	10.RNV		37	23	9.14	4.33	0.74	2.19	0.22	0.19	0.35
	10.WC		8	7	3.97	1.55	0.23	1.36	0.28	0.27	0.39
	10.WG		11	12	4.40	1.91	0.35	1.54	0.19	0.18	0.28
White (non-Hispanic)	Overall Test	24,030	83	61	27.40	9.41	0.86	3.52	0.24	0.20	0.34
	10.RLV		27	24	7.88	3.50	0.65	2.08	0.23	0.19	0.33
	10.RNV		37	23	10.15	4.55	0.76	2.23	0.24	0.20	0.37
	10.WC		8	8	4.21	1.53	0.24	1.33	0.30	0.27	0.39
	10.WG		11	12	4.75	1.97	0.38	1.56	0.21	0.18	0.28
Multi-racial	Overall Test	1,868	83	61	26.33	9.19	0.86	3.46	0.23	0.20	0.34
	10.RLV		27	24	7.58	3.50	0.65	2.07	0.22	0.19	0.33
	10.RNV		37	23	9.63	4.51	0.76	2.21	0.23	0.20	0.37
	10.WC		8	8	4.05	1.53	0.23	1.34	0.29	0.27	0.39
	10.WG		11	12	4.48	1.93	0.36	1.54	0.19	0.18	0.28

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hawaiian/Pacific Islander	Overall Test	24	83	44	27.58	8.33	0.81	3.58	0.24	0.21	0.34
	10.RLV		27	16	8.33	3.32	0.61	2.08	0.24	0.20	0.37
	10.RNV		37	17	10.17	3.67	0.63	2.24	0.24	0.21	0.34
	10.WC		8	7	4.13	1.30	0.13	1.21	0.29	0.27	0.36
	10.WG		11	9	4.96	1.92	0.43	1.45	0.22	0.20	0.28
Special Education	Overall Test	10,516	83	61	21.34	8.11	0.82	3.39	0.19	0.18	0.33
	10.RLV		27	19	6.32	3.13	0.58	2.02	0.18	0.17	0.33
	10.RNV		37	23	7.46	3.97	0.72	2.11	0.18	0.17	0.36
	10.WC		8	7	3.40	1.46	0.15	1.34	0.24	0.24	0.37
	10.WG		11	12	3.72	1.80	0.32	1.49	0.16	0.16	0.27
SES (High)	Overall Test	18,238	83	61	27.81	9.31	0.86	3.49	0.24	0.21	0.34
	10.RLV		27	24	7.97	3.49	0.65	2.08	0.23	0.20	0.33
	10.RNV		37	23	10.26	4.56	0.76	2.23	0.24	0.20	0.37
	10.WC		8	8	4.28	1.52	0.25	1.32	0.31	0.28	0.40
	10.WG		11	12	4.81	1.95	0.37	1.55	0.21	0.19	0.28
SES (Low)	Overall Test	22,713	83	61	25.40	8.91	0.85	3.44	0.22	0.20	0.33
	10.RLV		27	24	7.38	3.43	0.64	2.07	0.21	0.19	0.33
	10.RNV		37	23	9.12	4.33	0.74	2.19	0.22	0.19	0.36
	10.WC		8	8	3.92	1.53	0.22	1.35	0.28	0.26	0.39
	10.WG		11	12	4.34	1.90	0.34	1.54	0.19	0.17	0.27
ELL/LEP	Overall Test	3,296	83	61	22.22	8.16	0.83	3.36	0.19	0.18	0.33
	10.RLV		27	24	6.51	3.14	0.59	2.01	0.19	0.18	0.31
	10.RNV		37	23	7.55	3.90	0.70	2.13	0.18	0.17	0.34
	10.WC		8	8	3.58	1.58	0.22	1.39	0.26	0.25	0.40
	10.WG		11	12	3.85	1.85	0.29	1.56	0.17	0.16	0.27

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Section 504	Overall Test	1,078	83	61	27.07	9.32	0.86	3.53	0.24	0.20	0.34
	10.RLV		27	19	7.79	3.51	0.65	2.08	0.22	0.19	0.33
	10.RNV		37	23	10.04	4.51	0.75	2.23	0.24	0.20	0.37
	10.WC		8	7	4.17	1.51	0.23	1.32	0.30	0.27	0.39
	10.WG		11	12	4.66	1.92	0.35	1.55	0.20	0.18	0.28

Table 37. Summary Data Mathematics Grade 10 Winter 2018

Population	Reporting Category	N	Number of Points						Average p-value		
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
All	Overall Test	69,490	119	73	27.31	10.16	0.87	3.60	0.17	0.15	0.27
	10.DS		12	11	3.88	1.65	0.41	1.26	0.19	0.18	0.28
	10.GM		8	6	1.89	1.13	0.25	0.98	0.21	0.20	0.31
	10.LEI		47	25	10.30	4.25	0.71	2.28	0.18	0.17	0.30
	10.MP		8	9	1.66	1.39	0.47	1.01	0.09	0.08	0.19
	10.NSC		19	11	5.00	2.22	0.52	1.54	0.24	0.22	0.36
	10.QEF		13	9	1.99	1.38	0.30	1.15	0.09	0.08	0.16
	10.SEI		12	7	2.19	1.56	0.49	1.12	0.17	0.14	0.33
Male	Overall Test	35,360	119	73	26.80	10.35	0.88	3.58	0.17	0.15	0.28
	10.DS		12	11	3.84	1.69	0.44	1.27	0.19	0.18	0.28
	10.GM		8	6	1.87	1.13	0.26	0.97	0.21	0.20	0.31
	10.LEI		47	25	10.14	4.32	0.72	2.28	0.18	0.16	0.30
	10.MP		8	9	1.52	1.36	0.47	0.99	0.08	0.07	0.19
	10.NSC		19	11	4.94	2.29	0.55	1.54	0.24	0.22	0.37
	10.QEF		13	9	1.96	1.38	0.31	1.14	0.09	0.07	0.16
	10.SEI		12	7	2.06	1.55	0.49	1.11	0.16	0.13	0.33
Female	Overall Test	34,129	119	73	27.84	9.93	0.87	3.61	0.17	0.15	0.27
	10.DS		12	11	3.92	1.61	0.38	1.26	0.20	0.18	0.27
	10.GM		8	6	1.90	1.12	0.24	0.98	0.21	0.20	0.30
	10.LEI		47	25	10.48	4.18	0.70	2.29	0.19	0.17	0.30
	10.MP		8	9	1.80	1.40	0.46	1.02	0.10	0.08	0.20
	10.NSC		19	11	5.07	2.14	0.48	1.54	0.24	0.22	0.35
	10.QEF		13	9	2.01	1.39	0.30	1.16	0.09	0.08	0.16
	10.SEI		12	7	2.32	1.56	0.48	1.12	0.18	0.15	0.33

Population	Reporting Category	N	Number of Points						Average p-value		
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
American Indian / Alaska Native	Overall Test	153	119	51	27.36	9.09	0.85	3.53	0.17	0.16	0.27
	10.DS		12	8	3.87	1.50	0.34	1.22	0.19	0.19	0.26
	10.GM		8	4	1.84	1.09	0.20	0.98	0.20	0.19	0.28
	10.LEI		47	21	10.39	4.04	0.69	2.26	0.19	0.17	0.29
	10.MP		8	5	1.68	1.27	0.43	0.96	0.09	0.08	0.19
	10.NSC		19	10	5.14	2.13	0.47	1.55	0.24	0.23	0.35
	10.QEF		13	7	1.89	1.35	0.32	1.11	0.08	0.07	0.15
	10.SEI		12	6	2.04	1.51	0.48	1.09	0.16	0.14	0.31
African American	Overall Test	11,699	119	73	23.83	9.18	0.86	3.46	0.15	0.14	0.27
	10.DS		12	8	3.41	1.58	0.38	1.24	0.17	0.17	0.27
	10.GM		8	4	1.61	1.06	0.17	0.97	0.18	0.17	0.28
	10.LEI		47	25	8.95	4.00	0.69	2.23	0.16	0.15	0.29
	10.MP		8	9	1.29	1.25	0.45	0.93	0.07	0.06	0.19
	10.NSC		19	11	4.32	2.11	0.48	1.52	0.21	0.20	0.35
	10.QEF		13	9	1.76	1.30	0.26	1.12	0.08	0.07	0.16
	10.SEI		12	7	1.87	1.45	0.44	1.08	0.14	0.13	0.32
Asian	Overall Test	1,192	119	69	29.42	12.07	0.91	3.69	0.18	0.15	0.29
	10.DS		12	8	3.88	1.76	0.47	1.28	0.19	0.17	0.28
	10.GM		8	4	1.88	1.13	0.25	0.97	0.21	0.18	0.31
	10.LEI		47	25	11.07	5.06	0.79	2.32	0.20	0.16	0.32
	10.MP		8	9	1.95	1.52	0.49	1.09	0.11	0.08	0.20
	10.NSC		19	11	5.53	2.41	0.60	1.52	0.26	0.23	0.38
	10.QEF		13	9	2.36	1.68	0.43	1.27	0.10	0.08	0.18
	10.SEI		12	6	2.47	1.63	0.52	1.13	0.19	0.15	0.33

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hispanic	Overall Test	10,036	119	71	26.04	9.55	0.86	3.52	0.16	0.15	0.27
	10.DS		12	8	3.71	1.62	0.40	1.25	0.19	0.18	0.27
	10.GM		8	4	1.76	1.10	0.21	0.98	0.20	0.19	0.29
	10.LEI		47	25	9.84	4.13	0.70	2.27	0.18	0.16	0.30
	10.MP		8	9	1.52	1.30	0.44	0.97	0.08	0.07	0.19
	10.NSC		19	11	4.77	2.15	0.49	1.54	0.23	0.22	0.36
	10.QEF		13	9	1.85	1.32	0.28	1.12	0.08	0.07	0.15
	10.SEI		12	7	2.06	1.50	0.46	1.10	0.16	0.14	0.32
White (non-Hispanic)	Overall Test	43,042	119	73	28.55	10.26	0.87	3.64	0.18	0.16	0.28
	10.DS		12	11	4.06	1.65	0.41	1.27	0.20	0.19	0.28
	10.GM		8	6	2.00	1.14	0.26	0.97	0.22	0.20	0.31
	10.LEI		47	25	10.80	4.24	0.71	2.30	0.19	0.17	0.30
	10.MP		8	9	1.79	1.42	0.47	1.03	0.10	0.08	0.19
	10.NSC		19	11	5.25	2.21	0.51	1.55	0.25	0.23	0.36
	10.QEF		13	9	2.08	1.41	0.31	1.17	0.09	0.08	0.16
	10.SEI		12	7	2.30	1.59	0.50	1.13	0.18	0.15	0.33
Multi-racial	Overall Test	3,273	119	73	26.59	9.78	0.87	3.55	0.17	0.15	0.27
	10.DS		12	11	3.78	1.61	0.39	1.26	0.19	0.18	0.27
	10.GM		8	6	1.83	1.11	0.22	0.98	0.20	0.19	0.31
	10.LEI		47	25	10.03	4.17	0.70	2.27	0.18	0.16	0.29
	10.MP		8	9	1.56	1.35	0.47	0.98	0.09	0.07	0.19
	10.NSC		19	11	4.84	2.19	0.50	1.54	0.23	0.22	0.35
	10.QEF		13	9	1.95	1.33	0.28	1.13	0.08	0.08	0.16
	10.SEI		12	7	2.12	1.53	0.47	1.11	0.16	0.14	0.32

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hawaiian/Pacific Islander	Overall Test	42	119	51	28.17	10.75	0.88	3.69	0.18	0.15	0.28
	10.DS		12	6	3.76	1.38	0.14	1.28	0.19	0.17	0.25
	10.GM		8	4	1.90	1.08	0.15	0.99	0.21	0.20	0.25
	10.LEI		47	21	10.90	4.64	0.76	2.26	0.19	0.17	0.32
	10.MP		8	4	1.74	1.31	0.38	1.03	0.10	0.08	0.18
	10.NSC		19	10	5.40	2.46	0.61	1.54	0.26	0.23	0.38
	10.QEF		13	7	2.19	1.47	0.41	1.13	0.10	0.08	0.16
	10.SEI		12	5	2.26	1.58	0.52	1.09	0.17	0.14	0.31
Special Education	Overall Test	12,082	119	73	21.08	8.27	0.83	3.41	0.13	0.13	0.27
	10.DS		12	8	3.03	1.51	0.33	1.24	0.15	0.15	0.27
	10.GM		8	4	1.48	1.04	0.17	0.95	0.16	0.16	0.29
	10.LEI		47	25	8.16	3.63	0.64	2.19	0.15	0.14	0.30
	10.MP		8	9	0.94	1.11	0.42	0.84	0.05	0.05	0.17
	10.NSC		19	11	4.28	2.13	0.48	1.53	0.20	0.20	0.38
	10.QEF		13	9	1.44	1.11	0.15	1.02	0.06	0.06	0.14
	10.SEI		12	7	1.39	1.25	0.35	1.01	0.11	0.10	0.31
SES (High)	Overall Test	34,245	119	73	29.18	10.32	0.87	3.65	0.18	0.16	0.28
	10.DS		12	11	4.13	1.64	0.40	1.27	0.21	0.19	0.28
	10.GM		8	6	2.00	1.13	0.26	0.97	0.22	0.20	0.31
	10.LEI		47	25	11.00	4.28	0.71	2.30	0.20	0.17	0.30
	10.MP		8	9	1.86	1.43	0.46	1.05	0.10	0.08	0.19
	10.NSC		19	11	5.33	2.23	0.52	1.54	0.25	0.23	0.36
	10.QEF		13	9	2.15	1.43	0.31	1.18	0.09	0.08	0.16
	10.SEI		12	7	2.39	1.60	0.50	1.13	0.18	0.15	0.33

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
SES (Low)	Overall Test	34,373	119	73	25.40	9.49	0.86	3.53	0.16	0.15	0.27
	10.DS		12	11	3.63	1.61	0.39	1.25	0.18	0.17	0.27
	10.GM		8	6	1.77	1.11	0.23	0.97	0.20	0.19	0.30
	10.LEI		47	25	9.61	4.06	0.69	2.26	0.17	0.16	0.29
	10.MP		8	9	1.45	1.29	0.45	0.96	0.08	0.07	0.18
	10.NSC		19	11	4.68	2.15	0.49	1.54	0.22	0.21	0.36
	10.QEF		13	9	1.81	1.30	0.27	1.11	0.08	0.07	0.15
	10.SEI		12	7	1.98	1.49	0.46	1.10	0.15	0.14	0.32
ELL/LEP	Overall Test	3,377	119	73	22.14	9.12	0.86	3.41	0.14	0.13	0.28
	10.DS		12	11	3.06	1.55	0.36	1.25	0.15	0.15	0.26
	10.GM		8	6	1.48	1.05	0.17	0.95	0.16	0.16	0.29
	10.LEI		47	25	8.35	4.05	0.70	2.21	0.15	0.14	0.31
	10.MP		8	9	1.13	1.21	0.44	0.91	0.06	0.06	0.19
	10.NSC		19	11	4.26	2.15	0.50	1.52	0.20	0.20	0.38
	10.QEF		13	9	1.60	1.25	0.28	1.06	0.07	0.07	0.16
	10.SEI		12	7	1.61	1.34	0.39	1.04	0.12	0.12	0.31
Section 504	Overall Test	1,733	119	73	27.73	10.25	0.87	3.67	0.17	0.15	0.28
	10.DS		12	8	3.90	1.63	0.39	1.27	0.20	0.18	0.28
	10.GM		8	4	1.94	1.13	0.25	0.98	0.22	0.20	0.30
	10.LEI		47	25	10.53	4.23	0.70	2.30	0.19	0.17	0.30
	10.MP		8	9	1.73	1.39	0.46	1.02	0.10	0.08	0.19
	10.NSC		19	11	5.13	2.23	0.52	1.55	0.24	0.22	0.37
	10.QEF		13	9	2.03	1.40	0.28	1.19	0.09	0.08	0.16
	10.SEI		12	7	2.22	1.56	0.48	1.13	0.17	0.14	0.33

Note. * indicates that since some items were correlated for certain subgroups and reporting categories, missing results occurred.

Table 38. Summary Data E/LA Grade 10 Spring Retest 2019

Population	Reporting Category	N	Number of Points					Average p-value			
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
All	Overall Test	22,177	51	73	21.66	8.12	0.78	3.82	0.33	0.31	0.43
	10.RLV		18	14	4.57	2.47	0.59	1.59	0.18	0.18	0.25
	10.RNV		22	21	6.90	3.10	0.62	1.91	0.31	0.31	0.41
	10.WC		5	5	1.56	1.04	0.18	0.94	0.19	0.19	0.24
	10.WG		6	5	1.08	0.90	0.22	0.79	0.10	0.09	0.14
Male	Overall Test	12,841	51	73	21.48	8.26	0.78	3.86	0.33	0.31	0.45
	10.RLV		18	14	4.53	2.48	0.59	1.58	0.18	0.17	0.26
	10.RNV		22	21	6.87	3.09	0.62	1.90	0.31	0.31	0.42
	10.WC		5	5	1.56	1.06	0.20	0.94	0.20	0.19	0.25
	10.WG		6	5	1.04	0.89	0.21	0.79	0.09	0.09	0.14
Female	Overall Test	9,333	51	73	21.91	7.91	0.78	3.75	0.33	0.31	0.42
	10.RLV		18	14	4.64	2.45	0.58	1.59	0.19	0.18	0.25
	10.RNV		22	21	6.94	3.12	0.62	1.91	0.32	0.31	0.40
	10.WC		5	5	1.54	1.03	0.15	0.95	0.19	0.19	0.23
	10.WG		6	5	1.14	0.91	0.22	0.80	0.10	0.10	0.14
American Indian / Alaska Native	Overall Test	59	51	37	22.15	7.54	0.80	3.41	0.34	0.31	0.43
	10.RLV		18	9	4.69	2.40	0.56	1.58	0.19	0.19	0.24
	10.RNV		22	14	6.94	3.29	0.67	1.88	0.32	0.31	0.45
	10.WC		5	4	1.72	1.06	0.19	0.95	0.21	0.23	0.26
	10.WG		6	3	1.14	0.85	0.16	0.78	0.10	0.10	0.13
African American	Overall Test	5,237	51	52	19.85	7.25	0.75	3.60	0.30	0.29	0.40
	10.RLV		18	14	4.15	2.38	0.57	1.56	0.17	0.16	0.24
	10.RNV		22	18	6.34	2.92	0.59	1.88	0.29	0.28	0.38
	10.WC		5	4	1.46	1.02	0.16	0.93	0.18	0.18	0.22
	10.WG		6	5	0.96	0.83	0.14	0.77	0.09	0.08	0.13

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Asian	Overall Test	483	51	67	25.25	11.20	0.86	4.21	0.38	0.35	0.54
	10.RLV		18	14	5.25	2.89	0.69	1.61	0.21	0.20	0.30
	10.RNV		22	19	8.46	3.79	0.73	1.96	0.38	0.37	0.51
	10.WC		5	5	1.66	1.14	0.33	0.93	0.21	0.19	0.28
	10.WG		6	5	1.46	1.06	0.38	0.84	0.13	0.12	0.21
Hispanic	Overall Test	3,704	51	66	21.89	7.89	0.77	3.79	0.33	0.32	0.42
	10.RLV		18	14	4.54	2.42	0.57	1.58	0.18	0.18	0.24
	10.RNV		22	19	7.04	3.06	0.61	1.92	0.32	0.31	0.40
	10.WC		5	5	1.57	1.06	0.20	0.95	0.20	0.19	0.25
	10.WG		6	5	1.08	0.87	0.20	0.78	0.10	0.10	0.13
White (non-Hispanic)	Overall Test	11,624	51	73	22.32	8.30	0.78	3.86	0.34	0.32	0.44
	10.RLV		18	14	4.76	2.49	0.58	1.60	0.19	0.18	0.26
	10.RNV		22	21	7.06	3.12	0.63	1.91	0.32	0.31	0.41
	10.WC		5	5	1.59	1.05	0.18	0.95	0.20	0.19	0.25
	10.WG		6	5	1.12	0.92	0.24	0.80	0.10	0.10	0.15
Multi- racial	Overall Test	1,017	51	61	21.00	7.65	0.75	3.82	0.32	0.30	0.42
	10.RLV		18	14	4.49	2.42	0.57	1.59	0.18	0.17	0.25
	10.RNV		22	17	6.75	3.00	0.60	1.90	0.31	0.30	0.40
	10.WC		5	4	1.55	1.03	0.15	0.95	0.19	0.19	0.23
	10.WG		6	4	1.04	0.87	0.18	0.79	0.09	0.09	0.13
Hawaiian/Pacific Islander	Overall Test	16	51	41	23.19	9.12	0.84	3.66	0.35	0.31	0.48
	10.RLV		18	10	5.19	2.51	0.56	1.66	0.21	0.17	0.31
	10.RNV		22	15	8.25	3.38	0.66	1.96	0.38	0.32	0.53
	10.WC		5	4	1.56	1.21	0.48	0.87	0.20	0.16	0.31
	10.WG		6	4	1.31	1.14	0.39	0.89	0.12	0.09	0.20

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Special Education	Overall Test	7,188	51	57	19.47	7.18	0.73	3.72	0.29	0.29	0.41
	10.RLV		18	13	4.24	2.26	0.53	1.55	0.17	0.17	0.25
	10.RNV		22	18	6.34	2.83	0.55	1.89	0.29	0.29	0.39
	10.WC		5	4	1.49	1.02	0.15	0.94	0.19	0.19	0.24
	10.WG		6	5	0.93	0.84	0.15	0.78	0.08	0.08	0.13
SES (High)	Overall Test	8,379	51	73	22.83	8.74	0.80	3.90	0.35	0.32	0.45
	10.RLV		18	14	4.87	2.57	0.61	1.60	0.19	0.18	0.27
	10.RNV		22	21	7.22	3.24	0.65	1.91	0.33	0.32	0.42
	10.WC		5	5	1.61	1.06	0.20	0.95	0.20	0.20	0.25
	10.WG		6	5	1.17	0.94	0.26	0.81	0.11	0.10	0.15
SES (Low)	Overall Test	13,559	51	65	20.95	7.59	0.76	3.74	0.32	0.30	0.42
	10.RLV		18	14	4.40	2.39	0.56	1.58	0.18	0.17	0.24
	10.RNV		22	19	6.71	2.99	0.59	1.90	0.30	0.30	0.39
	10.WC		5	5	1.52	1.03	0.17	0.94	0.19	0.19	0.24
	10.WG		6	5	1.02	0.86	0.18	0.78	0.09	0.09	0.13
ELL/LEP	Overall Test	2,527	51	67	21.36	8.34	0.78	3.89	0.32	0.31	0.45
	10.RLV		18	14	4.40	2.43	0.58	1.57	0.18	0.17	0.26
	10.RNV		22	19	7.11	3.19	0.64	1.93	0.32	0.32	0.44
	10.WC		5	5	1.54	1.04	0.17	0.95	0.19	0.19	0.25
	10.WG		6	5	1.07	0.87	0.22	0.77	0.10	0.10	0.15
Section 504	Overall Test	654	51	65	22.12	7.93	0.77	3.84	0.34	0.32	0.43
	10.RLV		18	14	4.70	2.39	0.55	1.61	0.19	0.18	0.25
	10.RNV		22	16	6.98	3.02	0.60	1.90	0.32	0.31	0.40
	10.WC		5	4	1.61	1.03	0.13	0.96	0.20	0.20	0.23
	10.WG		6	4	1.11	0.91	0.21	0.80	0.10	0.09	0.15

*Note. This table includes only first-time testers.

Table 39. Summary Data Mathematics Grade 10 Spring Retest 2019

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
All	Overall Test	42,412	70	73	25.36	8.97	0.86	3.32	0.27	0.26	0.47
	10.DS		8	11	5.33	2.18	0.50	1.53	0.33	0.33	0.50
	10.GM		7	8	2.17	1.30	0.32	1.08	0.22	0.21	0.38
	10.LEI		26	23	8.63	3.50	0.63	2.14	0.30	0.29	0.50
	10.MP		4	9	1.04	1.01	0.33	0.82	0.12	0.11	0.27
	10.NSC		12	11	4.13	2.21	0.59	1.41	0.30	0.29	0.54
	10.QEF		4	4	1.10	0.93	0.26	0.80	0.18	0.18	0.38
	10.SEI		9	7	2.19	1.63	0.53	1.12	0.24	0.23	0.54
Male	Overall Test	21,618	70	73	24.70	9.17	0.87	3.32	0.27	0.25	0.47
	10.DS		8	11	5.22	2.26	0.53	1.54	0.33	0.33	0.51
	10.GM		7	8	2.15	1.36	0.36	1.09	0.21	0.21	0.40
	10.LEI		26	23	8.44	3.53	0.64	2.13	0.29	0.29	0.50
	10.MP		4	9	0.96	1.01	0.36	0.81	0.11	0.10	0.27
	10.NSC		12	11	4.01	2.22	0.60	1.40	0.29	0.28	0.55
	10.QEF		4	4	1.07	0.92	0.24	0.80	0.18	0.17	0.37
	10.SEI		9	7	2.02	1.58	0.52	1.10	0.22	0.21	0.53
Female	Overall Test	20,790	70	73	26.04	8.69	0.85	3.32	0.28	0.27	0.46
	10.DS		8	11	5.45	2.08	0.47	1.52	0.34	0.34	0.49
	10.GM		7	8	2.19	1.23	0.27	1.06	0.22	0.21	0.36
	10.LEI		26	23	8.83	3.46	0.62	2.15	0.30	0.30	0.50
	10.MP		4	9	1.13	0.99	0.31	0.83	0.13	0.12	0.26
	10.NSC		12	11	4.26	2.18	0.58	1.41	0.30	0.29	0.54
	10.QEF		4	4	1.14	0.94	0.28	0.80	0.19	0.18	0.38
	10.SEI		9	7	2.36	1.66	0.53	1.14	0.26	0.25	0.54

Population	Reporting Category	N	Number of Points						Average p-value		
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
American Indian / Alaska Native	Overall Test	95	70	49	24.83	8.75	0.87	3.20	0.27	0.25	0.46
	10.DS		8	11	4.89	2.18	0.52	1.52	0.31	0.31	0.54
	10.GM		7	6	2.15	1.24	0.25	1.08	0.21	0.22	0.30
	10.LEI		26	15	8.57	3.28	0.59	2.10	0.30	0.30	0.45
	10.MP		4	4	0.88	0.90	0.28	0.76	0.10	0.09	0.28
	10.NSC		12	10	4.13	2.20	0.58	1.43	0.29	0.29	0.57
	10.QEF		4	4	1.04	1.03	0.47	0.75	0.17	0.16	0.47
	10.SEI		9	6	2.10	1.60	0.52	1.11	0.23	0.23	0.52
African American	Overall Test	8,331	70	60	22.74	8.47	0.85	3.23	0.24	0.24	0.46
	10.DS		8	11	4.70	2.12	0.46	1.55	0.29	0.30	0.48
	10.GM		7	8	1.83	1.17	0.23	1.03	0.18	0.18	0.34
	10.LEI		26	20	7.79	3.39	0.62	2.10	0.27	0.27	0.50
	10.MP		4	7	0.84	0.90	0.30	0.75	0.09	0.09	0.23
	10.NSC		12	11	3.63	2.16	0.59	1.38	0.26	0.26	0.54
	10.QEF		4	4	1.05	0.90	0.24	0.79	0.17	0.17	0.38
	10.SEI		9	7	1.94	1.56	0.51	1.09	0.22	0.21	0.54
Asian	Overall Test	713	70	67	27.24	10.70	0.89	3.49	0.29	0.26	0.49
	10.DS		8	11	5.31	2.23	0.52	1.54	0.33	0.32	0.48
	10.GM		7	8	2.26	1.47	0.43	1.11	0.23	0.20	0.42
	10.LEI		26	22	9.18	3.86	0.69	2.17	0.32	0.29	0.53
	10.MP		4	6	1.24	1.12	0.37	0.89	0.14	0.12	0.30
	10.NSC		12	11	4.78	2.45	0.67	1.41	0.34	0.31	0.60
	10.QEF		4	4	1.49	1.06	0.38	0.83	0.25	0.22	0.46
	10.SEI		9	7	2.47	1.79	0.60	1.14	0.27	0.24	0.54

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hispanic	Overall Test	6,615	70	66	25.11	8.72	0.86	3.30	0.27	0.26	0.46
	10.DS		8	11	5.27	2.14	0.50	1.51	0.33	0.33	0.49
	10.GM		7	8	2.13	1.27	0.30	1.06	0.21	0.21	0.37
	10.LEI		26	21	8.55	3.44	0.62	2.13	0.29	0.29	0.49
	10.MP		4	7	1.04	0.97	0.31	0.81	0.12	0.11	0.26
	10.NSC		12	11	4.05	2.17	0.58	1.40	0.29	0.28	0.54
	10.QEF		4	4	1.15	0.94	0.27	0.80	0.19	0.18	0.39
	10.SEI		9	7	2.16	1.60	0.52	1.12	0.24	0.23	0.53
White (non-Hispanic)	Overall Test	24,558	70	73	26.32	8.98	0.86	3.34	0.28	0.27	0.47
	10.DS		8	11	5.59	2.16	0.51	1.51	0.35	0.35	0.50
	10.GM		7	8	2.31	1.33	0.33	1.09	0.23	0.22	0.38
	10.LEI		26	23	8.95	3.50	0.62	2.15	0.31	0.30	0.50
	10.MP		4	9	1.12	1.04	0.34	0.85	0.12	0.11	0.28
	10.NSC		12	11	4.32	2.20	0.59	1.42	0.31	0.30	0.54
	10.QEF		4	4	1.10	0.93	0.25	0.80	0.18	0.17	0.37
	10.SEI		9	7	2.27	1.65	0.53	1.13	0.25	0.24	0.54
Multi-racial	Overall Test	2,029	70	61	24.71	8.57	0.85	3.27	0.27	0.25	0.46
	10.DS		8	11	5.19	2.17	0.48	1.56	0.32	0.33	0.50
	10.GM		7	8	2.04	1.25	0.27	1.07	0.20	0.20	0.37
	10.LEI		26	22	8.47	3.41	0.61	2.14	0.29	0.29	0.49
	10.MP		4	6	0.97	0.96	0.32	0.80	0.11	0.10	0.25
	10.NSC		12	11	4.01	2.19	0.59	1.40	0.29	0.28	0.54
	10.QEF		4	4	1.05	0.91	0.25	0.79	0.18	0.17	0.37
	10.SEI		9	7	2.12	1.58	0.50	1.12	0.24	0.23	0.53

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hawaiian/Pacific Islander	Overall Test	30	70	45	26.53	9.28	0.85	3.54	0.29	0.26	0.45
	10.DS		8	9	5.43	2.03	0.39	1.58	0.34	0.32	0.48
	10.GM		7	5	2.17	1.26	0.38	1.00	0.22	0.19	0.38
	10.LEI		26	16	9.43	3.27	0.56	2.16	0.33	0.30	0.49
	10.MP		4	2	1.17	0.79	0.14	0.74	0.13	0.12	0.17
	10.NSC		12	8	4.73	2.13	0.57	1.40	0.34	0.30	0.57
	10.QEF		4	3	1.33	0.96	0.34	0.78	0.22	0.19	0.42
	10.SEI		9	6	2.27	1.62	0.48	1.16	0.25	0.22	0.44
Special Education	Overall Test	8,990	70	57	21.04	8.10	0.84	3.26	0.23	0.22	0.46
	10.DS		8	11	4.33	2.08	0.44	1.56	0.27	0.27	0.49
	10.GM		7	8	1.73	1.21	0.27	1.04	0.17	0.17	0.37
	10.LEI		26	19	7.52	3.18	0.57	2.09	0.26	0.26	0.49
	10.MP		4	6	0.70	0.87	0.33	0.71	0.08	0.07	0.25
	10.NSC		12	11	3.67	2.13	0.58	1.38	0.26	0.26	0.56
	10.QEF		4	4	0.91	0.85	0.20	0.76	0.15	0.15	0.37
	10.SEI		9	7	1.53	1.35	0.41	1.03	0.17	0.16	0.51
SES (High)	Overall Test	18,712	70	73	26.90	9.09	0.86	3.35	0.29	0.27	0.47
	10.DS		8	11	5.68	2.14	0.50	1.51	0.35	0.35	0.50
	10.GM		7	8	2.33	1.33	0.33	1.09	0.23	0.22	0.38
	10.LEI		26	23	9.09	3.55	0.63	2.15	0.31	0.30	0.50
	10.MP		4	9	1.15	1.05	0.35	0.85	0.13	0.12	0.27
	10.NSC		12	11	4.44	2.22	0.59	1.42	0.32	0.30	0.55
	10.QEF		4	4	1.16	0.95	0.27	0.81	0.19	0.18	0.38
	10.SEI		9	7	2.38	1.67	0.53	1.14	0.26	0.25	0.54

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
SES (Low)	Overall Test	23,393	70	65	24.16	8.65	0.86	3.28	0.26	0.25	0.46
	10.DS		8	11	5.08	2.16	0.49	1.54	0.32	0.32	0.49
	10.GM		7	8	2.05	1.26	0.29	1.06	0.20	0.20	0.37
	10.LEI		26	23	8.28	3.41	0.61	2.12	0.29	0.28	0.50
	10.MP		4	8	0.96	0.96	0.31	0.80	0.11	0.10	0.26
	10.NSC		12	11	3.90	2.16	0.58	1.40	0.28	0.27	0.54
	10.QEF		4	4	1.06	0.91	0.25	0.79	0.18	0.17	0.38
	10.SEI		9	7	2.04	1.58	0.51	1.10	0.23	0.22	0.53
ELL/LEP	Overall Test	2,970	70	67	22.31	8.68	0.85	3.32	0.24	0.23	0.47
	10.DS		8	11	4.49	2.04	0.44	1.52	0.28	0.28	0.45
	10.GM		7	8	1.83	1.23	0.27	1.05	0.18	0.18	0.38
	10.LEI		26	22	7.77	3.36	0.61	2.10	0.27	0.26	0.51
	10.MP		4	5	0.84	0.91	0.32	0.75	0.09	0.09	0.26
	10.NSC		12	11	3.77	2.20	0.61	1.38	0.27	0.26	0.57
	10.QEF		4	4	1.13	0.94	0.29	0.79	0.19	0.18	0.43
	10.SEI		9	7	1.80	1.51	0.50	1.07	0.20	0.19	0.56
Section 504	Overall Test	1,240	70	65	25.63	8.86	0.85	3.38	0.28	0.26	0.47
	10.DS		8	11	5.46	2.11	0.44	1.58	0.34	0.34	0.49
	10.GM		7	8	2.21	1.27	0.27	1.08	0.22	0.22	0.37
	10.LEI		26	21	8.76	3.45	0.62	2.14	0.30	0.29	0.50
	10.MP		4	8	1.09	1.04	0.35	0.84	0.12	0.11	0.28
	10.NSC		12	11	4.23	2.22	0.59	1.42	0.30	0.29	0.55
	10.QEF		4	4	1.07	0.96	0.30	0.80	0.18	0.17	0.39
	10.SEI		9	7	2.19	1.60	0.49	1.14	0.24	0.23	0.52

Note. * indicates that since some items were correlated for certain subgroups and reporting categories, missing results occurred.

Table 40. Summary Data E/LA Grade 10 Spring 2019

Population	Reporting Category	N	Number of Points					Average p-value			
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
All	Overall Test	73,564	45	60	36.84	11.77	0.91	3.48	0.60	0.40	0.73
	10.RLV		17	21	12.53	4.74	0.79	2.16	0.60	0.38	0.73
	10.RNV		16	20	11.81	4.49	0.80	1.98	0.59	0.37	0.73
	10.WC		6	9	6.20	1.95	0.57	1.28	0.69	0.51	0.80
	10.WG		6	11	6.31	2.00	0.53	1.37	0.57	0.42	0.67
Male	Overall Test	37,294	45	60	34.87	12.07	0.92	3.51	0.57	0.39	0.72
	10.RLV		17	21	11.79	4.87	0.80	2.17	0.56	0.36	0.73
	10.RNV		16	20	11.26	4.55	0.81	1.98	0.56	0.37	0.73
	10.WC		6	9	5.83	1.98	0.55	1.33	0.65	0.49	0.78
	10.WG		6	11	6.01	2.08	0.56	1.38	0.55	0.40	0.67
Female	Overall Test	36,268	45	60	38.87	11.08	0.90	3.42	0.64	0.42	0.74
	10.RLV		17	21	13.30	4.47	0.77	2.14	0.63	0.40	0.74
	10.RNV		16	20	12.38	4.35	0.79	1.98	0.62	0.37	0.74
	10.WC		6	9	6.58	1.84	0.57	1.20	0.73	0.54	0.82
	10.WG		6	11	6.62	1.87	0.50	1.32	0.60	0.44	0.68
American Indian / Alaska Native	Overall Test	148	45	55	37.22	10.49	0.89	3.48	0.61	0.42	0.72
	10.RLV		17	20	12.69	4.00	0.70	2.18	0.60	0.41	0.71
	10.RNV		16	20	11.94	4.21	0.78	1.99	0.60	0.39	0.71
	10.WC		6	9	6.16	1.87	0.54	1.27	0.68	0.48	0.79
	10.WG		6	11	6.43	1.99	0.57	1.31	0.58	0.41	0.68
African American	Overall Test	8,151	45	58	29.92	11.34	0.90	3.57	0.49	0.38	0.69
	10.RLV		17	21	10.29	4.70	0.79	2.17	0.49	0.36	0.71
	10.RNV		16	20	9.04	4.32	0.78	2.04	0.45	0.33	0.67
	10.WC		6	9	5.31	1.91	0.49	1.36	0.59	0.49	0.76
	10.WG		6	11	5.30	1.95	0.49	1.40	0.48	0.40	0.63

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Asian	Overall Test	1,901	45	60	39.95	12.39	0.92	3.41	0.65	0.39	0.77
	10.RLV		17	21	13.68	4.70	0.80	2.10	0.65	0.38	0.76
	10.RNV		16	20	12.68	4.76	0.84	1.93	0.63	0.33	0.76
	10.WC		6	9	6.73	1.99	0.61	1.24	0.75	0.50	0.85
	10.WG		6	11	6.85	2.18	0.57	1.43	0.62	0.41	0.71
Hispanic	Overall Test	8,528	45	59	33.13	11.50	0.90	3.57	0.54	0.39	0.70
	10.RLV		17	21	11.30	4.71	0.79	2.18	0.54	0.37	0.71
	10.RNV		16	20	10.41	4.38	0.78	2.04	0.52	0.35	0.70
	10.WC		6	9	5.66	1.92	0.49	1.37	0.63	0.50	0.77
	10.WG		6	11	5.78	1.98	0.50	1.39	0.53	0.41	0.64
White (non-Hispanic)	Overall Test	51,613	45	60	38.50	11.29	0.91	3.44	0.63	0.41	0.74
	10.RLV		17	21	13.07	4.60	0.78	2.14	0.62	0.38	0.74
	10.RNV		16	20	12.48	4.29	0.79	1.96	0.62	0.39	0.74
	10.WC		6	9	6.42	1.90	0.57	1.24	0.71	0.52	0.81
	10.WG		6	11	6.55	1.93	0.51	1.35	0.60	0.43	0.68
Multi-racial	Overall Test	3,139	45	60	35.85	11.85	0.91	3.51	0.59	0.40	0.73
	10.RLV		17	21	12.22	4.79	0.80	2.16	0.58	0.38	0.73
	10.RNV		16	20	11.45	4.51	0.80	2.00	0.57	0.36	0.72
	10.WC		6	9	6.03	1.97	0.57	1.29	0.67	0.50	0.79
	10.WG		6	11	6.17	2.01	0.53	1.39	0.56	0.42	0.67
Hawaiian/Pacific Islander	Overall Test	58	45	54	34.34	11.34	0.90	3.51	0.56	0.41	0.72
	10.RLV		17	20	11.74	4.18	0.73	2.18	0.56	0.41	0.71
	10.RNV		16	20	10.40	4.72	0.82	2.01	0.52	0.33	0.71
	10.WC		6	9	6.02	1.69	0.36	1.35	0.67	0.57	0.77
	10.WG		6	10	6.19	2.27	0.68	1.29	0.56	0.42	0.70

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Special Education	Overall Test	9,553	45	57	24.54	10.83	0.90	3.51	0.40	0.34	0.68
	10.RLV		17	21	8.16	4.42	0.77	2.12	0.39	0.32	0.69
	10.RNV		16	20	7.52	4.15	0.77	1.97	0.38	0.30	0.68
	10.WC		6	9	4.41	1.85	0.42	1.41	0.49	0.43	0.73
	10.WG		6	11	4.45	2.00	0.51	1.40	0.40	0.35	0.62
SES (High)	Overall Test	42,170	45	60	40.10	10.75	0.90	3.39	0.66	0.42	0.75
	10.RLV		17	21	13.63	4.42	0.77	2.13	0.65	0.39	0.74
	10.RNV		16	20	13.02	4.11	0.78	1.94	0.65	0.39	0.75
	10.WC		6	9	6.65	1.82	0.57	1.20	0.74	0.53	0.82
	10.WG		6	11	6.81	1.86	0.50	1.32	0.62	0.44	0.69
SES (Low)	Overall Test	31,029	45	60	32.51	11.63	0.91	3.56	0.53	0.39	0.70
	10.RLV		17	21	11.08	4.75	0.79	2.18	0.53	0.37	0.71
	10.RNV		16	20	10.20	4.45	0.79	2.03	0.51	0.35	0.70
	10.WC		6	9	5.60	1.95	0.52	1.35	0.62	0.49	0.77
	10.WG		6	11	5.65	1.98	0.51	1.39	0.51	0.40	0.64
ELL/LEP	Overall Test	2,853	45	54	22.71	8.82	0.84	3.58	0.37	0.34	0.64
	10.RLV		17	21	7.47	3.70	0.67	2.11	0.36	0.32	0.64
	10.RNV		16	20	6.67	3.30	0.63	2.00	0.33	0.30	0.62
	10.WC		6	9	4.31	1.80	0.32	1.49	0.48	0.45	0.73
	10.WG		6	10	4.27	1.90	0.40	1.48	0.39	0.36	0.59
Section 504	Overall Test	1,481	45	60	35.58	11.19	0.90	3.50	0.58	0.41	0.72
	10.RLV		17	21	12.03	4.63	0.78	2.17	0.57	0.39	0.72
	10.RNV		16	20	11.49	4.34	0.79	2.00	0.57	0.39	0.72
	10.WC		6	9	5.97	1.88	0.52	1.30	0.66	0.52	0.78
	10.WG		6	11	6.09	1.94	0.52	1.35	0.55	0.42	0.66

Table 41. Summary Data Mathematics Grade 10 Spring 2019

Population	Reporting Category	N	Number of Points						Average p-value		
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
All	Overall Test	73,791	57	73	35.74	14.35	0.94	3.59	0.49	0.37	0.71
	10.DS		8	12	7.65	2.44	0.66	1.42	0.64	0.54	0.83
	10.GM		5	5	2.84	1.37	0.50	0.97	0.57	0.46	0.78
	10.LEI		19	21	10.32	4.85	0.83	1.99	0.49	0.36	0.75
	10.MP		4	9	2.84	2.01	0.68	1.14	0.32	0.20	0.53
	10.NSC		11	12	6.08	2.72	0.70	1.48	0.51	0.39	0.73
	10.QEF		5	8	3.38	1.85	0.58	1.20	0.42	0.31	0.64
	10.SEI		5	6	2.63	1.53	0.61	0.96	0.44	0.31	0.68
Male	Overall Test	37,447	57	73	35.26	14.76	0.94	3.59	0.48	0.36	0.72
	10.DS		8	12	7.62	2.53	0.68	1.43	0.64	0.53	0.84
	10.GM		5	5	2.85	1.40	0.52	0.97	0.57	0.46	0.80
	10.LEI		19	21	10.17	4.92	0.84	1.99	0.48	0.35	0.75
	10.MP		4	9	2.70	2.01	0.68	1.13	0.30	0.19	0.52
	10.NSC		11	12	6.02	2.81	0.73	1.47	0.50	0.38	0.74
	10.QEF		5	8	3.36	1.90	0.60	1.20	0.42	0.31	0.65
	10.SEI		5	6	2.54	1.55	0.62	0.96	0.42	0.30	0.67
Female	Overall Test	36,342	57	73	36.23	13.88	0.93	3.59	0.50	0.38	0.71
	10.DS		8	12	7.69	2.35	0.64	1.40	0.64	0.55	0.81
	10.GM		5	5	2.82	1.33	0.47	0.96	0.56	0.46	0.76
	10.LEI		19	21	10.48	4.78	0.83	1.99	0.50	0.37	0.74
	10.MP		4	9	2.98	2.00	0.66	1.16	0.33	0.22	0.54
	10.NSC		11	12	6.14	2.61	0.67	1.49	0.51	0.40	0.72
	10.QEF		5	8	3.39	1.80	0.55	1.20	0.42	0.31	0.62
	10.SEI		5	6	2.73	1.51	0.59	0.97	0.45	0.33	0.68

Population	Reporting Category	N	Number of Points					Average p-value			
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
American Indian / Alaska Native	Overall Test	147	57	66	34.11	13.18	0.92	3.63	0.47	0.37	0.70
	10.DS		8	12	7.24	2.46	0.66	1.44	0.60	0.53	0.80
	10.GM		5	5	2.77	1.27	0.41	0.98	0.55	0.47	0.77
	10.LEI		19	21	9.79	4.52	0.80	2.03	0.47	0.36	0.73
	10.MP		4	8	2.71	1.96	0.72	1.04	0.30	0.21	0.53
	10.NSC		11	12	5.81	2.58	0.65	1.53	0.48	0.39	0.71
	10.QEF		5	8	3.33	1.83	0.57	1.19	0.42	0.32	0.65
	10.SEI		5	5	2.47	1.43	0.54	0.97	0.41	0.32	0.65
African American	Overall Test	8,217	57	70	26.47	11.39	0.90	3.60	0.36	0.32	0.67
	10.DS		8	12	6.13	2.29	0.55	1.54	0.51	0.48	0.78
	10.GM		5	5	2.10	1.24	0.33	1.01	0.42	0.38	0.71
	10.LEI		19	21	7.49	3.99	0.75	1.99	0.36	0.31	0.72
	10.MP		4	9	1.76	1.55	0.56	1.03	0.20	0.16	0.47
	10.NSC		11	12	4.55	2.32	0.59	1.49	0.38	0.34	0.68
	10.QEF		5	8	2.51	1.55	0.42	1.18	0.31	0.28	0.60
	10.SEI		5	6	1.94	1.39	0.51	0.97	0.32	0.28	0.65
Asian	Overall Test	1,903	57	73	43.05	16.84	0.96	3.47	0.59	0.38	0.77
	10.DS		8	12	8.41	2.68	0.75	1.34	0.70	0.52	0.85
	10.GM		5	5	3.14	1.44	0.61	0.90	0.63	0.42	0.81
	10.LEI		19	21	12.79	5.58	0.89	1.85	0.61	0.37	0.82
	10.MP		4	9	3.76	2.48	0.78	1.17	0.42	0.20	0.60
	10.NSC		11	12	7.42	2.97	0.78	1.39	0.62	0.41	0.79
	10.QEF		5	8	4.32	2.11	0.66	1.22	0.54	0.34	0.71
	10.SEI		5	6	3.21	1.60	0.69	0.89	0.54	0.32	0.72

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hispanic	Overall Test	8,541	57	71	30.52	12.46	0.92	3.62	0.42	0.35	0.68
	10.DS		8	12	6.86	2.34	0.60	1.48	0.57	0.52	0.80
	10.GM		5	5	2.44	1.31	0.41	1.00	0.49	0.43	0.74
	10.LEI		19	21	8.71	4.35	0.79	2.00	0.41	0.34	0.72
	10.MP		4	9	2.16	1.71	0.61	1.07	0.24	0.18	0.48
	10.NSC		11	12	5.23	2.44	0.62	1.51	0.44	0.37	0.69
	10.QEF		5	8	2.87	1.64	0.47	1.19	0.36	0.30	0.60
	10.SEI		5	6	2.25	1.44	0.54	0.98	0.37	0.30	0.65
White (non-Hispanic)	Overall Test	51,737	57	73	37.95	14.09	0.94	3.58	0.52	0.39	0.72
	10.DS		8	12	8.02	2.35	0.65	1.38	0.67	0.56	0.83
	10.GM		5	5	3.02	1.34	0.49	0.95	0.60	0.48	0.79
	10.LEI		19	21	10.99	4.79	0.83	1.98	0.52	0.37	0.75
	10.MP		4	9	3.11	2.01	0.66	1.16	0.35	0.22	0.53
	10.NSC		11	12	6.43	2.68	0.70	1.48	0.54	0.41	0.73
	10.QEF		5	8	3.58	1.85	0.58	1.20	0.45	0.32	0.64
	10.SEI		5	6	2.80	1.52	0.60	0.96	0.47	0.33	0.68
Multi-racial	Overall Test	3,157	57	72	33.59	14.07	0.93	3.60	0.46	0.36	0.71
	10.DS		8	12	7.38	2.41	0.64	1.44	0.61	0.53	0.82
	10.GM		5	5	2.65	1.35	0.46	0.99	0.53	0.44	0.77
	10.LEI		19	21	9.68	4.80	0.83	1.99	0.46	0.35	0.75
	10.MP		4	9	2.56	1.94	0.67	1.12	0.28	0.19	0.52
	10.NSC		11	12	5.69	2.68	0.69	1.49	0.47	0.38	0.72
	10.QEF		5	8	3.15	1.81	0.57	1.19	0.39	0.30	0.63
	10.SEI		5	6	2.47	1.53	0.60	0.97	0.41	0.30	0.68

			Number of Points						Average p-value		
Population	Reporting Category	N	No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
Hawaiian/Pacific Islander	Overall Test	61	57	67	34.25	15.73	0.95	3.61	0.47	0.35	0.74
	10.DS		8	12	7.23	2.85	0.75	1.42	0.60	0.49	0.86
	10.GM		5	5	2.39	1.57	0.66	0.91	0.48	0.32	0.83
	10.LEI		19	21	10.16	5.01	0.85	1.95	0.48	0.36	0.76
	10.MP		4	8	2.67	2.28	0.75	1.13	0.30	0.17	0.58
	10.NSC		11	11	6.08	2.72	0.71	1.46	0.51	0.39	0.76
	10.QEF		5	8	3.18	1.87	0.60	1.18	0.40	0.30	0.62
	10.SEI		5	5	2.52	1.58	0.64	0.95	0.42	0.30	0.68
Special Education	Overall Test	9,602	57	71	23.63	10.81	0.89	3.56	0.32	0.29	0.69
	10.DS		8	12	5.56	2.34	0.55	1.58	0.46	0.43	0.81
	10.GM		5	5	1.95	1.26	0.35	1.01	0.39	0.36	0.75
	10.LEI		19	21	6.76	3.61	0.70	1.97	0.32	0.29	0.72
	10.MP		4	9	1.29	1.41	0.56	0.94	0.14	0.12	0.46
	10.NSC		11	12	4.30	2.36	0.61	1.48	0.36	0.33	0.72
	10.QEF		5	8	2.21	1.47	0.40	1.14	0.28	0.25	0.61
	10.SEI		5	6	1.57	1.28	0.46	0.94	0.26	0.23	0.65
SES (High)	Overall Test	42,239	57	73	40.04	14.19	0.94	3.55	0.55	0.40	0.72
	10.DS		8	12	8.31	2.30	0.65	1.35	0.69	0.57	0.83
	10.GM		5	5	3.14	1.33	0.50	0.94	0.63	0.49	0.79
	10.LEI		19	21	11.66	4.85	0.84	1.96	0.56	0.38	0.76
	10.MP		4	9	3.37	2.05	0.67	1.18	0.37	0.23	0.54
	10.NSC		11	12	6.77	2.69	0.70	1.46	0.56	0.42	0.74
	10.QEF		5	8	3.81	1.88	0.59	1.21	0.48	0.33	0.65
	10.SEI		5	6	2.98	1.51	0.61	0.95	0.50	0.34	0.68

Population	Reporting Category	N	Number of Points						Average p-value		
			No. of Items	Max	Mean	SD	Alpha	SEM_Alpha	State	Did Not Pass	Pass
SES (Low)	Overall Test	31,197	57	72	30.03	12.40	0.91	3.62	0.41	0.35	0.68
	10.DS		8	12	6.79	2.36	0.60	1.49	0.57	0.51	0.80
	10.GM		5	5	2.44	1.32	0.41	1.01	0.49	0.43	0.75
	10.LEI		19	21	8.54	4.25	0.78	2.01	0.41	0.33	0.71
	10.MP		4	9	2.13	1.71	0.60	1.07	0.24	0.18	0.48
	10.NSC		11	12	5.15	2.46	0.63	1.50	0.43	0.37	0.69
	10.QEF		5	8	2.80	1.64	0.48	1.19	0.35	0.29	0.60
	10.SEI		5	6	2.18	1.44	0.54	0.98	0.36	0.29	0.65
ELL/LEP	Overall Test	2,858	57	65	22.22	9.37	0.86	3.55	0.30	0.29	0.66
	10.DS		8	12	5.21	2.16	0.46	1.59	0.43	0.42	0.74
	10.GM		5	5	1.78	1.14	0.19	1.02	0.36	0.34	0.69
	10.LEI		19	21	6.24	3.33	0.66	1.93	0.30	0.28	0.72
	10.MP		4	8	1.13	1.23	0.46	0.90	0.13	0.11	0.38
	10.NSC		11	12	4.08	2.14	0.52	1.49	0.34	0.32	0.71
	10.QEF		5	8	2.22	1.41	0.35	1.14	0.28	0.26	0.63
	10.SEI		5	5	1.57	1.24	0.42	0.95	0.26	0.24	0.63
Section 504	Overall Test	1,479	57	70	34.28	13.55	0.93	3.60	0.47	0.37	0.71
	10.DS		8	12	7.53	2.35	0.63	1.42	0.63	0.55	0.82
	10.GM		5	5	2.78	1.33	0.46	0.98	0.56	0.47	0.78
	10.LEI		19	21	9.78	4.57	0.81	2.01	0.47	0.36	0.74
	10.MP		4	9	2.65	1.87	0.64	1.13	0.29	0.20	0.52
	10.NSC		11	12	5.84	2.65	0.69	1.48	0.49	0.39	0.73
	10.QEF		5	8	3.25	1.83	0.57	1.19	0.41	0.31	0.65
	10.SEI		5	6	2.44	1.47	0.56	0.98	0.41	0.30	0.66

*Note. This table includes only first-time testers.

Appendix J: ISTEP+ Scale Score Descriptive Data and Distribution

Table 42. Scale Score Statistics E/LA Grade 10 Winter 2018

Subgroup	N	Mean	SD	Min	Max	Variance	Skew	Kurtosis	Alpha	SEM
All	43,308	208.80	46.05	100	400	2120.22	-0.24	0.00	0.86	17.26
Female	18,052	215.02	43.46	100	400	1889.16	-0.24	0.34	0.85	16.67
Male	25,249	204.34	47.31	100	400	2238.71	-0.21	-0.20	0.86	17.51
American Indian	107	212.69	41.69	109	304	1737.67	-0.42	-0.19	0.83	17.27
African American	8,740	199.97	44.07	100	382	1941.85	-0.29	-0.32	0.85	16.90
Asian	1,025	208.45	49.97	100	400	2496.61	-0.08	0.29	0.86	18.50
Hispanic	6,822	206.03	44.74	100	394	2001.71	-0.36	-0.10	0.85	17.07
White	24,586	212.69	46.46	100	400	2158.84	-0.24	0.09	0.86	17.40
Multiracial	1,935	207.51	45.56	100	400	2075.36	-0.18	-0.05	0.86	17.14
Hawaiian/ Pacific Islander	24	213.50	37.87	123	279	1434.00	-0.24	0.03	0.81	16.30
Special Education	10,880	182.30	43.02	100	400	1850.48	0.01	-0.50	0.82	18.00
SES (High)	18,748	214.96	45.64	100	400	2082.90	-0.30	0.14	0.86	17.11
SES (Low)	23,656	203.13	44.91	100	400	2016.79	-0.28	-0.17	0.85	17.33
LEP/ESL	3,461	188.77	44.06	100	400	1941.60	-0.15	-0.20	0.83	18.11
Section 504	1,104	211.28	45.44	100	400	2065.05	-0.21	0.05	0.86	17.20

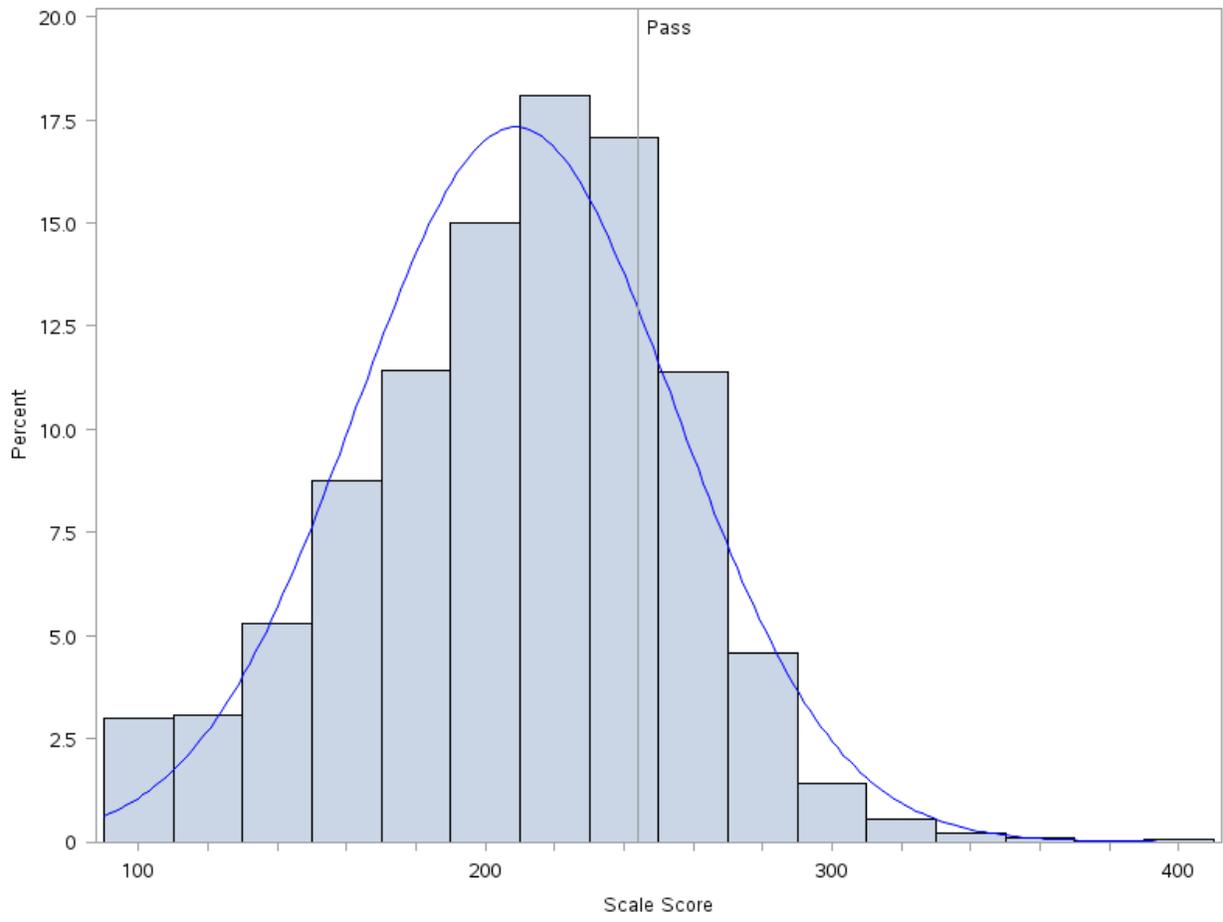


Figure 23. Scale Score Distribution E/LA Grade 10 Winter 2018

Table 43. Scale Score Statistics Mathematics 10 Winter 2018

Subgroup	N	Mean	SD	Min	Max	Variance	Skew	Kurtosis	Alpha	SEM
All	71,175	232.79	38.08	100	400	1450.06	-0.29	0.48	0.87	13.48
Female	34,814	235.55	36.31	100	400	1318.33	-0.29	0.46	0.87	13.21
Male	36,360	230.12	39.53	100	400	1562.81	-0.26	0.45	0.88	13.66
American Indian	158	233.18	36.39	100	310	1324.00	-0.99	2.21	0.85	14.11
African American	12,240	220.08	37.41	100	400	1399.21	-0.26	0.04	0.86	14.10
Asian	1,217	240.26	44.77	100	400	2004.57	-0.04	0.67	0.91	13.68
Hispanic	10,386	228.88	37.14	100	400	1379.14	-0.36	0.39	0.86	13.69
White	43,699	237.15	37.44	100	400	1401.93	-0.31	0.64	0.87	13.30
Multiracial	3,375	230.46	37.09	100	400	1375.39	-0.28	0.60	0.87	13.47
Hawaiian/ Pacific Islander	42	236.90	38.54	154	310	1485.50	-0.22	-0.44	0.88	13.23
Special Education	12,448	207.70	37.05	100	400	1372.75	-0.12	0.13	0.83	15.29
SES (High)	34,816	239.56	37.20	100	400	1384.11	-0.34	0.73	0.87	13.16
SES (Low)	35,431	225.94	37.20	100	400	1383.95	-0.32	0.23	0.86	13.82
LEP/ESL	3,557	213.01	39.42	100	400	1554.09	0.02	0.62	0.86	14.74
Section 504	1,765	234.63	37.55	100	400	1409.73	-0.29	0.65	0.87	13.44

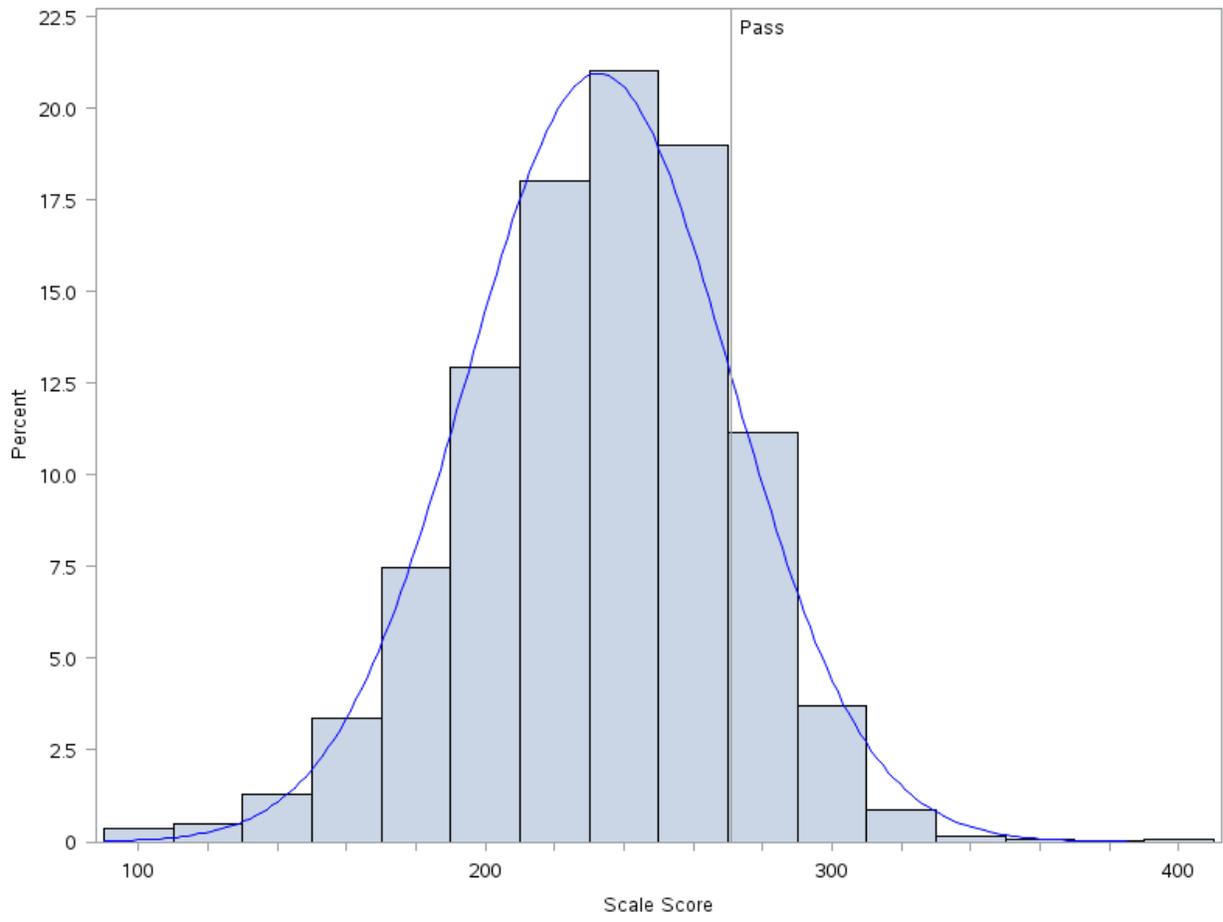


Figure 24. Scale Score Distribution Mathematics Grade 10 Winter 2018

Table 44. Scale Score Statistics E/LA Grade 10 Spring Retest 2019

Subgroup	N	Mean	SD	Min	Max	Variance	Skew	Kurtosis	Alpha	SEM
All	23,378	194.80	46.31	100	400	2144.86	-0.16	-0.33	0.78	21.76
Female	9,763	203.55	43.44	100	400	1886.72	-0.27	0.02	0.78	20.58
Male	13,611	188.44	47.29	100	400	2236.42	-0.05	-0.47	0.78	22.09
American Indian	64	202.38	49.42	100	300	2442.45	-0.44	-0.62	0.80	22.33
African American	5,665	189.50	45.64	100	342	2082.83	-0.21	-0.63	0.75	22.66
Asian	496	191.31	49.86	100	341	2486.17	-0.08	-0.65	0.86	18.73
Hispanic	3,912	194.93	45.88	100	399	2104.82	-0.26	-0.43	0.77	22.07
White	12,112	197.39	46.36	100	400	2148.92	-0.11	-0.15	0.78	21.57
Multiracial	1,072	192.88	46.40	100	329	2153.00	-0.23	-0.64	0.75	23.16
Hawaiian/ Pacific Islander	16	214.31	43.25	141	281	1870.76	-0.07	-0.94	0.84	17.38
Special Education	7,536	177.34	43.53	100	321	1894.50	0.05	-0.65	0.73	22.56
SES (High)	8,766	200.36	47.18	100	400	2226.17	-0.14	-0.03	0.80	21.05
SES (Low)	14,346	191.23	45.29	100	366	2050.81	-0.20	-0.58	0.76	22.29
LEP/ESL	2,662	181.23	45.08	100	321	2031.92	-0.12	-0.81	0.78	21.02
Section 504	683	198.61	46.37	100	327	2149.81	-0.28	-0.44	0.77	22.43

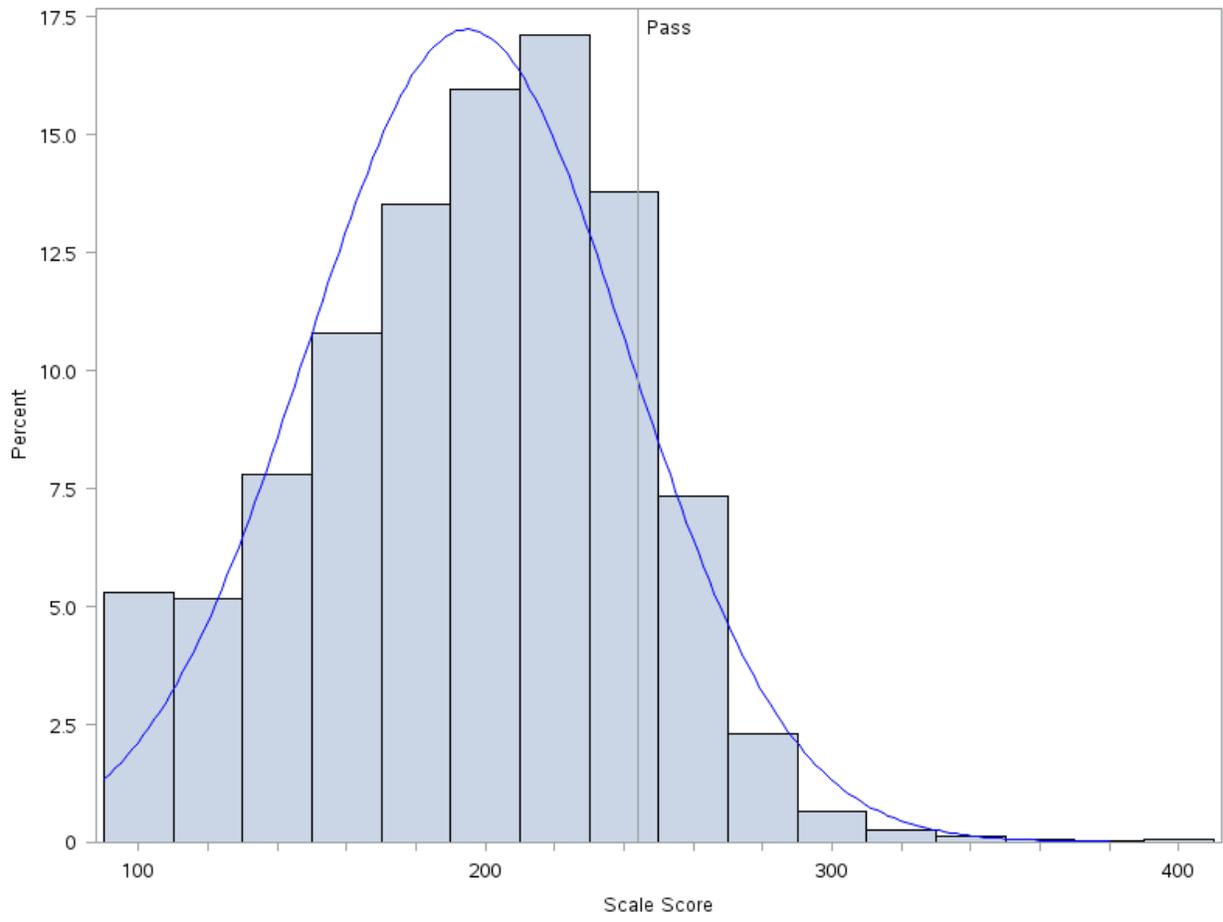


Figure 25. Scale Score Distribution E/LA Grade 10 Spring Retest 2019

Table 45. Scale Score Statistics Mathematics Grade 10 Spring Retest 2019

Subgroup	N	Mean	SD	Min	Max	Variance	Skew	Kurtosis	Alpha	SEM
All	44,435	222.59	36.61	100	400	1340.59	-0.51	0.54	0.86	13.55
Female	21,624	226.25	34.30	100	400	1176.52	-0.52	0.64	0.85	13.09
Male	22,805	219.07	38.38	100	400	1473.39	-0.46	0.38	0.87	13.88
American Indian	101	220.09	36.36	127	294	1322.09	-0.36	-0.30	0.87	13.30
African American	8,984	212.16	37.38	100	325	1397.43	-0.40	-0.03	0.85	14.25
Asian	734	229.39	41.69	100	376	1737.97	-0.29	0.47	0.89	13.61
Hispanic	6,939	221.84	36.12	100	361	1304.95	-0.58	0.54	0.86	13.66
White	25,467	226.40	35.60	100	400	1267.14	-0.55	0.84	0.86	13.25
Multiracial	2,134	219.93	36.16	100	329	1307.75	-0.58	0.34	0.85	13.80
Hawaiian/ Pacific Islander	30	228.73	33.05	153	288	1092.55	-0.24	-0.33	0.85	12.62
Special Education	9,449	203.38	38.26	100	320	1463.61	-0.26	-0.12	0.84	15.38
SES (High)	19,421	228.69	35.42	100	400	1254.43	-0.53	0.99	0.86	13.07
SES (Low)	24,673	217.88	36.71	100	356	1347.82	-0.52	0.24	0.86	13.93
LEP/ESL	3,139	209.57	39.49	100	376	1559.72	-0.24	0.23	0.85	15.09
Section 504	1,287	224.05	35.70	100	347	1274.25	-0.58	0.68	0.85	13.63

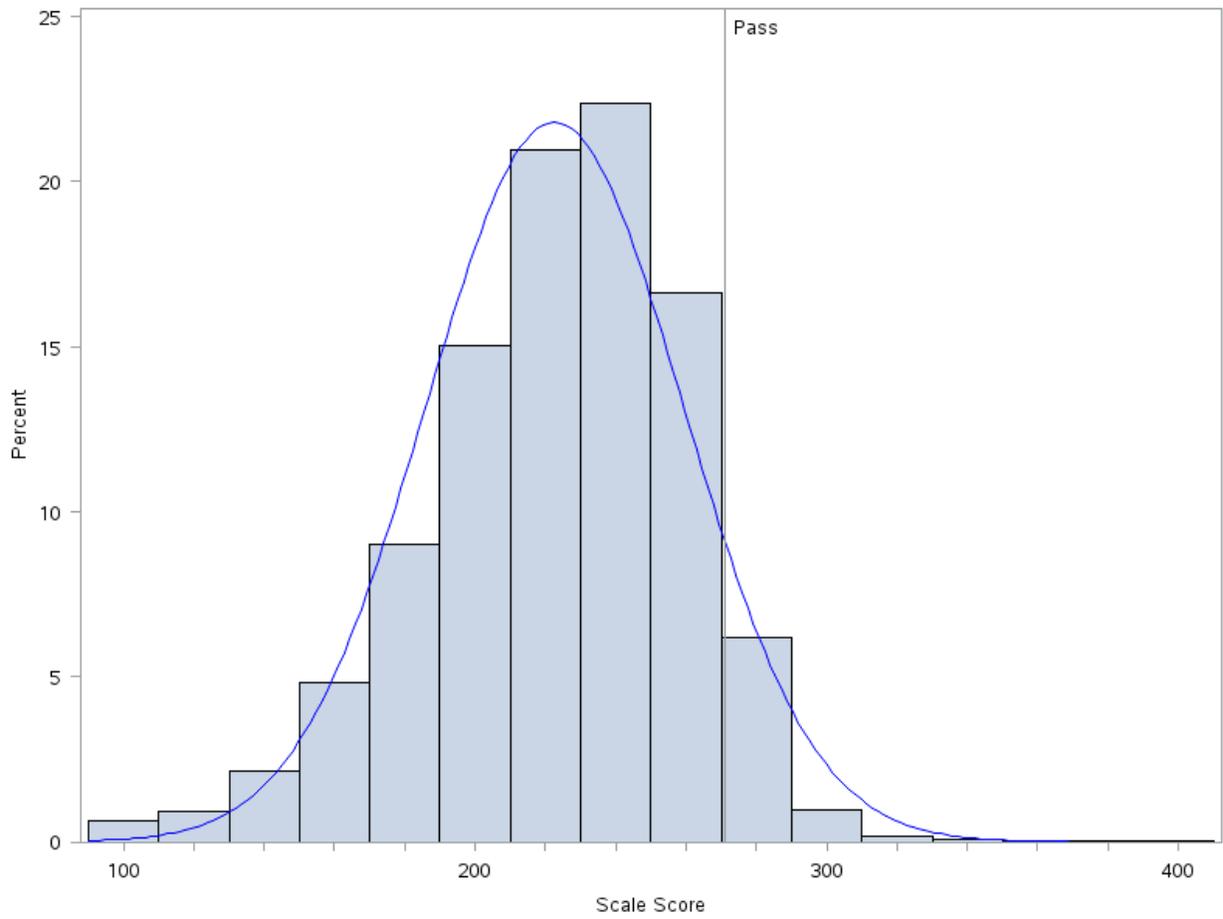


Figure 26. Scale Score Distribution Mathematics Grade 10 Spring Retest 2019

Table 46. Scale Score Statistics E/LA Grade 10 Spring 2019

Subgroup	N	Mean	SD	Min	Max	Variance	Skew	Kurtosis	Alpha	SEM
All	73,564	254.23	55.84	100	400	3118.13	-0.38	0.00	0.91	16.61
Female	36,268	264.21	52.54	100	400	2760.13	-0.37	0.19	0.90	16.31
Male	37,294	244.53	57.23	100	400	3275.17	-0.34	-0.20	0.91	16.75
American Indian	148	255.29	48.30	117	351	2332.96	-0.60	0.04	0.89	16.06
African American	8,151	221.95	52.81	100	400	2788.58	-0.16	-0.31	0.90	16.71
Asian	1,901	270.35	62.15	100	400	3862.65	-0.50	0.04	0.92	17.20
Hispanic	8,528	236.62	53.55	100	400	2867.22	-0.37	-0.07	0.90	16.71
White	51,613	261.96	53.83	100	400	2897.58	-0.45	0.20	0.91	16.48
Multiracial	3,139	249.52	56.17	100	400	3155.39	-0.29	-0.10	0.91	16.74
Hawaiian/ Pacific Islander	58	242.38	51.88	100	350	2691.47	-0.25	-0.22	0.90	16.19
Special Education	9,553	195.77	52.98	100	400	2806.78	0.24	-0.37	0.89	17.25
SES (High)	42,170	269.72	51.77	100	400	2680.57	-0.49	0.42	0.90	16.40
SES (Low)	31,029	233.65	54.25	100	400	2942.69	-0.26	-0.20	0.91	16.70
LEP/ESL	2,853	188.12	45.47	100	350	2067.61	-0.17	-0.54	0.83	18.50
Section 504	1,481	248.27	52.40	100	400	2745.84	-0.24	-0.01	0.90	16.52

*Note. This table includes only first-time testers.

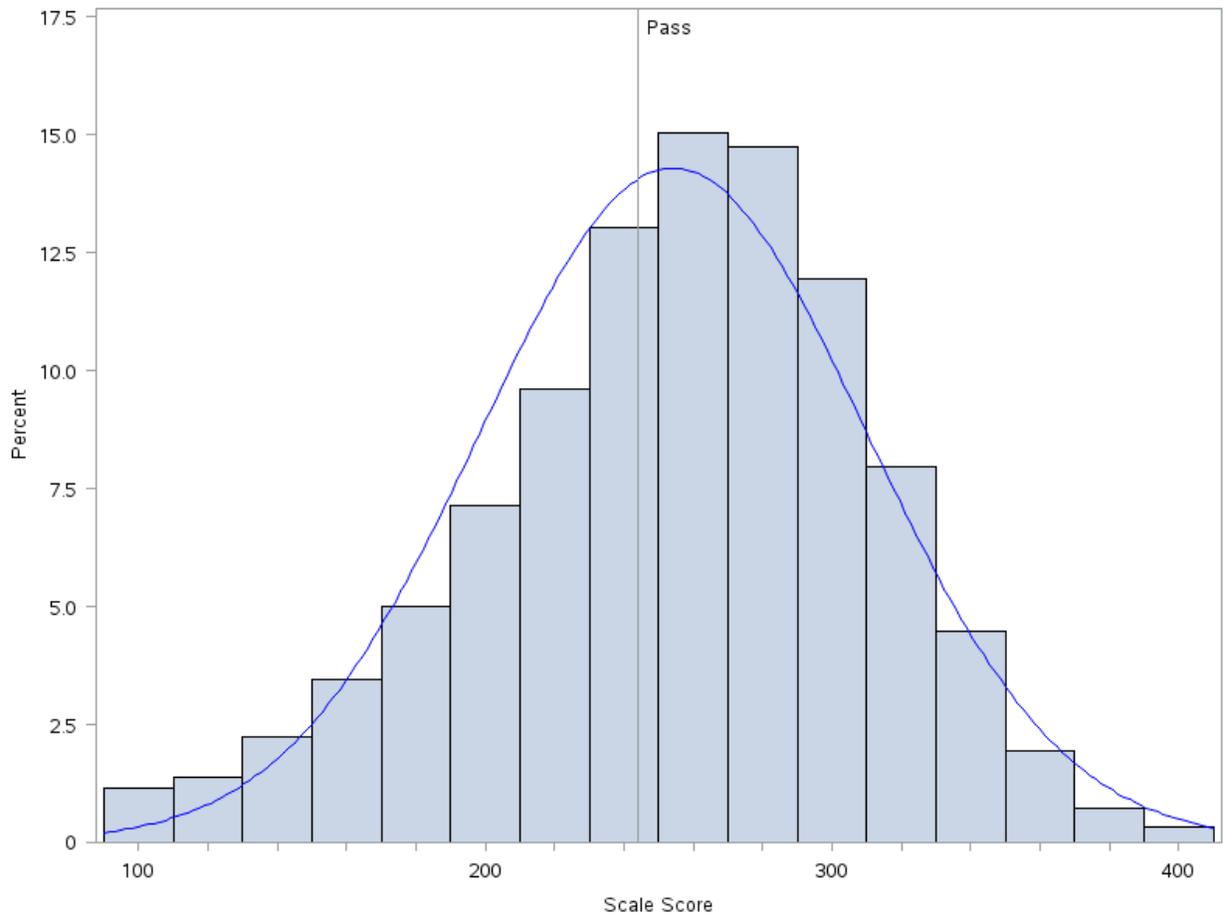


Figure 27. Scale Score Distribution E/LA Grade 10 Spring 2019

Table 47. Scale Score Statistics Mathematics Grade 10 Spring 2019

Subgroup	N	Mean	SD	Min	Max	Variance	Skew	Kurtosis	Alpha	SEM
All	73,791	251.25	51.12	100	400	2612.88	0.07	0.05	0.93	13.20
Female	36,342	253.59	48.55	100	400	2357.02	0.06	0.11	0.93	12.94
Male	37,447	248.98	53.39	100	400	2850.74	0.10	-0.02	0.94	13.40
American Indian	147	245.87	45.63	133	370	2082.26	0.17	-0.08	0.92	12.89
African American	8,217	218.20	44.53	100	400	1983.24	0.13	0.16	0.90	14.43
Asian	1,903	278.12	63.07	100	400	3977.43	-0.11	-0.32	0.95	13.50
Hispanic	8,541	233.29	46.19	100	400	2133.60	0.05	0.32	0.91	13.71
White	51,737	258.98	49.27	100	400	2427.32	0.02	0.09	0.93	12.88
Multiracial	3,157	243.79	50.64	100	400	2564.89	0.19	0.10	0.93	13.36
Hawaiian/ Pacific Islander	61	246.82	57.45	100	373	3300.65	0.30	0.20	0.94	13.60
Special Education	9,602	205.03	45.35	100	400	2056.55	0.35	0.54	0.89	15.24
SES (High)	42,239	266.39	49.50	100	400	2449.88	0.00	0.11	0.93	12.77
SES (Low)	31,197	231.19	45.92	100	400	2109.03	0.05	0.19	0.91	13.73
LEP/ESL	2,858	200.22	42.28	100	360	1787.90	0.08	0.33	0.85	16.16
Section 504	1,479	246.76	47.13	100	400	2221.24	0.21	0.18	0.92	12.96

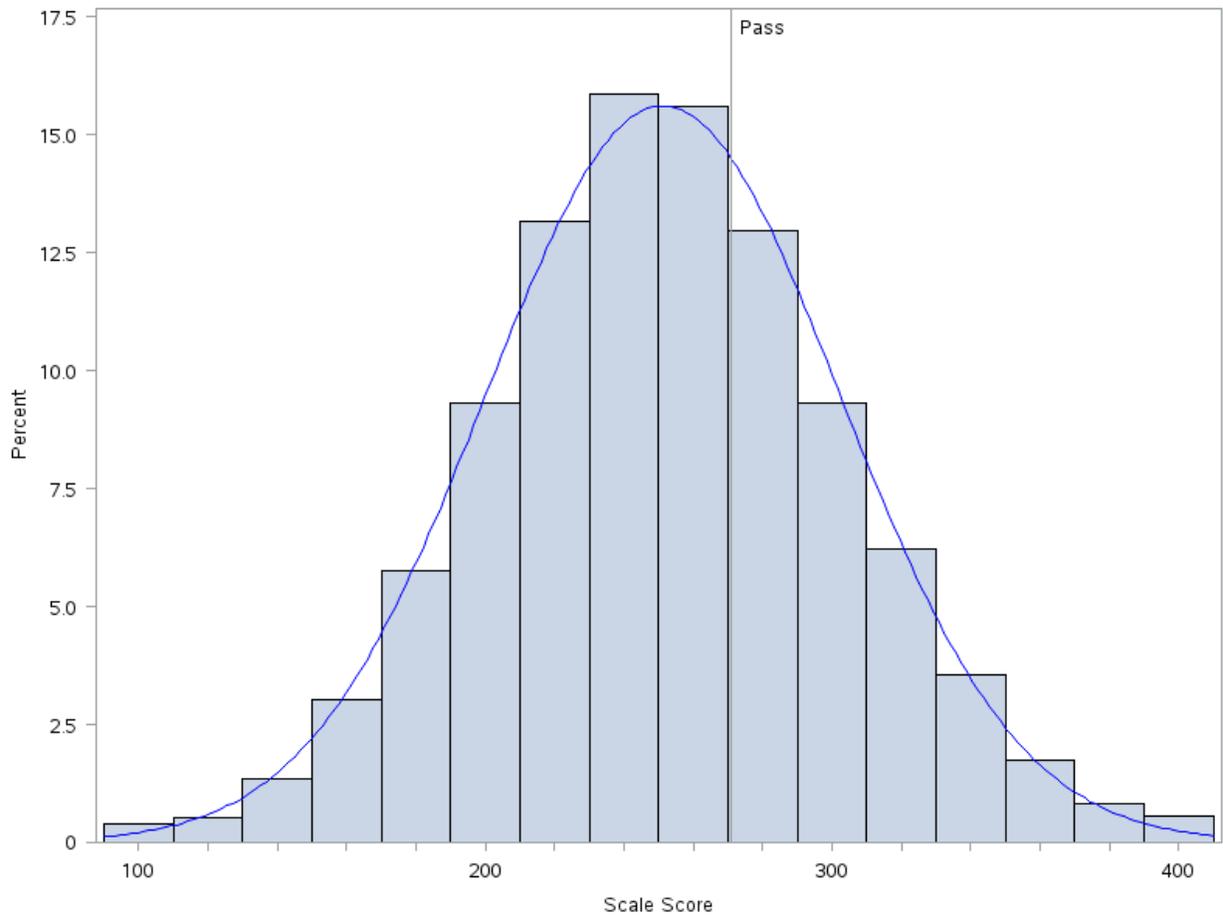


Figure 28. Scale Score Distribution Mathematics Grade 10 Spring 2019

Appendix K: ISTEP+ Performance Level Distribution

Table 48. Performance Level Distribution E/LA Grade 10 Winter 2018

Subgroup	N	Performance Level						Proficiency	
		Pass+		Pass		Did Not Pass		N	%
		N	%	N	%	N	%		
All	41786	904	2.16%	8717	20.86%	32165	76.98%	9621	23.02%
Female	17476	454	2.60%	3995	22.86%	13027	74.54%	4449	25.46%
Male	24304	450	1.85%	4722	19.43%	19132	78.72%	5172	21.28%
American Indian	104	2	1.92%	21	20.19%	81	77.88%	23	22.12%
African American	8180	70	0.86%	1234	15.09%	6876	84.06%	1304	15.94%
Asian	1005	35	3.48%	200	19.90%	770	76.62%	235	23.38%
Hispanic	6513	75	1.15%	1239	19.02%	5199	79.82%	1314	20.18%
White	24030	684	2.85%	5638	23.46%	17708	73.69%	6322	26.31%
Multiracial	1868	36	1.93%	371	19.86%	1461	78.21%	407	21.79%
Hawaiian/ Pacific Islander	24	0	0.00%	6	25.00%	18	75.00%	6	25.00%
Special Education	10516	30	0.29%	768	7.30%	9718	92.41%	798	7.59%
SES (High)	18238	518	2.84%	4505	24.70%	13215	72.46%	5023	27.54%
SES (Low)	22713	262	1.15%	3996	17.59%	18455	81.25%	4258	18.75%
LEP/ESL	3296	15	0.46%	293	8.89%	2988	90.66%	308	9.34%
Section 504	1078	26	2.41%	245	22.73%	807	74.86%	271	25.14%

Table 49. Performance Level Distribution Mathematics Grade 10 Winter 2018

Subgroup	N	Performance Level						Proficiency	
		Pass+		Pass		Did Not Pass		N	%
		N	%	N	%	N	%		
All	69493	184	0.26%	10460	15.05%	58849	84.68%	10644	15.32%
Female	34131	76	0.22%	5445	15.95%	28610	83.82%	5521	16.18%
Male	35361	108	0.31%	5015	14.18%	30238	85.51%	5123	14.49%
American Indian	153	0	0.00%	20	13.07%	133	86.93%	20	13.07%
African American	11700	6	0.05%	929	7.94%	10765	92.01%	935	7.99%
Asian	1192	26	2.18%	252	21.14%	914	76.68%	278	23.32%
Hispanic	10037	11	0.11%	1173	11.69%	8853	88.20%	1184	11.80%
White	43043	135	0.31%	7634	17.74%	35274	81.95%	7769	18.05%
Multiracial	3273	6	0.18%	440	13.44%	2827	86.37%	446	13.63%
Hawaiian/ Pacific Islander	42	0	0.00%	8	19.05%	34	80.95%	8	19.05%
Special Education	12083	3	0.02%	496	4.10%	11584	95.87%	499	4.13%
SES (High)	34247	121	0.35%	6639	19.39%	27487	80.26%	6760	19.74%
SES (Low)	34374	25	0.07%	3640	10.59%	30709	89.34%	3665	10.66%
LEP/ESL	3377	10	0.30%	207	6.13%	3160	93.57%	217	6.43%
Section 504	1733	2	0.12%	276	15.93%	1455	83.96%	278	16.04%

*Note. This table includes only first-time testers.

Table 50. Performance Level Distribution E/LA Grade 10 Spring Retest 2019

Subgroup	N	Performance Level						Proficiency	
		Pass+		Pass		Did Not Pass		N	%
		N	%	N	%	N	%		
All	22014	220	1.00%	2943	13.37%	18851	85.63%	18851	85.63%
Female	9257	111	1.20%	1494	16.14%	7652	82.66%	7652	82.66%
Male	12754	109	0.85%	1448	11.35%	11197	87.79%	11197	87.79%
American Indian	58	2	3.45%	9	15.52%	47	81.03%	47	81.03%
African American	5170	26	0.50%	553	10.70%	4591	88.80%	4591	88.80%
Asian	485	6	1.24%	67	13.81%	412	84.95%	412	84.95%
Hispanic	3666	23	0.63%	499	13.61%	3144	85.76%	3144	85.76%
White	11580	160	1.38%	1668	14.40%	9752	84.21%	9752	84.21%
Multiracial	1001	3	0.30%	138	13.79%	860	85.91%	860	85.91%
Hawaiian/ Pacific Islander	16	0	0.00%	4	25.00%	12	75.00%	12	75.00%
Special Education	7164	19	0.27%	440	6.14%	6705	93.59%	6705	93.59%
SES (High)	8346	139	1.67%	1319	15.80%	6888	82.53%	6888	82.53%
SES (Low)	13433	71	0.53%	1582	11.78%	11780	87.69%	11780	87.69%
LEP/ESL	2508	4	0.16%	195	7.78%	2309	92.07%	2309	92.07%
Section 504	651	8	1.23%	99	15.21%	544	83.56%	544	83.56%

Table 51. Performance Level Distribution Mathematics Grade 10 Spring Retest 2019

Subgroup	N	Performance Level						Proficiency	
		Pass+		Pass		Did Not Pass		N	%
		N	%	N	%	N	%		
All	42419	37	0.09%	2900	6.84%	39482	93.08%	39482	93.08%
Female	20795	14	0.07%	1525	7.33%	19256	92.60%	19256	92.60%
Male	21620	23	0.11%	1375	6.36%	20222	93.53%	20222	93.53%
American Indian	95	0	0.00%	6	6.32%	89	93.68%	89	93.68%
African American	8332	0	0.00%	336	4.03%	7996	95.97%	7996	95.97%
Asian	713	4	0.56%	94	13.18%	615	86.26%	615	86.26%
Hispanic	6617	3	0.05%	440	6.65%	6174	93.31%	6174	93.31%
White	24560	30	0.12%	1908	7.77%	22622	92.11%	22622	92.11%
Multiracial	2030	0	0.00%	112	5.52%	1918	94.48%	1918	94.48%
Hawaiian/ Pacific Islander	30	0	0.00%	4	13.33%	26	86.67%	26	86.67%
Special Education	8993	0	0.00%	281	3.12%	8712	96.88%	8712	96.88%
SES (High)	18717	29	0.15%	1636	8.74%	17052	91.10%	17052	91.10%
SES (Low)	23394	6	0.03%	1248	5.33%	22140	94.64%	22140	94.64%
LEP/ESL	2972	4	0.13%	143	4.81%	2825	95.05%	2825	95.05%
Section 504	1240	1	0.08%	92	7.42%	1147	92.50%	1147	92.50%

Table 52. Performance Level Distribution E/LA Grade 10 Spring 2019

Subgroup	N	Performance Level						Proficiency	
		Pass+		Pass		Did Not Pass		N	%
		N	%	N	%	N	%		
All	73564	19061	25.91%	26079	35.45%	28424	38.64%	28424	38.64%
Female	36268	11253	31.03%	13474	37.15%	11541	31.82%	11541	31.82%
Male	37294	7808	20.94%	12604	33.80%	16882	45.27%	16882	45.27%
American Indian	148	35	23.65%	61	41.22%	52	35.14%	52	35.14%
African American	8151	683	8.38%	2294	28.14%	5174	63.48%	5174	63.48%
Asian	1901	755	39.72%	589	30.98%	557	29.30%	557	29.30%
Hispanic	8528	1201	14.08%	2976	34.90%	4351	51.02%	4351	51.02%
White	51613	15658	30.34%	19039	36.89%	16916	32.77%	16916	32.77%
Multiracial	3139	717	22.84%	1093	34.82%	1329	42.34%	1329	42.34%
Hawaiian/ Pacific Islander	58	11	18.97%	18	31.03%	29	50.00%	29	50.00%
Special Education	9553	386	4.04%	1461	15.29%	7706	80.67%	7706	80.67%
SES (High)	42170	14818	35.14%	15930	37.78%	11422	27.09%	11422	27.09%
SES (Low)	31029	4208	13.56%	10069	32.45%	16752	53.99%	16752	53.99%
LEP/ESL	2853	14	0.49%	298	10.45%	2541	89.06%	2541	89.06%
Section 504	1481	303	20.46%	515	34.77%	663	44.77%	663	44.77%

*Note. This table includes only first-time testers.

Table 53. Performance Level Distribution Mathematics Grade 10 Spring 2019

Subgroup	N	Performance Level						Proficiency	
		Pass+		Pass		Did Not Pass		N	%
		N	%	N	%	N	%		
All	73791	3507	4.75%	21832	29.59%	48452	65.66%	48452	65.66%
Female	36342	1569	4.32%	11323	31.16%	23450	64.53%	23450	64.53%
Male	37447	1938	5.18%	10509	28.06%	25000	66.76%	25000	66.76%
American Indian	147	4	2.72%	38	25.85%	105	71.43%	105	71.43%
African American	8217	47	0.57%	905	11.01%	7265	88.41%	7265	88.41%
Asian	1903	343	18.02%	682	35.84%	878	46.14%	878	46.14%
Hispanic	8541	121	1.42%	1604	18.78%	6816	79.80%	6816	79.80%
White	51737	2861	5.53%	17811	34.43%	31065	60.04%	31065	60.04%
Multiracial	3157	126	3.99%	774	24.52%	2257	71.49%	2257	71.49%
Hawaiian/ Pacific Islander	61	5	8.20%	14	22.95%	42	68.85%	42	68.85%
Special Education	9602	57	0.59%	665	6.93%	8880	92.48%	8880	92.48%
SES (High)	42239	3124	7.40%	16258	38.49%	22857	54.11%	22857	54.11%
SES (Low)	31197	381	1.22%	5531	17.73%	25285	81.05%	25285	81.05%
LEP/ESL	2858	9	0.31%	125	4.37%	2724	95.31%	2724	95.31%
Section 504	1479	52	3.52%	372	25.15%	1055	71.33%	1055	71.33%

*Note. This table includes only first-time testers.

Appendix L: Quality Criteria Checklist

Quality Criteria Checklist- Items

Accessibility
<p>Does the item utilize Universal Design specifications? Is the item accessible for the following subpopulations?</p> <ul style="list-style-type: none"> ● Braille ● Visually impaired <p>Does the item utilize graphics or footnotes, if required or needed?</p>
<p>Does the item utilize</p> <ul style="list-style-type: none"> ● simple sentences? ● few clauses? ● grade-appropriate vocabulary? (<i>Note: Students are responsible for any vocabulary noted in the standards</i>) ● bulleted information as appropriate? ● sensitivity to linguistic complexity?
<p>Does the item provide context as needed to ensure student response? (<i>Context does not inhibit responses</i>)</p>

Sensitivity/Bias
<p>Sensitivity.</p> <ul style="list-style-type: none"> ● Are the items appropriate to use specifically in Indiana? ● Does the item avoid complex social issues, including <ul style="list-style-type: none"> ○ climate change / global warming? ○ deportation of immigrants? ○ the occult, witches, ghosts, or vampires? ○ religion? ○ sex? ○ disease? ○ death/dying? ○ politics? ○ gambling? ○ cancer? ○ drugs/Alcohol? ○ AIDS? ○ advocacy? ○ killing animals for sport? ○ other sensitive topics?

Bias
<ul style="list-style-type: none"> • Does the item avoid bias for gender and/or race? • Does the item avoid bias for idiomatic language?

Item Content
<p>Standards.</p> <ul style="list-style-type: none"> • Does the item align to one primary standard? <i>OR</i> • Does the item contain one content and one process standard if it is a dual-aligned item? (Math and Science only)
<p>Specifications</p> <ul style="list-style-type: none"> • Does the item align to limits of the item specification?
<p>Aligns to Depth of Knowledge (DOK)</p> <ul style="list-style-type: none"> • Does the item align to the DOK (<i>cognitive complexity</i>) defined by the standard?
<p>Content</p> <ul style="list-style-type: none"> • Is the content accurate? • Is the content grade level appropriate?
<p>Scoring</p> <ul style="list-style-type: none"> • Is there a single correct key for multiple choice items? • Does the Technology Enhanced Item scores correctly? (does not apply to Social Studies) • Does the metadata show the correct number of points? • For open ended items, is the rubric or top score response clear and complete?
<p>Functionality (TEI only)</p> <ul style="list-style-type: none"> • Does the TEI function the way you would expect it?

Universal Design Item Checklist

A. Precisely Designed Constructs

Definition: The item construct is clearly defined so that all irrelevant cognitive, sensory, emotional, and physical barriers are removed.

- The item does not add skills to those being measured (no extraneous skills tested).

B. Language Appropriateness

Definition: The item avoids words or phrases that are sexist, racist, or otherwise offensive, inappropriate, or negative to any subgroup. Language should be simple and clear.

- The item uses commonly used words—simpler is better.
- The item uses vocabulary appropriate for the grade level.
- Idiomatic speech and figurative language are avoided unless being measured.
- The item avoids technical terms unrelated to the content.
- The item contains no unnecessary words.
- The sentence complexity contained in the item is appropriate for the grade level.
- The item avoids ambiguous or multiple-meaning words (e.g., crane—the bird—can easily be confused with crane—heavy machinery).
- All pronouns have clear referents.
- The item avoids the use of proper names. (Such names may be unfamiliar or difficult for cultural subgroups.)
- The item avoids irregularly spelled words.

C. Gender Stereotypes

Definition: The item avoids stereotyping as results of associating genders with certain professions or activities. All groups of society should be portrayed accurately and fairly regarding gender.

- The item is free of content that might offend a gender subgroup.
- The item is free of content that might unfairly advantage or disadvantage a gender subgroup.

D. Ethnic Stereotypes

Definition: The item avoids unnecessary references to and uses the proper reference for ethnic, racial, or cultural groups.

- The item is free of content that might offend an ethnic subgroup.
- The item is free of content that might unfairly advantage or disadvantage an ethnic subgroup.
- The artwork included in an item adequately reflects the diversity of the student population.

E. Cultural Familiarity

Definition: Does not rely on an assumed shared experience that is class oriented or native English speaking oriented. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments.

- The item does not rely on an assumed shared experience that is class oriented or native English speaking oriented.
- The item is free from content that might offend a socioeconomic subgroup.

- The item is free of content that might unfairly advantage or disadvantage a socioeconomic subgroup.
- The item is free from unnecessary cultural references.
- The item is free from religious references.

F. Geographic Bias

Definition: All groups of society should be portrayed accurately and fairly regarding geographic setting.

- A particular geographic setting should not be used repeatedly, and urban, suburban, and rural settings should be represented across items.
- The item is free of content that might offend a geographic subgroup.
- The item is free of content that might unfairly advantage or disadvantage a geographic subgroup.

G. Disability Bias

Definition: All groups of society should be portrayed accurately and fairly regarding disability.

Stereotypes related to any particular disability should be avoided. No undue restrictions should exist in the item that would interfere with the ability of a student to comprehend or respond to the item.

- The item is free of content that might offend a disability subgroup.
- The item is free of content that might unfairly advantage or disadvantage a disability subgroup.
- A graphic representation is used in the items, as appropriate. The complexity of the graphic is appropriate to the purpose—simpler is better.
- The item avoids content that depends on sensory knowledge (such as references to movement, sound, smell, etc.) unless this is crucial to the overall item.
- The item could be put into Braille.
- The item avoids using both O and Q.
- Letter pairs can be easily distinguished when read. (S and T are okay; S and X are not).

H. Art Supports Text

Definition: Art is related to the item and supports the reader when possible. Item text and art are legible and accessible, and art is appropriately placed in the item to support the reader. Art does not distract the test taker and provides a scaffold to overall comprehension.

- All pictures relate to items.
- The item is free from pictorial clutter: All pictures are needed to answer the item.
- Graphics are clear and non-fuzzy.
- Any symbols used are highly distinguishable.
- Visual load requirements are reasonable for the grade level.
- Multi-dimensional graphics and complex shading are avoided.
- Tables have replaced any cluttered graphs.
- Labels read clockwise (as is easier for Braille readers).

I. Special Populations Considerations

Definition: Consideration must be given for maximum accessibility to all students including but not limited to English learners, limited sight, hearing impaired, cognitively challenged, etc. These considerations will assist all students.

- The item contains scaffolding techniques to support student understanding of what is being asked in the item.
- Text is replaced with graphic representations, when appropriate.
- The item is written with simplified text load.
- The item is written with simplified sentences.
- The item has as little extraneous information as possible.
- The item provides context, but it is simplified.
- The item uses smaller or less complicated numbers or expressions where not otherwise required.
- The item avoids negative phrasing or questions; for example, questions are not asked in the negative.