Technical Manual

ACT® QualityCore®

act.org/qualitycore

ACT endorses the *Code of Fair Testing Practices in Education* and the *Code of Professional Responsibilities in Educational Measurement*, guides to the conduct of those involved in educational testing. ACT is committed to ensuring that each of its testing programs upholds the guidelines in each *Code*.

Visit ACT's website at: www.act.org

v3.0

Contents

Figures ..... **iii**

Tables ..... **iv**

Preface ..... **v**

ACT's Mission **1**

Code of Fair Testing Practices in Education **and** ***Code of Professional Responsibilities in Educational Measurement*** **1**

Overview and Purpose of QualityCore® **2**
Foundations in Research **2**

QualityCore® End-of-Course Assessments **3**
Description of the QualityCore End-of-Course Assessments **3**
Test Development Procedures for the End-of-Course Assessments **4**
Selecting and Training Item Writers 4
Item Development 4
Content and Fairness Review of Items 5
Field-Testing 5
Training Sets (Exemplar Responses) 5
Field-Test Review and Operational Forms Construction 6
QualityCore Scores and Scoring **7**
Subscores 7
Scoring QualityCore Constructed-Response Items 7
Training Constructed-Response Item Scorers 9
Operational Training Materials for Constructed-Response Scoring 9
Estimated ACT® and PLAN® Scores and College Readiness Benchmarks 10
QualityCore Score Reports 11

Technical Characteristics of the QualityCore® End-of-Course Assessments **16**
National Study (2007–2008) **16**
Special Study (2009) **17**
Scaling **18**
Data 18
Scaling the Multiple-Choice Tests 19
Equating **28**
Data 28
Results 28
Comparability Study **32**
Data 32
Results 35

Estimated PLAN/ACT Test Score Ranges **49**
Data 49
Estimated PLAN/ACT Test Score Ranges 49
QualityCore National Administrations by Academic Year **56**

Validity Evidence for the QualityCore® End-of-Course Assessments **58**
On Course for Success **58**
ACT Course Standards **59**
High School Survey **60**
Expert Review **61**
Analysis of Content Experts' Review 62

Conclusion **63**

References **64**

Figures

Figure 1 Roster Report by Subscore....12

Figure 2 Roster Report by Gender....13

Figure 3 Roster Report by Ethnicity....13

Figure 4 Roster Multiple-Choice Only Scores....14

Figure 5 Overview Score Report....14

Figure 6 Student Report....15

Figure 7 Raw-to-Scale Score Transformations for QualityCore Base Forms....22

Figure 8 Scale Score Standard Error of Measurement, Graphed by True Scale Score for QualityCore Base Forms....25

Figure 9 Cumulative Percent of Raw Score for P&P and CBT Modes of QualityCore....37

Figure 10 Plots of Item P-values between P&P and CBT Modes of QualityCore....39

Figure 11 Cumulative Percent of Scale Score for P&P and CBT Modes of QualityCore....43

Figure 12 Plots of Item *b* Parameters between P&P and CBT Modes of QualityCore....45

Figure 13 Test Characteristic Curve Comparison of P&P and CBT Modes of QualityCore....47

Figure 14 Illustration of Equating Links for QualityCore MC + CR Forms....56

Tables

Table 1 College Readiness Benchmark Scores by Subject Test 10

Table 2 QualityCore Score Reports 11

Table 3 Overall Percent of Population and Sample by Stratum for the National Study (2007–2008) 17

Table 4 Overall Percent of Sample by Stratum for the Special Study (2009) 18

Table 5 Demographic Information for Scaling Samples 19

Table 6 Raw Test Score Summary Statistics and Fit of the Four Parameter Beta Compound Binomial Model for QualityCore Base Forms 21

Table 7 Demographic Information for New Forms 29

Table 8 Raw Score Summary Statistics for New Forms 29

Table 9 Scale Score Comparison of Two Equating Methods for Five QualityCore Courses 30

Table 10 Scale Score Summary of Equating Forms for Nine QualityCore Courses 31

Table 11 Demographic Information for Comparability Study Samples 34

Table 12 Raw Score Summary for P&P and CBT Modes of QualityCore 36

Table 13 Scale Score Summary for P&P and CBT Modes of QualityCore Using Separate Equatings 42

Table 14 Scale Score Summary of Test Scores Used to Establish the Estimated PLAN/ACT Score Range Study 51

Table 15 Estimated PLAN/ACT Score Intervals for QualityCore Scores 52

Table 16 Scale Score Summary of QualityCore Tests Administered by Academic Year 57

Table 17 Number of Surveys Returned by Course 61

Preface

The purpose of this manual is to document the technical characteristics of the QualityCore® End-of-Course (EOC) Assessments in light of their intended purposes. ACT regularly conducts research as part of the ongoing evaluation of its programs. The research is intended to ensure that the programs are technically sound.

This report describes the development of the assessments and provides validity and reliability information, results of scaling and equating activities, and interpretations of assessment results. Summary technical information is presented, including a description of the item analysis procedures for the multiple-choice and constructed-response items, and disaggregated test and strand level results for gender and ethnicity. Research studies conducted in support of the EOC assessments are also discussed in this manual, including linking studies that result in estimated PLAN® or ACT® subscores based upon QualityCore EOC scores.

The content of this manual complies with requirements of the testing industry as established in the *Code of Professional Responsibilities in Educational Measurement* (NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), and the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004). This manual is divided into four sections. These sections include the following information:

- An overview of QualityCore
- A description of the QualityCore End-of-Course Assessments
- QualityCore technical characteristics, such as reliability, scaling, and estimated ACT test score ranges
- Validity evidence for the QualityCore End-of-Course Assessments

We encourage individuals who want more detailed information on a topic discussed in this manual, or on a related topic, to contact ACT. Please direct comments or inquiries to ACT, QualityCore Research Services (31), P.O. Box 168, Iowa City, Iowa 52243-0168.

Iowa City, Iowa
January 2010

ACT's Mission

ACT is a nonprofit organization dedicated to helping people achieve education and workplace success.

Code of Fair Testing Practices in Education and *Code of Professional Responsibilities in Educational Measurement*

Since publication of the original edition in 1988, ACT has endorsed the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004), a statement of the obligations to test takers of those who develop, administer, or use educational tests and test data. The development of the *Code* was sponsored by a joint committee of the American Association for Counseling and Development, Association for Measurement and Evaluation in Counseling and Development, American Educational Research Association, American Psychological Association, American Speech-Language-Hearing Association, and National Council on Measurement in Education to advance, in the public interest, the quality of testing practices.

The *Code* sets forth fairness criteria in four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. Separate standards are provided for test developers and for test users in each of these four areas.

ACT's endorsement of the *Code* represents a commitment to vigorously safeguard the rights of individuals participating in its testing programs. ACT employs an ongoing review process whereby each of its testing programs is routinely reviewed to ensure that it upholds the standards set forth in the Code for appropriate test development practice and test use.

Similarly, ACT endorses and is committed to complying with the *Code of Professional Responsibilities in Educational Measurement* (NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), a statement of professional responsibilities for those who develop assessments; market and sell assessments; select assessments; administer assessments; interpret, use, and communicate assessment results; educate about assessment; and evaluate programs and conduct research on assessments.

Overview and Purpose of QualityCore®

Research by ACT, Inc., has demonstrated the critical need for rigorous college preparatory courses for all high school students. When students have the opportunity to take rigorous core courses and when their course achievement is reliably assessed, the prospects of their being ready for college and work dramatically increase (ACT, 2007).

QualityCore® is a research-based system of educator's resources, formative items, and end-of-course assessments designed to help schools better prepare all students for college and workforce training. It accomplishes this goal by ensuring that high school core courses are focused on the most essential postsecondary skills and by enabling educators to monitor progress toward, and attainment of, those skills. Through QualityCore, ACT hopes to raise the overall quality of high school core courses across the country. The goal of QualityCore is to help more students be ready for college and work after high school.

Foundations in Research

Beginning in 2003, ACT and The Education Trust collaborated on a study to determine the courses, level of rigor, and instructional practices that are most likely to lead to success for students. *On Course for Success* (2004) focused on schools with high minority and low-income student populations that produce graduates who meet or exceed the ACT College Readiness Benchmarks in proportions greater than those seen nationally.

Preliminary Course Standards developed during that study provided real answers to the questions of what a rigorous course looks like and how it can best be taught. The Course Standards were then reviewed by a group of teachers selected from the 300 highest ACT-performing high schools nationwide. Their recommendations and the resulting ACT Course Standards form the foundation for the test specifications of the QualityCore End-of-Course (EOC) Assessments.

QualityCore® End-of-Course Assessments

Description of the QualityCore End-of-Course Assessments

QualityCore® End-of-Course (EOC) Assessments measure the learning outcomes all students need to attain in order to succeed in college and in their careers. Each assessment includes problem-based items embedded in contexts that are accessible and relevant to high school students. These real-world problems require practical applications of concepts, theories, principles, and processes.

The EOC system is modular, consisting of either two 35–38 item multiple-choice components or one 35–38 item multiple-choice component combined with a constructed-response component. Each component requires 45 minutes of testing time. Students may take the full assessment across two 45-minute class periods or during one 90-minute block of time. The tests are offered on computer and as a paper-and-pencil administration, depending on test configuration and content area. (Tests that include a constructed-response component are not available online for mathematics, science, or social studies because of the complexity of the constructed-response items.)

QualityCore EOC assessments are aligned to rigorous, empirically derived ACT Course Standards (ACT, Inc., & The Education Trust, 2004). In developing the item pool from which the EOC assessments are assembled, each item is coded to a particular ACT standard. During the forms construction process, items are gathered into reporting categories that comprise clusters of the standards. (See *End-of-Course Test Blueprints*.)

Each item is also assigned a depth-of-knowledge (DOK) level (Webb, 2002) to describe the thinking processes measured on the EOC assessments. Webb developed descriptions of the DOK levels specifically for mathematics, science, social studies, and English language arts. The following text is an example of these descriptions as they apply to mathematics.

> Level 1: Recall requires the recall of information such as a fact, term, definition, or simple procedure. Students must demonstrate a rote response or perform a simple procedure.
>
> Level 2: Skill/Concept requires mental processing that goes beyond recalling or reproducing an answer. Students must make some decisions about how to approach a problem. The cognitive demands are more complex than in Level 1.
>
> Level 3: Strategic Thinking requires planning, thinking, explaining, justifying, using evidence, conjecturing, and postulating. The cognitive demands are complex and abstract, going beyond Level 2.

The *End-of-Course Test Blueprint* for each course shows the distribution of items for each reporting category and the distribution of items for each of the DOK levels as well as a more complete description of the thinking processes covered in the assessments.

Test Development Procedures for the End-of-Course Assessments

This section describes the procedures used to develop multiple-choice and constructed-response items for the QualityCore End-of-Course (EOC) Assessments in English, mathematics, science, and social studies.

Selecting and Training Item Writers

Item writers are chosen from a pool of well-qualified high school teachers who have extensive content knowledge, and who represent both genders as well as diverse ethnic backgrounds and geographic locations. These item writers include active high school teachers from a variety of schools, small to large, public and private. During initial development, content specialists at ACT work closely with QualityCore item writers, revising and refining items, to produce high-quality items designed to meet test specifications, to reflect diversity, and to meet fairness standards.

ACT provides multiple-choice item writers with detailed guidelines that specify the content, cognitive skill level, and format for each multiple-choice item. These guidelines also include specifications for nondiscriminatory subject matter and language usage.

Constructed-response item writers are given a comprehensive course-specific guide for writing constructed-response items to the specifications of the QualityCore program. These item writers must submit a writing sample before they are contracted to write items that may ultimately be selected for use on field tests or on operational forms.

Item Development

Items developed for the EOC tests assess content aligned to the research-based ACT Course Standards, which detail the material students should master by the end of each course. EOC items are contextualized with scenarios that relate to students' everyday activities. The items present problem-based tasks embedded in rich, authentic contexts.

Constructed-response items require students to explain, justify, critique, create, propose, produce, design, or otherwise demonstrate knowledge and understanding in ways that cannot be assessed by multiple-choice items. These items—like the multiple-choice items—are rigorous, so students can write thoughtful and detailed responses while making realistic and meaningful connections to real-world situations. When possible, context is provided to reinforce students' practical applications of concepts, theories, principles, and processes.

Students taking the English test respond to one item; those taking the social studies tests respond to two items; and students taking the mathematics and science tests respond to three items. Each constructed-response test is designed such that students should be able to respond in the allotted 45-minute testing time.

Content and Fairness Review of Items

After EOC items are developed and refined by ACT test development specialists, the items go through a rigorous review by external content experts. These experts ensure the multiple-choice and constructed-response items accurately assess the content standard assigned to the item, are free of factual and graphical errors, are clearly presented, and assess only what students should know and be able to do at the end of a rigorous course of study. Additionally, the constructed-response scoring criteria are reviewed to ensure they are correct and complete and contain only information relevant to the item.

All items are also reviewed by fairness experts who carefully consider each item to ensure that neither the language nor the content of the item will be offensive to a student, and that no item, through context or language, will disadvantage any student from any geographic, socioeconomic, or cultural background.

Field-Testing

New EOC items are field-tested every development cycle with students from rural and urban settings, small and large schools, and public and private schools. Multiple-choice items may also be embedded in operational forms. Schools are asked to administer field-test items to students who have taken the corresponding course in school. The field tests are administered in the spring at the end of the course. The goal is to have at least 500 responses for each test item.

Student responses to the constructed-response items are read and scored by trained ACT raters. Training for ACT field-test raters is very similar to training for operational raters. Training materials for field-testing of constructed-response test items are created during "range-finding." Range-finding involves test development specialists, other content experts, and expert raters previewing student responses to determine whether the content-specific scoring criteria for each item accurately reflect and encompass all of the acceptable student responses.

During range-finding, student responses are individually rated by multiple raters using the scoring criteria and the appropriate rubric. Responses that do not receive the same score by all raters are discussed by the entire group and a consensus score is reached. Additions and clarifications to the scoring criteria can be made at this time, if necessary. On assessments which do not employ scoring criteria, such as QualityCore English items, discussion centers on correctly identifying which responses best exemplify each score point on the scoring guide.

Training Sets (Exemplar Responses)

After test development specialists and expert raters have completed range-finding, responses are sorted and put together in training sets. Typically, for field-test training, one item is selected as the qualification item. Training materials for this item are more extensive than the training materials for other items, since this will typically be the initial training set used for new raters. Training materials for the qualification item include an anchor set, multiple practice sets, and at least two qualification sets that prospective raters must pass in order to score the field-test responses.

Responses chosen for the anchor set represent clear examples of each score point. If possible, these responses represent the full range of student responses typically seen during range-finding. Practice set and qualification set responses also include clear examples; however, these sets also contain responses that are not as perfectly aligned to the scoring criteria, falling either slightly high or low within each score point. The test development specialist tries to include particular kinds of responses that occur frequently and that will be challenging to score, to ensure that raters are qualified to handle these more difficult scoring decisions during live scoring.

Range finders may select additional responses to be used for recalibration sets if the actual scoring of the item is anticipated to take more than one day. Training materials for additional items also contain an anchor set and practice sets, but do not include qualification sets.

Each response included in the training materials is analyzed by the range-finding team. An articulation (rationale) explaining why a particular score was assigned accompanies the response. The articulation explains how the holistic rubric and scoring criteria were used to determine the score; citations from the exemplar response are included, where appropriate, to illustrate the claims made in the articulation.

Field-Test Review and Operational Forms Construction

Once the EOC-field test items are scored, they undergo statistical analysis focusing on difficulty, validity, and accessibility to determine whether they are suitable for operational use. The new EOC items are reviewed to ensure they adhere to rigorous QualityCore standards and are comparable to previous operational items. Items are then chosen to fulfill the test blueprint developed for each course.

Operational multiple-choice and constructed-response test forms are constructed by selecting eligible items from the item pool that align with content (blueprints) and statistical specifications. A good-faith effort is made to present diversity on every test form. The resulting operational forms are then subjected to three reviews. The first review, performed by ACT's content experts, editorial staff, and measurement staff ensures content accuracy, conformity to statistical and content test specifications, fairness, and adherence to best testing practices. The forms are then reviewed by two sets of external consultants who did not review the items prior to field-testing: content reviewers (including classroom teachers, college faculty, and curriculum specialists) and fairness reviewers. Items that do not pass a review are removed from the pool. Depending on the nature of the issue, such items are either discarded from the pool or are edited and placed into the development cycle anew.

The resulting QualityCore EOC assessments contain a wide array of analytical and problem-solving tasks, including in-depth constructed-response prompts. These assessments measure students' ability to apply the content knowledge and reasoning skills acquired in their course work to high-level tasks.

QualityCore Scores and Scoring

QualityCore End-of-Course (EOC) Assessment results can be used to identify student (both individual and group) achievement relative to ACT Course Standards and college readiness. In addition, the results can give administrators a sense of course quality and consistency across classrooms and schools.

A QualityCore scale score is provided for each QualityCore EOC assessment. For the multiple-choice segment(s) of the test, the raw score is the number of correct responses. Constructed-response raw scores are the number of points the student obtains. These raw scores are converted to scale scores through statistical scaling/equating procedures. The QualityCore scale ranges from 125 to 175, lowest to highest. Equating ensures that scores reported across test forms have a constant meaning, e.g., a scale score of 150 on English 11 Form 1 has the same meaning as a 150 on English 11 Form 2. It is not possible, however, to compare a score of 150 on the English 11 test to a 150 on the Algebra II test. Scale scores for different subjects are established independently, and the equating procedures are implemented only within the same subject.

Subscores

Subscores reflect student performance with respect to topical areas and groups of ACT standards. They are reported as raw scores (for multiple-choice, the number of questions a student answered correctly; for constructed-response, the points obtained). Three to five subscores are reported for each QualityCore EOC assessment. Summaries of the content standards for each subscore are displayed on the overview reports and in the *End-of-Course Test Blueprints*. Subscores are reported as the number of points the student obtained compared to the number of points possible for that subscore. There is no arithmetic relationship between the subscores and the QualityCore final score. Since the subscores are raw scores and are not equated, they cannot be compared to subscores obtained on different forms of the test. Subscores are provided to give teachers insight into the relative performance of their students and their classes in various content areas. Based on the subscores, teachers also can adapt their teaching strategies and content coverage.

Scoring QualityCore Constructed-Response Items

Each QualityCore EOC constructed-response item is scored by a human rater who uses an analytic or holistic rubric. Rubrics for the various assessments are provided in the *End-of-Course Test Blueprints*. Each blueprint shows the reporting category with which the constructed-response items are aligned.

The QualityCore EOC English constructed-response assessments consist of one item. The rater scores the student essay using an analytic rubric and gives four different scores based on the following domains: content, development, organization, and language use. Each domain is scored on a six-point scale, totaling to 24 possible points on the essay. Essay content is evaluated based on a student's demonstrated ability to use course-specific writing skills, including analysis of meaningful reflective experiences, analysis of literary elements, articulation of judgments, or analysis of complex issues that remain unresolved. Development is achieved through thorough explanation and a clear, consistent focus. Organization is demonstrated by organizing and

presenting ideas in a logical way. Language use is measured by clear communication and effective application of the conventions of standard written English. The student's essay result is reported as its own subscore.

Three constructed-response items are included on the QualityCore mathematics constructed-response tests. Raters score the items using a four-point holistic rubric. The raw scores for the constructed-response items are doubled, totaling 24 possible points. Student responses are evaluated based on the evidence of the student's demonstrated ability to address all parts of the question or problem correctly; to demonstrate efficient and accurate use of appropriate procedures, showing (as requested) all steps used to find the answer, including those performed on a calculator; to show a proficient understanding of mathematical concepts and principles in the explanation; and to provide a clear and coherent explanation. Responses are scored holistically on the basis of the overall impression created by all of these criteria—correctness of content, procedures, conceptual understanding, and communication. Subscores for the mathematics tests include the results from both the constructed-response items and the multiple-choice items. Constructed-response results are included in the appropriate subscore.

Student responses to the three items on the QualityCore science constructed-response tests are also scored using a four-point holistic rubric. The raw scores for the constructed-response items are doubled, totaling 24 possible points. Responses are evaluated based on the evidence they demonstrate of a student's ability to address all parts of the question or problem correctly and thoroughly; to demonstrate insight into scientific concepts or principles through elaboration, extension, and detail; and to clearly explain and enhance the answer by correctly using appropriate scientific terminology. Responses are scored holistically on the basis of the overall impression created by all of these criteria—correctness of content, conceptual understanding, and communication. Subscores for the science tests include the results from both the multiple-choice items and the related constructed-response items. Constructed-response results are included in the appropriate subscore.

Two constructed-response items are included on the QualityCore U.S. History constructed-response tests. Responses to the U.S. History analysis item are typically scored using a four-point holistic rubric, while the visual stimulus responses are scored according to a three-point holistic rubric. The raw scores for the constructed-response items are tripled, totaling 21 possible points. Responses are evaluated on the evidence they demonstrate of a student's ability to address all parts of the task correctly and thoroughly; to demonstrate insight into history concepts or principles; to effectively support logical conclusions through analysis, argument, problem solving and/or evaluation; and to communicate clearly and effectively. Responses are scored on the basis of the overall impression created by all of these criteria—correctness of content, conceptual understanding, and communication. Subscores for the U.S. History tests include the results from both the multiple-choice items and the related constructed-response items. Constructed-response results are included in the appropriate subscore.

Training Constructed-Response Item Scorers

All prospective raters, at a minimum, must have an undergraduate degree and many prospective raters also have graduate degrees. They are then assigned to score content area tests (English, mathematics, science, or social studies) that most closely match their degree. Every effort is made to hire prospective raters who are teaching or have taught in a high school or postsecondary setting.

Prospective raters receive intensive, on-site training by test development specialists at ACT who have experience with constructed-response tests. Considerable time is spent discussing the scoring rubric so raters understand the implications of overall holistic or individual analytic domain scoring and how the rubric criteria should be applied at each score point. Also, expectations for student performance are explained, and raters are shown how student performance varies at each score point. Course rigor is stressed, and the implications for these expectations and for scoring responses are discussed in detail. Raters learn how the rubric interacts with item-specific scoring criteria (the specific content knowledge and understanding that students must demonstrate in each response) and how both the rubric and the scoring criteria are used to evaluate each response. Raters also learn to evaluate performance in relation to all rubric criteria—not just correctness of content.

During training, prospective raters are presented with multiple learning opportunities to apply this newly acquired knowledge and understanding to actual student responses. First, they learn how the rubric was used by experienced raters to score exemplar student responses, and then they practice on their own. Once prospective raters have completed all learning opportunities and have worked with experienced raters to discuss questions and resolve their scores for responses they had difficulty scoring, prospective raters take a minimum of two qualification sets that they must pass at predetermined standards. Raters who pass are qualified to score operational responses; they are then monitored throughout the scoring session. Blind validity responses (responses with known scores) and recalibration training responses are used frequently to ensure and maintain consistent and accurate scoring.

Operational Training Materials for Constructed-Response Scoring

Operational training materials are constructed after a form, which can consist of one to three items, has been selected according to the course-specific blueprint. Training materials from the field test are retained for use when appropriate. However, more student responses are needed for operational training materials than those typically used for field-testing training materials. The test development specialist selects responses that meet the requirements of the operational training materials.

Operational training materials include a more robust anchor set and multiple practice sets for each item in the form. In addition, mathematics and science operational training materials include a mixed-item practice set and at least two mixed-item qualification sets. The operational training materials also include multiple recalibration sets. Recalibration sets are constructed from student responses that are typical and/or especially challenging to score. These responses may be

chosen during operational scoring as the test development specialist becomes aware of particular scoring issues that need to be addressed.

All responses chosen for operational training material sets conform to the same guidelines and specifications as the field-test training sets. Each response is articulated to explain the score assigned to each paper. The field-test responses initially chosen as operational training materials are eventually replaced with operational responses as scoring sessions continue and similar examples from live, operational testing conditions can be identified and articulated. Ultimately, operational training materials contain only operational student responses.

Raters participate in a full training session for each new form they are asked to score, but once they have qualified on a particular form, they only need to pass additional qualification sets to score that form at a later date; they do not need to repeat the full training session.

Estimated ACT® or PLAN® Scores and College Readiness Benchmarks

ACT has conducted research to compare QualityCore test scores to ACT® or PLAN® test scores and provides an estimated ACT or PLAN subject test score range on the student Overview Report, along with an indication of whether the student is or will be ready for college-level work. The estimated ACT or PLAN score range is an estimation of what a student's score on the ACT or PLAN subject test would be if that test were taken at the same time as the QualityCore test. These scores are only estimates, not guarantees, based upon national samples tested as part of a special research study. The width of the intervals for the ranges varies with the QualityCore score. In general, the ACT/PLAN ranges are wider in the middle of the QualityCore score scale than at the extremes. (See Table 15.)

ACT has identified, for each ACT subject test, the minimum score that indicates a 50% chance of obtaining a B or higher or a 75% chance of obtaining a C or higher in the corresponding first-year college course. This is called the College Readiness Benchmark Score. (See Table 1.) On the QualityCore student score report, students are given an estimated ACT or PLAN range that is compared to the relevant ACT College Readiness Benchmark.

Table 1
College Readiness Benchmark Scores by Subject Test

Subject Test	PLAN Test Score	ACT Test Score
English	15	18
Mathematics	19	22
Reading	18	22
Science	20	23

Some suggestions about how students might improve their college readiness, based upon ACT's research and College Readiness Standards®, are included on their QualityCore Student Report.

QualityCore Score Reports

QualityCore provides a variety of online reports designed to convey relevant score information to teachers and administrators. (See Table 2.) Teachers receive information about the performance of each of their students as well as information about their class as a group. Administrators have access to the student-level reports and to reports that roll up data into larger groups, allowing insight into school-, district-, and statewide performance.

Examples of the score reports are presented in Figures 1–6. See the *QualityCore Online User Guide* for more information about accessing and navigating score reports.

Table 2
QualityCore Score Reports

Report	Teacher Access	School Access	District Access
Roster Report by Subscore	Provides average class QualityCore (QC) score and student QC scores and subscores. Teacher's access is limited to assigned class rosters.	Provides average QC scores and subscores at the school, class, and student levels. User has ability to drill down to class and teacher reports.	Provides average QC scores and subscores at the district, school, class, and student levels. User has ability to drill down to school, class, and teacher reports.
Roster Report by Gender	Provides class average QC score and average QC score by gender at the class level. Teacher's access is limited to assigned class rosters.	Provides average QC score for total population and average QC score by gender at the school, class, and student levels. User has ability to drill down to class and teacher reports.	Provides average QC score for total population and average QC score by gender at the district, school, class, and student levels. User has ability to drill down to school, class, and teacher reports.
Roster Report by Ethnicity	Provides average QC score for total population and average QC score by ethnicity at the class level. Teacher's access is limited to assigned class rosters.	Provides average QC score for total population and average QC score by ethnicity at the school, class, and student levels. User has ability to drill down to class and teacher reports.	Provides average QC score for total population and average QC score by ethnicity at the district, school, class, and student levels. User has ability to drill down to school, class, and teacher reports.
Overview Score Report	Provides average QC score at the class level and average QC subscores. Teacher's access is limited to assigned class rosters.	Provides average QC score at the school and class levels. User has ability to drill down to class and teacher reports.	Provides average QC score at the district, school, and class levels. User has ability to drill down to school, class, and teacher reports.
Student Report	Provides QC score, % of students at school/district/state levels performing at or below the student's score, college readiness report, and subscores. Teacher's access is limited to assigned class rosters.	Contains QC score, % of students at the school/district/ state levels performing at or below the student's score, college readiness report, and subscores. School's access is limited to students enrolled in the school.	Contains QC score, % of students at the school/district/state levels performing at or below the student's score, college readiness report, and subscores. District's access is limited to students enrolled in the district.
Roster Multiple-Choice Only Scores Only available for multiple-choice/constructed-response administrations.	Provides multiple-choice raw data (percent correct) at the class and student levels. User has ability to drill down to class and teacher reports. Teacher's access is limited to assigned class rosters.	Provides multiple-choice raw data (percent correct) at the school, class, and student levels. User has ability to drill down to class and teacher reports.	Provides multiple-choice raw data (percent correct) at the district, school, class, and student levels. User has ability to drill down to school, class, and teacher reports.

Figure 1. Roster Report by Subscore

SCHOOL FINAL QUALITYCORE REPORT: Algebra I

Test Date: 12/2007-11/2009
Sub-district: Demo Sub-District
State: Demo State

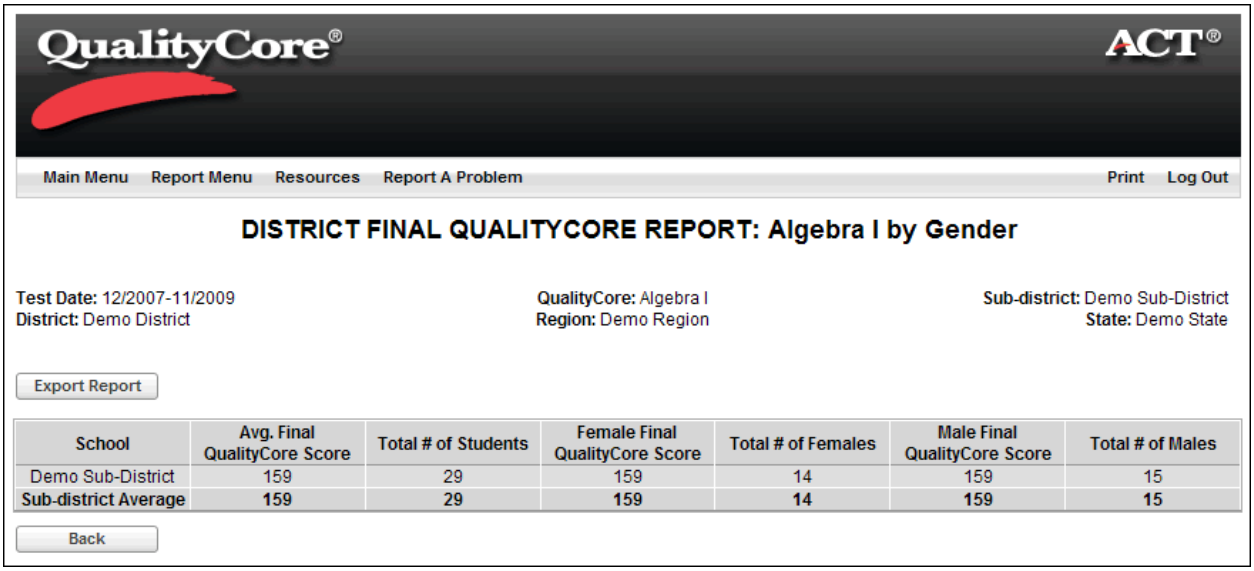
QualityCore: Algebra I
District: Demo District

School: Demo School #1
Region: Demo Region

Export Report

Teacher/Group /Student	Student ID	Avg. Final QualityCore Score	Exploring Expressions, Equations, and Functions in the First Degree	Exploring Other Nonlinear Equations and Functions	Exploring Quadratic Equations and Functions	Number Sense, Operation and Graph Skills	# of Students
Ortiz, Michelle							
Ortiz Alg I School 1 Period One							
Baldwin, Michael		165	27	11	10	15	1
Chang, Sharon		155	13	11	13	10	1
Cortez, Lorna		164	26	11	11	14	1
Nguyen, Hiep		160	18	12	14	12	1
Perez, Miguel		159	21	12	14	8	1
Smith, Edward		145	4	8	12	4	1
Tapahonso, Sheri		150	9	10	12	7	1
Torrence, Shauna		171	27	12	14	15	1
Group Average		**159**	**18**	**11**	**13**	**11**	**8**
Overall Group Average		**159**	**18**	**11**	**13**	**11**	**8**
Chu, Amy							
Chu Alg I School 1 S1							
Baldwin, Michael		153	26	8	9	0	1
Chang, Sharon		169	25	12	14	16	1
Cortez, Lorna		150	11	6	4	16	1
Nguyen, Hiep		158	20	9	13	11	1
Perez, Miguel		149	18	5	7	5	1
Smith, Edward		167	25	11	13	16	1
Tapahonso, Sheri		162	19	12	14	14	1
Torrence, Shauna		148	12	8	5	9	1
Group Average		**157**	**20**	**9**	**10**	**11**	**8**

Figure 2. Roster Report by Gender



QualityCore® ACT®

Main Menu Report Menu Resources Report A Problem Print Log Out

DISTRICT FINAL QUALITYCORE REPORT: Algebra I by Gender

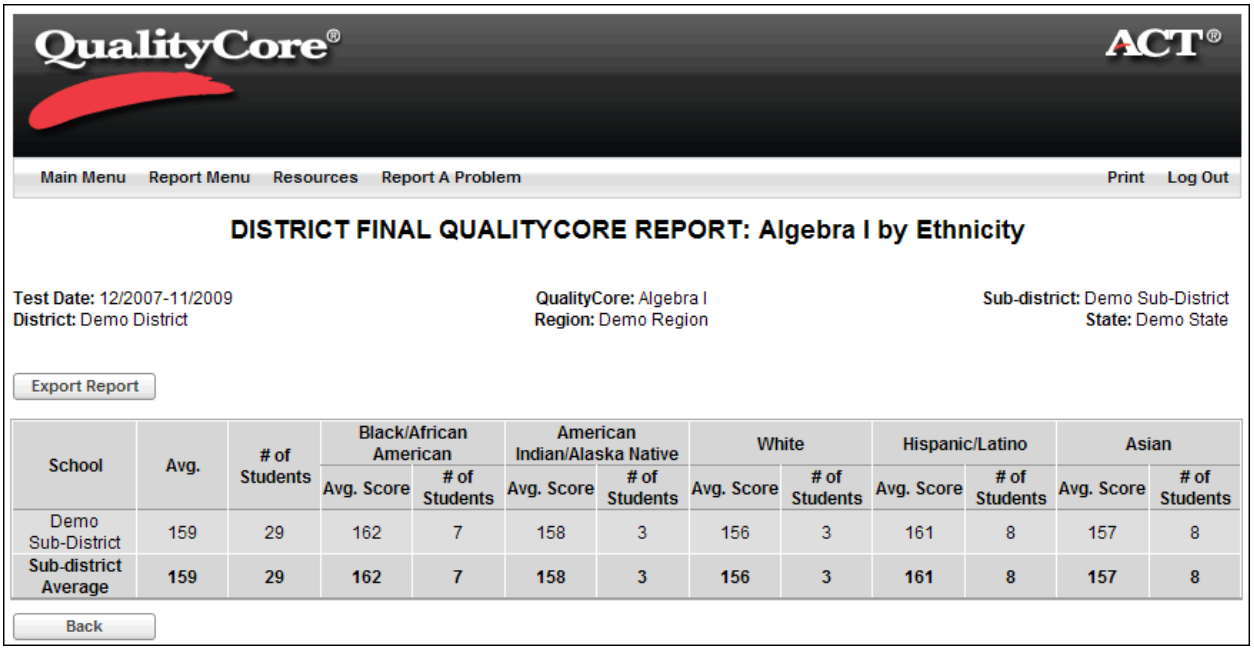
Test Date: 12/2007-11/2009
District: Demo District
QualityCore: Algebra I
Region: Demo Region
Sub-district: Demo Sub-District
State: Demo State

Export Report

School	Avg. Final QualityCore Score	Total # of Students	Female Final QualityCore Score	Total # of Females	Male Final QualityCore Score	Total # of Males
Demo Sub-District	159	29	159	14	159	15
Sub-district Average	**159**	**29**	**159**	**14**	**159**	**15**

Back

Figure 3. Roster Report by Ethnicity



QualityCore® ACT®

Main Menu Report Menu Resources Report A Problem Print Log Out

DISTRICT FINAL QUALITYCORE REPORT: Algebra I by Ethnicity

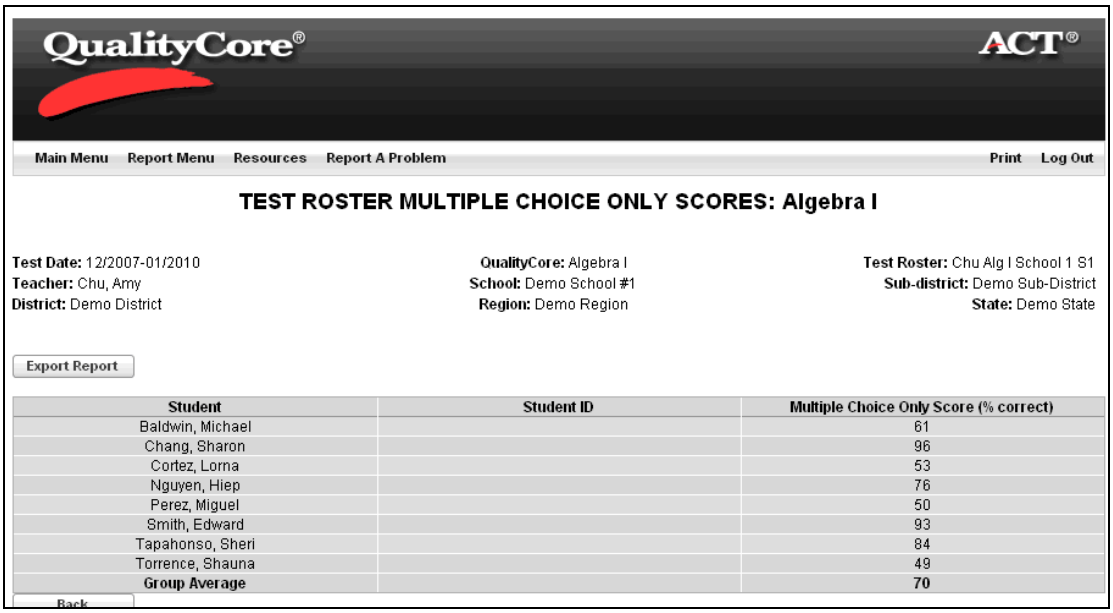
Test Date: 12/2007-11/2009
District: Demo District
QualityCore: Algebra I
Region: Demo Region
Sub-district: Demo Sub-District
State: Demo State

Export Report

School	Avg.	# of Students	Black/African American		American Indian/Alaska Native		White		Hispanic/Latino		Asian	
			Avg. Score	# of Students	Avg. Score	# of Students	Avg. Score	# of Students	Avg. Score	# of Students	Avg. Score	# of Students
Demo Sub-District	159	29	162	7	158	3	156	3	161	8	157	8
Sub-district Average	**159**	**29**	**162**	**7**	**158**	**3**	**156**	**3**	**161**	**8**	**157**	**8**

Back

Figure 4. Roster Multiple-Choice Only Scores



QualityCore® ACT®

Main Menu Report Menu Resources Report A Problem Print Log Out

TEST ROSTER MULTIPLE CHOICE ONLY SCORES: Algebra I

Test Date: 12/2007-01/2010
Teacher: Chu, Amy
District: Demo District

QualityCore: Algebra I
School: Demo School #1
Region: Demo Region

Test Roster: Chu Alg I School 1 S1
Sub-district: Demo Sub-District
State: Demo State

Export Report

Student	Student ID	Multiple Choice Only Score (% correct)
Baldwin, Michael		61
Chang, Sharon		96
Cortez, Lorna		53
Nguyen, Hiep		76
Perez, Miguel		50
Smith, Edward		93
Tapahonso, Sheri		84
Torrence, Shauna		49
Group Average		**70**

Back

Figure 5. Overview Score Report

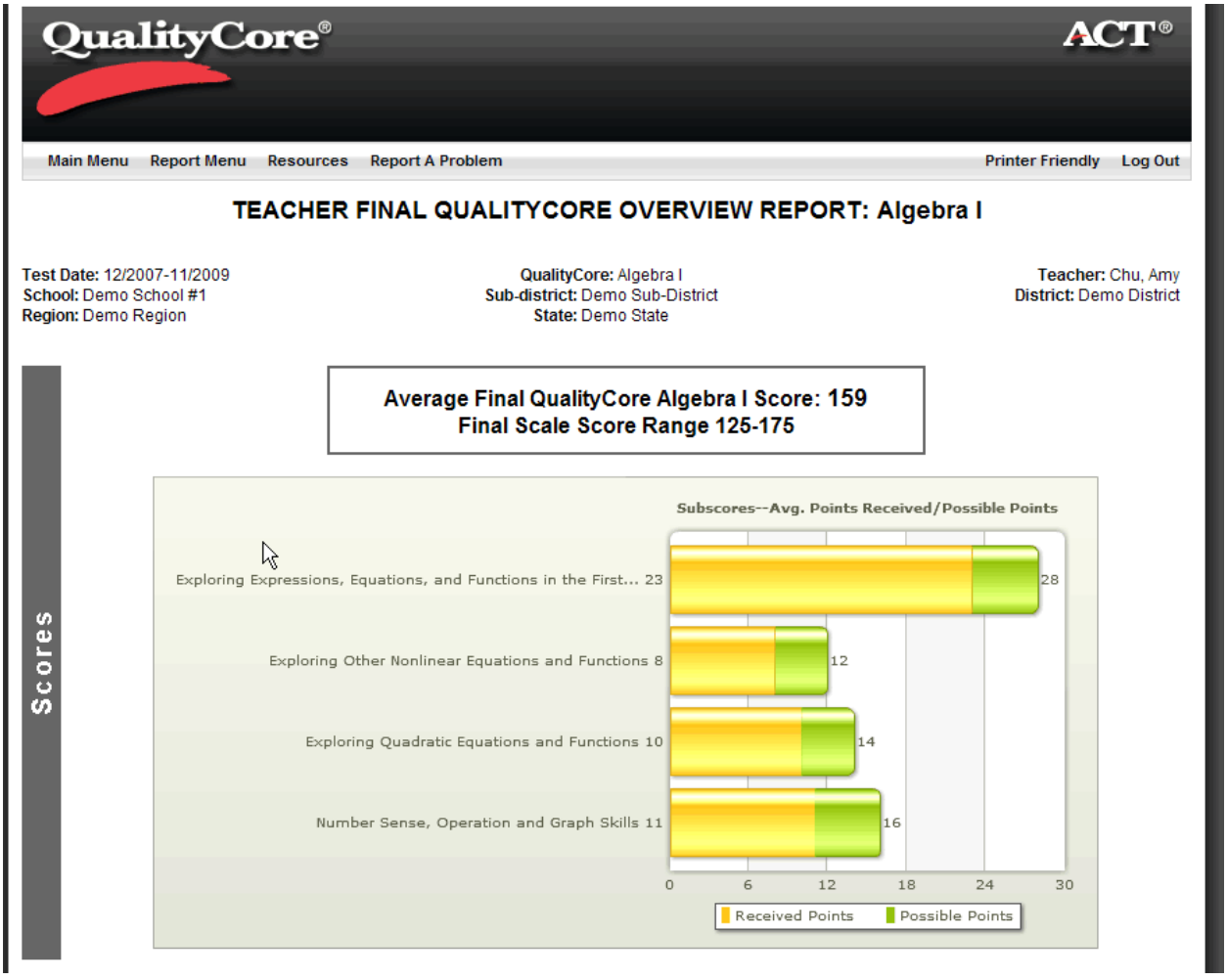
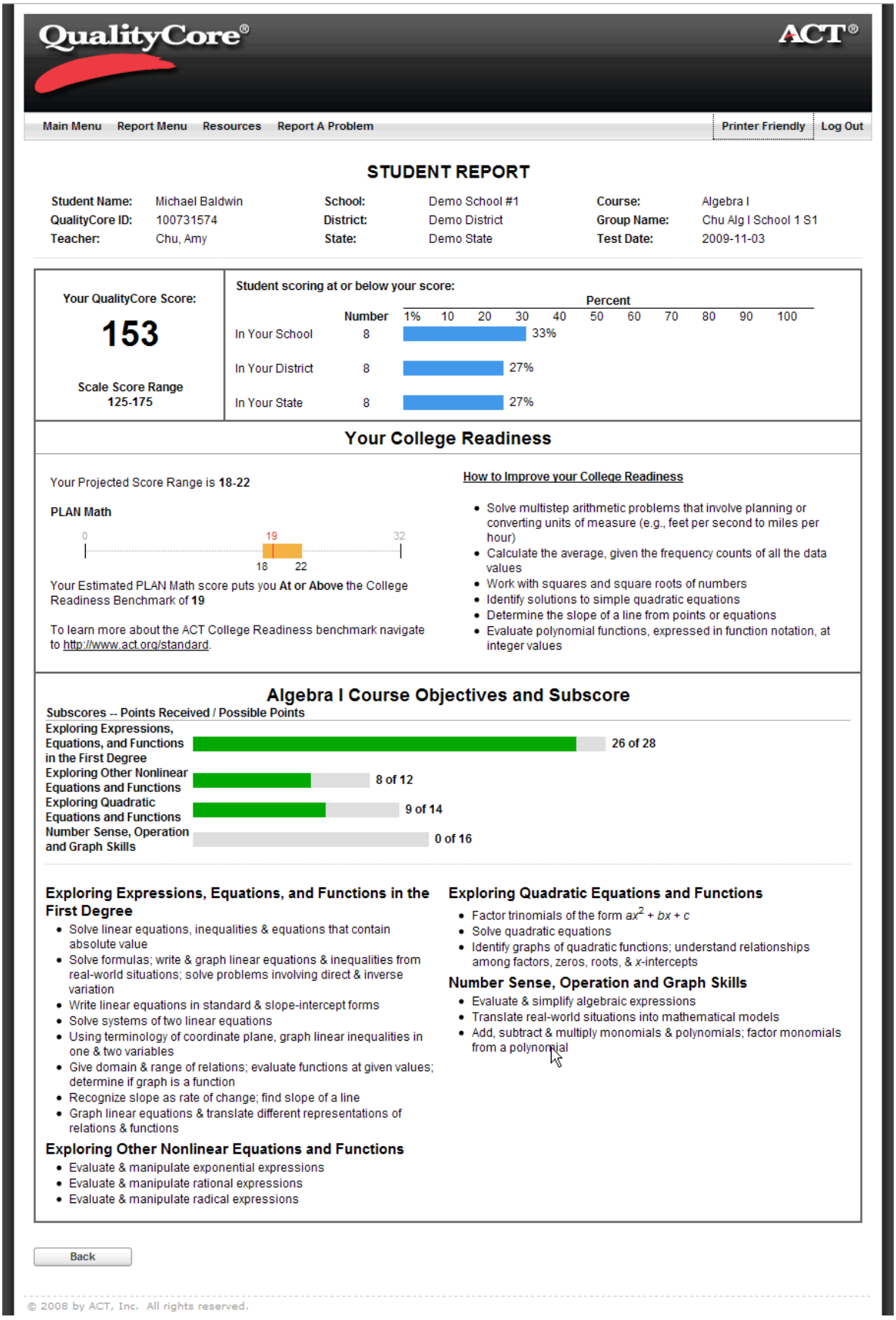


Figure 6. Student Report



Technical Characteristics of the QualityCore® End-of-Course Assessments

This section describes the technical characteristics of the 12 QualityCore® End-of-Course (EOC) Assessments administered in 2008 and 2009. More specifically, it summarizes the setting of scales for English 9, English 10, English 11, English 12, Algebra I, Algebra II, Geometry, Precalculus, Biology, Chemistry, Physics, and U.S. History; equating across test forms; the comparability study of paper-and-pencil (P&P) test forms and computer-based test (CBT) forms; and determination of an estimated PLAN® or ACT® test score range for each of the eight QualityCore scale scores (one scale score per course). Data from a national study and a comparability study cond99ucted in the 2007–2008 academic year were used in the above analyses for eight courses: English 10, English 11, English 12, Algebra I, Algebra II, Geometry, Biology, and Chemistry. Data from a special study conducted in 2009 was used in scaling and equating (English 9 only), and linking to PLAN or the ACT test of four courses: English 9, Precalculus, Physics, and U.S. History. The national study and the special study are described below.

National Study (2007–2008)

The national study was conducted during the 2007–2008 academic year. Participating schools were asked to test students enrolled in courses that matched the QualityCore content areas (English 10, English 11, English 12, Algebra I, Algebra II, Geometry, Biology, and Chemistry). The EOC test and a corresponding PLAN or ACT subject test were administered at the end of the course. The purpose of the study was to provide data to be used to scale the tests and to link them to the PLAN or ACT.

The target population consisted of students in grades 10 through 12 in public and private U.S. schools. The sample was selected to be nationally representative. A stratified random-sampling design was used in which schools were stratified by size, type (public/private), and geographical location. School size and type were explicit strata. Small schools (less than 100 students per grade) were excluded to ensure stable results for scaling, equating, and estimated PLAN/ACT test score linking studies. Within the fixed strata, the lists of schools were ordered by ZIP code and a systematic sample was drawn. This implicit stratification was to obtain a geographic representation of schools.

To attain a reasonable spectrum of school performance, the number of schools per course was set at a minimum range of 50–60. To meet this minimum, approximately 90 schools per course were invited to participate. Some schools tested in more than one subject area. During recruitment, the number of schools in each stratum was carefully monitored to maintain the representativeness of the sample.

Table 3 lists characteristics of the participating schools, the percentage of the target population in each stratum, as well as the percentage of participation in the national study by stratum. The characteristics of the subset of schools included in the scaling study are also provided. Compared to the target population, larger proportions of private schools and schools in the Midwest participated, while smaller proportions of public schools and schools in the East region participated.

Table 3
Overall Percent of Population and Sample by Stratum for the National Study (2007–2008)

Stratum	Target Population	National Study	Scaling Sample
Type			
Public	92	88	67
Private	8	12	33
Size			
400–799	39	43	42
800 plus	61	57	58
Region			
East	46	38	27
Midwest	24	41	49
Southwest	10	14	10
West	19	7	14

Special Study (2009)

The special study of four QualityCore courses (English 9, Precalculus, Physics, and U.S. History) was conducted in the spring of 2009. Participating schools were asked to test students enrolled in the courses being studied. In this study, a QualityCore test and a corresponding PLAN or ACT subject test were administered. The purpose of the study was to obtain data to use in scaling the tests and to link them to PLAN or the ACT.

The target population consisted of students in Grades 9–12 in public and private U.S. schools. The sample was nationally representative. Very large and very small schools were excluded.

Table 4 lists the demographics of the schools that were recruited for, and participated in, the special study. Compared to the recruited schools, larger proportions of private, rural, and Midwest schools participated, while smaller proportions of public, suburban, and schools in the East participated.

Table 4
Overall Percent of Sample by Stratum for the Special Study (2009)

Stratum	Recruited	Participated
Type		
Public	86	76
Private	14	24
Size		
1–799	58	61
800 plus	42	39
Region		
East	35	29
Midwest	26	41
Southwest	15	11
West	24	19
Metro Code		
Rural/Non-Metro	42	54
Suburban	36	25
Urban	22	21

Scaling

Scaling is a process of setting up a rule of correspondence between the observed raw scores and the scale score values assigned to them for reporting purposes. The usefulness of a score scale depends on whether it can facilitate meaningful interpretations and minimize misinterpretation and unwarranted inferences (Petersen, Kolen & Hoover, 1989). This section summarizes how the score scales for the QualityCore English 9, English 10, English 11, English 12, Algebra I, Algebra II, Geometry, Precalculus, Biology, Chemistry, Physics, and U.S. History EOC tests were established using data from the national and special studies.

Data

Table 5 provides demographic information for the final samples used in the analyses. Student records with any of the following issues were excluded from all analyses:

- Missing form identification code
- Missing ID number and last name
- Missing raw responses (A, B, C, D) and scored responses (0, 1) for all items
- Missing raw total score or raw total score equal to zero
- Having 75% (26 items) or more omitted items per section

After the data were cleaned based on the above rules, additional student records were removed from analyses if students were in nontraditional grade levels for a particular course. For English 9, 10,

11, and 12, only current 9th-, 10th-, 11th-, and 12th-grade students, respectively, were included in the sample. For Algebra I and Biology, only current 9th- or 10th-grade students were included. For Algebra II and Chemistry, only current 10th- or 11th-grade students were included. For Geometry and U.S. History, only current 9th-, 10th-, or 11th-grade students were included. For Precalculus, only current 11th- or 12th-grade students were included. For Physics, only current 10th-, 11th-, or 12th-grade students were included.

Table 5
Demographic Information for Scaling Samples

	English 9	English 10	English 11	English 12	Algebra I	Algebra II
Sample size	2,226	2,873	2,667	1,439	1,832	1,986
Gender						
Male	1,075	1,432	1,379	700	920	936
Female	1,099	1,406	1,264	726	884	1,021
Not specified	52	35	24	13	28	29
Race/Ethnicity[a]						
White		1,960	1,884	927	1,258	1,507
African American		278	227	137	190	156
Hispanic	N.A.[b]	419	376	270	278	220
Asian/Pacific Islander		134	110	64	50	54
American Indian		57	59	23	41	39
Not specified/ Other		247	213	112	147	143

	Geometry	Precalculus	Biology	Chemistry	Physics	U.S. History
Sample size	2,838	1,561	2,757	2,535	1,890	2,228
Gender						
Male	1,315	764	1,334	1,217	1,072	1,083
Female	1,490	789	1,387	1,283	793	1,127
Not specified	33	8	36	35	25	18
Race/Ethnicity[a]						
White	2,016		1,873	1,834		
African American	209		260	196		
Hispanic	384	N.A.[b]	442	340	N.A.[b]	N.A.[b]
Asian/Pacific Islander	97		106	110		
American Indian	50		62	49		
Not specified/ Other	252		238	197		

[a] Some students selected more than one race/ethnicity.

[b] N.A.: Race/Ethnicity information was not collected in the Spring 2009 special study.

Scaling the Multiple-Choice Tests

The score scales for the QualityCore tests were developed using the equal standard error of measurement method developed by Kolen (1988). This method was also used to establish scale scores for the ACT (Brennan, 1989), PLAN, and EXPLORE® tests.

The number of scale score points was selected to be 51 by considering the number of raw score points (i.e., 70 for tests with two multiple-choice [MC] components and 59 for tests with one multiple-choice component and one constructed-response [CR] component) and potential guessing effect (guessing equals 1/4 of the maximum multiple-choice score). Total raw score minus guessing is approximately 53 for the MC + MC test and 50 for the MC + CR test. Thus, 51 was selected as the number of possible scale score points. The scale score range of 125–175 was chosen to avoid any potential confusion with the ACT scale (scale score range = 1–36), other established scales, and percent correct (or percentile) scores.

The initial criteria used for constructing the score scale for the QualityCore EOC tests were (a) possible scale scores should cover the entire score range of 125–175; (b) there should not be too many gaps in the raw-to-scale score conversion tables, and there should not be too many raw scores converting to the same scale score; (c) the standard error of measurement should be as constant as possible across the scale score range; and (d) the standard error of measurement should be as close to 2 points as possible, so an approximate 68% confidence interval could be formed by adding ±2 points to students' scale scores.

Table 6 provides summary statistics for the distributions of raw test scores and for the distribution fit using the four parameter beta compound binomial model. The bottom portion of Table 6 summarizes the fitted distribution. (For an interpretation, please refer to Kolen and Hanson [1989].) For each test, the raw scores were transformed using the arcsine transformation, computed by $c(i) = 1/2(\sin^{-1}\sqrt{i/(k+1)} + \sin^{-1}\sqrt{(i+1)/(k+1)})$, where i = raw score, $\sin^{-1}$ is the arcsine function, and k = number of items. The transformed arcsine values were then linearly transformed to a score scale using $s = A \times c(i) + B$, where s is the scale score, A is the slope, and B is the intercept. $A = (s_{\max} - s_{\min})/[c(i)_{\max} - c(i)_{\min}]$ and $B = s_{\max} - A \times c(i)_{\max}$ with the target mean and standard error of measurement. These linearly transformed scores were rounded, truncated, and adjusted (where necessary) to stay within the reported 125–175 score range.

Figure 7 depicts the raw-to-scale score transformations for all 12 courses. In this figure, the points refer to the final raw-to-scale score conversion that resulted after rounding, truncating, and adjusting, where necessary.

As indicated earlier, one of the major goals of the scaling process was to produce scale scores that would have approximately equal standard errors of measurement along most of the score scale. If this goal was met, then the same standard error of measurement could be used for all students, regardless of their scores. In Figure 8, the scale score standard error of measurement is presented by the true scale score.[1] In this figure, note that the true scale scores do not go down to a scale score of 125. The minimum true scale score was approximately 132 for all 12 courses. This minimum was indirectly defined by the l parameter of the four parameter beta distribution of true scores. For most of the true scale score range, the scale score standard error of measurement was reasonably constant on all eight tests, at approximately 2.1.

[1] The true scale score is defined as the expected value of the student's test scores over many repeated testings with the same test.

Table 6

Raw Test Score Summary Statistics and Fit of the Four Parameter Beta Compound Binomial Model for QualityCore Base Forms

	English 9	English 10	English 11	English 12	Algebra I	Algebra II
No. of items	70	70	70	70	70	70
Sample size	2,226	2,873	2,667	1,439	1,832	1,986
Observed						
Mean	43.61	40.29	37.28	32.83	28.06	28.02
SD	13.82	14.07	15.45	14.01	8.15	8.73
Skewness	-0.43	–0.29	0.09	0.30	0.48	0.43
Kurtosis	2.21	2.01	1.81	1.97	3.12	2.97
Reliability	0.93	0.93	0.94	0.93	0.79	0.82
SEM	3.62	3.63	3.64	3.74	3.77	3.75
Fitted						
Mean	43.61	40.29	37.28	32.83	28.06	28.02
SD	13.82	14.07	15.45	14.01	8.15	8.73
Skewness	-0.43	–0.29	0.09	0.30	0.48	0.43
Kurtosis	2.21	2.01	1.81	1.97	3.01	2.90
Lord's *K*	2.04	3.06	2.32	2.05	4.11	4.01
No. of moments fit	4	4	4	4	3	3
Beta-a	1.55	1.20	0.79	0.72	2.54	2.83
Beta-J3	0.92	0.85	0.90	1.04	7.65	7.25
Beta-*l*	0.16	0.17	0.20	0.20	0.20	0.17
Beta-*u*	0.90	0.86	0.91	0.86	1.00	1.00
Pearson χ^2 (*df*)	110.38 (66)	104.06 (66)	147.75 (66)	96.93 (66)	273.05 (67)	133.27 (67)

	Geometry	Precalculus	Biology	Chemistry	Physics	U.S. History
No. of items	70	70	70	70	70	70
Sample size	2,838	1,561	2,757	2,535	1,890	2,228
Observed						
Mean	27.15	28.50	33.93	28.82	29.11	27.20
SD	8.19	9.32	11.51	10.64	9.29	9.43
Skewness	0.52	0.86	0.24	0.61	0.86	0.74
Kurtosis	3.12	3.85	2.28	2.82	3.78	3.15
Reliability	0.78	0.83	0.89	0.87	0.83	0.84
SEM	3.82	3.81	3.78	3.81	3.78	3.82
Fitted						
Mean	27.15	28.50	33.93	28.82	29.11	27.20
SD	8.19	9.32	11.51	10.64	9.29	9.43
Skewness	0.52	0.86	0.24	0.61	0.86	0.74
Kurtosis	3.07	3.50	2.28	2.82	3.50	3.15
Lord's *K*	2.85	2.96	3.28	2.29	3.71	2.13
No. of moments fit	3	3	4	4	3	4
Beta-a	2.34	0.91	1.48	1.15	0.88	0.99
Beta-J3	7.54	3.79	2.15	2.98	3.67	3.25
Beta-*l*	0.20	0.27	0.21	0.21	0.28	0.23
Beta-*u*	1.00	1.00	0.89	0.93	1.00	0.90
Pearson χ^2 (*df*)	64.85 (67)	101.07 (67)	116.84 (66)	74.05 (66)	124.20 (67)	52.66 (66)

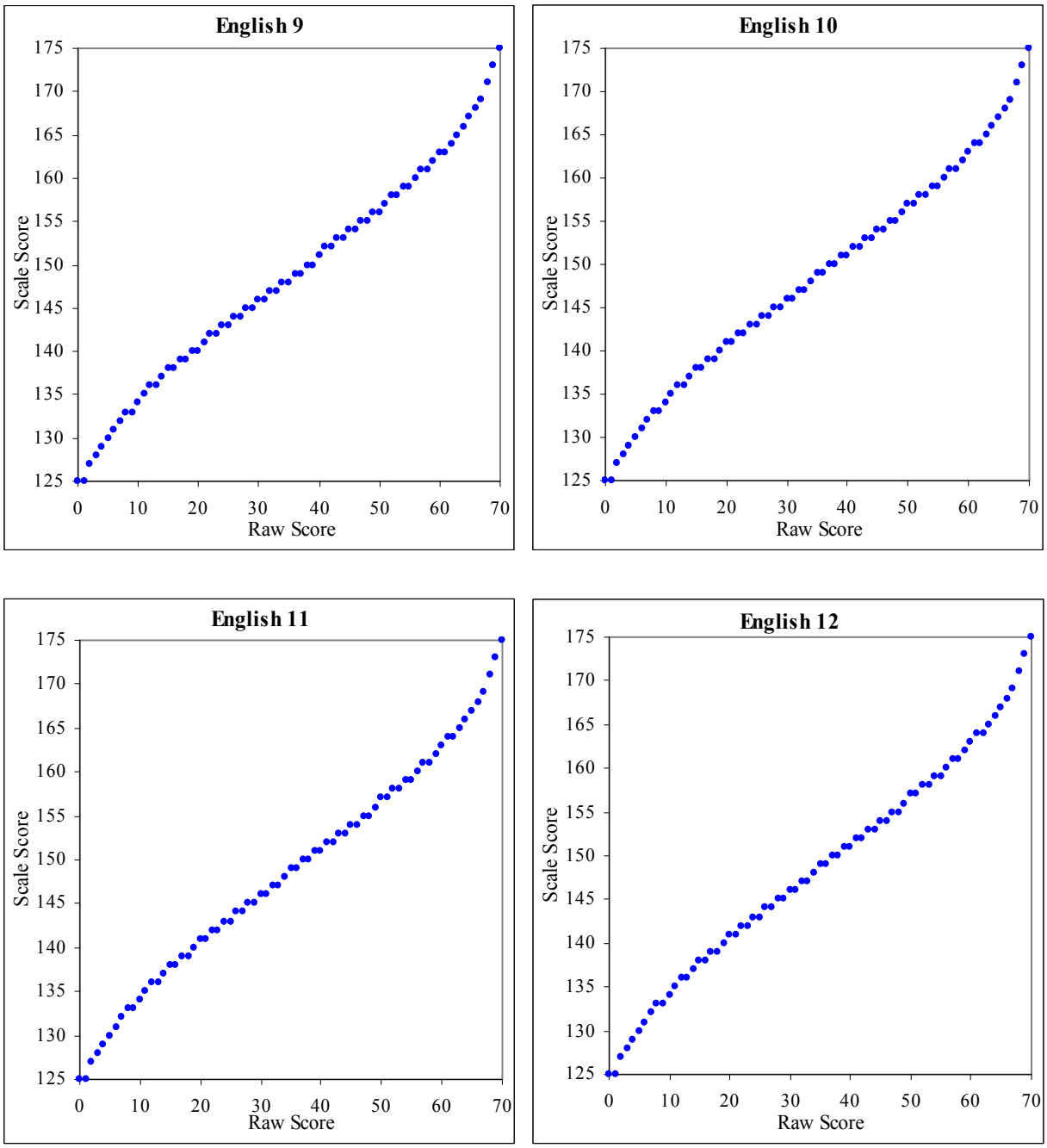


Figure 7. Raw-to-Scale Score Transformations for QualityCore Base Forms

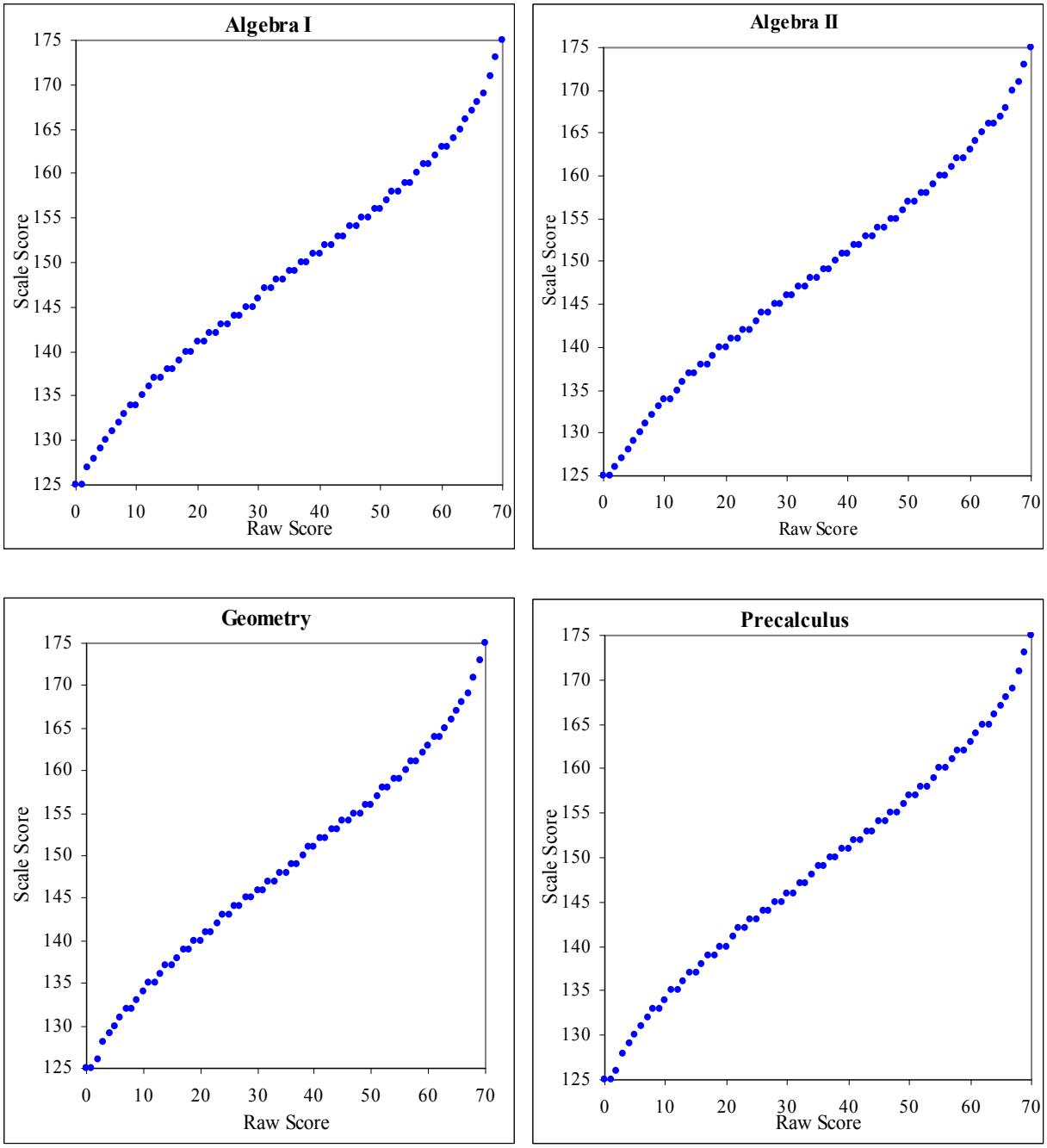


Figure 7. (continued) Raw-to-Scale Score Transformations for QualityCore Base Forms

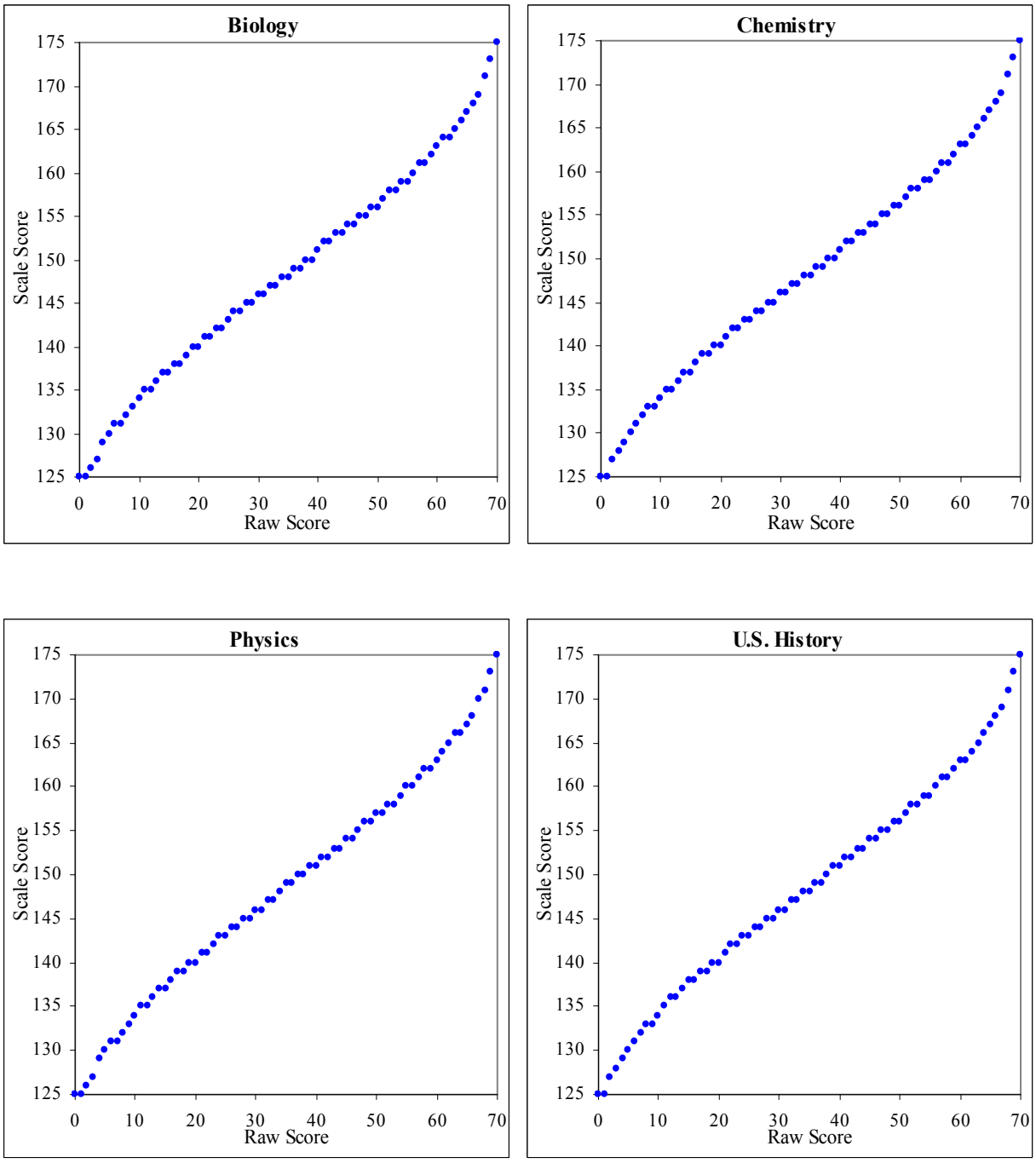


Figure 7. (continued) Raw-to-Scale Score Transformations for QualityCore Base Forms

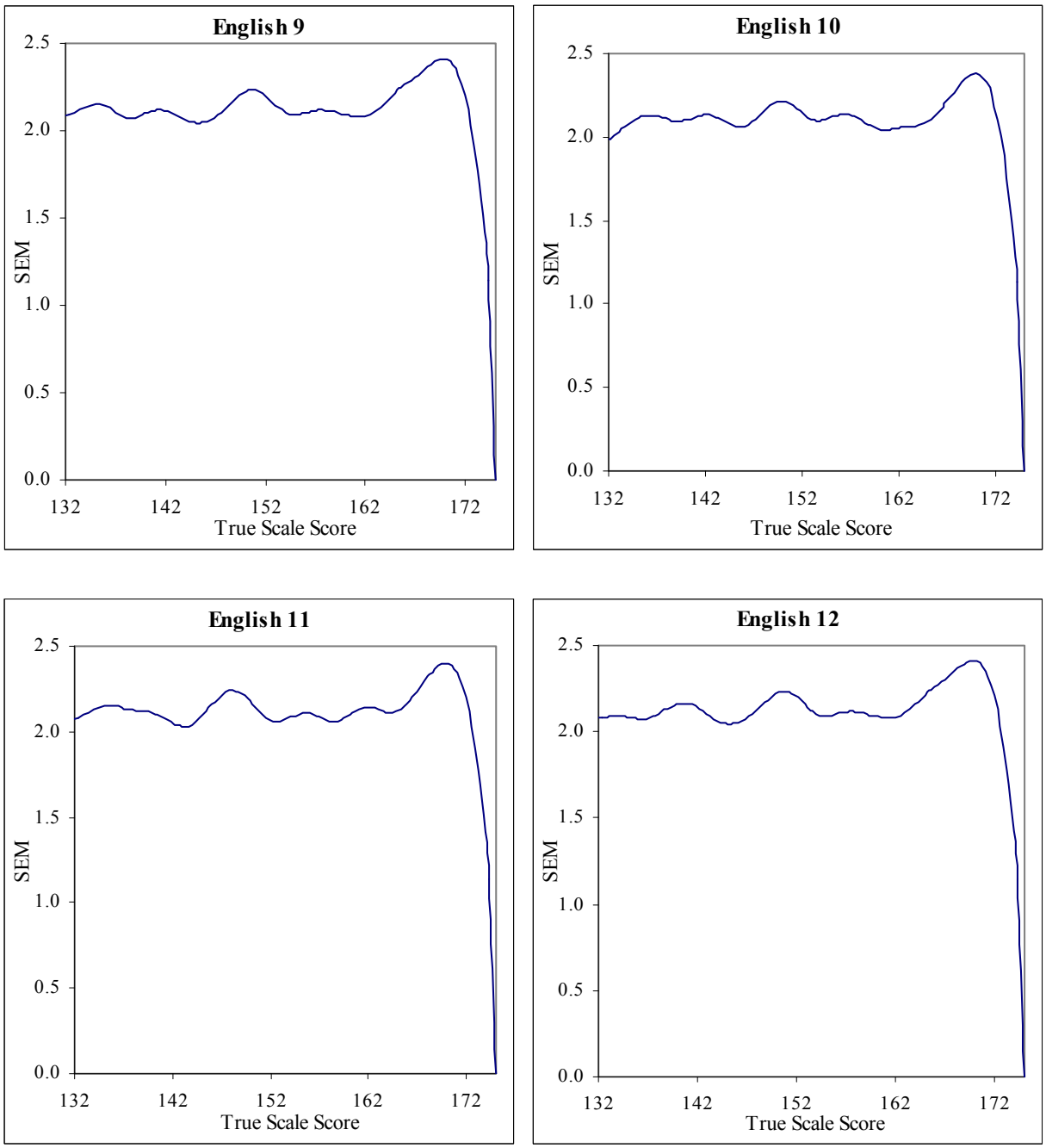


Figure 8. Scale Score Standard Error of Measurement, Graphed by True Scale Score for QualityCore Base Forms

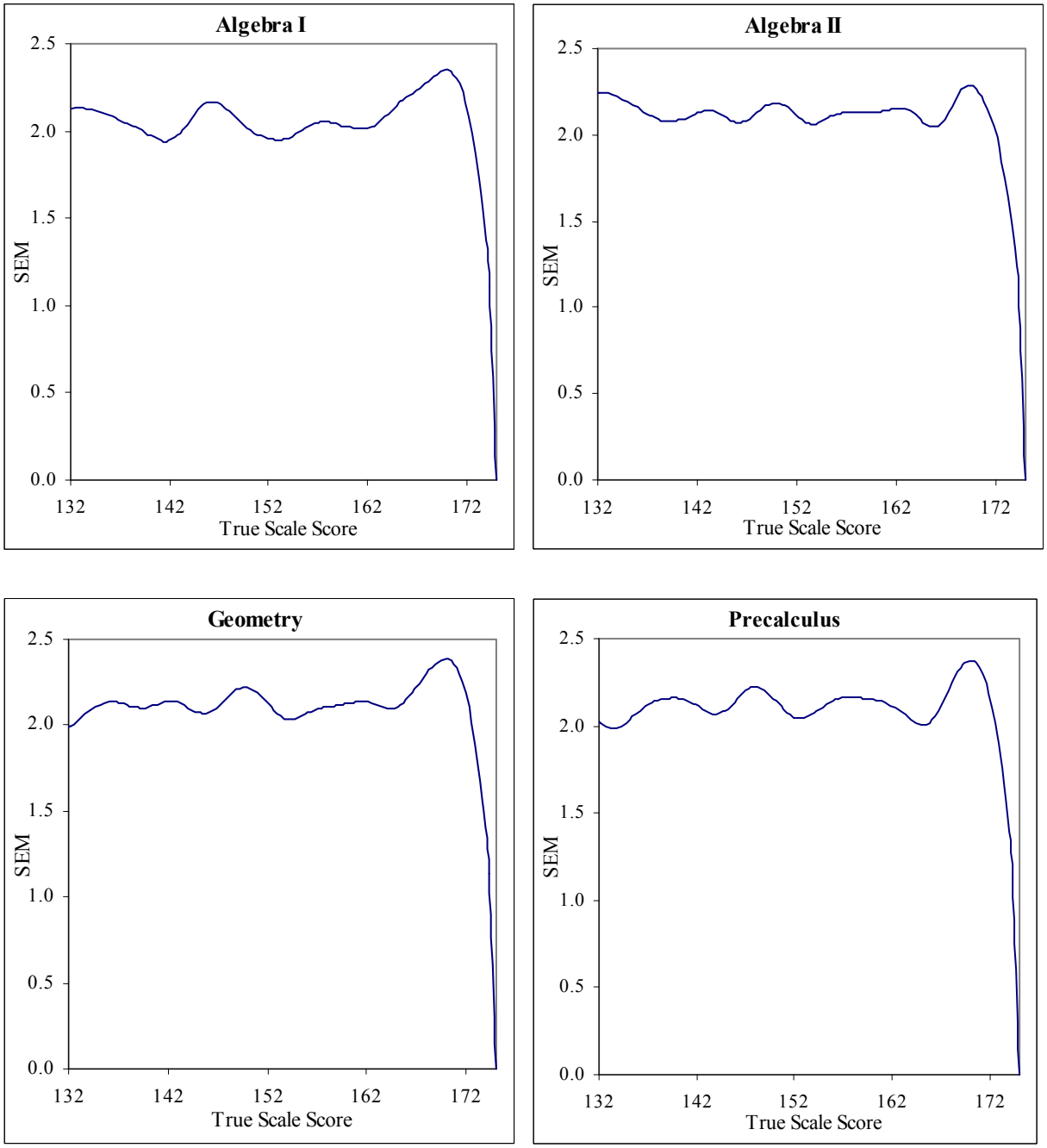


Figure 8. (continued) Scale Score Standard Error of Measurement, Graphed by True Scale Score for QualityCore Base Forms

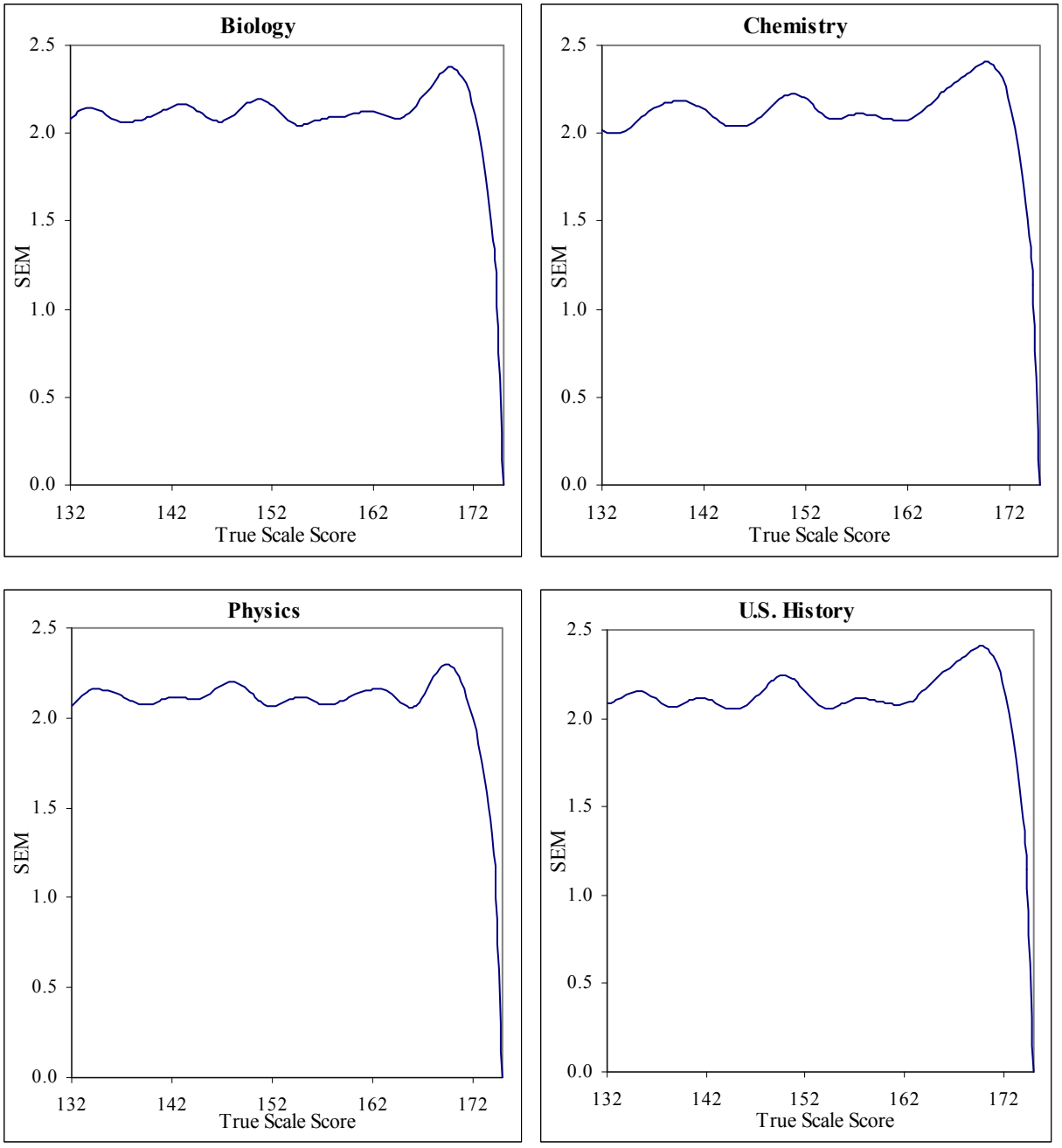


Figure 8. (continued) Scale Score Standard Error of Measurement, Graphed by True Scale Score for QualityCore Base Forms

Equating

Though each QualityCore form is constructed to adhere to the same content and statistical specifications for each subject, the form difficulty may vary slightly. To control for this variation, scores on all forms within a subject are equated to the same scale so that when they are reported as scale scores, the equated scale scores have the same meaning regardless of the form administered. For this reason, scale scores are comparable across test forms; however, they are not comparable across different QualityCore subject tests. For instance, scale scores of 150 on English 10 form A and form B are comparable, but a scale score of 150 on English 10 is not comparable to a scale score of 150 on English 11 or to a scale score of 150 on Algebra I.

This section summarizes how the equating method for the QualityCore tests was determined. It also describes the equating of the English 9, English 10, English 11, English 12, Algebra I, Algebra II, Geometry, Biology, and Chemistry new forms to the base forms used in the national study (2007–2008) and the special study (2009), described earlier.

Data

Student records with any of following issues were excluded from all analyses:

- Missing form identification code
- Missing ID number and last name
- Missing raw responses (A, B, C, D) or scored responses (0, 1) for all items
- Missing raw total score or raw total score equal to zero
- Having 75% (26 items) or more omitted items per section

After the data were cleaned based on the above rules, additional student records were removed from analyses if students were in nontraditional grade levels for a particular course by adopting the same criteria that were applied to scaling. The final sample sizes, therefore, for the base forms were the same as those that were used in scaling. Table 7 provides demographic information for all nine new forms, by course, for the final samples. Table 8 summarizes the raw score statistics for the new forms of the nine QualityCore courses.

Results

A study using the national data was conducted in 2008 to decide which common item equating method was appropriate for the QualityCore tests. Data from five tests—English 10, English 11, English 12, Algebra I, Algebra II—were examined to determine which common item equating method would produce results most comparable to random group equipercentile equating. Note that there are more than 20% common items between the forms. The forms were spiraled when administered so that a random group equating design could be used as the criterion to investigate which common item equating method best fit the data.

For random groups equating, equipercentile equating with no smoothing and smoothing (degree = .10) were run. For common item nonequivalent groups equating, eight equating methods—Tucker mean, Levine mean, Braun/Holland mean, Tucker linear, Levine linear, Braun/Holland linear, unsmoothed and smoothed (degree = .10) equipercentile equating methods—were considered as alternatives. The common item equating method that produced

results closest to the random groups equatings in terms of conversions and scale score means was selected as the operational equating method for the QualityCore tests.

Table 7
Demographic Information for New Forms

	English 9	English 10	English 11	English 12	Algebra I	Algebra II	Geometry	Biology	Chemistry
Sample size	2,236	2,841	2,606	1,414	1,747	1,951	2,786	2,728	2,469
Gender									
Male	1,051	1,445	1,299	721	877	899	1,337	1,337	1,158
Female	1,127	1,367	1,274	679	846	1,037	1,418	1,358	1,279
Not specified	58	29	33	14	24	15	31	33	32
Race/Ethnicity[a]									
White	N.A.[b]	1,914	1,869	902	1,194	1,478	1,995	1,863	1,801
African American		297	222	138	207	145	227	282	180
Hispanic		402	356	275	280	226	384	415	361
Asian/Pacific Islander		113	98	63	37	54	86	106	101
American Indian		85	56	31	48	34	52	62	61
Not specified/ Other		271	214	117	134	142	232	224	198

[a] Some students selected more than one race/ethnicity.
[b] N.A.: Race/Ethnicity information was not collected in the Spring 2009 special study.

Table 8
Raw Score Summary Statistics for New Forms

	English 9	English 10	English 11	English 12	Algebra I	Algebra II [a]	Geometry [b]	Biology	Chemistry
Sample size	2,236	2,841	2,606	1,414	1,747	1,951	2,786	2,728	2,469
Raw score									
Mean	42.65	38.84	33.88	28.11	26.79	27.88	27.88	33.52	29.28
SD	13.62	13.41	14.74	11.64	8.37	8.71	8.13	11.12	10.92
Skewness	-0.37	–0.11	0.29	0.63	0.84	0.43	0.66	0.21	0.63
Kurtosis	2.21	2.03	1.92	2.61	3.81	2.84	3.56	2.37	2.79
Reliability	0.92	0.92	0.93	0.89	0.77	0.79	0.76	0.87	0.87
SEM	3.77	4.06	4.00	4.68	3.87	3.96	4.02	4.13	3.85

[a] Two items were not scored.
[b] One item was not scored.

The common item equipercentile equatings with post-smoothing produced the smallest differences from the random groups equating with both smoothing and no smoothing options in the raw-to-raw

score conversions and the raw-to-scale score conversions. Table 9 summarizes the scale scores of the base forms and the new forms of the five courses. For the new forms, scale scores resulting from the random groups and common item equipercentile equatings with post-smoothing option are reported in the table. The scale score differences between random groups and common item equipercentile equating with post-smoothing methods were between .50 and .02 for all five courses. Therefore, it can be concluded that the common item equipercentile equating with post-smoothing method is appropriate as the operational equating method for the QualityCore tests.

Table 9
Scale Score Comparison of Two Equating Methods for Five QualityCore Courses

	Mean	SD	Skewness	Kurtosis
English 10				
Base Form	151.44	7.87	–0.22	2.11
New Form				
Random group equipercentile equating	151.40	7.90	–0.19	2.14
Common item equipercentile equating	151.32	7.73	–0.21	2.18
English 11				
Base Form	149.91	8.63	0.19	2.05
New Form				
Random group equipercentile equating	149.96	8.71	0.18	2.03
Common item equipercentile equating	149.39	8.75	0.29	2.10
English 12				
Base Form	147.23	7.68	0.31	2.12
New Form				
Random group equipercentile equating	147.25	7.63	0.32	2.17
Common item equipercentile equating	146.89	7.59	0.36	2.21
Algebra I				
Base Form	144.92	4.36	0.40	3.18
New Form				
Random group equipercentile equating	144.88	4.34	0.50	3.77
Common item equipercentile equating	144.85	4.31	0.45	3.78
Algebra II				
Base Form	144.56	4.86	0.36	3.14
New Form				
Random group equipercentile equating	144.54	4.81	0.29	3.06
Common item equipercentile equating	144.56	4.75	0.28	3.09

Table 10 summarizes scale score statistics of equating forms (base and new forms) for the nine courses administered between 2007 and 2009. The degree of smoothing for a common item equipercentile equating was decided based on the standard error band of the raw-to-raw score equivalents for the post-smoothing with various smoothing degrees and their scale score moments. Table 10 also presents scale score reliability and standard error of measurement.

Table 10

Scale Score Summary of Equating Forms for Nine QualityCore Courses

	Mean	SD	Skewness	Kurtosis	Reliability	SEM
English 9 (S = .05)						
Base Form	153.26	7.77	-0.27	2.30	0.92	2.21
New Form	153.24	7.89	-0.29	2.28	0.92	2.26
English 10 (S = .10)						
Base Form	151.44	7.87	-0.22	2.11	0.92	2.23
New Form	151.32	7.73	-0.21	2.18	0.91	2.28
English 11 (S = .05)						
Base Form	149.91	8.63	0.19	2.05	0.93	2.21
New Form	149.43	8.79	0.28	2.08	0.93	2.40
English 12 (S = .05)						
Base Form	147.23	7.68	0.31	2.12	0.92	2.20
New Form	146.88	7.58	0.35	2.20	0.88	2.66
Algebra I (S = .05)						
Base Form	144.92	4.36	0.40	3.18	0.75	2.19
New Form	144.85	4.31	0.45	3.79	0.73	2.22
Algebra II (S = .05)						
Base Form	144.56	4.86	0.36	3.14	0.78	2.27
New Form	144.56	4.75	0.28	3.09	0.78	2.23
Geometry (S = .10)						
Base Form	144.17	4.46	0.44	3.20	0.75	2.23
New Form	144.28	4.50	0.42	3.29	0.74	2.28
Biology (S = .05)						
Base Form	147.78	6.33	0.27	2.51	0.87	2.24
New Form	147.76	6.39	0.26	2.51	0.87	2.33
Chemistry (S = .05)						
Base Form	145.07	5.78	0.58	3.01	0.85	2.21
New Form	145.23	5.68	0.57	3.13	0.86	2.16

Note. S = smoothing degree

Comparability Study

This section summarizes the findings of a comparability study that investigated whether paper-and-pencil (P&P) and computer-based (CBT) modes of test administration produced comparable scores for the QualityCore End-of-Course Assessments. The purpose of the study was to compare the P&P and CBT testing modes to determine if there was a mode effect, and if so, whether the effect was significant enough to require separate conversions/adjustments to ensure comparable scores and, if necessary, to make adjustments for the first set of forms.

Data

The base forms consisted of all multiple-choice items; the same test form was administered in both modes for the comparability study.

To accurately assess the testing modes, it was important that the students taking the P&P test were similar to those taking the CBT test in terms of demographic characteristics such as gender, race/ethnicity, course grades, computer proficiency, socioeconomic status, and ability. If students in the two testing modes were similar with respect to these characteristics, then any difference in scores could be attributed to testing mode.

One of the best methods for obtaining similar groups is to randomly assign students or classrooms to testing mode. Random assignment ensures each student or classroom has the same chance of being in the P&P or CBT testing mode. Random assignment can be conducted at the student- or classroom-level. Site administrators were provided instructions for randomly assigning students/classes to the testing modes.

For student-level random assignment, site administrators were advised to determine the number of participating students and to make a list of all students' names. Site administrators were asked to randomly select a student for each testing mode and to document which testing mode was assigned. Random selection could be performed by flipping a coin for each student and assigning heads to either P&P or CBT and tails to the other mode or by placing the name of each student in a bowl and drawing a name for each testing mode.

Classroom-level random assignment involved identifying pairs or groups of classrooms that were similar with respect to some or all of the following demographic characteristics: academic performance, gender and racial/ethnic distributions, socioeconomic status, and class size. Classrooms that were matched by demographic characteristics were then randomly assigned to different testing modes. Site administrators were asked to complete a survey regarding the method of random assignment and their opinion about the similarity of the groups. These surveys were used to eliminate students and/or sites that did not meet the random assignment requirements.

Sites were included in the analyses if data were received for both the P&P and CBT modes. Within each site, student demographics were compared across modes to ensure the samples had similar

characteristics. Also, student records with any of the following issues were excluded from all analyses:

- Missing form identification code
- Missing ID number and last name
- Missing gender
- English as the student's second language
- Missing raw responses (A, B, C, D) or scored responses (0, 1) for all items
- Missing raw total score or raw total score equal to zero
- Having 75% (26 items) or more omitted items per section

Once the data were cleaned based on the above rules, additional student records were removed from analyses if students were in nontraditional grade levels for a particular course. For English 10, 11, and 12, only current 10th-, 11th-, and 12th-grade students, respectively, were included in the sample. For Algebra I and Biology, only current 9th- or 10th-grade students were included. For Algebra II and Chemistry, only current 10th- or 11th-grade students were included. For Geometry, only current 9th-, 10th-, or 11th-grade students were included. Table 11 summarizes the sample sizes by demographic variables.

Table 11
Demographic Information for Comparability Study Samples

	English10		English11		English12		Algebra I	
Mode	CBT	P&P	CBT	P&P	CBT	P&P	CBT	P&P
Sample size	418	500	547	510	97	84	373	382
Gender								
Male	205	249	272	263	41	44	172	187
Female	213	251	275	247	56	40	201	195
Race/Ethnicity								
White	350	410	452	425	67	61	287	275
African American	25	40	47	45	20	18	17	19
Hispanic	23	27	33	25	8	2	57	74
Asian/Pacific Islander	13	15	7	8	1	1	8	12
American Indian	1	1	1	1	–	–	2	–
Not specified/ Other	6	7	7	6	1	2	2	2

	Algebra II		Geometry		Biology		Chemistry	
Mode	CBT	P&P	CBT	P&P	CBT	P&P	CBT	P&P
Sample size	590	702	661	662	755	691	699	731
Gender								
Male	271	345	304	306	343	317	337	356
Female	319	357	357	356	412	374	362	375
Race/Ethnicity								
White	502	570	559	554	628	592	569	630
African American	25	33	34	48	63	41	55	41
Hispanic	46	74	39	39	37	36	45	38
Asian/Pacific Islander	11	22	12	11	20	11	18	14
American Indian	–	–	–	3	–	2	2	1
Not specified/ Other	6	3	17	7	7	9	10	7

Results

Table 12 summarizes the raw score statistics. The raw score mean differences between modes were less than .9 except for Algebra I and Geometry. The six courses also shared similar variances. A *t*-test also showed no significant difference between P&P and CBT scores for the six courses. However, Algebra I and Geometry showed statistically significant differences. Figure 9 shows the cumulative percent of raw scores for P&P and CBT. Algebra I and Geometry showed a slightly higher cumulative percent by raw score point for the CBT mode.

The correlation between percent correct (*P*) values on items across modes is .944, .935, .872, .963, .979, .969, .975, and .968 for English 10, English 11, English 12, Algebra I, Algebra II, Geometry, Biology and Chemistry, respectively. Even though English 12 had a correlation less than .9, the higher correlations imply comparable rank order of item difficulty ties between the two modes. Figure 10 plots the relationship between item *P*-values for P&P and CBT modes for the eight QualityCore tests.

In addition, a differential item functioning (DIF) analysis was implemented to determine whether any item had different difficulties across the two modes. The results showed only 4, 3, 1, 2, 0, 0, and 0 of 70 items with significant Mantel-Haenszel indices for English 10, English 11, Algebra I, Algebra II, Geometry, Biology and Chemistry, respectively. Note that due to small sample size, DIF analysis of English 12 was not implemented. In general, if the Mantel-Haenszel index is greater than 2 or lower than 0.5, the items are flagged as DIF items.

Table 12
Raw Score Summary for P&P and CBT Modes of QualityCore

	English10		English11		English12		Algebra I	
Mode	CBT	P&P	CBT	P&P	CBT	P&P	CBT	P&P
No. of items	70	70	70	70	70	70	70	70
Sample size	418	500	547	510	97	84	373	382
Raw score								
Mean	45.26	44.81	43.92	43.48	40.56	39.69	30.01	27.79
SD	10.04	10.54	12.29	12.69	13.42	13.58	6.61	7.17
Minimum	19	13	10	11	13	13	16	8
Maximum	68	65	68	67	64	62	58	58
t-statistic (raw means)	0.66 ($df = 916$, $p = 0.51$)		0.57 ($df = 1055$, $p = 0.57$)		0.43 ($df = 179$, $p = 0.67$)		4.42 ($df = 753$, $p < .01$)	
	Algebra II		Geometry		Biology		Chemistry	
Mode	CBT	P&P	CBT	P&P	CBT	P&P	CBT	P&P
No. of items	70	70	70	70	70	70	70	70
Sample size	590	702	661	662	755	691	699	731
Raw score								
Mean	29.69	30.13	30.98	28.94	38.35	38.59	34.64	33.76
SD	6.79	7.39	7.68	7.68	10.53	11.18	10.77	10.70
Minimum	9	11	13	11	13	11	12	12
Maximum	52	56	57	58	69	66	69	65
t-statistic (raw means)	–1.10 ($df = 1290$, $p = 0.27$)		4.83 ($df = 1321$, $p < .01$)		–0.42 ($df = 1444$, $p = 0.67$)		1.55 ($df = 1428$, $p = 0.12$)	

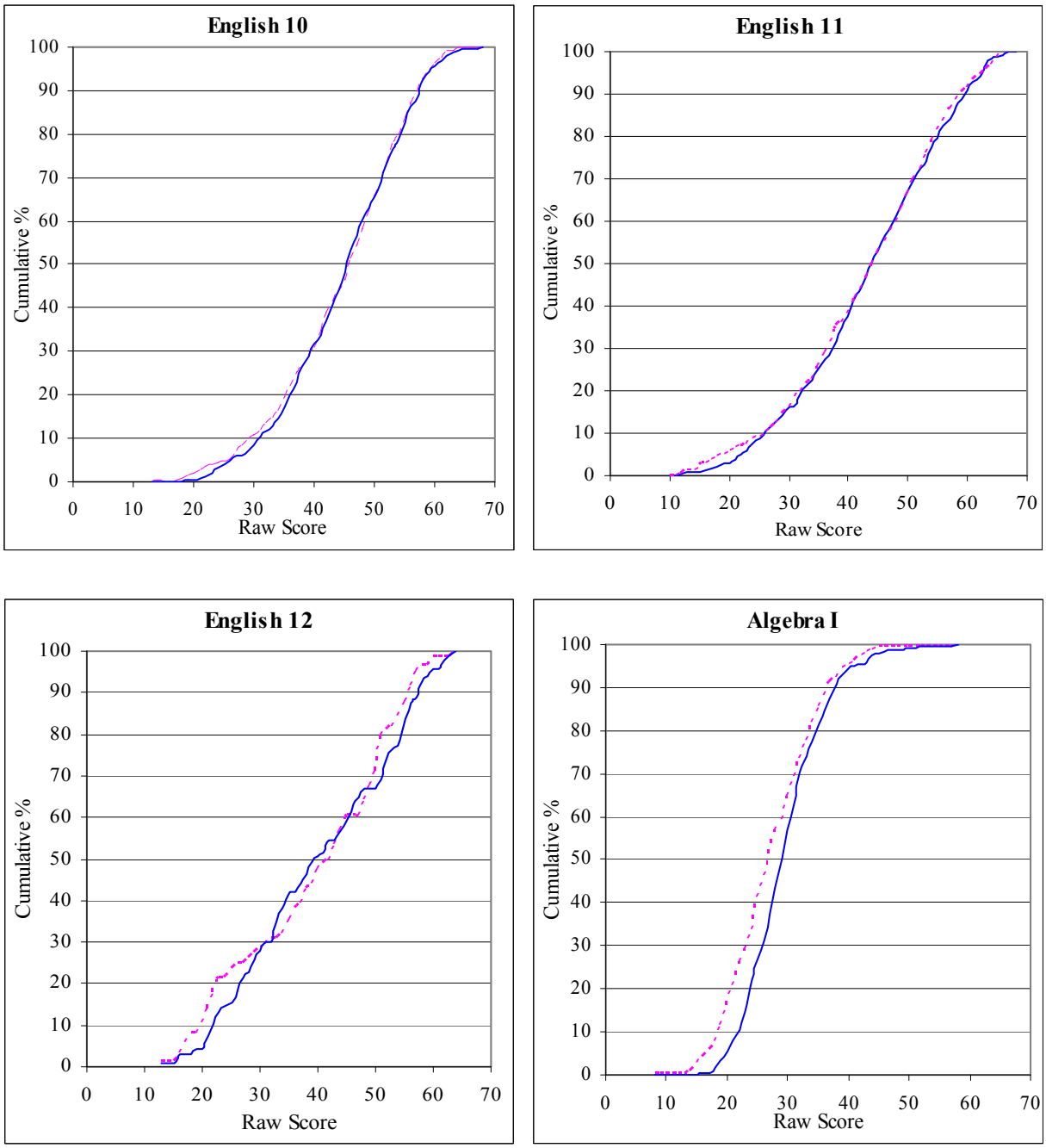


Figure 9. Cumulative Percent of Raw Score for P&P and CBT Modes of QualityCore

Note. P&P ---------- CBT ————

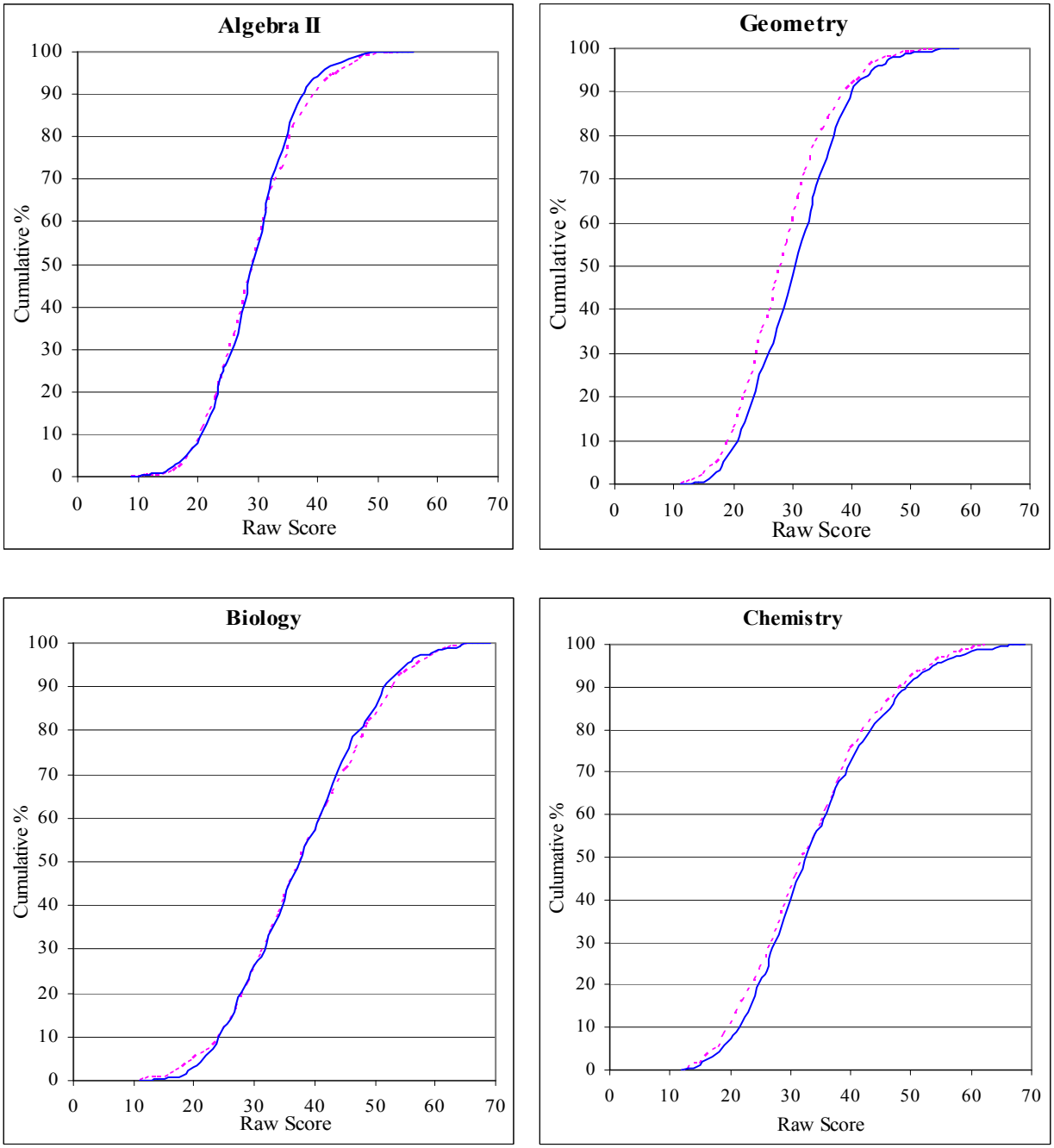


Figure 9. (continued) Cumulative Percent of Raw Score for P&P and CBT Modes of QualityCore

Note. P&P ---------- CBT ————

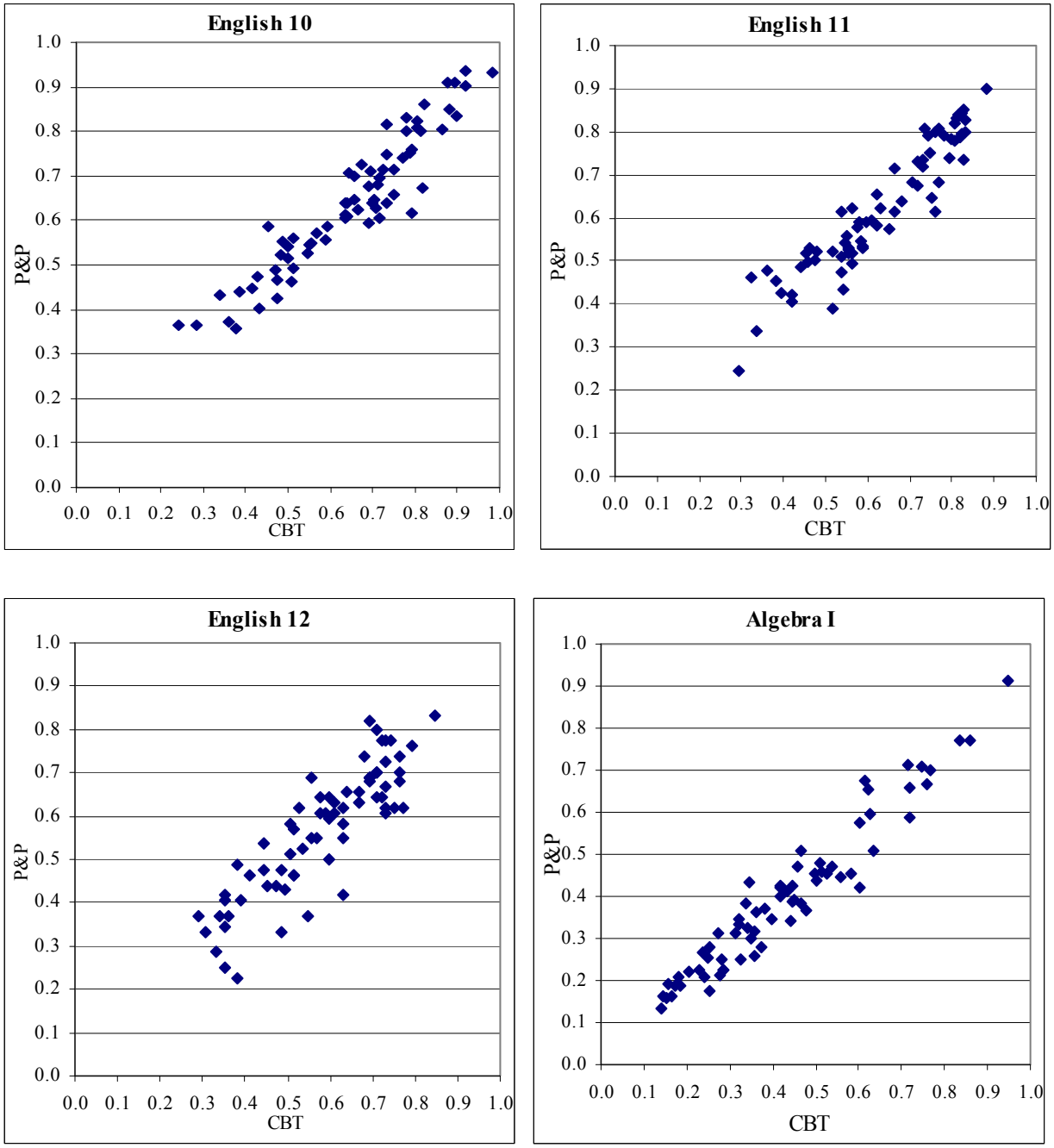


Figure 10. Plots of Item P-values between P&P and CBT Modes of QualityCore

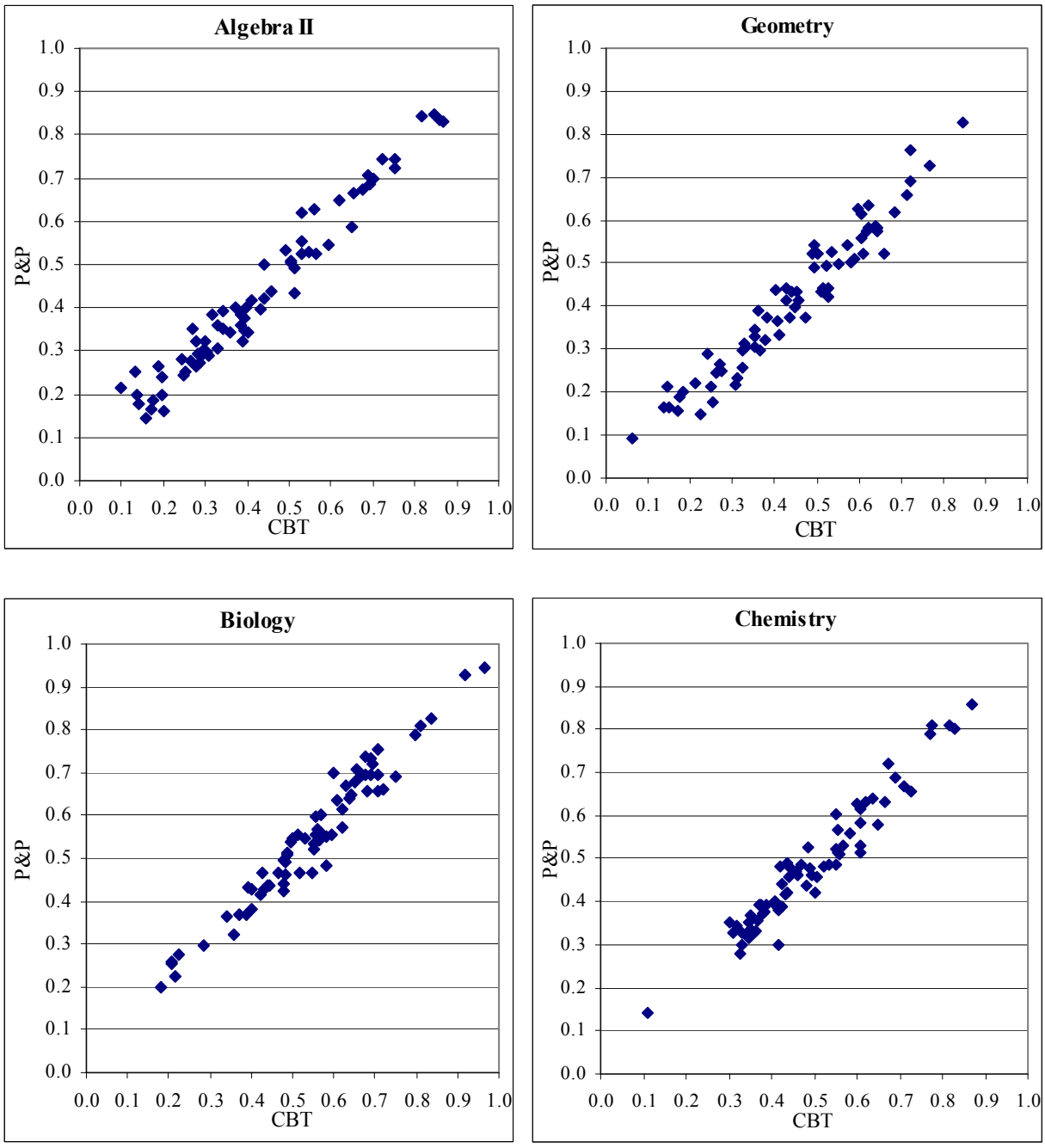


Figure 10. (continued) Plots of Item P-values between P&P and CBT Modes of QualityCore

Separate common item equipercentile equatings (there were more than 20% common items between forms) of P&P and CBT versions were implemented to compare scale score distributions between P&P and CBT. Table 13 summarizes scale score statistics for the P&P and CBT modes. Note that equating was not implemented for English 12 and Algebra I due to small sample size. Scale score means and standard deviations (SDs) for P&P and CBT tests did not show large differences. English 11, Algebra II, and Chemistry did not show significant difference between P&P and CBT scores, while English 10, Geometry, and Biology did. Figure 11 presents the cumulative percents of scale scores for the two modes. The cumulative percents of scale scores are very similar except for Geometry.

Finally, comparisons based on item response theory (IRT) were investigated. The biggest differences in the average *b*-parameters between the two modes among the seven courses were .359 for Algebra I and .271 for Geometry (concurrent calibration with one parameter logistic model). Other courses showed very similar *b*-parameter averages. Note that due to small sample size, IRT analysis of English 12 was not implemented. The correlations between *b*-parameter estimates of P&P and CBT were .936, .935, .966, .973, .966, .976, and .970 for English 10, English 11, Algebra I, Algebra II, Geometry, Biology, and Chemistry, respectively. Figure 12 plots the relationship between *b*-parameter estimates for P&P and CBT. Also, test characteristic curves (TCCs with Rasch model) were compared between P&P and CBT. Figure 13 shows that TCCs from the two modes were almost identical.

Table 13

Scale Score Summary for P&P and CBT Modes of QualityCore Using Separate Equatings

	English10		English11		Algebra II	
Mode	CBT	P&P	CBT	P&P	CBT	P&P
No. of items	70	70	70	70	70	70
Sample size	418	500	547	510	590	702
Scale score						
Mean	154.32	153.45	153.19	152.62	145.34	145.27
SD	6.12	6.60	7.46	7.64	4.02	4.27
Minimum	136	134	134	134	133	133
Maximum	171	168	171	170	158	160
t-statistic (scale score means)	2.05 ($df = 916$, $p < 0.05$)		1.22 ($df = 1055$, $p = 0.22$)		0.30 ($df = 1290$, $p = 0.77$)	

	Geometry		Biology		Chemistry	
Mode	CBT	P&P	CBT	P&P	CBT	P&P
No. of items	70	70	70	70	70	70
Sample size	661	662	755	691	699	731
Scale score						
Mean	145.66	144.72	149.83	150.48	147.77	147.46
SD	4.51	4.54	6.09	6.08	5.84	5.83
Minimum	134	134	135	134	134	134
Maximum	161	163	171	166	171	166
t-statistic (scale score means)	3.77 ($df = 1321$, $p < 0.01$)		–2.03 ($df = 1444$, $p < 0.05$)		1.02 ($df = 1428$, $p = 0.31$)	

Note. Equatings were not implemented for English 12 and Algebra I due to small sample sizes.

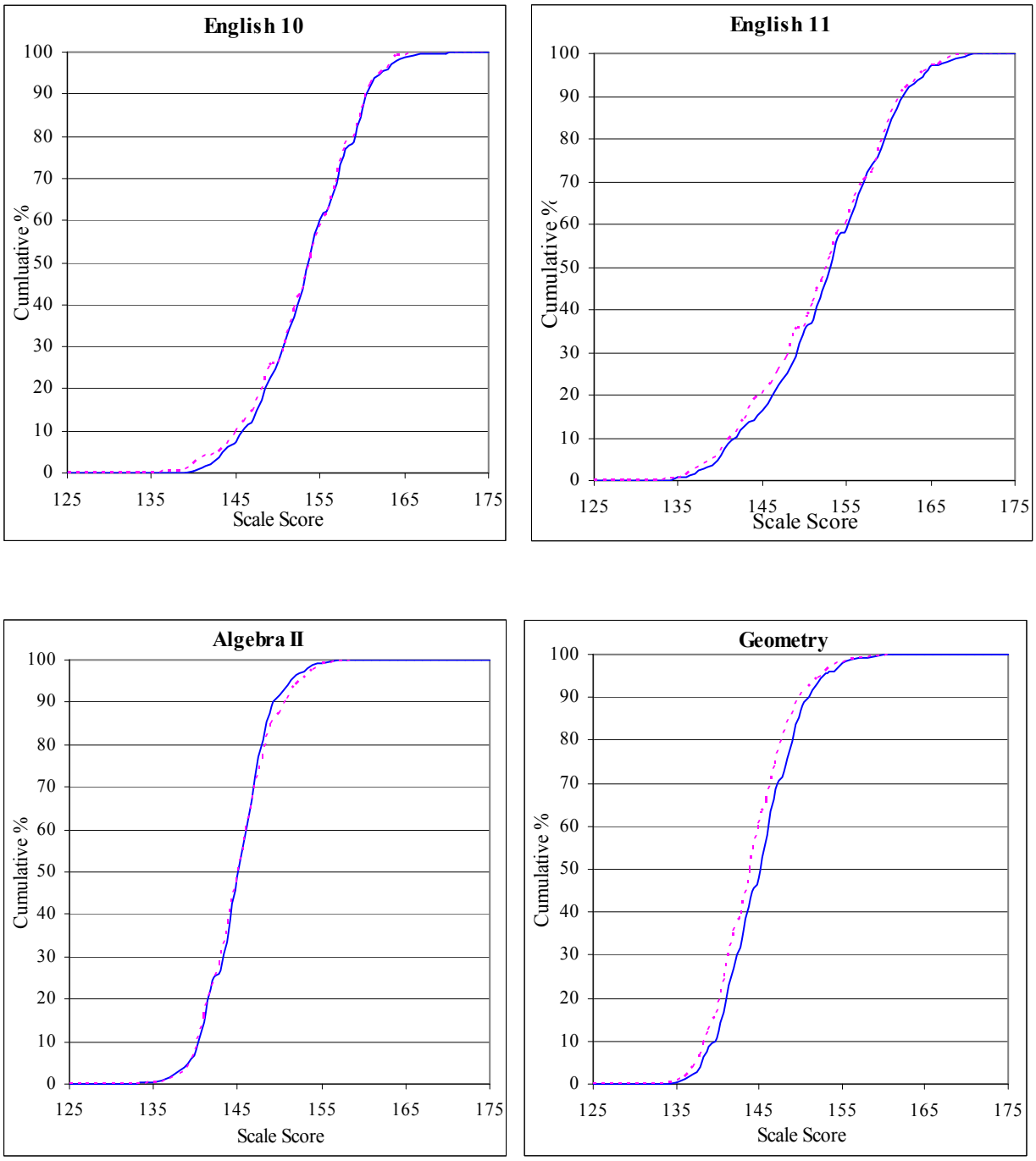


Figure 11. Cumulative Percent of Scale Score for P&P and CBT Modes of QualityCore (English 12 and Algebra I are not included.)

Note. P&P -------- CBT ————

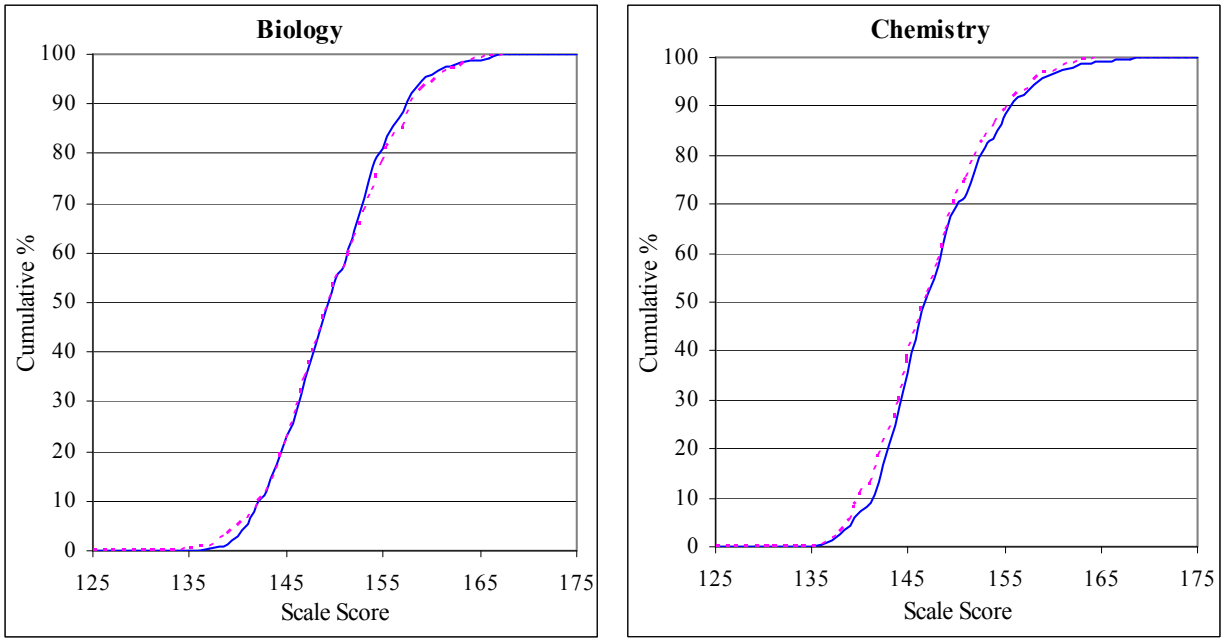


Figure 11. (continued) Cumulative Percent of Scale Score for P&P and CBT Modes of QualityCore (English 12 and Algebra I are not included.)

Note. P&P -------- CBT ————

Overall, the raw score mean differences were significant for Algebra I and Geometry, and scale score mean differences were significant for English 10, Geometry, and Biology. However, the correlation between *P*-values on items across modes were .944, .963, and .975 for English 10, Geometry, and Biology, respectively, which implied comparable rank order of item difficulties between the two modes. In addition, the correlation between *b* parameters on items across modes were 936, .966, .966, and .976 for English 10, Algebra I, Geometry, and Biology, respectively, which implied comparable rank order of IRT item difficulties between the two modes.

The DIF analyses on items across modes showed that only 1 item out of 70 items for Algebra I had a significant Mantel-Haenszel index value. Geometry and Biology did not have any items with significant Mantel-Haenszel values. Also, TCCs from the two modes were almost identical for these subjects.

In conclusion, there is insufficient evidence to support the use of different conversions for P&P and CBT modes of the QualityCore tests (i.e., adjustment of the CBT conversion). P&P and CBT comparability will be revisited when additional data are available. Until then, the same conversion is being applied to the two modes of the QualityCore tests.

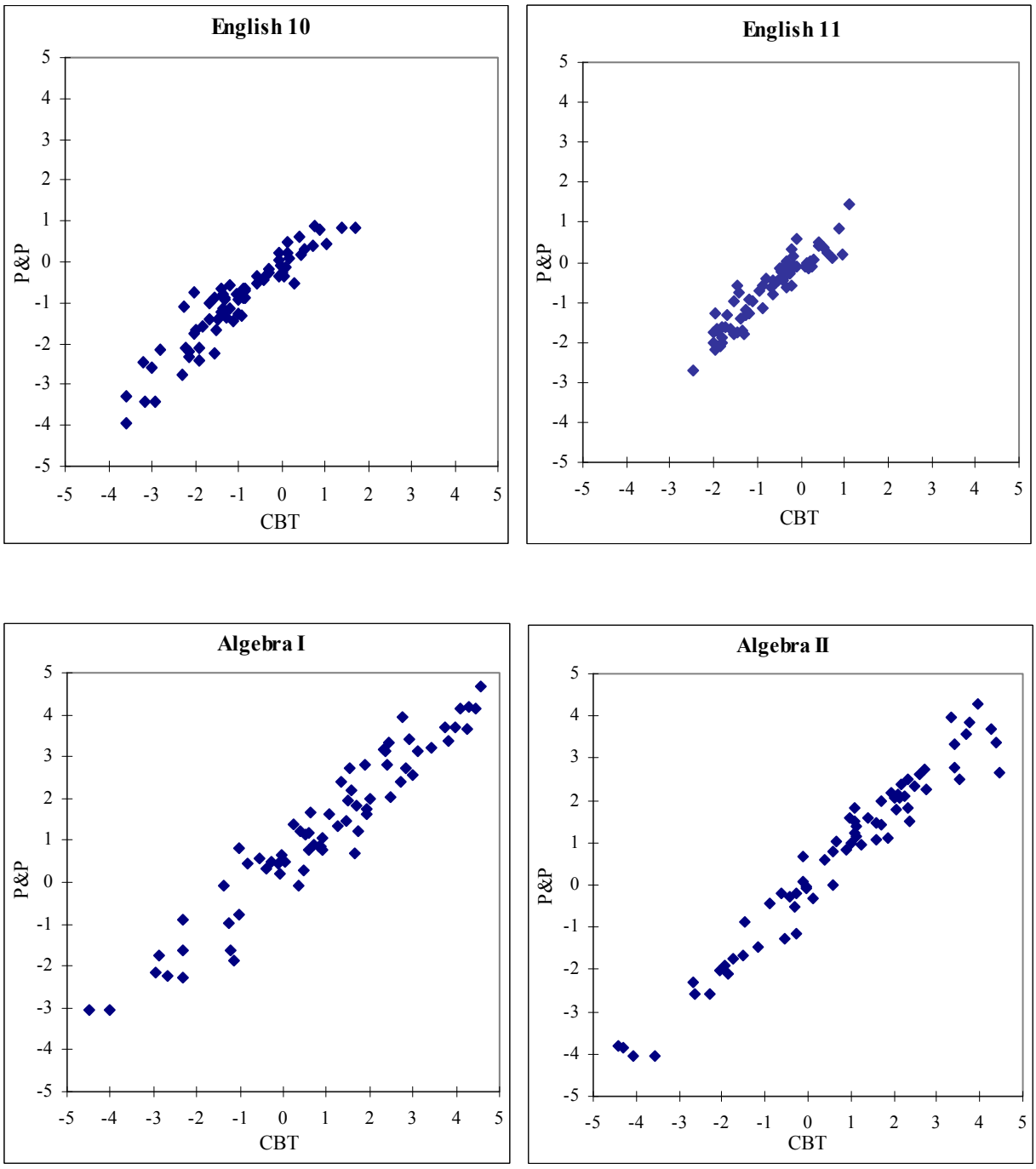


Figure 12. Plots of Item *b* Parameters between P&P and CBT Modes of QualityCore (IRT analysis was not implemented for English 12 due to small sample size.)

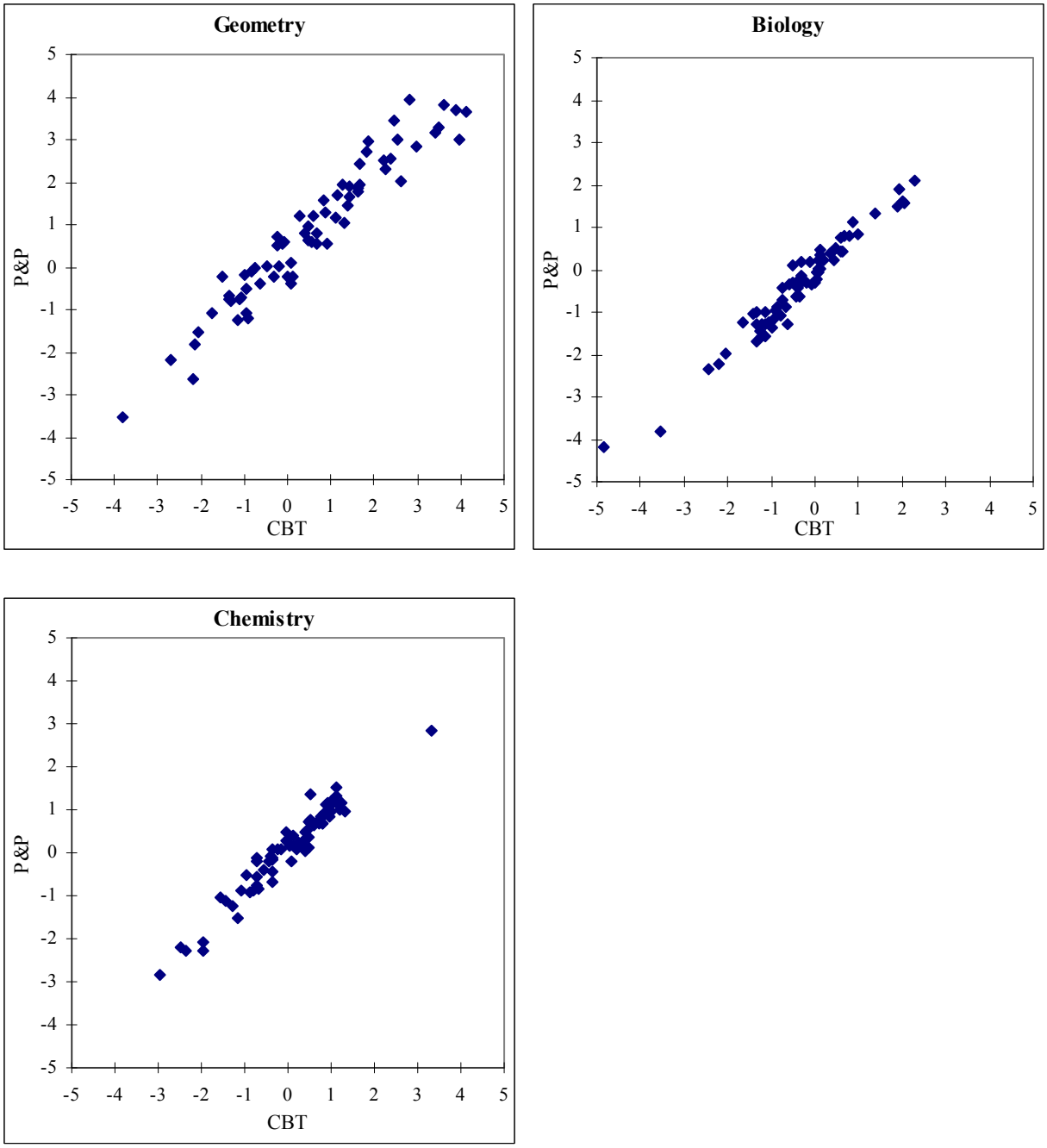


Figure 12. (continued) Plots of Item *b* Parameters between P&P and CBT Modes of QualityCore (IRT analysis was not implemented for English 12 due to small sample size.)

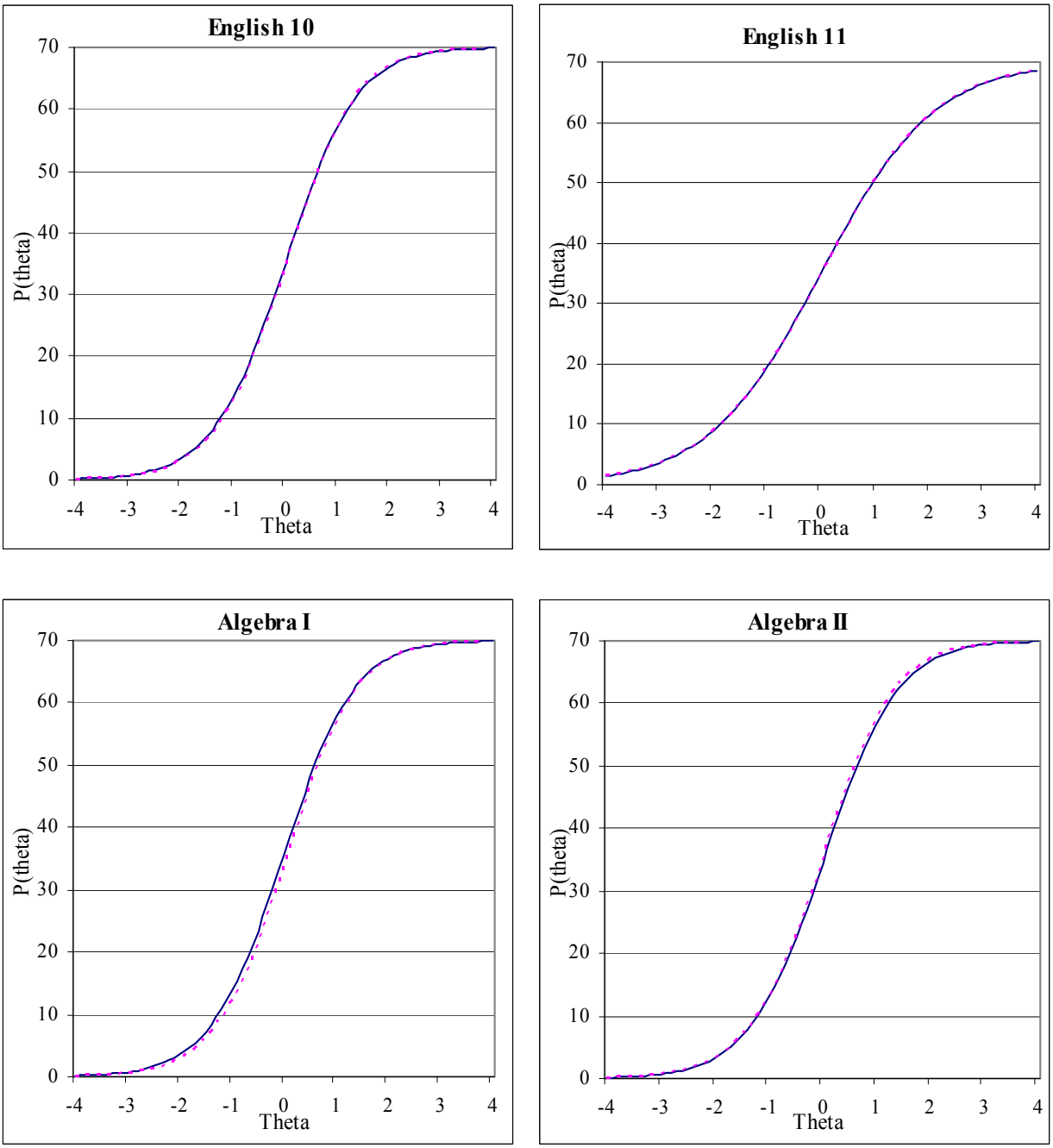


Figure 13. Test Characteristic Curve Comparison of P&P and CBT Modes of QualityCore (IRT analysis was not implemented for English 12 due to small sample size.)

Note. P&P -------- CBT ————

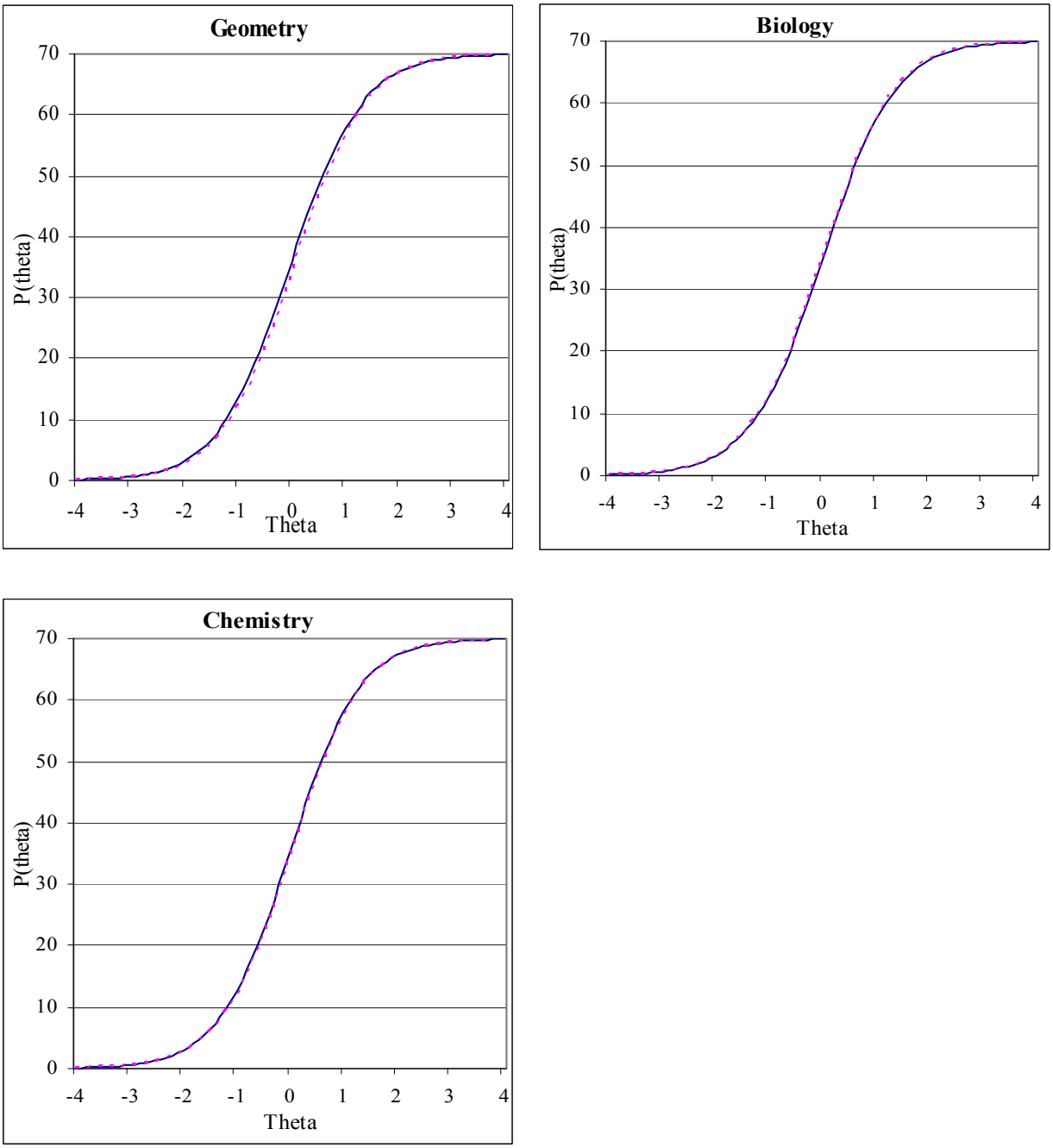


Figure 13. (continued) Test Characteristic Curve Comparison of P&P and CBT Modes of QualityCore (IRT analysis was not implemented for English 12 due to small sample size.)

Note. P&P -------- CBT ————

Estimated PLAN or ACT Test Score Ranges

All students who take QualityCore operational End-of-Course Assessments receive corresponding estimated PLAN or ACT test score ranges. This section describes how the estimated PLAN or ACT score ranges were determined using the national (2007-2008) and special (2009) study samples that took both QualityCore and the PLAN or ACT tests. QualityCore English 9–10, Algebra I, and Biology test takers took the PLAN English, Mathematics, and Science Tests, respectively. QualityCore English 11 and English 12 test takers took the ACT English Test. QualityCore Algebra II, Geometry, and Precalculus test takers took the ACT Mathematics Test. QualityCore Chemistry and Physics test takers took the ACT Science Test. QualityCore U.S. History test takers took the ACT Reading Test. Note that because QualityCore and the PLAN/ACT subject tests do not assess the same scope of content, the estimated PLAN or ACT score is represented as a range rather than as a score point.

Data

Student records for the ACT subject tests that exhibited any of the following issues were excluded from all analyses:

- Missing form identification code
- Missing ID number and last name
- Missing raw responses (A, B, C, D) or scored responses (0, 1) for all items
- Missing raw total score or raw total score equal to zero
- Having less than one response in the raw response string

After the data were cleaned based on the above rules, additional student records were excluded from analyses if students were in nontraditional grade levels for a particular course. For English 9, 10, 11, and 12, only current 9th-, 10th-, 11th-, and 12th-grade students, respectively, were included in the sample. For Algebra I and Biology, only current 9th- or 10th-grade students were included. For Algebra II and Chemistry, only current 10th- or 11th-grade students were included. For Geometry and U.S. History, only current 9th-, 10th-, or 11th-grade students were included. For Precalculus, only current 11th- or 12th-grade students were included. For Physics, only current 10th-, 11th-, or 12th-grade students were included.

Finally, students who took a PLAN or ACT Test and any form of the QualityCore tests were selected. This resulted in a matched sample of 3,725 students for English 9; 4,579 students for English 10; 4,367 students for English 11; 2,021 students for English 12; 3,091 students for Algebra I; 3,554 for Algebra II; 4,602 students for Geometry; 1,348 students for Precalculus; 4,546 students for Biology; 4,230 students for Chemistry; 1,462 students for Physics; and 1,762 students for U.S. History.

Estimated PLAN/ACT Test Score Ranges

The estimated PLAN or ACT score range is a prediction of how a student is likely to score on the PLAN or ACT test if the PLAN or ACT is taken at the same time as a QualityCore EOC test. Note that these scores are only estimates, not guarantees.

Table 14 summarizes the scale scores for the 12 QualityCore tests and the PLAN or ACT tests. The estimated PLAN/ACT score ranges were constructed based on the bivariate distribution of QualityCore scale scores and PLAN/ACT scale scores, with some hand adjustment. The goal was to obtain an overall probability of 0.67 that a student's score would fall within the estimated PLAN or ACT score range. For any particular student, the score obtained on the PLAN or ACT test may or may not fall within the range reported on the QualityCore score report.

The final intervals are reported in Table 15. For example, for a QualityCore English 11 score of 159, the lower limit of the estimated ACT English score range is 20 and the upper limit is 26. Consequently, an estimated ACT English score range of 20–26 is reported for students with a QualityCore English 11 score of 159. The width of the range varies with the QualityCore score. The ranges are wider for QualityCore English 10–12, Precalculus, Chemistry, Physics, and U.S. History (5 to 10 scale score points) than for English 9, Algebra I, Algebra II, Geometry, and Biology (4 to 6 scale score points).

To confirm that the ranges estimated the PLAN or ACT score ranges appropriately, the most recent operational ACT English (N = 4,800) and Mathematics (N = 2,804) scale scores were obtained for those students in one state who took the QualityCore EOC English 11 and Algebra II tests, respectively. The estimated ACT English score ranges obtained from the study covered 67.0% of the operational ACT English scale scores. The estimated ACT Mathematics Test score ranges obtained from the study covered 66.5% of the operational ACT Mathematics Test scale scores.

If a student's estimated PLAN or ACT score range falls below the College Readiness Benchmark Score (see note on Table 15), it is likely that the student does not meet the College Readiness Benchmark. If a student's estimated PLAN or ACT score range is above the College Readiness Benchmark Score, it is likely that the student meets the College Readiness Benchmark. And if a student's estimated PLAN or ACT score range includes the College Readiness Benchmark Score, the student may or may not meet the College Readiness Benchmark. In the QualityCore student score report, the students whose estimated PLAN or ACT score range includes the College Readiness Benchmark Score are grouped with the students whose estimated PLAN or ACT score ranges exceed the College Readiness Benchmark.

The estimated PLAN or ACT score range was set based on a national sample but not under operational conditions—the full ACT test was not administered—and without confirming if all participating students took a rigorous course and/or had appropriate motivation. The estimated ACT score ranges will be monitored and refined as more data becomes available.

Table 14

Scale Score Summary of Test Scores Used to Establish the Estimated PLAN/ACT Score Range Study

	QC English 9	PLAN English	QC English 10	PLAN English
Sample size	3,725		4,579	
Scale score				
Mean	153.87	16.04	151.77	16.58
SD	7.55	4.29	7.75	5.14
Skewness	-0.33	0.16	–0.26	0.16
Kurtosis	2.36	3.05	2.20	2.80
	QC English 11	ACT English	QC English 12	ACT English
Sample size	4,367		2,021	
Scale score				
Mean	150.09	15.70	147.79	14.66
SD	8.76	6.79	7.56	6.54
Skewness	0.20	0.57	0.21	0.77
Kurtosis	2.04	2.64	2.11	2.98
	QC Algebra I	PLAN Mathematics	QC Algebra II	ACT Mathematics
Sample size	3,091		3,554	
Scale score				
Mean	145.13	16.06	144.64	18.59
SD	4.32	3.58	4.81	4.37
Skewness	0.45	0.17	0.31	0.66
Kurtosis	3.43	5.06	3.12	2.83
	QC Geometry	ACT Mathematics	QC Precalculus	ACT Mathematics
Sample size	4,602		1,348	
Scale score				
Mean	144.38	17.01	145.11	22.48
SD	4.44	3.32	5.20	4.87
Skewness	0.44	0.96	0.84	0.23
Kurtosis	3.24	4.47	4.05	2.36
	QC Biology	PLAN Science	QC Chemistry	ACT Science
Sample size	4,546		4,230	
Scale score				
Mean	147.95	18.03	145.43	18.01
SD	6.35	3.58	5.74	4.77
Skewness	0.26	0.54	0.53	0.15
Kurtosis	2.52	3.84	3.02	2.77
	QC Physics	ACT Science	QC U.S. History	ACT Reading
Sample size	1,462		1,762	
Scale score				
Mean	145.64	21.18	144.62	17.33
SD	5.23	5.84	5.08	5.74
Skewness	0.83	0.04	0.66	0.67
Kurtosis	3.97	2.70	3.14	2.96

Note. QualityCore scale score range of 125–175; ACT scale score range of 1–36; PLAN scale score range of 1–32.

Table 15
Estimated PLAN/ACT Score Intervals for QualityCore Scores

English 9 QualityCore Score	PLAN English Intervals			English 10 QualityCore Score	PLAN English Intervals			English 11 QualityCore Score	ACT English Intervals		
	Low Score	High Score	Width		Low Score	High Score	Width		Low Score	High Score	Width
125	1	5	5	125	1	5	5	125	1	7	7
126	1	5	5	126	1	5	5	126	2	8	7
127	2	6	5	127	2	6	5	127	3	9	7
128	2	6	5	128	2	6	5	128	4	10	7
129	3	7	5	129	3	7	5	129	5	10	6
130	3	7	5	130	4	8	5	130	5	10	6
131	4	8	5	131	5	9	5	131	5	10	6
132	4	8	5	132	6	10	5	132	5	10	6
133	5	9	5	133	7	11	5	133	6	11	6
134	6	10	5	134	8	12	5	134	6	11	6
135	6	10	5	135	8	12	5	135	6	11	6
136	7	10	4	136	8	12	5	136	7	12	6
137	8	11	4	137	8	13	6	137	8	13	6
138	8	12	5	138	8	13	6	138	8	13	6
139	8	12	5	139	9	14	6	139	8	13	6
140	8	12	5	140	9	14	6	140	8	13	6
141	9	13	5	141	9	14	6	141	8	13	6
142	9	13	5	142	10	15	6	142	8	14	7
143	9	14	6	143	10	15	6	143	9	16	8
144	10	14	5	144	11	17	7	144	9	16	8
145	11	15	5	145	11	17	7	145	9	16	8
146	11	15	5	146	12	17	6	146	10	17	8
147	11	16	6	147	12	18	7	147	10	17	8
148	12	16	5	148	13	18	6	148	10	17	8
149	12	16	5	149	13	18	6	149	11	18	8
150	12	16	5	150	14	18	5	150	11	18	8
151	12	16	5	151	14	19	6	151	13	20	8
152	12	16	5	152	14	19	6	152	15	22	8
153	12	16	5	153	15	20	6	153	15	22	8
154	13	17	5	154	16	20	5	154	15	22	8
155	14	18	5	155	16	21	6	155	15	22	8
156	14	18	5	156	16	21	6	156	16	23	8
157	15	19	5	157	16	21	6	157	17	24	8
158	15	19	5	158	17	22	6	158	18	25	8
159	16	20	5	159	18	23	6	159	20	26	7
160	16	20	5	160	18	23	6	160	20	26	7
161	17	21	5	161	19	24	6	161	21	28	8
162	18	22	5	162	20	26	7	162	21	28	8
163	18	23	6	163	20	26	7	163	23	30	8
164	19	23	5	164	22	28	7	164	24	31	8
165	20	23	4	165	23	29	7	165	24	31	8
166	20	25	6	166	24	30	7	166	25	31	7
167	20	25	6	167	25	31	7	167	26	33	8
168	21	26	6	168	26	31	6	168	28	35	8
169	23	28	6	169	26	31	6	169	28	35	8
170	23	28	6	170	26	31	6	170	29	35	7
171	24	28	5	171	26	31	6	171	30	35	6
172	25	29	5	172	27	32	6	172	31	36	6
173	26	30	5	173	27	32	6	173	31	36	6
174	27	31	5	174	28	32	5	174	31	36	6
175	28	32	5	175	28	32	5	175	31	36	6

Note. ACT's College Readiness Benchmark Scores for English are 15 and 18 for PLAN and the ACT, respectively.

Table 15 (continued)
Estimated PLAN/ACT Score Intervals for QualityCore Scores

English 12 QualityCore Score	ACT English Intervals			Algebra I QualityCore Score	PLAN Mathematics Intervals			Algebra II QualityCore Score	ACT Mathematics Intervals		
	Low Score	High Score	Width		Low Score	High Score	Width		Low Score	High Score	Width
125	1	7	7	125	6	10	5	125	6	10	5
126	2	8	7	126	7	11	5	126	7	11	5
127	3	9	7	127	8	12	5	127	8	12	5
128	4	10	7	128	9	13	5	128	9	13	5
129	4	10	7	129	9	13	5	129	10	14	5
130	5	11	7	130	10	14	5	130	10	14	5
131	5	11	7	131	10	14	5	131	11	15	5
132	6	12	7	132	11	15	5	132	12	16	5
133	6	12	7	133	11	15	5	133	13	16	4
134	8	13	6	134	12	16	5	134	13	17	5
135	8	13	6	135	12	16	5	135	13	17	5
136	8	14	7	136	12	16	5	136	14	17	4
137	8	14	7	137	12	16	5	137	14	17	4
138	8	14	7	138	13	17	5	138	14	17	4
139	8	14	7	139	13	17	5	139	14	17	4
140	8	14	7	140	13	17	5	140	14	17	4
141	8	14	7	141	14	17	4	141	15	19	5
142	8	14	7	142	14	17	4	142	15	19	5
143	9	14	6	143	14	17	4	143	16	20	5
144	9	15	7	144	15	18	4	144	16	20	5
145	9	15	7	145	15	18	4	145	16	21	6
146	9	16	8	146	15	18	4	146	17	22	6
147	10	16	7	147	16	19	4	147	17	22	6
148	11	18	8	148	16	19	4	148	18	23	6
149	11	18	8	149	16	19	4	149	20	25	6
150	12	19	8	150	16	19	4	150	21	26	6
151	14	21	8	151	16	20	5	151	22	27	6
152	14	21	8	152	17	21	5	152	22	27	6
153	14	21	8	153	18	22	5	153	23	27	5
154	15	22	8	154	18	23	6	154	24	28	5
155	15	22	8	155	19	24	6	155	24	28	5
156	16	23	8	156	20	25	6	156	24	28	5
157	18	25	8	157	21	26	6	157	24	28	5
158	19	26	8	158	22	27	6	158	25	29	5
159	20	27	8	159	23	28	6	159	25	30	6
160	20	27	8	160	24	29	6	160	25	30	6
161	21	28	8	161	25	29	5	161	26	31	6
162	23	30	8	162	25	29	5	162	26	31	6
163	25	32	8	163	26	30	5	163	27	31	5
164	27	34	8	164	26	30	5	164	28	32	5
165	28	34	7	165	26	30	5	165	29	33	5
166	28	34	7	166	26	30	5	166	29	33	5
167	28	34	7	167	26	30	5	167	30	34	5
168	29	35	7	168	27	31	5	168	30	34	5
169	29	35	7	169	27	31	5	169	30	34	5
170	29	35	7	170	27	31	5	170	31	35	5
171	29	35	7	171	27	31	5	171	31	35	5
172	30	36	7	172	28	32	5	172	31	35	5
173	30	36	7	173	28	32	5	173	32	36	5
174	31	36	6	174	28	32	5	174	32	36	5
175	31	36	6	175	28	32	5	175	32	36	5

Note. ACT's College Readiness Benchmark Score for English is 18 for the ACT, and ACT's College Readiness Benchmark Scores for Mathematics are 19 and 22 for PLAN and the ACT, respectively.

Table 15 (continued)
Estimated PLAN/ACT Score Intervals for QualityCore Scores

Geometry QualityCore Score	ACT Mathematics Intervals			Precalculus QualityCore Score	ACT Mathematics Intervals			Biology QualityCore Score	PLAN Science Intervals		
	Low Score	High Score	Width		Low Score	High Score	Width		Low Score	High Score	Width
125	6	9	4	125	1	5	5	125	6	10	5
126	7	10	4	126	2	6	5	126	7	11	5
127	8	11	4	127	3	7	5	127	8	12	5
128	9	12	4	128	5	9	5	128	8	12	5
129	10	13	4	129	6	10	5	129	9	13	5
130	11	14	4	130	7	11	5	130	9	13	5
131	12	15	4	131	9	13	5	131	10	14	5
132	12	15	4	132	10	14	5	132	10	14	5
133	12	15	4	133	11	15	5	133	11	15	5
134	13	16	4	134	13	17	5	134	11	15	5
135	13	16	4	135	14	17	4	135	12	16	5
136	13	16	4	136	14	17	4	136	13	16	4
137	13	16	4	137	14	18	5	137	13	16	4
138	14	17	4	138	15	19	5	138	13	16	4
139	14	17	4	139	15	20	6	139	13	16	4
140	14	17	4	140	15	20	6	140	13	16	4
141	14	17	4	141	16	22	7	141	14	17	4
142	15	18	4	142	16	22	7	142	14	17	4
143	15	18	4	143	16	22	7	143	14	17	4
144	15	18	4	144	19	25	7	144	15	18	4
145	16	19	4	145	19	25	7	145	15	19	5
146	16	19	4	146	21	25	5	146	15	19	5
147	16	19	4	147	21	27	7	147	16	19	4
148	17	20	4	148	23	28	6	148	16	20	5
149	18	21	4	149	24	28	5	149	17	21	5
150	18	22	5	150	24	29	6	150	17	21	5
151	19	23	5	151	24	29	6	151	18	21	4
152	20	24	5	152	25	29	5	152	18	21	4
153	21	25	5	153	25	30	6	153	19	22	4
154	22	26	5	154	25	31	7	154	19	22	4
155	23	26	4	155	26	31	6	155	19	22	4
156	24	27	4	156	27	31	5	156	19	23	5
157	25	28	4	157	28	32	5	157	20	24	5
158	26	29	4	158	28	33	6	158	21	24	4
159	26	29	4	159	29	33	5	159	21	25	5
160	27	30	4	160	30	34	5	160	22	26	5
161	27	30	4	161	30	35	6	161	22	26	5
162	28	31	4	162	31	35	5	162	23	27	5
163	28	31	4	163	31	36	6	163	24	28	5
164	29	32	4	164	31	36	6	164	24	28	5
165	29	32	4	165	31	36	6	165	24	28	5
166	30	33	4	166	31	36	6	166	26	30	5
167	30	33	4	167	31	36	6	167	27	31	5
168	31	34	4	168	31	36	6	168	28	32	5
169	31	34	4	169	31	36	6	169	28	32	5
170	32	35	4	170	32	36	5	170	28	32	5
171	32	35	4	171	32	36	5	171	28	32	5
172	32	35	4	172	32	36	5	172	29	32	4
173	33	36	4	173	32	36	5	173	29	32	4
174	33	36	4	174	32	36	5	174	29	32	4
175	33	36	4	175	32	36	5	175	29	32	4

Note. ACT's College Readiness Benchmark Score for Mathematics is 22 for the ACT. ACT's College Readiness Benchmark Score for Science is 20 for PLAN.

Table 15 (continued)
Estimated PLAN/ACT Score Intervals for QualityCore Scores

Chemistry QualityCore Score	ACT Science Intervals			Physics QualityCore Score	ACT Science Intervals			U.S. History QualityCore Score	ACT Reading Intervals		
	Low Score	High Score	Width		Low Score	High Score	Width		Low Score	High Score	Width
125	6	11	6	125	1	6	6	125	1	6	6
126	6	11	6	126	2	7	6	126	1	6	6
127	7	12	6	127	3	8	6	127	2	7	6
128	7	12	6	128	4	9	6	128	3	8	6
129	8	13	6	129	5	10	6	129	4	9	6
130	8	13	6	130	6	11	6	130	5	10	6
131	9	14	6	131	7	12	6	131	6	11	6
132	9	14	6	132	8	13	6	132	7	12	6
133	10	15	6	133	9	14	6	133	8	13	6
134	11	15	5	134	9	15	7	134	8	14	7
135	11	15	5	135	10	17	8	135	8	14	7
136	11	17	7	136	11	17	7	136	8	14	7
137	11	17	7	137	11	18	8	137	9	14	6
138	12	17	6	138	12	19	8	138	9	14	6
139	12	18	7	139	12	19	8	139	10	15	6
140	12	18	7	140	12	20	9	140	11	16	6
141	12	18	7	141	13	21	9	141	11	16	6
142	13	19	7	142	13	21	9	142	11	17	7
143	14	20	7	143	14	22	9	143	11	19	9
144	15	21	7	144	16	24	9	144	13	20	8
145	15	21	7	145	17	25	9	145	13	20	8
146	16	22	7	146	18	25	8	146	13	21	9
147	17	22	6	147	19	26	8	147	14	23	10
148	18	23	6	148	19	26	8	148	16	24	9
149	19	25	7	149	19	26	8	149	18	26	9
150	19	25	7	150	20	27	8	150	19	28	10
151	20	25	6	151	21	29	9	151	19	28	10
152	20	26	7	152	21	29	9	152	19	28	10
153	21	26	6	153	23	29	7	153	21	29	9
154	21	26	6	154	24	31	8	154	21	29	9
155	21	26	6	155	26	33	8	155	22	30	9
156	21	26	6	156	26	34	9	156	23	30	8
157	21	27	7	157	26	34	9	157	24	30	7
158	22	28	7	158	26	34	9	158	24	31	8
159	23	29	7	159	26	34	9	159	24	31	8
160	24	29	6	160	27	35	9	160	25	32	8
161	25	30	6	161	28	35	8	161	25	33	9
162	25	30	6	162	29	35	7	162	26	33	8
163	25	30	6	163	29	35	7	163	27	33	7
164	25	30	6	164	30	36	7	164	28	33	6
165	26	31	6	165	30	36	7	165	28	33	6
166	27	32	6	166	31	36	6	166	28	33	6
167	28	32	5	167	31	36	6	167	29	34	6
168	29	33	5	168	31	36	6	168	29	34	6
169	29	33	5	169	31	36	6	169	29	34	6
170	30	34	5	170	31	36	6	170	30	35	6
171	30	34	5	171	32	36	5	171	30	35	6
172	31	35	5	172	32	36	5	172	30	35	6
173	31	35	5	173	32	36	5	173	31	36	6
174	32	36	5	174	32	36	5	174	31	36	6
175	32	36	5	175	32	36	5	175	31	36	6

Note. ACT's College Readiness Benchmark Score for Science is 23 for the ACT. ACT's College Readiness Benchmark Score for Reading is 22 for the ACT.

QualityCore National Administrations by Academic Year

This section presents information about the QualityCore End-of-Course (EOC) assessments that were administered nationally during the 2007–2008 and 2008–2009 academic years. It also includes information on the equating process for QualityCore multiple-choice (MC) and constructed-response (CR) forms, sample sizes, and scale score summary statistics. (Note that the QualityCore tests have two sections: The first section consists of MC items, and the second section consists of either MC or CR items. These sections are administered as either MC + MC or MC + CR forms.)

To ensure comparable scale scores for the MC + CR and MC + MC forms, the linking of MC + CR forms to MC + MC forms was implemented using common item equipercentile equating with post-smoothing. Note that all of the MC items in the first section (i.e., MC1 in Figure 14) were used as the common items between the two different types of forms. Doing so put the scale scores for the two different types of QualityCore forms on the same scale, so the scale scores from the two types of forms are comparable. Figure 14 illustrates the linking for QualityCore MC + CR and QualityCore MC + MC forms.

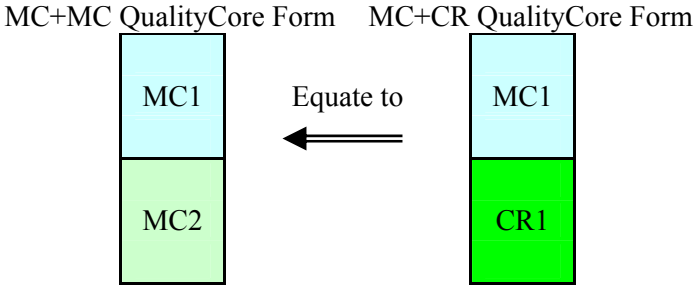


Figure 14. Illustration of Equating Links for QualityCore MC + CR Forms

Student records with any of the following issues were excluded from the equating process:

- Missing form identification code
- Missing ID number and last name
- Missing raw responses (A, B, C, D) and scored responses (0, 1) for MC items
- Missing raw total score
- Raw total MC or CR score equal to zero
- Having 75% (26 items) or more omitted items in the MC section
- Accommodations, special education, and limited English proficiency
- Repeat tester
- Irregularity

Data for QualityCore English courses were analyzed by grade (9, 10, 11, or 12), while all grades were included for other courses.

Since the MC + CR forms have fewer raw score points than MC + MC forms, the raw-to-scale score conversions of the MC + CR forms may not cover the entire score scale. When the sample

size is very small, the raw score range is limited. As a result, there may be gaps in the raw-to-scale score conversion tables and/or multiple raw scores converted to the same scale score. Hand adjustments or linear equating conversions were used to adjust the conversions developed using equipercentile equating with small sample sizes. For the test forms with small sample sizes, the equating will be reimplemented when larger sample sizes are available. The conversions will be finalized after sufficient sample sizes are obtained.

ACT delivered QualityCore EOC tests nationally for eight courses in the 2007–2008 academic year and for 12 courses in the 2008–2009 academic year. Overall, 97,090 tests (including both MC + MC and MC + CR forms) were administered in 2007–2008, and 57,147 tests (also including both MC + MC and MC + CR forms) were administered in 2008–2009. Table 16 presents the scale score summary statistics for the 2007–2008 and 2008–2009 academic years.

Table 16
Scale Score Summary of QualityCore Tests Administered by Academic Year

Course	Academic Year	Sample size	Mean	SD	Skewness	Kurtosis
English 9	2008–2009	7,202	154.44	7.62	–0.36	2.48
English 10	2007–2008	324	153.37	6.19	–0.21	2.56
	2008–2009	7113	152.80	7.43	–0.42	2.49
English 11	2007–2008	54,302	152.14	7.75	0.04	2.27
	2008–2009	6,069	152.37	7.97	–0.07	2.17
English 12	2007–2008	20	153.10	7.51	0.26	2.46
	2008–2009	21	149.24	8.77	0.08	2.00
Algebra I	2007–2008	1,149	146.60	4.45	0.45	3.13
	2008–2009	7,954	145.87	4.48	0.23	3.21
Algebra II	2007–2008	40,671	145.91	4.34	0.33	3.42
	2008–2009	6,267	146.14	4.83	0.14	3.03
Geometry	2007–2008	59	144.95	4.07	0.54	3.65
	2008–2009	6,650	144.95	4.62	0.38	3.45
Precalculus	2008–2009	2,682	146.90	5.35	0.51	3.70
Biology	2007–2008	549	147.41	5.70	0.49	2.84
	2008–2009	7,293	150.13	6.57	0.05	2.51
Chemistry	2007–2008	16	145.56	5.10	–0.10	1.98
	2008–2009	5,217	147.15	6.20	0.42	2.85
Physics	2008–2009	401	143.97	4.73	0.12	3.29
U.S. History	2008–2009	278	145.54	5.54	0.33	2.62

Validity Evidence for the QualityCore® End-of-Course Assessments

More and more high school graduates are going to college. Nearly three out of four new high school graduates are pursuing postsecondary education within two years of leaving high school (Berkner & Chavez, 1997). Many more students follow over the course of their working lives. However, many of these students must begin college by taking remedial rather than credit-bearing courses. Nearly a quarter of those who start in four-year colleges—and nearly half of those who start in two-year colleges—do not make it to the second year (Cambiano, Denny, & Devore, 2000).

Although high schools have worked hard to increase the numbers of students who go to college, the challenge now is to determine what can be done to improve students' chances of success once they get there. Adelman (2006) reports that the most reliable predictors of college success are the quality and intensity of the high school curriculum. Students who take a complete college-preparatory sequence of courses not only do better on college admissions examinations, such as the ACT test, but also are more likely to succeed once they are admitted.

Identifying a course of study is the easy part. Studies have determined the minimum high school course work essential for college success to be four or more years of English and three or more years each of mathematics, social sciences, and natural sciences (Barth, 2003; ACT, 2004). Defining quality is more difficult.

Some high schools—including high schools serving significant low-income and minority student populations—are successfully defining and ensuring quality. Students in these schools are performing significantly better on the ACT test than their counterparts in other high schools. What do classes in these high-performing schools look like?

On Course for Success

During the 2003–2004 academic year, ACT and The Education Trust sought to answer this question by collaborating on a study to determine the courses, the level of rigor, and the instructional practices that most likely lead to student success in college. *On Course for Success* (2004) focused on 10 high-performing high schools where the student population was at least 40% minority and/or at least 50% low-income. Each of those schools also had a significant proportion of graduates who had met or exceeded the ACT College Readiness Benchmarks for English, mathematics, and science, which predict at least a C grade in first-year college courses in English Composition, Algebra, and Biology. The study addressed the question: What components of high school courses prepare students for successful entry into postsecondary education without the need for remediation?

The study team surveyed 69 teachers—whose students scored at or above the College Readiness Benchmarks—about their educational experience, teaching philosophy, and instructional practices. The study team reviewed and evaluated three weeks of these teachers' classroom

lesson plans and instructional materials and also obtained state-, district-, and building-level course standards that pertained to the teachers' work. Finally, the teachers were observed in their classrooms, and an hour-long interview with each teacher was conducted, taped, and subsequently transcribed. The principal of each school was interviewed as well. Results of the survey supported research which has shown that providing the following key academic resources has a positive effect on student learning:

- **High-level college-oriented content**

Successful students were enrolled in college-preparatory courses in their high schools and were learning the skills needed to be ready for college-level work. The content of these courses put students on a trajectory toward college from Grade 9 through Grade 12.

- **Well-qualified teachers**

Teachers of successful high school courses were qualified to teach their academic discipline in high school, and many held advanced degrees.

- **Flexible pedagogical styles**

The teachers commanded flexible pedagogical styles, allowing rapport with students. To assist in the comprehension of difficult concepts, the teachers made connections to previous learning, to current events, to popular culture, and across the curriculum.

- **Tutorial support**

In the 10 schools and 69 courses studied, both the schools and the teachers of the courses supported students with tutorial help, formally and informally.

ACT Course Standards

The instructional materials that were collected and analyzed as part of *On Course for Success* were the basis for the ACT Course Standards produced for key courses in English, mathematics, and science. Sets of rigorous Course Standards have been developed for English 9, English 10, English 11, English 12, Algebra I, Geometry, Algebra II, Precalculus, Biology I, Biology II, Chemistry, and Physics. The ACT Course Standards are a synthesis of the materials—course objectives and standards, course syllabi, pacing charts, instructional materials—submitted by the 69 teachers who participated in *On Course for Success* (31 English teachers, 17 mathematics teachers, and 21 science teachers). These rigorous ACT Course Standards begin to provide real answers to the questions: What does rigor look like? How can it best be incorporated into teaching?

The materials submitted by the teachers participating in *On Course for Success* were examined for content (disciplinary coverage) and for clarity and specificity (quality). The standards subsequently chosen as ACT Course Standards were evaluated from the perspective of content, clarity, and organizational structure.

Content

- The ACT Course Standards cover the major concepts of the field, the essential ideas that students must master if they are to have a grasp of the field.

- The knowledge and skills appear at the appropriate levels, i.e., in the appropriate course.
- The content, taken as a whole, represents what is manageable for instruction.

Clarity and Specificity
- The language is clear and free of jargon; technical terms are defined.
- What is included as an example and what represents material students must learn are clear.
- What is expected of students by the end of the course is clear.
- The ACT Course Standards are measurable; they employ strong verbs such as analyze, compare, demonstrate, describe, evaluate, explain, identify, illustrate, locate, make, trace, and construct.
- The Course Standards are of increasing intellectual difficulty for each course and/or at each higher educational level. The standards represent increasing intellectual sophistication and higher levels of abstraction, as well as the skills required to deal with increasingly complex arrays of information, across the sequence of courses.

Organization
- The ACT Course Standards work as organizing statements of the discipline.
- The Course Standards are grouped in categories reflecting coherent bodies of scholarship.
- The organization aids lesson plan and unit development.

High School Survey

A validity study was conducted in January–February 2005 by surveying high school English, mathematics, and science teachers about the importance in their courses of the course standards identified in the *On Course for Success* study. The purpose of the course standards survey was to determine which standards a majority of the respondents would deem essential in preparing students for college course work. Participating teachers were chosen, using information from Market Data Retrieval, from a list of the top 300 high-performing high schools, based on ACT test results from the 2004 graduating class. Surveys were sent to 200 English language arts teachers, 195 mathematics teachers, and 219 science teachers.

The teachers were told the survey would help researchers identify "rigorous" knowledge and skills taught in college-preparatory courses. Rigorous Course Standards were described as:

- Being specific enough to allow teachers to effectively plan and teach a particular course
- Being written in clear, precise language
- Including challenging content of appropriate breadth and depth
- Focusing on essential concepts and skills
- Requiring students to demonstrate increasingly sophisticated thinking skills, such as application, analysis, synthesis, and evaluation
- Being easy to measure

The survey asked teachers to ". . . indicate whether the knowledge and skills are ESSENTIAL (must be taught and mastered to prepare students for entry-level college course work) and/or OPTIONAL (may be taught to enrich the course)." If teachers deemed the skills and knowledge

unimportant or not applicable (e.g., taught in another course, already mastered), they were asked not to respond. Survey return data are provided in Table 17.

Table 17
Number of Surveys Returned by Course

Course	No. of Surveys Returned
English 9	29
English 10	29
English 11	30
Algebra I	33
Geometry	33
Algebra II	32
Precalculus	32
Biology I	36
Biology II	34
Chemistry	39
Physics	41

Various "cutoff" scores (70%, 75%, 80%, and 90%) were considered for identifying the corresponding number of course standards that should be omitted. Across the surveys, the 70% agreement rate seemed the most viable.

Expert Review

In a second validity study conducted in Spring 2005, the results of the high school survey and the ACT Course Standards were sent to a panel of content experts—three each for English 9, 10, and 11; Algebra I, Algebra II, Precalculus, and Geometry; Biology I, Biology II, and Chemistry; and two for Physics. These content experts were asked to rate their level of agreement for each standard by circling a "1," "2," or "3," which denoted No Change Necessary, Revision Needed, or Delete, respectively.

Each content expert was also asked to review the high school survey results and respond to these questions: To what degree does the survey data confirm your belief about what is occurring in high school classrooms? And, to what degree do the ACT Course Standards represent what should be occurring in high school classrooms? Additional questions were related to specific standards and/or types of standards, for example: Would it make sense to avoid requiring students to develop proficiency in public speaking? What do you make of the fact that so few teachers found it essential to discuss issues relating to media in their classes? Finally, the content experts were asked to evaluate the degree to which the ACT Course Standards represent an increasing level of rigor.

Analysis of Content Experts' Review

Upon receiving the content experts' reviews, two types of tables were prepared to record the information: color-coded tables to record the reviewers' rankings of a "1," "2," or a "3" and tables to record their comments.

If the content reviewers rated a course standard a "1" and 70% of the high school teachers deemed it an "Essential" skill, the decision was straightforward. Decisions related to course standards receiving a "3" were easily dealt with as well; these standards tended to be too difficult for the course under study, i.e., needed to be moved to the next course in the sequence and/or were not easily measurable. The course standards that received a "2," however, needed to be retained, but they required some type of revision. The majority of these revisions were minor wording changes, given a recognized need to stay as true as possible to the course standards submitted by the *On Course for Success* teachers.

Conclusion

This QualityCore® Technical Manual has documented the technical characteristics of the QualityCore End-of-Course (EOC) Assessments in light of their intended purposes. ACT regularly conducts research as part of the ongoing evaluation of its programs. This research is intended to ensure that the programs are technically sound.

This report described the development of the EOC assessments and provided validity and reliability information, results of scaling and equating activities, and interpretations of assessment results. Summary technical information was presented, including a description of the item analysis procedures for the multiple-choice and constructed-response items, and disaggregated test and strand level results for gender and ethnicity. Research studies conducted in support of the EOC assessments were also discussed in this manual, including linking studies that resulted in estimated PLAN® or ACT® domain-score ranges based upon QualityCore EOC scores.

This manual will be updated as additional technical information becomes available and/or additional research studies are completed. Individuals who want more detailed information on a topic discussed in this manual, or on a related topic, are encouraged to contact ACT. Please direct comments or inquiries to ACT, QualityCore Research Services (31), P.O. Box 168, Iowa City, Iowa 52243-0168.

References

ACT. (2004). *ACT national data release.* Iowa City, IA: Author.

ACT. (2007). *Rigor at risk: Reaffirming quality in the high school core curriculum.* Iowa City, IA: Author.

ACT & The Education Trust. (2004). *On course for success: A close look at selected high school courses that prepare all students for college*. Iowa City, IA: Authors.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Barth, P. (2003). A common core curriculum for the new century. *Thinking K–16*, *7*(2), 3–19. Washington, DC: The Education Trust.

Berkner, L. K., & Chavez, L. (1997). *Access to postsecondary education for the 1992 high school graduates*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Brennan, R. L. (Ed.). (1989). *Methodology used in scaling the ACT Assessment and P-ACT+*. Iowa City, IA: ACT, Inc.

Cambiano, R., Denny, G., & Devore, J. (2000, Winter) College student retention at a midwest university: A six-year study. *Journal of College Admission, 166*, 22–29.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education.* Washington, DC: Author.

Kolen, M. J. (1988). Defining score scales in relation to measurement error. *Journal of Educational Measurement*, *25*, 97–110.

Kolen, M. J., & Hanson, B. A. (1989). Scaling the ACT Assessment. In R. L. Brennan (Ed.), *Methodology used in scaling the ACT Assessment and P-ACT+* (pp. 35–55). Iowa City, IA: ACT, Inc.

NCME Ad Hoc Committee on the Development of a Code of Ethics. (1995). *Code of professional responsibilities in educational measurement.* Washington, DC: National Council on Measurement in Education.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–262). New York: American Council on Education, and Macmillan.

Webb, Norman L. (2002). Depth-of-Knowledge Levels for Four Content Areas. Retrieved from http://facstaff.wcer.wisc.edu/normw /All%20content%20areas%20%20DOK%20levels%2032802.doc