



CLT

2018 TECHNICAL REPORT
The Classic Learning Test



2018 TECHNICAL REPORT
The Classic Learning Test

CONTRIBUTORS

Technical Report Lead

Noah Tyler

Psychometricians

Hong Jiao Liru Zhang

Chapter Leads

Stephanie Bucklin Conor McManus

Eric Newman Connor Hocking

Joseph Loomis Brittany Higdon

Writers/Reviewers

Kent Anhari Molly Barnes

Mary Trent Ashley Thorne

Leadership Reviewers

Jeremy Tate David Wagner

Graphic Designer

Meg Prom

Copyright © 2018 Classic Learning Initiatives, LLC. All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of Classic Learning Initiatives, LLC.

Letter from the President

I can't say I've always dreamed of starting a standardized testing company. As a boy, I hoped one day to become a football player, a cowboy, or a police officer. These professions seemed to offer adventure and fulfillment whereas my experience in public school, grades K-12, felt tiresome and stifling. Not once did a teacher or book present any of the questions that could have made school more meaningful.

By the time I entered school in the mid 1980s, any question that carried moral or ethical implications, or any question about the purpose of life, sacred responsibilities, or where to find human happiness, had been removed from the classroom. The education I experienced had been designed with purely utilitarian ends in mind. Any transcendent idea had been gutted from the curriculum and as a result, like most of my classmates, I was painfully bored. It wasn't until graduate school that I came to appreciate the holistic education previous Americans had received. The founding fathers of the United States revived my imagination. They were deeply interested in philosophy, human nature, political theory, and the pursuit of happiness. The education they received was aimed, most fundamentally, at making a person more fully human.

As I questioned how such a beautiful concept of education had been lost, I came to the conclusion that high-stakes testing, especially the SAT and ACT, were partially to blame. Not once since the launch of CLT has someone refuted the idea that high-stakes testing drives secondary curriculum. David Coleman, CEO of the College Board, has stated publicly that "teachers will teach towards the test. There is no force on this earth strong enough to prevent that." If teaching to the test is an inescapable reality, then shouldn't the most important test engage students with some of the most important ideas, texts, and subjects? CLT was born in response to this question. We hope that by offering a new standard that puts students in front of the thinkers and questions that have most meaningfully shaped our culture for the past two millennia, we can be a catalyst for renewal in education nationwide.

Jeremy Tate

Jeremy Tate,
CLT President

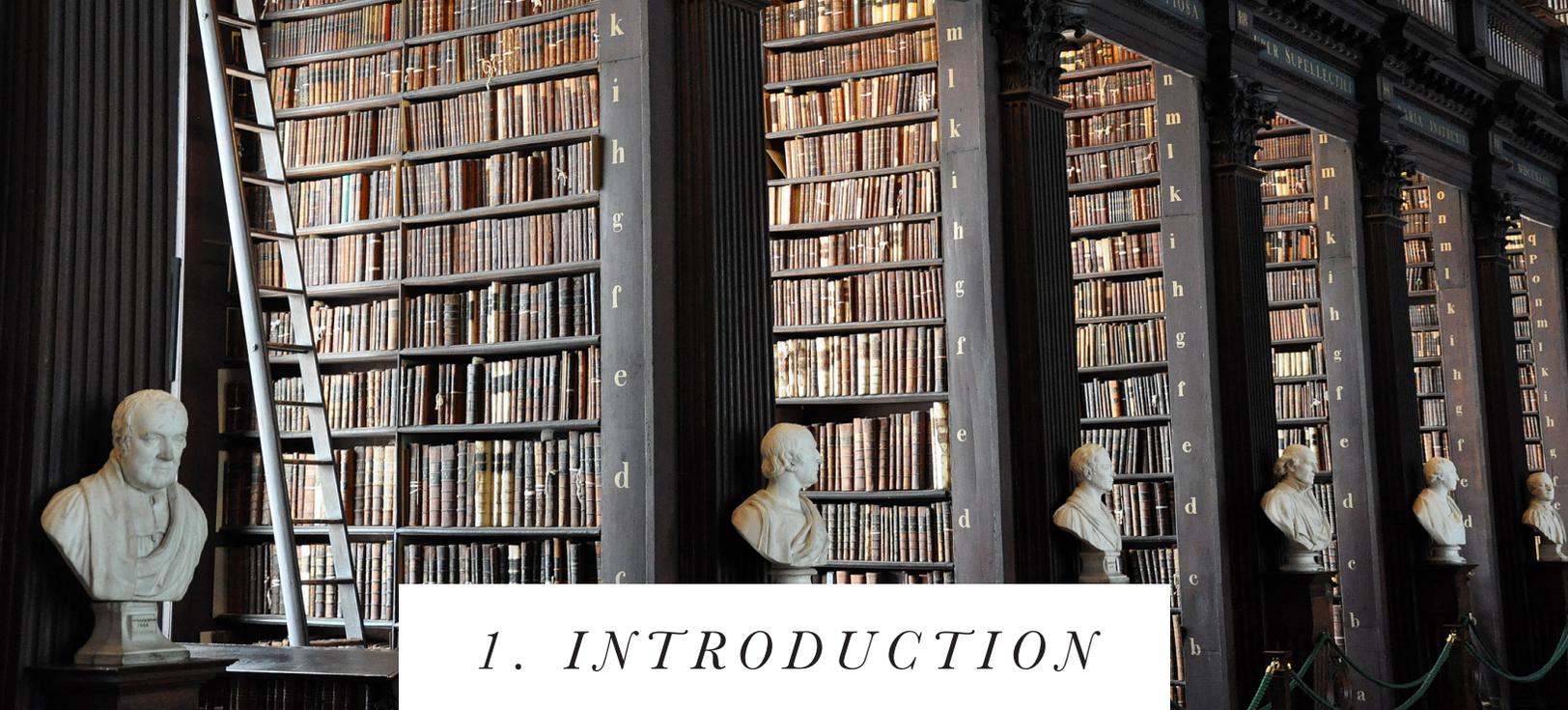


Table of Contents

1. Introduction	1
Improving Students' Test-Taking Experience	
Motivating Positive Change in Assessment and Education	
Contents of the Technical Report	
2. Standards and Content Coverage	5
Overview	
Verbal Reasoning Test	
Grammar/Writing Test	
Author Bank	
Quantitative Reasoning Test	
Calculator Policy	
Optional Essay	
Difficulty Levels	
3. Test Development	15
Selecting and Training Item Writers	
Design, Review, and Uploading	
Item Review	
4. Test Administration	17
Public vs. In-School Testing	
Proctors	
Test Day Schedules	
Test Scores	
Testing Formats	
Test Day CLT Support	

5. Test Security	21
Data Security	
Data Privacy and Acceptable Use	
Access Control	
Monitoring and Auditing	
Incident Management and Response	
Test Day Processes and Procedures	
6. Fairness and Accessibility	25
Fairness During the Testing Process	
Fairness in Score Interpretations for Intended Uses	
Fairness in Test Accessibility	
Accommodations and Requests	
7. Test Results	27
Student Score Reports	
College Score Reports	
Secondary School Score Reports	
Forthcoming Reporting Features	
8. Concordance Relationship	31
Overview	
Study Design	
Results	
Summary	

9. Item Analysis	39
Introduction to Psychometric Evaluation	
Summary of Item Analysis Results for CLT Tests	
10. Reported Scores	45
Overview	
Summary of Scores by Test Form and Testing Mode	
Summary of CLT Scores by Subgroup	
Summary of CLT Scores by Subgroup and Testing Mode	
Summary of CLT Scores by School Type and Family Income	
11. Standard Error of Measurement	55
Overview	
Reliability and Standard Error of Measurement	
12. Validity	65
Overview	
Evidence Based on Internal Structure	
Evidence Based on Content	
Quality Control of Psychometric Analysis	77
Appendix	79



1. INTRODUCTION

The Classic Learning Test (CLT) launched in December 2015 as a long-awaited alternative to the SAT and ACT. As of December 2018, more than 15,000 students at schools across the United States have taken an exam from the CLT suite of assessments,¹ and over 140 colleges and universities have adopted it as an admissions test.² The CLT is a different kind of standardized college entrance exam. It aims to dramatically improve students' test-taking experience, and to motivate positive change in assessment and education.

Improving Students' Test-Taking Experience

For students, the CLT is refreshingly user-friendly, clear-cut, and modern. It was designed with the goal of providing the best possible test-taking experience, and it includes the following features:

- » Online platform, on students' own devices
- » Clean, consistent format
- » Straightforward scoring—120 questions, 120 possible points
- » Shorter test-taking time—120 minutes
- » Same-day scores

ONLINE PLATFORM

The CLT is primarily administered online (84% of total test administration in 2018), though a paper version is an option for in-school testing. The online platform is more natural for contemporary students than a pencil and paper format, and it reduces the possibility of confusion and unnecessary mistakes. Students can select and change their answers with one click without having to fill in Scantron bubbles or take time to erase.

Students take the test on their own devices. Using an unfamiliar device for a high-stakes test can be frustrating since every desktop, laptop, and tablet has its own subtle differences. Allowing students to use a device they are already familiar with reduces the possibility that the device itself will impair the student's score.

1 The CLT suite of assessments is comprised of the CLT, a college entrance exam; the CLT10, a preparatory exam for the CLT offered to 9th and 10th graders; and the CLT8, an end-of-grade assessment tool designed for 8th-grade students as they prepare for high school.
2 The full list of colleges which have adopted the CLT as an admissions exam is provided at <https://www.cltxam.com/colleges>.

CLEAN, CONSISTENT FORMAT

The CLT is designed with simplicity and balance. Each of the three sections has 40 questions. Each Verbal Reasoning and Grammar/Writing section has exactly four (4) reading passages, and each passage has exactly ten (10) questions. Knowing what to expect frees students from anxiety that can come from a less predictable test design.

Each section loads into a single browser window, so that students can scroll to any part of that section without changing pages.

The test aesthetic is clean and free from distraction. It uses a white background and a readable serif font, and the reading questions line up side by side with the passage.

A progress bar is provided at the top of the page, with a percentage number, to give students a visual sense of how close they are to completion.

STRAIGHTFORWARD SCORING

On the CLT there are 120 scored questions, each worth one point, for a total of 120 points possible. There is no penalty for incorrect answers. The 120-point scale allows the test to be divided into three equally valuable sections with 40 questions each. This straightforward scoring method is yet another means by which CLT provides transparency and a better test experience.

SHORTER TEST-TAKING TIME

The CLT is 120 minutes long—2 hours—not counting the optional 30-minute essay. In contrast, the ACT is 2 hours and 45 minutes long, and the SAT is 3 hours long. The CLT was designed to be shorter than those tests due to evidence that the scores for many students can be negatively affected by the mental exhaustion they may experience after lengthy testing. CLT aims to provide colleges with a more accurate snapshot of students' academic ability by using a shorter test.

SAME-DAY SCORES

Students can access their scores on the same day they take the test. CLT operations releases scores after the students and proctor have left the last test site, and students may log in to their account to view their results. By contrast, wait time for SAT multiple choice scores is 13-23 days in the 2018-2019 academic year.³ For the ACT, "Multiple choice scores are normally available within two weeks after each national test date, but it can sometimes take up to eight weeks."⁴ Same-day CLT test results eliminate the long wait and help to reduce student anxiety.

Students may also share their CLT scores with colleges of their choice using their secure, online student account. For both the SAT and ACT, students may send score reports to only four colleges for free, after which there is a fee per each additional college. But for students taking the CLT, score sharing is free to an unlimited number of colleges. This significantly enhances students' ability to reach the colleges they care about, and it alleviates part of the financial burden of the college application process.

Motivating Positive Change in Assessment and Education

Besides offering a logistically better and more humane test experience, the Classic Learning Test sets itself apart by its content. CLT aims to change the landscape of assessment, and education generally, by providing a standard that is intellectually richer and more rigorous than other college entrance exams.

3 College Board, Getting Scores, When to Expect Scores, <https://collegereadiness.collegeboard.org/sat/scores/getting-scores>.

4 ACT Test Scores, "How Can I See My Scores?" <http://www.act.org/content/act/en/products-and-services/the-act/scores.html>.

The SAT and ACT are designed to reflect public school curricula, specifically the Common Core State Standards framework. Those standards rely more on recent writings and “informational texts” than on classic books and great stories. Their alignment with Common Core also pigeonholes the SAT and ACT to test only one kind of education.

The CLT is different. Hallmarks of the CLT’s content include the following:

- » Assesses both aptitude and achievement
- » Is more rigorous
- » Provides great reading passages
- » Promotes good education

APTITUDE AND ACHIEVEMENT

The CLT aims to assess not only students’ achievement, but also their aptitude. This is important as an aptitude test defers to the diversity of curriculum within American education. An achievement test, by definition, has to assume a particular body of content or set of academic standards to master.

Of course, students must draw upon the teaching they have received in order to demonstrate what they have learned. “Achievement” of knowledge is a vital and necessary aspect of assessment, and it is relevant for the CLT. Colleges are rightly interested in the educational content of the students who apply for admission. But entrance exams have historically sought to determine not only what students know but how they think. At its inception in 1926, the SAT’s initials stood for Scholastic Aptitude Test. However, the SAT has since rebranded itself and renounced its former identity as an aptitude exam. It now tests students based on their familiarity with the Common Core curriculum.

The CLT, instead of evaluating how much students have crammed for the test, considers students’ intellectual capacity. The CLT assesses skills students can develop through a variety of education types, such as their ability to communicate clearly, to understand metaphors, to think logically, and to solve puzzles. Thus the test is not dependent on a one-size-fits-all curriculum.

MORE RIGOR

CLT also seeks to address the problem of test score inflation. The SAT and ACT no longer differentiate among top students. Even a perfect 1600 on the SAT or a perfect 36 on the ACT does not set students apart from others applying for admission to Harvard, MIT, or other top universities. What is needed is a more rigorous standard that can truly differentiate students at the upper tail of the distribution. The Scoring Comparison for CLT vs. SAT vs. ACT shows that the CLT is more rigorous than the competing tests, and that students can set themselves apart with a strong showing on the CLT⁵.

GREAT READING PASSAGES

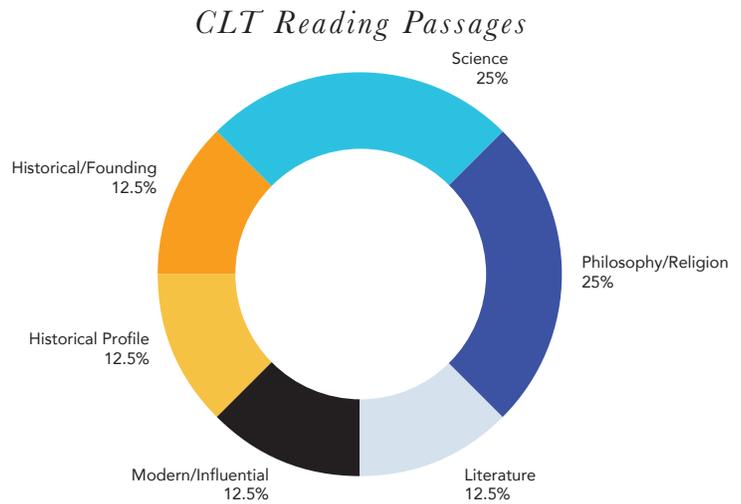
For two-thirds of its reading passages, the CLT uses an author bank of more than 100 men and women whose writings have had a lasting influence on culture and society. While the author bank is not an exhaustive list of every important thinker, the CLT intentionally acknowledges the particular significance of certain authors. The full author bank is listed in Chapter 2.

An important reason CLT uses older sources is that these texts have stood the test of time. Shakespeare was relevant 300 years ago, 50 years ago, and today, and it’s likely he will be relevant 100 years from now. The authors used on the CLT are truly worthy of student’s time and attention. Using older texts also gives the CLT healthy neutrality on contemporary political matters. The SAT and ACT, by drawing primarily from contemporary sources, tend to tacitly favor certain political opinions. Rather than favoring either side of the political spectrum, the CLT draws from works that are grounded and established in tradition.

5 CLT vs. SAT vs. ACT. <https://www.cltxam.com/comparison>.

The CLT’s reading passages also cover subjects that other standardized tests neglect. A study of the eight available SAT practice tests found that only 2% of the passages (1 out of 48) related to ethics, and that none of the passages related to religion. Only 6% of the readings came from classic literature. By far the largest SAT reading passage category was science, and of those passages, twice as many were written by bloggers or journalists as those written by scientific thinkers themselves.⁶

By contrast, the CLT has a balanced distribution of subject categories. On every test, out of eight reading passages, two (25%) are in Philosophy/Religion. One (12.5%) of the passages is drawn from Literature, two (25%) are in Science, one (12.5%) is an excerpt from Historical/Founding Documents, one (12.5%) is a Historical Profile, and one (12.5%) is drawn from Modern/Influential Thinkers.



PROMOTING GOOD EDUCATION

Ultimately the CLT seeks to enhance the way young people are educated. Students are no longer limited to the SAT and ACT. With the introduction of the CLT, schools and homeschooling families have a fresh incentive to focus on enduring ideas.

The test preparation process for the CLT is vastly different than for other tests. The SAT and ACT require extensive—and expensive—test prep in order to navigate them, and the process lends itself to student anxiety rather than encouraging a love of learning. In contrast, the authors students read in advance of and during the CLT are soul-enriching. Students preparing for the CLT gain training in creative problem solving and thinking skills that will benefit them their whole lives.

While “college and career readiness”—the objective of the Common Core State Standards—is important, it should not be the main goal of secondary education. Schools should instead aim to form men and women who are intellectually curious, think deeply, reason well, and live with integrity. The CLT promotes education that helps students on their way to becoming productive adults, but such qualities are also ends in themselves.

By elevating excellent writing and including students in the Great Conversation that began centuries ago, the CLT seeks to raise the standard of schooling in America and contribute to a movement in which education is meaningful and lasting.

Contents of the Technical Report

This technical report is a guide explaining the details of how the CLT exam works. Chapters 1-6 describe the design and administration of the CLT, and Chapters 7-12 explain and analyze the test’s metrics.

Chapter 2 presents the content of the test itself, including sample questions, the author bank, and information on how test questions are organized by difficulty level. Chapter 3 outlines the steps CLT takes to develop, edit, and prepare each test for administration. Chapters 4, 5, and 6 explain how the CLT is administered and describe the measures taken to ensure the test’s security and fairness.

Chapters 7 and 8 provide information on how CLT scores are reported to students and colleges and how these scores compare to scores on the SAT and ACT. Chapters 9 and 10 present demographic statistics on CLT test questions and reported scores, and Chapters 11 and 12 quantify the test’s reliability and validity.

⁶ “What’s Really on the SAT?” CLT, 2018.



2.

STANDARDS AND CONTENT COVERAGE

Overview

The Classic Learning Test (CLT) was created in the context of a national movement to renew the foundations of great education. “Classic” here simply means an assessment that reflects tried and true ideas rather than contemporary experiments. CLT is based on a liberal arts education model, which trains students in language arts and mathematics as a path “to make the acquisition of all later studies more simple and effective.”¹ Clark and Jain (2013) write, “Recovering the primacy of both the language arts and the mathematical arts is a pivotal piece of this paradigm. Together they train the student not just in what to think but in how to think.”²

Whereas other standardized tests rely on what is currently popular and recently legislated in American education, the CLT focuses on ideas that matter on a much grander scale. These include perennial questions about human nature and the physical world; lessons from history; and universal mathematical concepts.

Each CLT exam consists of three mandatory sections, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning, as well as an optional Essay. These are similar to the sections in the SAT and are recognizable to students taking standardized tests. But the content of the test is distinct from other standardized tests in two main ways.

First, instead of the majority of reading passages coming from contemporary sources, the CLT’s two English sections primarily use selections from time-tested authors who have shaped history, literature, and philosophy in foundational ways through the centuries. The CLT thus provides an opportunity for students to interact with important thinkers whose voices have made a profound difference in the world of ideas.

Second, the Quantitative Reasoning section assesses students’ ability to solve problems and to think in a logical and orderly manner. The test balances assessing mathematical reasoning capacity with testing specific mathematical skills and knowledge.

The SAT and ACT are designed around public school standards which change according to educational trends and legislative actions. In contrast, the CLT is based on enduring concepts accessible to students from a variety of educational backgrounds.

¹ Clark, Kevin and Ravi Jain. *The Liberal Arts Tradition: A Philosophy of Christian Classical Education*. Classical Academic Press, 2013.

² *Ibid.*

Verbal Reasoning Test

The Verbal Reasoning section tests a student’s ability to understand and analyze a text. Students are asked to interact with a variety of texts in different subject areas and are tested on their ability to comprehend the text and synthesize ideas within that text. They must be able to understand concepts such as how different phrases and words are used in context, the author’s purpose in a particular section or in the passage overall, how a text is structured, and what could be reasonably inferred based on the information in the text.

VERBAL REASONING TEST BLUEPRINTS

On the Verbal Reasoning section, questions are broken down into two types: Comprehension and Analysis. Comprehension questions include the subdomains “Passage as a Whole,” “Passage Details,” and “Passage Relationships.” Analysis questions include the subdomains “Textual Analysis” and “Interpretation of Evidence.”

Of the 40 Verbal Reasoning questions, 13 fall under Analysis and 27 fall under Comprehension.

One of the Interpretation of Evidence questions always refers to a figure accompanying the Science passage, which is always the second passage of the four.

SECTION	DOMAIN	SUBDOMAIN
Verbal Reasoning (40 questions)	Comprehension (27 questions)	Passage as a Whole
		Passage Details
		Passage Relationships
	Analysis (13 questions)	Textual Analysis
		Interpretation of Evidence

Two questions per passage in the Verbal Reasoning section test analogies based on the passage, for a total of eight analogy questions per set. The SAT removed analogies in 2005, but the CLT includes them based on the understanding that analogies require a high order of logical reasoning and synthesis. Whereas the SAT’s analogies were unattached to reading passages and were criticized for using difficult vocabulary that made the analogy impossible to understand without knowing the words’ definitions, CLT’s analogies refer to concepts within a passage and use terms students are likely to know already. These analogies require students to be able to connect high-level concepts within a passage and to make connections between ideas and terms in a passage.

VERBAL REASONING TEXT TYPES

Each Verbal Reasoning section consists of four passages: three full passages and one passage composed of two shorter excerpts presented together. They always follow a particular order:

- » Literature
- » Science
- » Philosophy/Religion
- » Historical/Founding Documents (2 shorter excerpts presented together)

Tests are calibrated so that each Verbal Reasoning passage fits narrowly within a word count range of 500-650 words. The total must be between 2,275-2,325, for an average of 2,300 words total.

VERBAL REASONING SAMPLE QUESTIONS

Below is one sample question for each subdomain in the Verbal Reasoning section.

Passage as a Whole

Overall, the passage can be best described as

- A) a subtle exploration of the rivalry between two colleagues.
- B) a whimsical tale of a fantastic beast.
- C) a cogent story about an attempt to seek out novelty.
- D) a meandering account of the sale of a crocodile.

Passage adapted from Fyodor Dostoevsky's "The Crocodile," 1865.

Passage Details

According to the passage, what is a hallmark of Mr. Pecksniff's character?

- A) Suspicion of conventional morality
- B) Affection for eloquent language
- C) Fear of the unknown
- D) Disinterest in the lives of his children

Passage adapted from Charles Dickens' Life and Adventures of Martin Chuzzlewit, 1844.

Passage Relationships

medicine : body ::

- A) exercise : spirit
- B) philosophy : soul
- C) politics : philosophy
- D) love : friends

Passage adapted from Plutarch's "On Education" in Moralia, first century AD.

Textual Analysis

In Passage 1, Philosophy indicates she believes Socrates was put to death primarily because

- A) his philosophy was ill-formed and only partial.
- B) he traveled to a distant, violent land filled with barbaric tribes.
- C) his allies, Anaxagoras and Zeno, did not support him.
- D) he lived an upright, ethical life in contrast to those around him.

Passage adapted from The Consolation of Philosophy by Boethius, sixth century AD.

Interpretation of Evidence

Which lines in the passage provide the best evidence in support of the answer to the previous question?

- A) Paragraph 4, Sentence 1 ("And this . . . reality")
- B) Paragraph 4, Sentence 2 ("The great . . . fertilize")
- C) Paragraph 5, Sentence 2 ("But the . . . tendency")
- D) Paragraph 6, Sentence 1 ("Consequently . . . study")

Passage adapted from Christopher Dawson's Religion and the Rise of Western Culture: The Classic Study of Medieval Civilization, 1950.

Grammar/Writing Test

The Grammar/Writing section tests a student’s ability to edit and improve a text. Students are asked to interact with a variety of texts in different subject areas and are tested on their ability to correct errors within that text and to improve its readability and flow. The section assesses students on their ability to use punctuation correctly, to convey points precisely and concisely, to make appropriate transitions, to choose the correct part of speech, to match verb tense, and to make other grammatically well-formed choices.

SECTION	DOMAIN	SUBDOMAIN
Grammar/Writing (40 questions)	Grammar (20 questions)	Agreement
		Punctuation and Sentence Structure
	Writing (20 questions)	Structure
		Style
		Word Choice

GRAMMAR/WRITING TEST BLUEPRINTS

On the Grammar/Writing section, questions are broken down into two types: Grammar and Writing. Grammar questions include the subdomains “Agreement” and “Punctuation and Sentence Structure.” Writing questions include the subdomains “Structure,” “Style,” and “Word Choice.”

Of the 40 Grammar/Writing questions, 20 are on Grammar and 20 are on Writing.

Grammar questions specifically test a student’s ability to correct agreement, punctuation, structure, and other errors. Writing questions test a student’s ability to improve upon a text’s style, flow, and word choice.

GRAMMAR/WRITING TEXT TYPES

Each Grammar/Writing section consists of four passages. They always follow a particular order:

- » Philosophy/Religion
- » Historical Profile
- » Science
- » Modern/Influential Thinker

Tests are calibrated so that each Grammar/Writing passage fits narrowly within a word count range of 460-565 words. The total must be between 2,075-2,125 words, for an average of 2,100 words total.

GRAMMAR/WRITING SAMPLE QUESTIONS

Below is one sample question for each subdomain in the Grammar/Writing section.

Agreement

caring decisions

- A) NO CHANGE
- B) caringly decisions
- C) careful decisions
- D) carefully decisions

Passage adapted from Hilaire Belloc's *The French Revolution, 1911*.

Alighieri, Dante	Dawson, Christopher	Kierkegaard, Søren	Ptolemy
Adams, John	Descartes, René	King, Martin Luther, Jr.	Pushkin, Alexander
Adler, Mortimer	Dickens, Charles	Lewis, C. S.	Rousseau, Jean-Jacques
Angelou, Maya	Dostoevsky, Fyodor	Locke, John	Sartre, Jean-Paul
Arendt, Hannah	Douglass, Frederick	Luther, Martin	Sayers, Dorothy
Aristotle	Du Bois, W. E. B.	Machiavelli, Niccolò	Schaeffer, Francis
Averroes	Edwards, Jonathan	Madison, James	Shakespeare
Avicenna	Einstein, Albert	Marx, Karl	Shaw, George Bernard
Bacon, Francis	Epicurus	Melville, Herman	Smith, Adam
Baldwin, James	Eliot, George	Mendel, Gregor	Sophocles
Beauvoir, Simone de	Eliot, T. S.	Merton, Thomas	St. Augustine of Hippo
Belloc, Hilaire	Fabre, J. Henri	Milton, John	St. Teresa of Ávila
Bentham, Jeremy	Fitzgerald, F. Scott	Montaigne, Michel de	St. Teresa Benedicta of the Cross
Boethius	Franklin, Benjamin	Montesquieu, Charles	St. Thomas Aquinas
Bonaparte, Napoleon	Galilei, Galileo	Newman, John Henry	St. Thomas More
Boyle, Robert	Hamilton, Alexander	Newton, Isaac	Thucydides
Bonhoeffer, Dietrich	Hegel, Georg	Nietzsche, Friedrich	Tolkien, J. R. R.
Buber, Martin	Wilhelm Friedrich	Novak, Michael	Tolstoy, Leo
Calvin, John	Heidegger, Martin	O'Connor, Flannery	Twain, Mark
Cavendish, Margaret	Hobbes, Thomas	Orwell, George	Undset, Sigrid
Caesar, Julius	Homer	Pascal, Blaise	Virgil
Chesterton, G. K.	Hume, David	Planck, Max	Voltaire, François
Chekhov, Anton	Hughes, Langston	Plato	von Hildebrand, Dietrich
Churchill, Winston	James, Henry	Pope St. John Paul II (Karol Wojtyła)	Washington, Booker T.
Cicero	James, William	Pope Benedict XVI (Joseph Ratzinger)	Waugh, Evelyn
Copernicus, Nicolaus	Jefferson, Thomas	Pope Leo XIII	Weber, Max
Confucius	Joyce, James	Percy, Walker	Weil, Simone
Cooper, Anna Julia	Kant, Immanuel	Plutarch	Wollstonecraft, Mary
Darwin, Charles	Kepler, Johannes		

Punctuation and Sentence Structure

in the National Government—in the Congress and in the States—to

- A) NO CHANGE
- B) in the National Government; in the Congress; and in the States—to
- C) in the National Government, in the Congress and in the States to
- D) in the National Government, in the Congress, and in the States to

Passage adapted from John F. Kennedy's "Address to the Nation on the State of the U.S. Economy," 1962.

Structure

The author wants to add a sentence to the end of this paragraph. Which option fits best in the passage?

- A) Pell never solved the ancient problems of Diophantos, however.
- B) By 1800, independent projects had listed the primes up to 1 million.
- C) Unfortunately, most of these numbers were incorrect.
- D) Pell would have been able to create two million primes had he a computer.

Passage adapted from Martin H. Weissman's "Why prime numbers still fascinate mathematicians, 2,300 years later," 2018.

Style

Of course, from the hearts of human beings, laws will not eliminate prejudice from them.

- A) NO CHANGE
- B) Of course, from human beings' hearts, prejudice will not be eliminated by human laws they create.
- C) Of course laws will not eliminate prejudice from the hearts of human beings.
- D) Laws of the hearts of human beings are not eliminated by prejudice, of course.

Passage adapted from Shirley Chisholm's "For the Equal Rights Amendment," 1970.

Word Choice

permeated

- A) NO CHANGE
- B) persisted
- C) persecuted
- D) persevered

Passage adapted from St. Teresa of Ávila's *The Way of Perfection*, 1583.

Author Bank

The CLT draws on sources that have helped shape the course of Western intellectual thought, including authors of diverse backgrounds and philosophies. Whereas the SAT and ACT use primarily passages from recent decades—many of the passages are younger than the students taking the test—the CLT looks to writings by time-honored authors writing from c. 400 B.C. to the present day. As of 2018, two-thirds of CLT Verbal Reasoning and Grammar/Writing passages are drawn from the list of authors on the opposite page. The other one-third of passages are drawn from modern scientific writings, modern influential thinkers, or historical figures. Authors are periodically added to the list.

Quantitative Reasoning Test

The Quantitative Reasoning section tests students' ability to think logically, use and manipulate symbols, and understand shapes. Students are asked to complete a variety of questions of various subtypes in order to assess their logic and reasoning ability across different domains.

SECTION	DOMAIN	SUBDOMAIN
Quantitative Reasoning (40 Questions)	Algebra (10 questions)	Arithmetic and Operations
		Algebraic Expressions and Equations
	Geometry (14 Questions)	Coordinate Geometry
		Properties of Shapes
		Trigonometry
	Mathematical Reasoning (16 Questions)	Logic
		Word Problems

QUANTITATIVE REASONING TEST BLUEPRINTS

On the Quantitative Reasoning section, questions are broken down into three types: Algebra, Geometry, and Mathematical Reasoning. Algebra questions include the subdomains “Arithmetic and Operations” and “Algebraic Expressions and Equations.” Geometry questions include the subdomains “Coordinate Geometry,” “Properties of Shapes,” and “Trigonometry.” Mathematical Reasoning questions include the subdomains “Logic” and “Word Problems.”

Of the 40 Quantitative Reasoning questions, there are 10 Algebra questions, 14 Geometry questions, and 16 Mathematical Reasoning questions, as of 2018. There also are 5 figures in each Quantitative Reasoning section.

QUANTITATIVE REASONING SAMPLE QUESTIONS

Below is one sample question for each subdomain in the Quantitative Reasoning section.

Arithmetic and Operations

100. The expression $2^7 + 2^7$ is equivalent to which of the following?

- A) 2^8
- B) 2^9
- C) 2^{14}
- D) 2^{49}

Algebraic Expressions and Equations

113. What are the x -coordinates of the points of intersection of the parabola $y = x^2 - 7$ and the line $y = x - 1$?

- A) $x = 1$, $x = \sqrt{7}$, and $x = -\sqrt{7}$
- B) $x = 1$ and $x = 3$
- C) $x = -2$ and $x = -3$
- D) $x = -2$ and $x = 3$

Coordinate Geometry

96. Line L is parallel to the line $2y - 3x = 7$. Which of the following is perpendicular to line L ?

- A) $y = \frac{3}{2}x - 7$
- B) $y = -\frac{1}{6}x + 7$
- C) $y = -\frac{2}{3}x + 7$
- D) $y = \frac{3}{2}x - \frac{1}{7}$

Properties of Shapes

89. The perimeter of one face of a cube is 20 cm. What is the surface area of the cube?

- A) 25 cm^2
- B) 50 cm^2
- C) 150 cm^2
- D) 600 cm^2

Trigonometry

110. Which of the following is equivalent to the expression $\frac{\sin x \sec x}{\sin^2 x + \cos^2 x}$?

- A) $\sin x$
- B) $\cos x$
- C) $\tan x$
- D) $\sin x \cos x$

Logic

112. A student has invented the following rule for right triangles:

All right triangles have side lengths in the ratio of 3:4:5.

Which of the following is a counterexample that disproves the above statement?

- A) A triangle with side lengths 2, 3, and 4.
- B) A triangle with side lengths 5, 12, and 13.
- C) A triangle with side lengths 6, 8, and 10.
- D) A triangle with side lengths 7, 7, and 10.

Word Problems

114. At a gift store, candles are sold in packages of 4, chocolates are sold in packages of 10, and thank-you cards are sold in packages of 3. Miranda is putting together gift bags, each of which contains one candle, one chocolate, and one card. What is the smallest number of gift bags she can make such that she doesn't have any items left over?

- A) 20
- B) 30
- C) 60
- D) 120

Calculator Policy

Calculators are not allowed on the CLT, including on the Quantitative Reasoning section. Questions are designed to be solvable without the use or need of a calculator.

The CLT is meant to test students' logical reasoning abilities and their ability to understand and simplify complex topics, rather than testing students' ability to make complicated calculations. This policy also secures test integrity and simplifies the test by avoiding the need to specify and monitor which calculator models are permitted.

Difficulty Levels

Reading passages in the Verbal Reasoning and Grammar/Writing sections are calibrated to fit narrowly within a consistent difficulty level. The test developers use TextEvaluator™, a passage calibration software with grade level ratings, to analyze the difficulty level of each passage and ensure it falls within a range of TE 10-12, with an average of 11.

Difficulty levels of questions are scored on a scale of 1 through 5: each section of the test contains 8 questions at each difficulty level, for a total of 24 questions at each difficulty level. On the Verbal Reasoning and Grammar/Writing section, difficulty levels are distributed evenly throughout each passage. Each passage, for which there are 10 questions, has 2 questions of each difficulty level. On the Quantitative Reasoning section, questions increase in difficulty as they progress.

Level 1 questions are the least difficult, and require straightforward reasoning, basic logic, and a minimal number of steps to answer. Level 5 questions are the most difficult, and require more complex reasoning, high-level thinking, and the ability to synthesize difficult concepts. The breakdown of difficulty levels is perfectly balanced, with 20% of questions falling in each of the five levels.

Optional Essay

In Fall 2018, CLT added an optional unscored essay section, in order to enable students to provide colleges with a sample of their writing ability under a time limit. Students have 30 minutes to answer one prompt. The text of their written response may be included with their test results when students send their scores to colleges. The following are sample essay prompts:

Sample Essay 1: Describe what you believe a community to be. What defines it? How large is it? What are its boundaries, and what determines who is inside and out of it? You can draw on contemporary, historical, or literary examples to support your claims.

Sample Essay 2: The Stoic philosophers were deeply concerned by emotion and its tendency to overwhelm. Can emotion be a good thing? Is it a threat to reason, or can it aid reason? Provide examples from history or literature to support your claims.

Sample Essay 3: Are there any situations in which censorship of works is appropriate? If so, explain in what context and why. If not, explain why not. Use examples to support your claims.



3. TEST DEVELOPMENT

The CLT writes and edits each test according to a specific set of parameters. The Test Development, Editorial, and Operations Teams work together in the test preparation process. They follow a schedule of development, review, and uploading, so that every test undergoes quality control and is ready on schedule. CLT analyzes the results of each previous exam and uses that data in the creation of future tests.

Selecting and Training Item Writers

CLT chooses item writers based on their qualifications and demonstrated ability in particular subject areas. Many have experience in fields such as teaching and tutoring. New item writers are supervised by experienced members of the test writing team, and are trained on the breakdown of question types, difficulty levels, and house style of the CLT exam. Their work then goes through multiple rounds of revision and editing to ensure that each section maintains the high standards of each CLT exam, and is consistent, clear, and accurate.

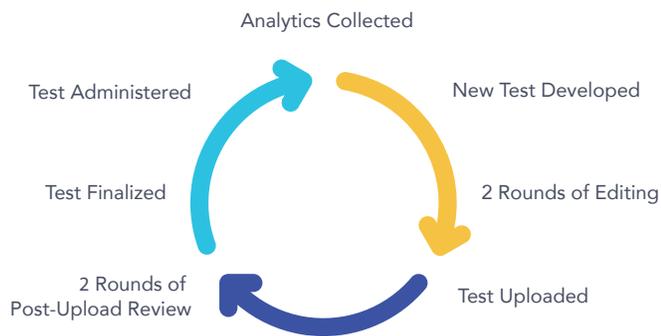
Design, Review, and Uploading

For each test administered, CLT adheres to a schedule for test development, proofreading, and preparation. This schedule is cyclical because each new test takes into account the analytics from previous tests.

DEVELOPMENT AND EDITORIAL PROCESS

After a test date, once the analytics from a test become available, the Test Development Team meets to discuss the performance of the test form, sub-sections and items and to re-calibrate as necessary. The Test Development Team looks at question performance within each specific difficulty level, and especially any questions that perform outside of expectations. At the end of each school year, the Test Development Team meets to review all feedback from that academic year's tests and creates revised guidelines going into the upcoming academic year.

On the basis of this analysis of past exams, and in conformity with the test blueprints laid out in Chapter 2, the Test Development Team creates a new test and answer key for every test date. After initial development, the test goes through two rounds of developmental editing to check for adherence to CLT standards, accuracy, and clarity. Editors rotate between rounds to avoid 'question fatigue.' Each test contains a consistent number of questions within each domain and subdomain (see the "Test Blueprints" sections in Chapter 2). The Editorial



Team confirms that question categories are accurate, difficulty levels are well-calibrated, and questions meet the CLT quality standards. In-house editors help guide the test through the editorial process. The Editorial Team may, at their discretion, edit passages for length during the test composition process. Independent editors and proofreaders perform additional reviews of the test’s accuracy and validity, under the oversight of the Editorial Team.

As part of the test development and editorial process, proofreaders and editors simulate taking the full test during each review, which includes a further check of the answer key and test content as well as proofreading.

The Rights and Permissions Team secures rights for any passages or images under copyright at least 8 weeks in advance of each exam.

UPLOADING PROCESS

After the Editorial Team subjects the test to these rounds of editing and proofreading, they send the test document to the Operations Team for uploading. The Operations Team uploads the exam and all passages and images through the website User Interface, which includes infrastructure for test management. The constituent components of test data in the website User Interface are test questions, passages, and images.

The digital infrastructure for test questions includes variable fields for question numbers (1-120), the text of the question itself, the URL associated with images, the uploaded passage with which the question is associated, the text of answers A, B, C, and D, the correct answer (A, B, C, or D), the difficulty of the question (1-5), and the question type (e.g., “Comprehension—Passage Relationships”). The Operations Team replicates data in each field exactly as it is represented in the test document drafted by the Editorial Team.

POST-UPLOAD REVIEW PROCESS

Once all of the passages, images, and test questions are replicated in the website, the Operations Team links each question to its associated image or passage and proofreads the final upload for completeness, mechanics, and faithfulness to the original test document. The Editorial Team then performs two rounds of post-upload reviews of the test. The Operations Team provides technical support to make any needed changes to the online version. The final reviewer is completely new to the test. In these two rounds of post-upload review, the Editorial Team performs initial and final checks to make sure that the answer key online corresponds to the original answer key (that no errors were introduced), to finalize the formatting of the answers, and to ensure that there are no duplicate answer choices. Once any last changes are made to the uploaded test, final checks are performed and the test is considered complete.

If a paper version of the test is required, the editorial team will create and format the paper document using the final version of the uploaded test. The paper test will then be reviewed in its entirety by a new editor, with a particular focus on formatting, formulas, and other types of errors which might be introduced with the new test mode.

Item Review

After each test administration, CLT completes a thorough item analysis of the test. CLT examines p-values for each item, and at different levels of aggregation. CLT analyzes factors that are correlated to item performance, including item difficulty, the passage connected with each item, subject, academic domain and subdomain, rank of item on the test, and the percentage of students who answered each answer option. These data are analyzed to identify trends and screen out certain types of content from future test administrations, creating a feedback loop between item performance and item creation. Detailed item analysis of this kind is provided in Chapter 9 for two tests administered in April 2018.



4. TEST ADMINISTRATION

Public vs. In-School Testing

The CLT is offered multiple times per year as a public test, open to all students, and as an in-school test, for students at particular institutions. In the 2018-2019 academic year, CLT will offer a public test date four times, twice in the fall and twice in the spring. In the 2018-2019 academic year CLT will offer an in-school date twice, once in the fall and once in the spring.

The public CLT is open to any registrant who is interested in taking the CLT, whether for personal assessment, practice, or for use as a college entrance exam. The only intrinsic limitation is that the student be able to access the testing site. Public test dates are always on Saturdays.

In-school tests are offered for specific U.S. secondary schools which contract with CLT to administer the test to a selection of their students. This test may be offered online or on paper. In-school test dates are always on Wednesdays. The administration process is identical to the public CLT, except that slightly more flexibility is permitted for the start and end time.

PUBLIC TEST SITES

Public CLT test sites are set up at a variety of locations. These sites may be colleges, schools, libraries, testing centers, and any private institutions that meet the criteria laid out in CLT's test site and proctor requirements document.

In order to register a location as a test site, a point of contact representing the location must complete a test site contract. Once these schools are in CLT's system, they may opt in or out of being registered as a test site for each testing date.

REGISTERING STUDENTS FOR IN-SCHOOL TESTS

For in-school testing, instead of individual students registering themselves, CLT creates a school administrator account in which a school may register students through a bulk import process. This process is outlined under the school's administrator profile and consists of uploading a batch of student names and user IDs into CLT's system.

Proctors

Proctors are often, but do not have to be, connected with the institution that is serving as a testing site. CLT carefully selects proctors using a set of proctor selection guidelines. CLT proctors must be twenty-one years of age or older and may not be related to any of the students they are proctoring. Athletic coaches may proctor exams, provided that at least one of the students they are proctoring is not a current or former player for that coach. Once proctors are selected, they must complete a proctor contract in which they provide contact and payment information. CLT proctors are compensated a standard amount, which varies dependent on whether or not there are accommodations.

CLT staff members will coordinate with proctors ahead of time to provide them with basic proctor training in the form of a manual. Proctors must become familiar with the proctor manual, which carefully outlines the procedures for proctoring the CLT.¹

Before the test begins, the proctor asks the students to indicate whether or not they plan to complete the optional essay. This determines whether or not the proctor will stay an additional 30 minutes. Students who do not choose to complete the essay are dismissed after they finish the multiple choice sections of the test.

While the test is in session, proctors are instructed to position themselves where they can view every screen in the room. If a student opens a new browser or a previously saved document, proctors are required to notify CLT immediately as such a testing violation results in a voided test score.

Test Day Schedules

The schedule is dependent on the time zone of the test location, whether or not any student is taking the optional essay, and whether or not the student has been granted extra time for accommodations.

Students are required to arrive 15 minutes before the scheduled test time. At the scheduled test time, 20 minutes are allotted for the proctor to provide instructions.

Students have 40 minutes for the Verbal Reasoning section, 35 minutes for Grammar/Writing, and 45 minutes for Quantitative Reasoning, for a total test-taking time of 120 minutes, not including the optional 30-minute essay. There is a 10-minute restroom break between the Grammar/Writing and Quantitative Reasoning sections.

Please request the CLT Official Test Timecodes for a complete breakdown of the different possible schedules for both proctors and students dependent on all possible variables.

Test Scores

One of CLT's distinctive features is that it releases scores of the online test to students on the same day on which the test is taken. Scores are released after the last test site has sent in their results, usually early evening of the test day.

To see their scores, students simply log into their account. From there, they are able to share their scores with as many colleges as they would like, free of charge.

Students taking an in-school test see their scores in same way as those taking the public CLT. School administrators can see the scores of the students whose names and user IDs they uploaded, as well as advanced analytics and metrics on how those students performed. Schools see this through their CLT school administrator account. Only persons known to CLT to work at the school—usually the CLT's direct contact—may be granted access to this account.

1 2018-2019 Public CLT Proctor Manual.

Testing Formats

The CLT is administered in two modes: primarily online through <http://cltexam.com>, and occasionally with paper and pencil; the latter only for in-school testing and by special request of the school. During the 2017-2018 academic year, 84% of the total CLT administration was online, and 16% was on the paper version.

ONLINE ADMINISTRATION

Students who take the CLT as a public test do so online on their own devices, which they bring to the testing site. This device may be a desktop computer, a laptop computer (including a chromebook) or a tablet. The test will work on most modern devices and browsers. It requires a reliable internet connection with Javascript enabled.

Questions in the Quantitative Reasoning portion of the exam may include mathematical notation. Mathematical notation is scripted in HTML, and thus will be visible regardless of the browser used to take the exam.

On occasion certain testing sites are able to provide devices for all of their students. In this case, students are not required to bring their own devices. This situation is the exception and is discussed individually with the testing site point of contact and students to ensure standardization.

PAPER OPTION

The CLT may also be administered on paper upon special request from the schools. This option is available for in-school test dates only, and not for public test dates. Parameters are the same as those for in-school tests, with the following changes:

The exam booklets and answer sheets are mailed to the school through FedEx or UPS a minimum of one week ahead of the test date. They are sent to the attention of the school's primary point of contact. Proctors at test sites administering the exam on paper receive three (3) additional copies of the test and three (3) extra answer sheets as a fail-safe. As with the regular CLT, proctors are expected to follow a strict process, outlined in the paper test manual.

Students complete a paper answer sheet which proctors are instructed to submit, by mail or by email, to the CLT Operations Team.

Students need CLT accounts in order to share scores with colleges. Schools register their students before test day to allow CLT time to give the schools their students' user IDs, which is a required field on the answer sheet. This can be done individually or through bulk import, as described above.

There is currently no essay on the paper test.

Students receive their scores within 30 days of the test date. CLT converts their answer sheets into a .csv spreadsheet using InspiroScan. The Operations Team then cross-references the spreadsheet with the original answer sheets to ensure that each student's answers are faithfully represented therein. CLT then uploads these documents into the students' CLT accounts.

Test Day CLT Support

For each test day, CLT has a dedicated team of customer service representatives who are available to answer questions from schools, proctors, and parents. This team includes the members of the operations/implementation team, members of the technology development team, as well as the regular customer service team. For this purpose, CLT uses Zendesk Chat/Zopim to field all questions. This platform allows CLT to respond quickly and directly to many different requests via chat or phone. For both public and in-school CLT tests, test-day support is available from before the proctor arrives until after the proctor leaves for every test site and time zone.

5. TEST SECURITY

CLT test security is designed to ensure fairness and equality among test-takers and is broken down into general data security and test day processes and procedures.

Data Security

CLT trains all its employees on the sensitivity levels of CLT data, including the access and use of confidential data such as personally identifiable information. CLT requires each employee to acknowledge and sign internal policies regarding the acceptable use of CLT data.

DATA PRIVACY AND ACCEPTABLE USE

CLT considers all student data confidential, including collected identifiable information (email and profile data) as well as test results. CLT employees may not share any student's data with a third party without that student's express consent.

Students may opt to share their profile and test results with specific colleges of their interest and/or opt into CLT's partnership program in which CLT shares students' data with partner institutions. Students who opt in may also opt out of the program at any time by logging into the CLT web application and editing their profile.

Proctors can view limited student data on test day to facilitate the test and verify attendance. Proctors do not have access to a student's full profile, test history, or any other data. Proctors are not permitted to share any student information with any third parties.

ACCESS CONTROL

CLT data may be accessed either through the web application or through the database directly. All users must be authenticated to access CLT data, and authorization is based on security level.

Web Application Access – The CLT web application security is role-based. By default, all users who register for an account receive the same level of access as students, the most minimal access level.

SUPPORT ACCESS

CLT employees are granted a support role in order to access necessary information to support customers. Users in a support role can view test registrations and view student data, but they cannot access the test

management section of the application.

PRIVILEGED ACCESS

A limited number of CLT employees have privileged access that allows them access to write, review, and modify test data in advance of test dates. This includes the ability to add tests, add and edit questions and answers in existing tests, change test dates and deadlines, and deactivate tests. Privileged access users are required to sign an additional policy regarding test integrity and the acceptable use of test data. Privileged access may be granted only by the Chief Technology Officer.

DATABASE/NETWORK ACCESS

Accessing the database directly falls under privileged access and is limited to the development and analytics teams. Network traffic to access the database is restricted by IP address. Each privileged user is granted two accounts, one read-only and one administrative account. Users use their read-only account unless a critical change is required. Some users, such as those on the Analytics team, may be granted only a read-only account.

Physical Access – all CLT data is stored in a secure cloud environment that is not accessible by CLT employees. The third-party cloud provider ensures the highest level of security and access.

MONITORING AND AUDITING

All activities are logged when changes are made in the software, database, and infrastructure. Logging is monitored on a regular basis to identify breaches, risks, or unexpected behavior. User roles are also monitored on a regular basis to ensure that users have not been inappropriately granted access to data.

INCIDENT MANAGEMENT AND RESPONSE

The CLT Executive Team manages all incidents, including data breaches and/or unacceptable use of data. In the event that user data is compromised, the issue will be immediately remediated and the affected parties will be contacted.

Test Day Processes and Procedures

Students may take the CLT only under secure, supervised conditions. These conditions are as follows:

TEST SITES

Students must take the exam at an approved CLT test site that has submitted a test site contract with CLT. The majority of CLT test sites are secondary schools.

PROCTORS

CLT proctors must be twenty-one years of age or older and may not be related to any of the students they are proctoring. Athletic coaches may proctor exams, provided that at least one of the students they are proctoring is not a current or former player for that coach. CLT creates CLT accounts for all proctors and provides them with the CLT proctor manual.

ADMITTING STUDENTS INTO THE TESTING ROOM

On test day, proctors have the final list of CLT students for their specific test site on their CLT accounts. The manual instructs proctors to verify students' identity before admitting them into the testing room, using any of the following types of approved photo ID:

- » Passport
- » Driver’s license or permit (if photo included)
- » State ID
- » Military ID
- » High school ID (current year only)
- » HSLDA student ID (current year only)
- » CLT Student ID Form

Proctors then assign seats for every admitted student.

WHAT STUDENTS MAY BRING INTO THE TESTING ROOM

Students may bring the following into the testing room:

- » Testing device
- » ID
- » Writing utensil (for scratch paper that the proctor provides)
- » A watch that does not have internet availability, the ability to communicate with other students, or a calculator, and that is set to silent mode

WHAT STUDENTS MAY NOT USE DURING THE EXAM

Students are not permitted to use any of the following during the exam:

- » Calculator
- » Reference material of any kind

CELL PHONES

These must be turned off and stowed in a backpack or purse or left with the proctor.

PASSWORD

In order to take the exam on test day, students must enter the proctor password specific to the exam in question. The proctor password is displayed on proctors’ CLT accounts on test day—it is never communicated to them by email—and the manual instructs proctors to provide their students with this password once all authorized students have been admitted and seated and the preliminary instructions have been read.

TIMING

One of the proctors’ primary duties is to ensure that all students adhere to the designated time lengths for each of the exam’s sections. Once the allotted time for a given section has elapsed, proctors are instructed to notify students of this, have them remove their hands from their devices, ensure all students have complied, and then begin reading the instructions for the next section. To aid the proctor in determining at a glance whether all the students are working on the appropriate section of the exam, each section is color-coded. A similar aid is available to proctors of paper exams: the names of the first, second, and third sections are printed in bold at the top-left, center, and right of the pages, respectively.

START TIMES

Test start times are staggered across US times zones to minimize instances of students in an earlier time zone taking the exam, then communicating about it with students in a later time zone who have not yet begun. The start time is 10:00 AM in the Eastern and Central Time Zones, 9:00 AM in the Mountain, Pacific, and Alaska Time Zones, and 8:00 AM in the Hawaii Time Zone.

ANOMALIES

Proctors must submit the Administration and Anomaly Report to CLT before exiting the testing room. They are instructed to note any testing anomalies on this report. The proctor manual outlines procedures regarding anomalies in section 2.4.¹ Instructions for potential testing anomalies that are to be noted on the report include:

- » Students Who Don't Arrive to an Exam
- » Students Who Arrive Late to an Exam
- » Students Who Leave During an Exam
- » Students Who Use an Additional Device or Open an Additional Page
- » Students Who Become Ill During an Exam
- » Questions Asked During an Exam
- » Disturbances During an Exam
- » Emergency Evacuations
- » Power Failure
- » WiFi Failure
- » Device Failure
- » Site Failure
- » Copying Test Materials

Summary

CLT ensures test security by limiting and carefully monitoring access levels to test data, and by setting up and enforcing specific parameters for test administration. By committing to these rules and training schools and proctors to uphold them, CLT provides an exam that is fair, reliable, and accurate.

¹ 2018-2019 Public CLT Proctor Manual.



6. *FAIRNESS & ACCESSIBILITY*

Overview

CLT is committed to providing every student with a good and fair test-taking experience. The CLT therefore offers a range of accommodations for students with documented learning or physical disabilities in order to give them equal opportunity for success, while maintaining the rigor and integrity of the exam. The CLT adheres to the Individuals with Disabilities Education Act (IDEA) and the Americans with Disabilities Act (ADA). In compliance with these laws and in keeping with its efforts to open the door to all students to fulfill their potential on the test, the CLT seeks to promote diversity and minimize bias.

Fairness During the Testing Process

All CLT testing takes into account learning differences and disabilities in accordance with the Standards for Educational and Psychological Testing. The CLT also considers fairness in testing a top concern, and it thus works to minimize bias, facilitate accessibility, and ensure universal design.

Fairness in Score Interpretations for Intended Uses

The purpose of the CLT exams are both an internal secondary school metric as well as a college entrance exam. A given score does not guarantee admission to a given college.

Fairness in Test Accessibility

CLT provides testing accommodations to students with documented disabilities to make testing accessible to all. CLT's accommodations apply to those whose disabilities are cited by a licensed professional as hindering education and whose accommodations are currently documented as being used in their educational setting.

Accommodations and Requests

First-time CLT test-takers are required to fill out and submit the CLT Testing Accommodations Request Form, which is available on the CLT website.¹ Students who have been granted testing accommodations on the official CLT and/or who have been granted a accommodations on the CLT8 and/or CLT10 in the past five (5) years who are interested in testing accommodations for additional exams may fill out and submit CLT's Repeat Testing Accommodations Request Form.

All accommodations request forms must be submitted on behalf of individual students. Accommodations request forms submitted for more than one student will not be considered.

The disabilities for which CLT provides accommodations include:

- » Specific Learning Disorders
- » Autism Spectrum Disorders
- » Communication Disorders
- » Psychiatric Disorders
- » Visual Impairment and Blindness
- » Hearing Impairment and Deafness
- » Physical Disorders
- » Traumatic Head Injuries
- » Tic Disorders

CLT makes the following accommodations available to students who provide appropriate documentation.

- » MP3 audio test format
- » Braille exam (includes all graphs and figures)
- » Large print exam (14-, 16-, 18-, or 20-point font)
- » Reader
- » Written instructions
- » Signed instructions
- » Extended time (50% and 100% allotments in both uniform and selective formats)
- » Preferential seating
- » Food breaks (must be contained in a clear plastic bag and remain unspoiled in room temperature for at least three hours)
- » Medication

To ensure the timely fulfillment of accommodations requests, requests for accommodations must be submitted to CLT at least eight weeks before the test date. The Testing Accommodations Request Form lists the current academic year's accommodations request deadlines.

The following two (2) types of documentation must be uploaded as PDF or Word documents to the CLT Accommodations Request Form.

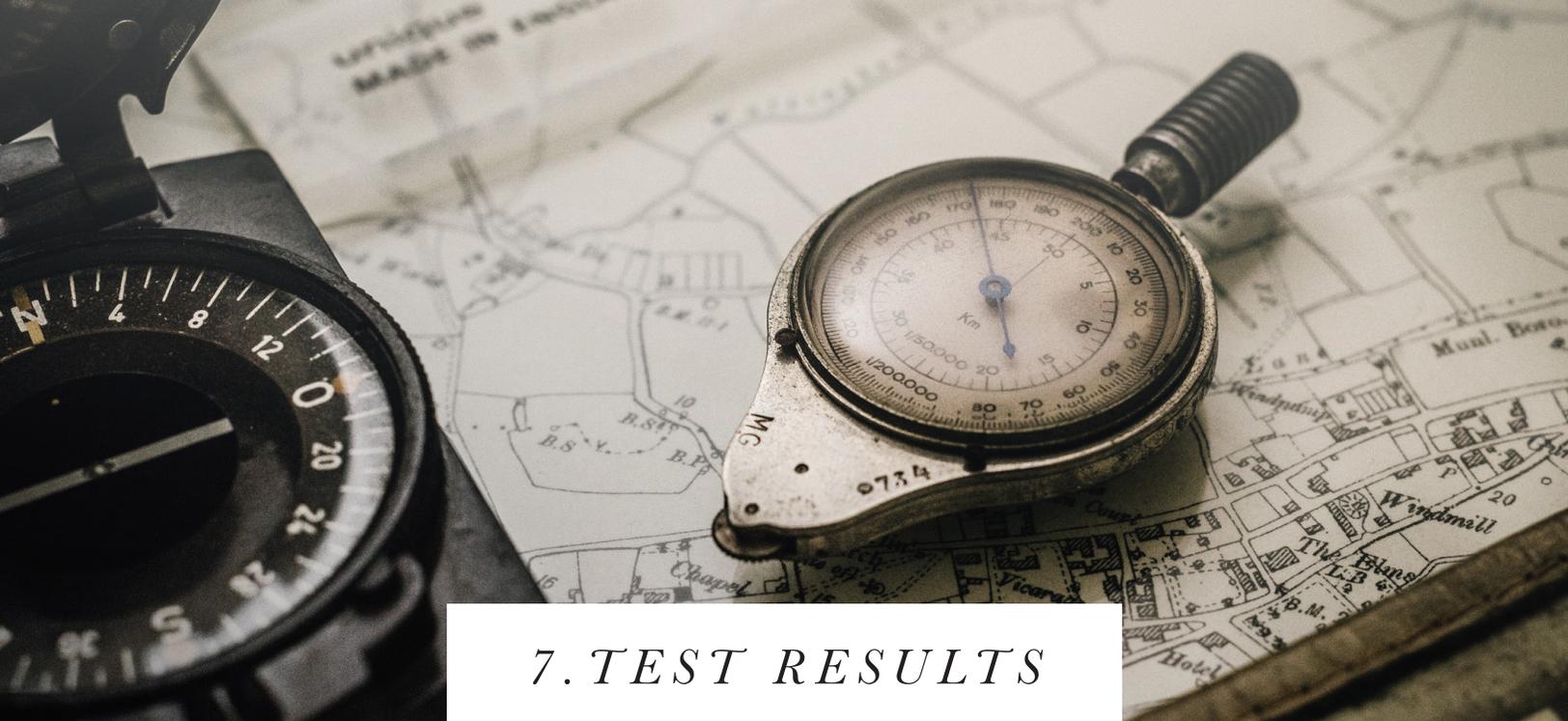
- » Student's IEP, 504, or school-issued official accommodations plan
- » One (1) of the following:
 - » Student's psychological evaluation including diagnostic code
 - » Diagnosis by a certified relevant physician accompanied by test results and diagnostic code

Students requesting repeat CLT accommodations are not required to re-submit this documentation.

CLT will notify the student regarding the student's approval status within two (2) weeks of reception of the pertinent accommodations request form. CLT will contact students by email with any matters that require clarification.

CLT will do its best to honor requests for accommodation. If a student's desired accommodation is not yet included in what CLT offers, students may submit certification of their disability and make a special request, and CLT will seek to fulfill this petition while maintaining the integrity of the exam.

¹ CLT Testing Accommodations Request Form, <https://www.cltprep.com/accommodations-request-form.html>.



7. TEST RESULTS

Overview

CHAPTER SEVEN

CLT has a unique philosophy behind its score reporting. The growth of higher education and the industrialization of standardized testing have led the world of education to become hyper-focused on metrics pertaining to standardized test scores. Scores are used to inform a myriad of decisions such as funding allocation and scholarship awards. While test scores do indicate students' academic ability levels, such high-stakes testing should itself be always under scrutiny. CLT recognizes this tension and thus works to communicate scores that are indicative of both achievement and aptitude, while using content that is intellectually richer and more rigorous than other standardized tests.

CLT provides tailored score reporting to students, colleges, and secondary schools (and home school organizations). For example, CLT provides secondary schools with advanced analytics on both student and school performance. This analytic reporting gives schools the ability to decipher specific areas of strength and weakness, both at the aggregate school level and for each individual student.

Student Score Reports

Students receive same-day test results as part of a score report which compiles their personal demographic, school, college preference, and test information. This is available to students through the student's online account on the CLT website. An example is provided in Image 7.1. The student has the option to share this report with as many colleges as he or she chooses at no additional cost. If the student completes the optional essay section, he or she may also choose whether or not to share the text of the essay with colleges.

Image 7.1 Example Student Score Report



CLT

Official Score Report

Student Information		CLT Scores	
Name	John Doe	Test Date	May 15, 2018
DOB	12/12/2000	Verbal Reasoning	35 / 40
Address	123 Main Street, Apt 456, New York, NY 10001	Grammar/Writing	35 / 40
Email	john.doe@email.com	Quantitative Reasoning	31 / 40
Phone	(123) 456-7890	Total Score	101 / 120
Gender	Female		

School Information		Additional Information	
GPA	4.2	Highest SAT	1470 *
High/Home School	Wilson High School	Highest ACT	33 *
High/Home School Type	Private	Highest PSAT	1400
Graduation Class	2017	Intended Major	Pre-Law or comparable major
		Financial Aid Interest	Must Have

* Projected based on concordance with CLT score

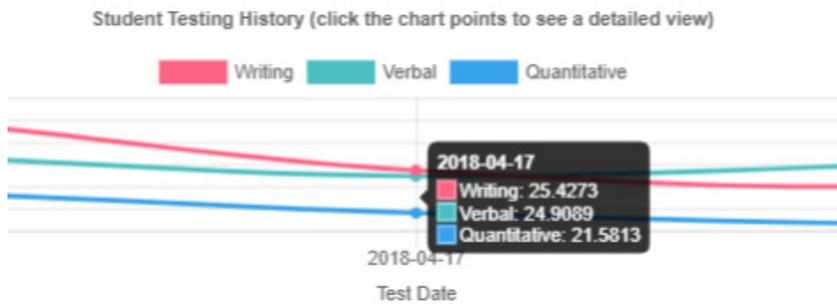
College Score Reports

When students opt to send their score reports and optional essays to colleges of their choice, admissions staff at these colleges receive those students' CLT score information, as well as the ACT and SAT concordance.

Secondary School Score Reports

Administrators of secondary schools and home school organizations also receive same-day test results for their students. These administrators have access to their students' scores through CLT's online "School Admin" portal. In this portal, school administrators can see how their school has performed on tests over time (by section, see Image 7.2), as well as examine specific test results.

Image 7.2 Test History



Test results are available at two levels: basic and premium. Basic analytics are provided free of cost, and include historical average scores for the school, as well as scores and CLT percentiles for each student, per test (Image 7.3). CLT percentiles are user-referenced and indicate how a student performed on the test as compared to their user group.

The current CLT user group consists of students in three main school types: Home School, Private School, and Charter School.

Image 7.3 Basic Test Analytics

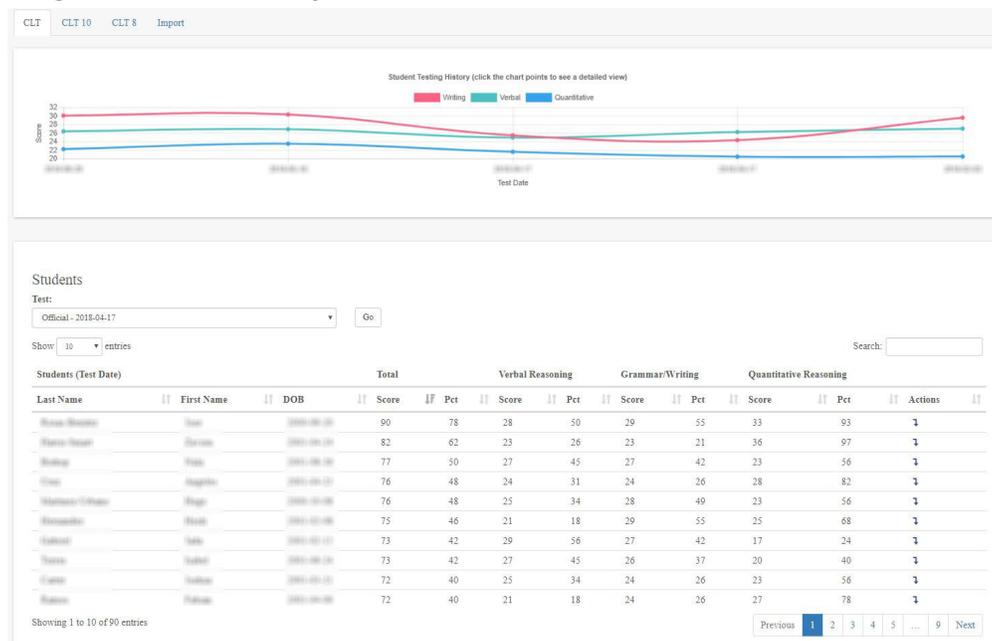
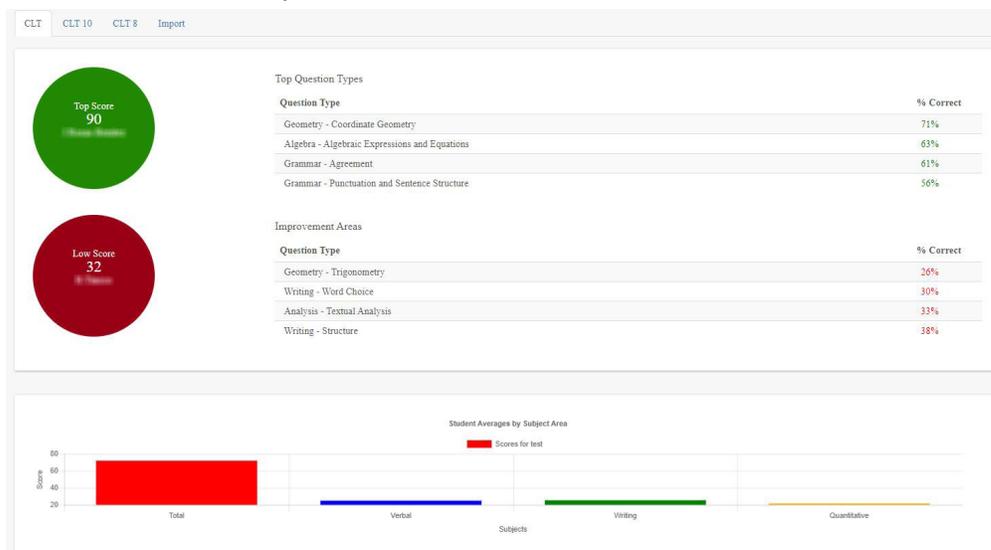


Table 7.1 CLT Sections, Domains, and Subdomains

SUBJECT SECTION	VERBAL REASONING		GRAMMAR/WRITING		QUANTITATIVE REASONING		
	Domain	Analysis	Comprehension	Grammar	Writing	Algebra	Geometry
Subdomain	Interpretation of Evidence	Passage as a whole	Agreement	Structure	Algebraic Expressions and Equations	Coordinate Geometry	Logic
	Textual Analysis	Passage Details	Punctuation and Sentence Structure	Style	Arithmetic and Operations	Properties of Shapes	Word Problems
		Passage Relationships		Word Choice			

Image 7.4

Premium Analytics—Top and Bottom Four Domain-Subdomain Pairings



Premium analytics give schools and home school organizations access to more detailed student and school-level analytics. These come in the form of metrics on how students performed on certain areas of the test, which are reported by academic domain and academic subdomain. These domains and subdomains are listed in Table 7.1.

Premium analytics are delivered to schools at two levels: school/home school organization-level, and student-level. For each level, a percent correct metric is given for each section, domain, and subdomain. At the school-level, this percent correct metric displays the percentage of questions the students at that school got correct within the specified category, for the specified test. School administrators can see the top and bottom four domain-subdomain pairings (in terms of performance), as well as a breakdown of how the school performed on each section, domain, and subdomain (Images 7.4 and 7.5).

Schools with premium analytics can also see any specific student’s test performance by section, domain, and subdomain. This report appears similar to Image 7.5, but includes the student’s identifying information at the top of the page.

Image 7.5

Premium Analytics—School Performance by Section, Domain, and Subdomain

Performance by Subject & Question Type

Verbal Reasoning %	48%
Analysis	35%
Interpretation of Evidence	38%
Textual Analysis	33%
Comprehension	43%
Passage as a Whole	46%
Passage Details	43%
Passage Relationships	42%
Grammar / Writing %	51%
Grammar	59%
Agreement	61%
Punctuation and Sentence Structure	56%
Writing	38%
Structure	38%
Style	42%
Word Choice	30%
Quantitative Reasoning %	50%
Algebra	50%
Algebraic Expressions and Equations	63%
Arithmetic and Operations	41%
Geometry	50%
Coordinate Geometry	71%
Properties of Shapes	51%
Trigonometry	26%
Mathematical Reasoning	44%
Logic	46%
Word Problems	42%

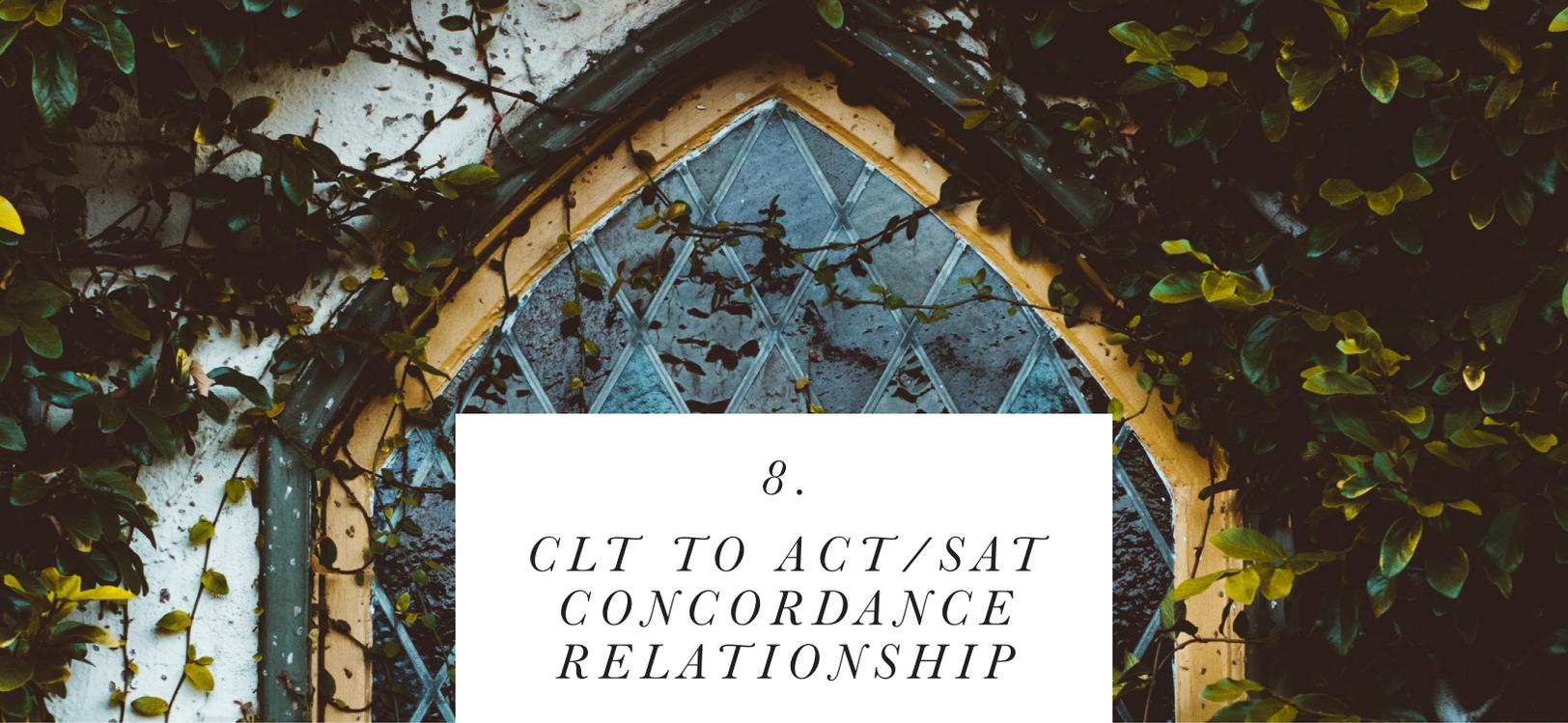
Forthcoming Reporting Features

The score and analytic reporting capability of the CLT continues to grow. By February 2018, CLT will also provide to students an equivalent to the premium school analytics. These analytics will feature student performance by section, domain, and subdomain, and will be available for purchase for students who take the CLT10, but included for students who take the CLT or CLT8.

Another feature in development is a “school associations portal.” On this portal, CLT will provide to schools the ability to compare its results with other schools and associations. This will be done by providing school-aggregate percentile metrics, as well as visualizations and metrics on how the school performed vs. other schools, organizations/associations, and schools of similar (and dissimilar) educational curricula. This feature will be available by the end of the 2018-2019 academic year.

Furthermore, CLT will build into its website a college data portal. This will provide the opportunity for a college to more easily download the information and scores from students who have chosen to share scores with that college.

These reporting capabilities, while still in development, give CLT the opportunity to continue to provide insightful scores and metrics to a wide variety of stakeholders—students, educators, secondary school and home school administrators, and colleges. As more and more students take the test and the user base grows, CLT will continue to innovate and expand its test result reporting features.



8.

CLT TO ACT/SAT CONCORDANCE RELATIONSHIP

Overview

Metrics are meaningless in a vacuum. This is true for any metric, including test scores. However, with the proper context, test scores—specifically, CLT scores—can provide students, parents, educators, and colleges an accurate and meaningful insight into a student’s academic aptitude and achievement.

The CLT aims to provide context for its scores in a number of ways, both qualitative and quantitative. The CLT collects certain qualitative information about its examinees, such as demographic data, secondary school information, and college preferences. These indicators can provide context to colleges that are interested in particular groups of students (e.g. students with an interest in liberal arts colleges, home school students, students with a classical education). Furthermore, the optional essay section provides a genuine writing sample to colleges, which can display a student’s thinking and writing abilities.

Quantitatively, the CLT provides two main measures to give context to its scores: CLT percentiles and ACT and SAT concordance scores. Percentiles are user-normed: that is, they are based on the students who take the test. These percentiles display how the examinee performed on the test as compared to his or her peers. CLT examinees are typically comprised of three main groups: home school students, private-school students, and charter school students. While work remains to be done to establish the national norms for the CLT, the user norm can be regarded as stable. Petersen et. al (1989) write that “this stability is sufficient to allow the norms to be used for meaningful profile comparisons and to allow an examinee to assess, reasonably accurately, where she or he stands in relation to a group of college-bound students” (p. 238).¹

In order to give context to CLT scores outside the CLT user group, CLT has developed a set of concordance tables. According to Holland and Dorans (2006), “Concordances are examples of scalings that produce direct links between the scores” on two different tests (p. 193).² CLT provides a concordance table that links CLT total scores to ACT Composite scores and SAT total scores, as well as concordance tables that link CLT section scores to certain ACT and SAT section scores.

These concordance tables enable students, parents, educators, and colleges that meaningfully compares scores between the three tests.

1 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: Macmillan.

2 Holland, P. W. & Dorans, N. J. (2006). Linking and Equating. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 187-220). Westport: American Council on Education and Praeger.

Study Design

The CLT concordance tables were developed by a study that was completed in two phases. The first phase, completed in June 2018, produced a concordance table between CLT total scores, ACT Composite scores, and SAT total scores. The second phase was completed in August 2018 and produced concordance tables that compared CLT section scores to comparable sections on the ACT and SAT.

Each phase of the study consisted of similar parts: data collection, data validation, linear regression, combining the data, quality control, and publication. The data collection for each phase was concurrent, while the other parts of the study were completed separately.

The data were collected from June 2016 to May 2018 by collecting ACT and SAT score information from students who took the CLT. CLT total scores were concorded to ACT Composite and SAT total scores. CLT section scores were concorded to sections on the ACT and SAT according to Table 8.1 above. The table also shows item counts for each section.

Table 8.1 Respective sections of concordance between the CLT, ACT, and SAT

CLT	ACT	SAT
Total (120 Items)	Composite (215 Items)	Total (154 Items)
Verbal Reasoning (40 Items) Grammar/Writing (40 Items)	English (75 Items) Reading (40 Items)	Evidence-Based Reading and Writing (96 Items)
Quantitative Reasoning (40 Items)	Mathematics (60 Items)	Math (58 Items)

Table 8.2 Descriptive Statistics of the Samples for Concordance Analysis, CLT and ACT

CLT Section	N	Correlation	CLT				ACT			
			Mean	SD	Min	Max	Mean	SD	Min	Max
Total	587	0.86	84.71	15.06	46	118	26.82	4.98	15	36
Verbal Reasoning + Grammar/Writing	35	0.87	66.70	7.78	48	77	61.37	9.87	35	72
Quantitative Reasoning	35	0.75	25.27	7.81	6	37	25.83	5.15	16	36

One obstacle that had to be addressed was the difference in population groups between the three tests. To date, the population group of CLT examinees differs from the population group of ACT and SAT examinees. This is due to the new nature of the test, as well as the CLT's specific appeal to students, families, and schools in the world of non-public education. To overcome this obstacle, the method of linear regression was chosen over an equipercentile linking method. This served to minimize the effect of any difference in the population groups and also accounted for an attenuated range of collected ACT and SAT scores. The results from these regressions were then combined to produce the final concordance tables.

Table 8.3 Descriptive Statistics of the Samples for Concordance Analysis, CLT and SAT

CLT Section	N	Correlation	CLT				SAT			
			Mean	SD	Min	Max	Mean	SD	Min	Max
Total	463	0.83	86.86	14.09	45	118	1284.82	154.81	740	1600
Verbal Reasoning + Grammar/Writing	37	0.77	68.88	5.46	51	79	708.11	58.87	570	790
Quantitative Reasoning	37	0.81	26.89	7.07	12	39	651.89	97.83	410	800

Results

Tables 8.2 and 8.3 below present the n -counts of the sample, the correlation between each test and test section, and descriptive statistics on each test and test section. Within the CLT-ACT concordance sample, the correlation between [CLT Verbal Reasoning + Grammar/Writing] and [ACT English + Reading] was higher than the total concordance and concordance between [CLT Quantitative Reasoning] and [ACT Mathematics]. However, within the SAT concordance sample, the correlation between CLT total and SAT total was higher than the correlation between [CLT Quantitative Reasoning] and [SAT Math], which was in turn higher than the correlation between [CLT Verbal Reasoning + Grammar/Writing] and [SAT Evidence-Based Reading and Writing]. Each sample, however, displayed high correlation between the CLT and the ACT and SAT, to include the correlation between the CLT and its respective sections on the other two tests. The scatter plots indicate linear relationships between each of the tests and sections concorded to one another. These are displayed in Figures 8.1 through 8.2.

Figure 8.1 Scatter Plots of CLT Scores vs. ACT Scores

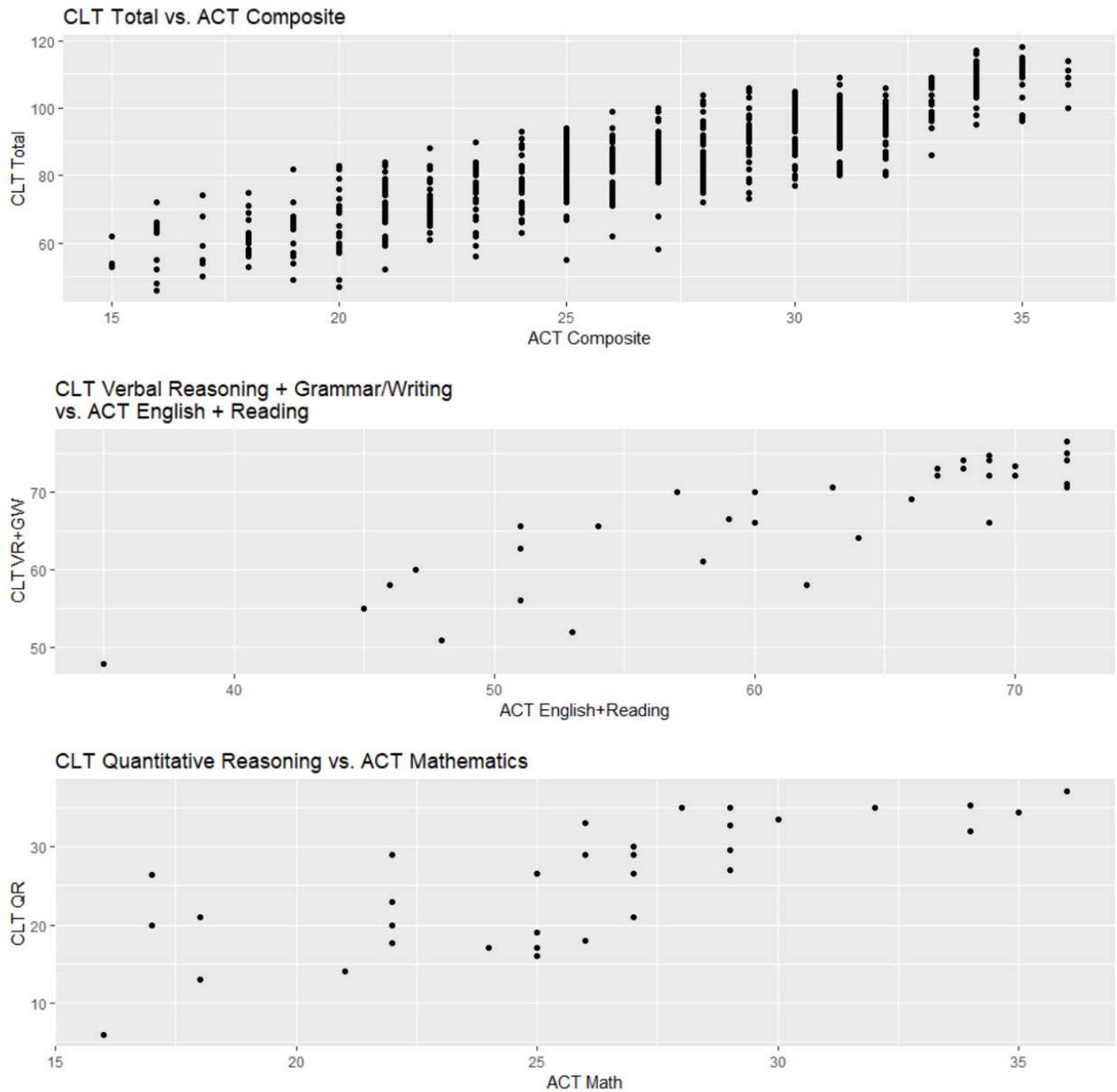
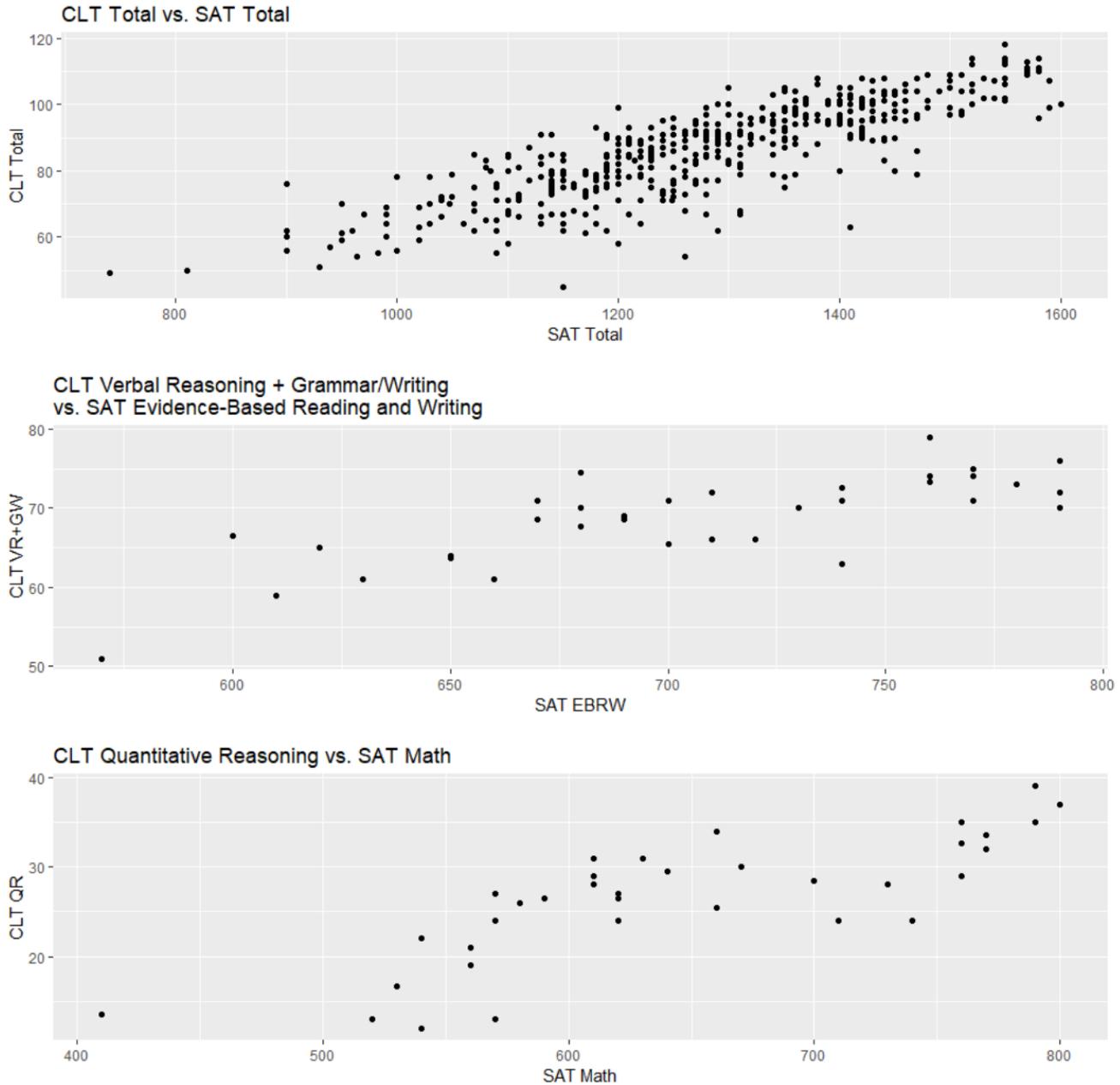


Figure 8.2 Scatter Plots of CLT Scores vs. SAT Scores



Summary

The two sets of correlations (CLT and ACT, CLT and SAT) were examined and then blended with the other to ensure consistency both with CLT's correlations and with the published concordance tables between the ACT and SAT.³ Top ACT and SAT scores are displayed as "N/A" because there are no equivalent ACT or SAT scores which concord with top CLT scores. The CLT provides an opportunity for further differentiation between students who score perfectly on both the ACT and SAT. These concordance tables (on the following pages) can be interpreted as displaying CLT scores in the context of either ACT or SAT scores.

³ Guide to the 2018 ACT® /SAT® Concordance, The College Board, ACT, Inc., 2018, <http://www.act.org/content/dam/act/unsecured/documents/ACT-SAT-Concordance-Information.pdf>.

Table 8.4 CLT Total Score Concordance Table

<i>CLT Total</i>	<i>ACT Composite</i>	<i>SAT Total</i>
120	N/A	N/A
119	N/A	N/A
118	N/A	N/A
117	N/A	N/A
116	N/A	N/A
115	N/A	N/A
114	36	1600
113	36	1590
112	36	1580
111	36	1570
110	35	1560
109	35	1550
108	35	1540
107	35	1530
106	34	1520
105	34	1510
104	34	1500
103	34	1490
102	33	1480
101	33	1470
100	33	1460
99	32	1440
98	32	1430

<i>CLT Total</i>	<i>ACT Composite</i>	<i>SAT Total</i>
97	32	1420
96	31	1410
95	31	1400
94	31	1390
93	30	1380
92	30	1370
91	29	1350
90	29	1340
89	29	1330
88	28	1320
87	28	1310
86	28	1300
85	27	1280
84	27	1260
83	26	1250
82	26	1240
81	26	1230
80	25	1220
79	25	1210
78	25	1200
77	24	1180
76	24	1160
75	23	1140

Table 8.4 CLT Total Score Concordance Table (Continued)

<i>CLT Total</i>	<i>ACT Composite</i>	<i>SAT Total</i>
74	23	1130
73	22	1110
72	22	1100
71	21	1090
70	21	1080
69	21	1070
68	21	1060
67	20	1040
66	20	1030
65	19	1010
64	19	990
63	18	980
62	18	970
61	18	960
60	17	950
59	17	940
58	17	930
57	17	920
56	16	900
55	16	880
54	15	870
53	15	850
52	15	830

<i>CLT Total</i>	<i>ACT Composite</i>	<i>SAT Total</i>
51	14	820
50	14	800
49	14	780
48	13	770
47	13	750
46	13	730
45	12	720
44	12	700
43	11	680
42	11	670
41	11	660
40	11	650
39	10	640
38	10	630

Table 8.5 CLT Verbal Reasoning+Grammar/Writing Concordance Table

<i>CLT VR+GW</i>	<i>ACT E+R</i>	<i>SAT EBRW</i>
80	N/A	N/A
79	N/A	N/A
78	N/A	N/A
77	N/A	N/A
76	72	800
75	72	790
74	72	790
73	72	790
72	71	780
71	71	780
70	71	780
69	71	770
68	71	770
67	70	760
66	70	750
65	68	730
64	66	710
63	64	700
62	61	680
61	60	670
60	58	660
59	57	650
58	55	640

<i>CLT VR+GW</i>	<i>ACT E+R</i>	<i>SAT EBRW</i>
57	54	630
56	51	610
55	48	590
54	45	570
53	44	560
52	43	550
51	42	540
50	40	530
49	39	520
48	38	510
47	38	510
46	37	500
45	37	500
44	35	490
43	35	490
42	34	480
41	33	470
40	26	400
39	25	390
38	24	380
37	23	370
36	22	360
35	21	350
34	20	340
33	19	330
32	18	320
31	17	310
30	16	300
29	14	280

Table 8.6 CLT Quantitative Reasoning Concordance Table

<i>CLT QR</i>	<i>ACT Math</i>	<i>SAT Math</i>
40	N/A	N/A
39	N/A	N/A
38	N/A	N/A
37	N/A	N/A
36	36	800
35	35	780
34	34	760
33	33	740
32	32	720
31	31	710
30	30	700
29	29	680
28	28	660
27	27	640
26	27	630
25	26	620
24	26	610
23	25	600
22	25	590
21	24	570
20	23	550
19	21	530
18	19	510

<i>CLT QR</i>	<i>ACT Math</i>	<i>SAT Math</i>
17	18	500
16	18	490
15	17	480
14	17	470
13	17	460
12	16	450
11	16	440
10	16	420
9	15	400
8	15	380
7	13	340
6	13	320
5	12	300
4	11	280
3	10	260



9. ITEM ANALYSIS

Introduction to Psychometric Evaluation

Item analysis is often conducted to evaluate the psychometric properties of individual items on a test form. In accordance with classical test theory (CTT), items are evaluated in terms of item difficulty and item discrimination.¹ Option or distractor analysis is often conducted to further investigate the discrimination power of the item. Further, differential item functioning is examined to flag items with potential bias.

ITEM DIFFICULTY

In CTT, item difficulty, also known as item p-value, is often quantified in terms of the mean item score or the percentage of students who answer the item correctly. It can be computed using the following equation

$$(9.1) \quad p\text{-value}_j = \frac{\sum_{i=1}^I X_{ij}}{N} =$$

where X_{ij} is the item score for student j on item I , and N is the total number of students answering the item. Theoretically, p-values range from 0 to 1. A higher p-value means that more students have answered the item correctly, and thus the item is judged to be easier. Vice versa, a lower p-value means that fewer students have answered the item correctly, and thus the item is judged to be more difficult. Usually the p-value falls within the range of 0.25 to 0.9. An item which falls outside this range could be very easy ($p\text{-value} > 0.9$) or very difficult ($p\text{-value} < 0.25$). The p-value is group dependent. It may vary as the group ability changes. The p-value may be higher for a high ability group whereas it may get lower for a low ability group. This index is not comparable across different test administrations.

ITEM DISCRIMINATION

Item discrimination is another measure that is often used to quantify the item property with the CTT framework. In general, items that can distinguish high ability from low ability students are considered good

¹ Novick, M.R. (1966) The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3,1-18.

items. Item discrimination can be quantified in terms of the item-total correlation. This index describes the relationship between students' performance on an item and their performance on the overall test. The point-biserial correlation is a specific type of item-total correlation for dichotomous items. The point-biserial correlation can be computed as follows.

$$(9.2) \quad p - bs_j = \text{Corr}(X_{ij}, X_i) = \frac{\text{Cov}(X_{ij}, X_i)}{SD(X_{ij})SD(X_i)}$$

The numerator in the equation is the covariance between the total test scores and the item scores, while the denominator is the product of the standard deviation of the item scores and the test scores respectively. High point-biserial correlations indicates that students who answer an item correctly also have high total test scores and vice versa. There is no consensus related to the cut value for the point-biserial correlation to evaluate item quality. Usually items with point-biserial correlations larger than 0.25 are considered acceptable. Items with a point-biserial correlation smaller than 0.1 are often deemed as not discriminating enough to distinguish between the high and low achievers. Items with a point-biserial correlation between 0.1 and 0.25 are often flagged for further investigation.

OPTION/DISTRACTOR ANALYSIS

Option/distractor analysis shows the percentage of students choosing each option in a dichotomous item. In addition, often the percentage of missing answers is also included in the analysis to get a general idea of the percentage of students who may skip the item. If the missing percentage is larger than 5%, caution should be exercised to scrutinize the item more carefully. This percentage is computed for all students and the subgroups of students. Usually three subgroups are used in such an analysis by dividing students into three categories of ability levels. The total test scores can be used to categorize students into low, middle, and high ability groups. For an item with good discriminating power, it is expected that more high ability students will choose the correct option while the low ability students would be attracted by the distractors which often represent different types of misconceptions.

The following two tables (Tables 9.1 and 9.2) contrast two items. The option highlighted in red indicates the key of the specific item (the correct answer for that question). The numbers in the option point-biserial are the highest values. It is expected that the key option should have the highest option point-biserial correlation. The point-biserial correlation with the item is 0.58 for Item 1 in Table 9.1, while that for Item 2 in Table 9.2 is 0.05. For Item 1, it is evident that the majority of the high and middle ability students chose Option D, which is the key of this item, while the majority of the low ability students chose Options B and C. The p-value of this item is 0.64.

Table 9.1 Option Analysis for an Item with Good Discrimination.

		Option-A	Option-B	Option-C	Option-D	Missing
Item 1	Total	11%	18%	7%	64%	0%
	Low Ability	20%	40%	35%	5%	0%
	Middle Ability	16%	27%	11%	46%	0%
	High Ability	4%	3%	1%	92%	0%
	Option Point-Biserial	-0.17	-0.33	-0.27	0.52	

Table 9.2 Option Analysis for an Item with Low Discrimination.

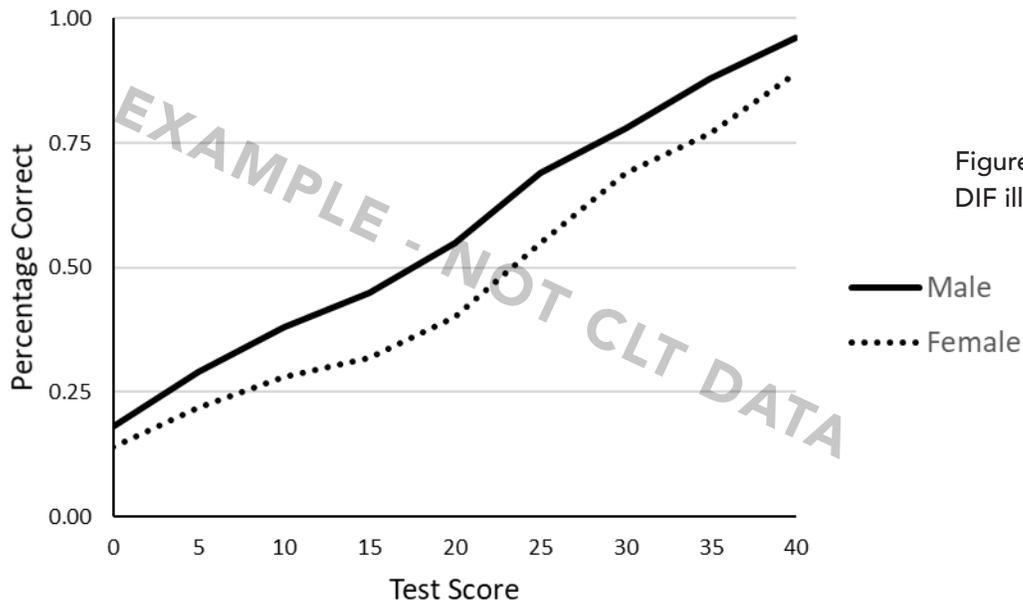
Item		Option-A	Option-B	Option-C	Option-D	Missing
2	Total	21%	22%	51%	5%	1%
	Low Ability	31%	19%	33%	15%	1%
	Middle Ability	23%	23%	49%	5%	0%
	High Ability	13%	22%	62%	2%	1%
Option Point-Biserial		-0.16	-0.02	0.20	-0.13	

On the other hand, the point-biserial correlation for Item 2 is low, thus this item has little discrimination. Regardless of the ability levels, all students are attracted by Option C while Option B is the key. This is reflected in that Option B is the key but has a negative option point-biserial correlation, but Option C has the highest positive option point-biserial correlation. This item is a difficult item with a p-value of 0.22. Such occurrences are rare but not unexpected.

Further, the option point-biserial correlation may present more information about the quality of an item. When computing the option point-biserial correlation, each option would be treated as the correct answer while all other options would be scored as incorrect regardless of the key. It is expected that the correct option will have positive option point-biserial correlation while other options or distractors will have negative option point-biserial correlations. In Table 9.1, the option point-biserial correlations for distractor options A, B, and C are negative while that for the key is positive, as expected. For Item 2 in Table 9.2, the option point-biserial correlation for the correct option is close to 0 while the option point-biserial for distractor Option C is positive, which raises a flag for further investigation.

DIFFERENTIAL ITEM FUNCTIONING

To further investigate the item psychometric properties, differential item functioning (DIF) is conducted. DIF refers to the unexpected differences in students' performance on an item between two groups after they are matched on their ability measured in the test. This conditional comparison is often conducted between two groups; one is called the reference group while the other is called the focal group. The reference group refers to individuals for which the test was expected to favor, most often related to the white or male in educational assessment while a focal group refers to individuals who are at risk or being disadvantaged by the test, most often related to the non-white or female. The following graph illustrates the presence of DIF in an item.



Mantel-Haenszel statistics and ETS categories are two commonly used DIF indexes. Mantel-Haenszel alpha can be computed using equation

$$(9.3) \quad I = \frac{\sum_{k=1}^m \frac{B_k C_k}{T_k} \frac{A_k D_k}{B_k C_k}}{\sum_{k=1}^m \frac{B_k C_k}{T_k}} = \frac{\sum_{k=1}^m \frac{A_k D_k}{T_k}}{\sum_{k=1}^m \frac{B_k C_k}{T_k}}$$

where cell frequencies represented by A, B, C and D in the above equation can be related to Table 9.3.

Table 9.3 Cross-Tabulation Table for Item Responses and Group Membership.

		Item Response Y		
		1	0	Total
Group	Reference	A _k	B _k	nR _k
	Focal	C _k	D _k	nF _k
	Total	n1 _k	n0 _k	T _k

Further, $\hat{\lambda}_{MH} = \ln(\hat{\alpha}_{MH})$ and $D-DIF = -2.35\hat{\lambda}_{MH}$. Items are classified into one of the three DIF categories

Category A: Negligible DIF, no contrast group favored;

Category B: Moderate DIF, one contrast group is slightly favored by the studied item;

Category C: Large DIF, one contrast group is strongly favored by the studied item.

The presence of DIF indicates only that the students with equal ability from different subgroups have an unequal probability of responding to an item correctly. Items in category B and C are flagged and should be carefully examined for potential bias against a particular group.

DIF does not necessarily mean that an item is biased. An item is biased if it measures attributes irrelevant to the intended construct or is somehow a less acceptable measure of the construct for one subgroup. The results of DIF analyses provide a convenient starting point for the study of item bias. Statistical bias does not imply the item is unfair. Expert review of item content is needed.

Summary of Item Analysis Results for CLT Tests

All analyses introduced in the above section were conducted for the April 2018 CLT administrations including test forms 1517 and 1618.

ITEM DIFFICULTY

Item p-values were computed for each item in both forms. Table 9.4 presents the descriptive statistics of p-values for all items in these two forms. The mean p-values ranged from 0.44 to 0.63, with the Quantitative Reasoning test in Form 1618 having the lowest mean p-values. The number of items with p-values smaller than 0.25 are 2 for each section of Form 1517, while those for Form 1618 are 0, 2, and 6 for the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning respectively. On the other hand, there are 4 items in the Grammar/Writing section of Form 1517 and 1 item in the Grammar/Writing section of Form 1618 that have p-values larger than 0.9. All other sections in Form 1517 and 1618 do not have items with p-values larger than 0.9. The p-value for individual items can be found in Appendix A1.

Table 9.4 Descriptive statistics of P-values.

Form	Subject	N	Mean	SD	Minimum	Maximum	P<0.25	p>0.9
1517	Verbal Reasoning	40	0.56	0.16	0.22	0.86	2	0
	Grammar/Writing	40	0.62	0.24	0.13	0.95	2	4
	Quantitative Reasoning	40	0.52	0.17	0.23	0.85	2	0
1618	Verbal Reasoning	40	0.63	0.16	0.30	0.87	0	0
	Grammar/Writing	40	0.58	0.21	0.16	0.94	2	1
	Quantitative Reasoning	40	0.44	0.18	0.15	0.86	6	0

Note: The last two columns represent the number of items in each section in each form that fall in the range of the indicated values.

ITEM DISCRIMINATION

The point-biserial correlations are summarized in Table 9.5. The mean point-biserial correlations ranged from 0.34 to 0.44 with the largest value of 0.62 in the Grammar/Writing section in Form 1618 and the smallest value of 0.01 in the Quantitative Reasoning section in Form 1618. As items with values ranging from 0.1 to 0.25 raise a warning flag, and those with values less than 0.1 also raise a red flag, such information is summarized as well. Only 4 items had a point-biserial correlation smaller than 0.1 with the lowest value of 0.01 in the Quantitative Reasoning section of Form 1618. The point-biserial for individual items can be found in Appendix A2.

Table 9.5 Descriptive statistics of point-biserial correlation.

Form	Subject	N	Mean	SD	Minimum	Maximum	pbs<0.1	0.1<pbs<0.25
1517	Verbal Reasoning	40	0.34	0.10	0.05	0.48	2	5
	Grammar/Writing	40	0.34	0.10	0.09	0.58	1	5
	Quantitative Reasoning	40	0.39	0.09	0.11	0.53	0	2
1618	Verbal Reasoning	40	0.44	0.11	0.20	0.59	0	4
	Grammar/Writing	40	0.41	0.11	0.19	0.62	0	4
	Quantitative Reasoning	40	0.37	0.14	0.01	0.57	1	7

Note: The last two columns represent the number of items in each section in each form that fall in the range of the indicated values.

OPTION/DISTRACTOR ANALYSIS

As illustrated above, option/distractor analysis further demonstrates item performance in different ability groups. For test security reasons, this information cannot be summarized in this document. In general, the information collected in these analyses further cross-validated what has been observed and summarized above and provided more detailed information about which option may be the potential cause for the low discrimination in items that have been flagged. Option/distractor analysis and option point biserial correlations for each individual item can be provided to relevant stakeholders of CLT upon request with the signing of a confidentiality agreement.

DIFFERENTIAL ITEM FUNCTIONING

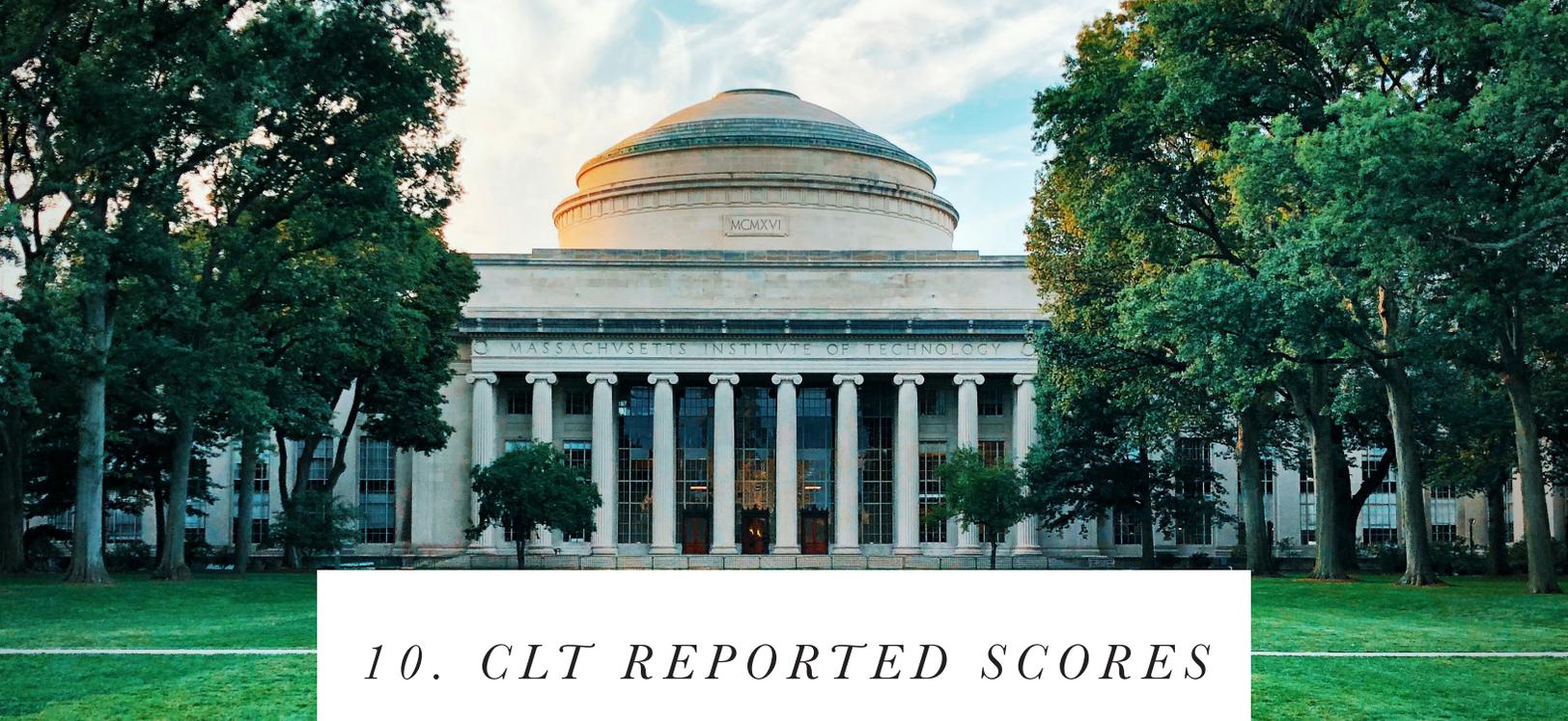
Two types of DIF analyses were conducted, one for gender and the other for race. For gender DIF analyses, the male group was designated as the reference group while the female group as the focal group. For the race

DIF analysis, the White student group was treated as the reference group while the Non-White student group was treated as the focal group. Students with missing group indicators were excluded from DIF analyses. *difR* package (Magis, Beland, Tuerlinckx, & De Boeck, 2010) was used with all the default settings.² The number of items flagged with DIF for each section of the CLT is summarized in Table 9.6. In general, the majority of the items in each section were classified with Category A DIF, which is negligible. Usually, items with Category C DIF require further scrutiny of item content. More items were flagged Category C DIF in Form 1618. Some items favored the reference groups while other items favored the focal groups. The detailed information about DIF analysis results for individual items can be found in Appendix A3 to A8.

Table 9.6 Differential Item Functioning

Form	Subject	Grouping variable	N	Number of Items in Each ETS Category		
				A	B	C
1517	Verbal Reasoning	Gender	40	34	3	3
		Race	40	35	3	2
	Grammar/Writing	Gender	40	33	2	5
		Race	40	29	8	3
	Quantitative Reasoning	Gender	40	35	5	0
		Race	40	28	7	5
1618	Verbal Reasoning	Gender	40	24	6	10
		Race	40	25	10	5
	Grammar/Writing	Gender	40	29	9	2
		Race	40	21	10	9
	Quantitative Reasoning	Gender	40	25	8	7
		Race	40	25	8	7

² Magis, D., Beland, S., Tuerlinckx, F. & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous different item functioning. *Behavior Research Methods*, 42, 847-862.



10. CLT REPORTED SCORES

Overview

The CLT contains three multiple-choice tests, Verbal Reasoning (VR), Grammar/Writing (GW), and Quantitative Reasoning (QR). Test scores on each of the three subject tests as well as the composite score (CLT) of the three tests are reported.

Each subject test consists of 40 multiple-choice items. For each item, one score point is awarded for a correct response; zero score points are given for an incorrect response. The raw score is the number of correct responses; the adjusted score is the reported score that takes item/test difficulty into consideration for adjustment across test forms. The adjusted scores are in the same score range of the raw scores.

The adjusted scores are reported to examinees on a 0-120 scale, and are determined with a method which uses external anchor scores, repeat-person scores, item analysis data, and score distribution data. CLT adjusts each subject section by examining the current test administration item and person scores. These raw scores are then compared with external anchors (ACT and SAT scores), as well as repeat-person scores, taking into account the distributional shapes of the raw scores as compared to previous test administrations. From this information, an adjusted score is produced which places the raw CLT score (section scores and total score) onto a scale that is consistent across test administrations and adjusted for test difficulty.

In this section, descriptive statistics are summarized using adjusted scores by test form, testing mode, subgroup (gender and race), school type, and student family income on the two test administrations in April 2018, Form 1517 and Form 1618. Identifications of gender, race/ethnicity, and family income are based on available self-reported information from examinees. Due to the small number of students in some racial groups (e.g., Hispanic, African American, and American Indian), the analyses are based on two general categories, White and Non-White.

Summary of CLT Scores by Test Form and Testing Mode

Descriptive statistics in Table 10.1 suggest that the average CLT scores are 73.52 (SD=16.73) in Form 1517 and 67.88 (SD=19.92) in Form 1618. The average scores for the three subtests are VR 25.58, GW 25.99, and QR 21.95 (SD=5.77-7.31) in Form 1517, and are VR 25.10, GW 23.05, and QR 19.74 (SD= 7.07-7.92) in Form 1618 for slightly lower average scores and larger SD than Form 1517.

Each test form was delivered in two testing modes, the online and the paper-and-pencil test (PPT) version. Table 10.2 shows the analysis results by form and testing mode. In Form 1517, the average CLT score is higher on the paper version (Mean=75.83) than that on the online version (71.92) with a similar SD (16.49 vs. 16.72). In Form 1618, the average CLT score is higher on the online version with a smaller SD (Mean=71.31; SD=18.64) than that on the paper version (Mean=60.30; SD=20.66). The average scores are slightly different over the three subtests in Form 1517, but more variations are observed in Form 1618.

Table 10.1. Descriptive Statistics of Test Scores by Test Form

Form	Test	N	Range	Min.	Max.	Mean	SD
1517	CLT	780	87	31	118	73.52	16.73
	Verbal Reasoning		31	9	40	25.58	6.32
	Grammar/Writing		33	7	40	25.99	5.77
	Quantitative Reasoning		35	5	40	21.95	7.31
1618	CLT	276	100	8	108	67.88	19.92
	Verbal Reasoning		37	3	40	25.10	7.92
	Grammar/Writing		38	0	38	23.05	7.25
	Quantitative Reasoning		37	2	39	19.74	7.07

Summary of CLT Scores by Subgroup

The results by gender summarized in Table 10.3 suggest a similar performance between females and males in Form 1517 on the average CLT scores (Females Mean=74.68 and SD=16.29; Males Mean=74.73 and SD=17.24), while female students scored higher (Mean=69.06 and SD=17.17) than male students (Mean=66.06; and SD=22.64) in Form 1618. On both forms, female students performed slightly better than male students on Verbal Reasoning and Grammar/Writing, but male students performed better on Quantitative Reasoning.

Table 10.2. Descriptive Statistics of Test Scores by Test Form and Testing Mode

Form	Mode	Test	N	Range	Min.	Max.	Mean	SD
1517	Online	CLT	461	87	31	118	71.92	16.72
		Verbal Reasoning		31	9	40	24.91	6.28
		Grammar/Writing		33	7	40	25.43	5.85
		Quantitative Reasoning		34	5	39	21.58	7.15
	Paper	CLT	319	84	32	116	75.83	16.49
		Verbal Reasoning		28	12	40	26.55	6.26
Grammar/Writing		29		10	39	26.81	5.55	
	Quantitative Reasoning		33	7	40	22.48	7.50	
1618	Online	CLT	190	87	21	108	71.31	18.64
		Verbal Reasoning		35	3	38	26.35	7.45
		Grammar/Writing		33	5	38	24.43	6.68
		Quantitative Reasoning		34	5	39	20.53	6.78
	Paper	CLT	86	99	8	107	60.30	20.66
		Verbal Reasoning		34	6	40	22.34	8.27
Grammar/Writing		38		0	38	19.99	7.55	
	Quantitative Reasoning		36	2	38	17.98	7.42	

Table 10.3. Descriptive Statistics of Test Scores by Test Form and Gender

Form	Gender	Test	N	Range	Min.	Max.	Mean	SD
1517	Female	CLT	311	82	32	114	74.68	16.29
		Verbal Reasoning		30	10	40	26.08	6.19
		Grammar/Writing		29	10	39	27.03	5.44
		Quantitative Reasoning		31	7	38	21.57	7.16
	Male	CLT	270	87	31	118	74.73	17.24
		Verbal Reasoning		29	11	40	25.86	6.33
		Grammar/Writing		33	7	40	25.84	6.08
		Quantitative Reasoning		31	8	39	23.04	7.41
1618	Female	CLT	82	74	32	106	69.06	17.17
		Verbal Reasoning		29	9	38	26.18	6.87
		Grammar/Writing		30	8	38	24.27	6.39
		Quantitative Reasoning		26	7	33	18.61	5.82
1618	Male	CLT	133	100	8	108	66.06	22.64
		Verbal Reasoning		37	3	40	24.04	8.65
		Grammar/Writing		38	0	38	22.01	7.83
		Quantitative Reasoning		37	2	39	20.02	8.16

Table 10.4 summarizes the analysis results by race/ethnicity, White and Non-White students. In Form 1517, for example, the average CLT score is 75.69 (SD=16.04) for White students and 70.85 (SD=18.23) for Non-White students; the former scored higher than the latter in both test forms. On the three subtests in Form 1618, the mean scores for White Students were VR 28.77, GW 26.64, and QR 20.04. For Non-White students in Form 1618, the mean scores on the subtests were VR 20.83, GW 18.97, and QR 16.74.

Summary of CLT Scores by Subgroup and Testing Mode

Descriptive statistics are computed based on adjusted scores by subgroup (e.g., gender and race) and testing mode for each test form. The analysis results summarized in Table 10.5 suggest that, in Form 1517, the average CLT score is higher on the PPT version (Mean=76.99 and SD=16.44) than that on the online version (Mean=73.17 and SD=16.05), as well as on the three subtests, for females; the average CLT scores and the corresponding SD are similar between testing modes for males (Mean=74.43 and SD=17.45; Mean=75.30 and SD=16.91). In Form 1618, statistics are not available due to a very small sample size for females' PPT version. The average CLT score and the three subtests for the online version are higher (Mean=72.14 and SD=22.97) than those on the PPT version (Mean=60.25 and SD=20.89) for the male students.

Table 10.6 presents the analysis results by test form, race, and testing mode. In Form 1517, the average CLT scores (Mean=71.63 and SD=18.58) and the average scores of the three subtests are higher for Non-White students in the online version than those in the PPT version; while the average CLT and subtest scores (Mean=77.78 and SD=15.56) are higher for White students in the PPT version than the online version. Similar results are found for Non-White students in Form 1618. Statistics are not available for White students on the PPT version due to a very small sample size.

Table 10.4. Descriptive Statistics of Test Scores by Race/Ethnicity and Test Form

Form	Race	Test	N	Range	Min.	Max.	Mean	SD
1517	Non-White	CLT	109	82	32	114	70.85	18.38
		Verbal Reasoning	109	29	11	40	23.96	6.66
		Grammar/Writing	109	29	10	39	24.90	6.29
	White	Quantitative Reasoning	109	32	7	39	21.99	7.85
		CLT	444	87	31	118	75.69	16.04
		Verbal Reasoning	444	30	10	40	26.48	6.06
		Grammar/Writing	444	33	7	40	26.88	5.53
		Quantitative Reasoning	444	32	7	39	22.34	7.17
		CLT	100	98	8	106	56.54	18.72
1618	Non-White	Verbal Reasoning	100	34	3	37	20.83	8.07
		Grammar/Writing	100	38	0	38	18.97	6.65
		Quantitative Reasoning	100	35	2	37	16.74	6.41
	White	CLT	107	69	39	108	77.45	17.35
		Verbal Reasoning	107	28	12	40	28.77	5.99
		Grammar/Writing	107	27	11	38	26.64	6.04
Quantitative Reasoning	107	33	6	39	22.04	7.36		

Table 10.5. Descriptive Statistics of Test Scores by Test Form, Gender, and Mode

Form	Gender Mode	Test	N	Range	Min.	Max.	Mean	SD
1517	Female Online	CLT	188	74	40	114	73.17	16.05
		Verbal Reasoning	188	30	10	40	25.37	6.08
		Grammar/Writing	188	27	11	38	26.63	5.37
	Female Paper	Quantitative Reasoning	188	31	7	38	21.17	7.04
		CLT	123	79	32	111	76.99	16.44
		Verbal Reasoning	123	28	12	40	27.17	6.21
	Male Online	Grammar/Writing	123	29	10	39	27.64	5.52
		Quantitative Reasoning	123	31	7	38	22.18	7.33
		CLT	176	87	31	118	74.43	17.45
	Male Paper	Verbal Reasoning	176	29	11	40	25.97	6.23
		Grammar/Writing	176	33	7	40	25.57	6.08
		Quantitative Reasoning	176	31	8	39	22.89	7.50
	Female Online	CLT	94	69	40	109	75.30	16.91
		Verbal Reasoning	94	27	13	40	25.65	6.53
		Grammar/Writing	94	28	10	38	26.33	6.09
	Male Paper	Quantitative Reasoning	94	29	8	37	23.32	7.28
		CLT	79	74	32	106	68.46	16.85
		Verbal Reasoning	79	29	9	38	26.01	6.84
Female Online	Grammar/Writing	79	30	8	38	23.99	6.27	
	Quantitative Reasoning	79	26	7	33	18.46	5.72	
	CLT	65	87	21	108	72.14	22.97	
Male Online	Verbal Reasoning	65	35	3	38	25.83	8.90	
	Grammar/Writing	65	31	6	37	24.43	7.35	
	Quantitative Reasoning	65	34	5	39	21.88	8.39	
Male Paper	CLT	68	99	8	107	60.25	20.89	
	Verbal Reasoning	68	34	6	40	22.32	8.11	
	Grammar/Writing	68	38	0	38	19.69	7.61	
1618	Male Paper	Quantitative Reasoning	68	36	2	38	18.24	7.57

Table 10.6. Descriptive Statistics of Test Scores by Test Form, Race/Ethnicity, and Testing Mode

Form	Race/Mode	Test	N	Range	Min.	Max.	Mean	SD	
1517	Non-White Online	CLT	63	82	32	114	71.63	18.58	
		Verbal Reasoning	63	29	11	40	24.08	6.83	
		Grammar/Writing	63	29	10	39	25.11	6.42	
	Non-White Paper	Quantitative Reasoning	63	32	7	39	22.44	8.00	
		CLT	46	73	38	111	69.78	18.24	
		Verbal Reasoning	46	24	13	37	23.80	6.48	
	White Online	Grammar/Writing	46	23	15	38	24.61	6.18	
		Quantitative Reasoning	46	31	7	38	21.37	7.69	
		CLT	283	87	31	118	74.51	16.22	
	White Paper	Verbal Reasoning	283	30	10	40	26.09	5.99	
		Grammar/Writing	283	33	7	40	26.43	5.53	
		Quantitative Reasoning	283	31	8	39	21.99	7.15	
	1618	Non-White Online	CLT	161	79	32	111	77.78	15.56
			Verbal Reasoning	161	28	12	40	27.16	6.13
			Grammar/Writing	161	29	10	39	27.66	5.45
		Non-White Paper	Quantitative Reasoning	161	31	7	38	22.96	7.18
			CLT	59	81	21	102	58.00	18.36
			Verbal Reasoning	59	34	3	37	21.20	8.35
White Online		Grammar/Writing	59	29	6	35	19.95	6.07	
		Quantitative Reasoning	59	32	5	37	16.85	6.09	
		CLT	41	98	8	106	54.44	19.25	
White Paper	Verbal Reasoning	41	29	6	35	20.29	7.73		
	Grammar/Writing	41	38	0	38	17.56	7.26		
	Quantitative Reasoning	41	31	2	33	16.59	6.92		
White Online	CLT	84	65	43	108	78.40	16.30		
	Verbal Reasoning	84	22	16	38	29.15	5.40		
	Grammar/Writing	84	25	13	38	27.08	5.58		
1618	White Paper	Quantitative Reasoning	84	30	9	39	22.17	7.22	

Summary of CLT Scores by School Type and Family Income

In this section, descriptive statistics are computed based on the adjusted scores by school type and family income of each test form for the three multiple-choice based tests, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning, and the composite CLT scores. Because of small sample size, statistics are reported only for the type of schools with a reasonable number of students taking the exam. Reported school types are Home School, Private School, and Charter School.¹ For the same reason, self-reported family incomes are reported in three general categories: \$50,000 or lower, \$50,001—\$125,000, and \$125,001—\$225,000 or higher. It is important to note that small sample size may result in statistics with large sampling error. Caution should be taken in the interpretations of those statistics, particularly in comparison between or across subgroups.

Table 10.7 presents the descriptive statistics of CLT scores and scores on the three subtests by school type. In Form 1517, the average CLT score is slightly higher for Home School (Mean=74.78 and SD=16.57) than Private School (Mean=73.72 and SD=16.74), but about 10 score points higher than Charter School (Mean=64.27 and SD=17.54). Similar patterns of the CLT scores are observed for the three subtests. In Form 1618, the average CLT score is higher for Home School (Mean=83.77 and SD=17.47) than that for Private School (Mean=66.29 and SD=20.18), as the same is true with the three subtests.

Descriptive statistics are presented in Table 10.8 by family income category of each test form. Among the categories, the family income from \$125,001 to \$225,000 or higher has the highest average CLT score and scores of subtests (Mean=79.01 and SD=15.27), followed by the family income from \$50,001 to \$125,000 in Form 1517. Similar patterns are observed in Form 1618.

¹ Distributions from each school type vary from test to test, but examinees on Form 1618, for example, were distributed as follows: 61% Private School, 28% Charter School, 11% Home School.

Table 10.7. Descriptive Statistics of Test Scores by School Type and Test Form

Form	Type	Test	N	Range	Min.	Max.	Mean	SD
1517	Home School	CLT	86	70	39	109	74.78	16.57
		Verbal Reasoning	86	27	12	39	26.95	6.03
		Grammar/Writing	86	25	13	38	27.23	5.50
	Private School	Quantitative Reasoning	86	30	8	38	20.59	7.09
		CLT	641	87	31	118	73.72	16.74
		Verbal Reasoning	641	31	9	40	25.45	6.33
		Grammar/Writing	641	33	7	40	25.91	5.84
		Quantitative Reasoning	641	33	7	40	22.36	7.27
		CLT	74	87	21	108	64.27	17.54
Charter School	Verbal Reasoning	74	35	3	38	23.30	7.75	
	Grammar/Writing	74	27	9	36	22.15	5.86	
	Quantitative Reasoning	74	34	5	39	18.82	6.40	
1618	Home School	CLT	30	58	49	107	83.77	17.47
		Verbal Reasoning	30	24	16	40	31.80	6.07
		Grammar/Writing	30	22	16	38	29.70	5.68
	Private School	Quantitative Reasoning	30	28	10	38	22.27	7.77
		CLT	164	100	8	108	66.29	20.18
		Verbal Reasoning	164	32	6	38	24.45	7.71
Private School	Grammar/Writing	164	38	0	38	22.12	7.54	
	Quantitative Reasoning	164	35	2	37	19.72	7.13	

Table 10.8. Descriptive Statistics of Test Scores by Family Income and Test Form

Family Income of \$50,000 or Lower							
Form	Test	N	Range	Min.	Max.	Mean	SD
1517	CLT	58	76	32	108	70.02	17.21
	Verbal Reasoning	58	30	10	40	24.55	7.37
	Grammar/Writing	58	26	10	36	25.16	5.67
	Quantitative Reasoning	58	29	7	36	20.31	6.99
1618	CLT	39	85	21	106	61.36	16.85
	Verbal Reasoning	39	30	8	38	23.03	7.20
	Grammar/Writing	39	32	6	38	21.41	6.29
	Quantitative Reasoning	39	23	7	30	16.92	5.61
Family Income of \$50,001 - \$125,000							
1517	CLT	181	76	38	114	76.22	15.74
	Verbal Reasoning	181	28	12	40	26.85	5.51
	Grammar/Writing	181	27	12	39	26.72	5.70
	Quantitative Reasoning	181	31	7	38	22.64	7.15
1618	CLT	53	74	33	107	76.57	19.53
	Verbal Reasoning	53	33	7	40	28.17	7.32
	Grammar/Writing	53	23	12	35	25.66	6.57
	Quantitative Reasoning	53	28	10	38	22.74	7.72
Family Income of \$125,001 - \$225,000 or Higher							
1517	CLT	109	66	45	111	79.01	15.27
	Verbal Reasoning	109	27	13	40	26.90	6.11
	Grammar/Writing	109	21	15	36	27.56	4.89
	Quantitative Reasoning	109	31	8	39	24.55	7.55
1618	CLT	22	66	42	108	77.36	17.42
	Verbal Reasoning	22	20	18	38	29.27	6.40
	Grammar/Writing	22	27	9	36	26.73	6.17
	Quantitative Reasoning	22	24	13	37	21.36	6.59



11.

RELIABILITY & STANDARD ERROR OF MEASUREMENT

Overview

Reliability refers to the consistency in test scores. Reliability coefficients quantify the level of consistency of test scores. Tests with high reliability coefficients provide stable test scores between test forms across occasions. Reliability is a necessary condition for the quality of a test. It is important to establish reliability of test scores through empirical studies so that sound judgments can be made. The reliability of test scores is a function of test content, test length, item difficulty, standard deviation, and student motivation, as well as the procedure for test development, test administration, and scoring.

Reliability coefficients typically range from zero to one, with the values near one indicating high consistency, and values near zero indicating low or no consistency. In classical test theory, reliability is defined as the ratio of the true score variance to the observed score variance, assuming the error variance is the same along the score scale. Reliability coefficients are usually estimated on a single test administration by calculating the inter-item covariances.

Cronbach alpha (1951) is one of the most widely used estimates to test reliability, which can be computed using the formula below:

$$(11.1) \quad \alpha_k = \frac{k}{k-1} \left(1 - \frac{\sum \text{var}(Y_i)}{\text{var}(Y_{tot})} \right)$$

where

k = number of items on the test

$\text{var.}(Y_i)$ = variance of item i

$\text{var.}(Y)$ = total test score variance

Standard error of measurement (SEM) provides another indicator of the accuracy of test scores, which summarizes the amount of errors or inconsistency in test scores of a test. SEM can be computed using the formula below:

$$(11.2) \quad SEM = SD(Y_{tot})\sqrt{1 - \text{reliability}}$$

where

$SD(Y)$ = Standard deviation of the test

Reliability and Standard Error of Measurement by Test Form and Testing Mode

The estimated reliability and standard error of measurement (SEM) for the three multiple-choice tests, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning, and the composite CLT scores are reported in this section with descriptive statistics of minimum and maximum scores, mean, and standard deviation (SD) derived from raw scores.

Table 11.1 provides the reliability coefficients and SEMs for the two test forms operated in April of the 2017-2018 school year, Form 1517 and Form 1618. The reliability is 0.92 with the SEM of 4.76 for CLT for Form 1517; the reliability is 0.95 with the SEM of 4.67 for Form 1618. The reliability coefficients for the three subtests are 0.80-0.86 with the SEMs of 2.56-2.84 for Form 1517; while the reliability coefficients are 0.85-0.89 with the SEMs of 2.63-2.76 for Form 1618. Both forms provide highly reliable CLT scores and moderately highly reliable subtest scores.

Each test form was delivered in two testing modes, the online and the paper-and-pencil test (PPT) versions. Table 11.2 provides descriptive statistics of test scores by test form and testing mode. The reliability coefficients are reasonably high for the CLT scores in the range of 0.92-0.95 with the SEM of 4.64-4.79 across test forms and testing modes. The magnitude of reliability coefficients is in 0.80s for the three subtests, 0.80-0.89 for Verbal Reasoning with the SEM of 2.60-2.84; 0.80-0.89 for Grammar/Writing with the SEM of 2.53-2.68; 0.85-0.87 for Quantitative Reasoning with the SEM of 2.74-2.80.

The results of analyses provide evidence to support highly reliable CLT scores and moderately highly reliable scores for the three subtests. The results also indicate the comparability of test scores in terms of accuracy between the online and the PPT modes within each form.

Table 11.1. Reliability and Standard Error of Measurement of Unadjusted Test Scores by Test Form

Test	Score	N	Min.	Max.	Mean	SD	Reliability	SEM
1517	CLT	780	26	113	68.52	16.73	0.92	4.76
	Verbal Reasoning		6	39	22.58	6.33	0.80	2.84
	Grammar/Writing		6	39	24.99	5.76	0.80	2.56
	Quantitative Reasoning		4	40	20.95	7.30	0.86	2.77
1618	CLT	276	6	106	65.88	19.92	0.95	4.67
	Verbal Reasoning		3	40	25.10	7.92	0.89	2.64
	Grammar/Writing		0	38	23.05	7.25	0.87	2.63
	Quantitative Reasoning		0	37	17.74	7.07	0.85	2.76

Table 11.2. Descriptive Statistics of Unadjusted Test Scores by Test Form and Testing Mode

Test	Mode	Score	N	Min.	Max.	Mean	SD	Reliability	SEM
1517	Total	CLT	780	26	113	68.52	16.73	0.92	4.76
		Verbal Reasoning		6	39	22.58	6.33	0.80	2.84
		Grammar/Writing		6	39	24.99	5.76	0.80	2.56
		Quantitative Reasoning		4	40	20.95	7.30	0.86	2.77
	Online	CLT	461	26	113	66.92	16.72	0.92	4.79
		Verbal Reasoning		6	39	21.92	6.30	0.80	2.85
		Grammar/Writing		6	39	24.43	5.85	0.81	2.58
		Quantitative Reasoning		4	38	20.57	7.14	0.85	2.79
	Paper	CLT	319	27	111	70.83	16.49	0.92	4.72
		Verbal Reasoning		9	38	23.55	6.27	0.80	2.82
		Grammar/Writing		9	38	25.80	5.55	0.79	2.52
		Quantitative Reasoning		6	40	21.49	7.51	0.87	2.74
Total	CLT	276	6	106	65.88	19.92	0.95	4.67	
	Verbal Reasoning		3	40	25.10	7.92	0.89	2.64	
	Grammar/Writing		0	38	23.05	7.25	0.87	2.63	
	Quantitative Reasoning		0	37	17.74	7.07	0.85	2.76	
1618	Online	CLT	190	19	106	69.31	18.64	0.94	4.64
		Verbal Reasoning		3	38	26.35	7.45	0.88	2.60
		Grammar/Writing		5	38	24.43	6.68	0.85	2.61
		Quantitative Reasoning		3	37	18.53	6.78	0.83	2.77
Paper	CLT	86	6	105	58.30	20.66	0.95	4.76	
	Verbal Reasoning		6	40	22.34	8.27	0.89	2.71	
	Grammar/Writing		0	38	19.99	7.55	0.87	2.63	
	Quantitative Reasoning		0	36	15.98	7.42	0.86	2.74	

Reliability and Standard Error of Measurement by Subgroup

The estimated reliability and SEM are computed by subgroup for the three multiple-choice based tests, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning, and the composite CLT scores. It is important to note that the reliability coefficient may not be stable from case to case due to small sample size (e.g., N<300). When the variance for some items becomes zero, those items are removed from the analysis.

Table 11.3 provides the analysis results by gender and test form. Gender identifications are based on the available self-reported information from examinees. Missing identifications are excluded from analysis. The reliability coefficients of the CLT scores are 0.92 with the SEMs of 4.72-4.75 by gender in Form 1517, and 0.93-0.96 with the SEMs of 4.64-4.70 by gender in Form 1618. The reliability coefficients of the three subtests are in the range of 0.79-0.85 for females with the SEMs of 2.50-2.82 in Form 1517, while the reliability coefficients are in the range of 0.80-0.86 for males with the SEMs of 2.50-2.84 in Form 1618. The results of analyses show highly reliable CLT scores and moderately highly reliable subtest scores for both male and female students by test form.

Table 11.3. Descriptive Statistics of Test Scores by Test Form and Gender

Test	Gender	Test	N	Min.	Max.	Mean	SD	Reliability	SEM
1517	Female	CLT	311	27	109	69.68	16.29	0.92	4.72
		Verbal Reasoning		7	38	23.08	6.20	0.79	2.82
		Grammar/Writing		9	38	26.03	5.44	0.79	2.50
		Quantitative Reasoning		6	37	20.57	7.16	0.85	2.78
	Male	CLT	270	26	113	69.73	17.24	0.92	4.75
		Verbal Reasoning		8	39	22.87	6.35	0.80	2.84
		Grammar/Writing		6	39	24.83	6.08	0.82	2.56
		Quantitative Reasoning		7	38	22.03	7.40	0.86	2.73
1618	Female	CLT	82	30	104	67.06	17.17	0.93	4.64
		Verbal Reasoning		9	38	26.18	6.87	0.86	2.59
		Grammar/Writing		8	38	24.27	6.39	0.84	2.59
		Quantitative Reasoning		5	31	16.61	5.82	0.77	2.78
	Male	CLT	133	6	106	64.06	22.64	0.96	4.70
		Verbal Reasoning		3	40	24.04	8.65	0.91	2.65
		Grammar/Writing		0	38	22.01	7.83	0.89	2.64
		Quantitative Reasoning		0	37	18.02	8.16	0.89	2.73

Table 11.4 provides reliability coefficients and SEMs by racial/ethnicity group and test form. Racial/ethnicity identifications are based on the available self-reported information from examinees. Missing identifications are excluded from analysis. Due to the small number of students in some racial groups, such as Hispanic, African American, and American Indian, the analyses are based on two general categories, White and Non-White. In Form 1517, the reliability is 0.91 of CLT scores with the SEM of 4.73 for White students and the reliability of 0.93 with the SEM of 4.79 for Non-White students. In Form 1618, the reliability is 0.93 of CLT scores with the SEM of 4.49 for White students and the reliability is 0.93 with the SEM of 4.81 for Non-White students. For the three subtests, the reliability coefficients for Form 1517 are 0.78-0.85 with the SEMs of 2.51-2.82 for White students and 0.82-0.88 with SEMs in 2.59-2.87 for Non-White students. The reliability coefficients for Form 1618 are 0.82-0.86 with the SEMs of 2.50-2.77 for White students and 0.81-0.89 with SEMs in 2.72-2.79 for Non-White students.

The results of analyses provide evidence to support highly reliable CLT scores and moderately highly reliable subtest scores for both White and Non-White students. The results also indicate the comparability of test scores for internal consistency and accuracy of test scores between the two racial/ethnicity groups across testing modes within each form.

Table 11.4. Descriptive Statistics of Test Scores by Race/Ethnicity and Test Form

Test	Race	Test	N	Min.	Max.	Mean	SD	Reliability	SEM
1517	Non-White	CLT	109	27	109	65.85	18.38	0.93	4.79
		Verbal Reasoning	109	8	39	20.99	6.73	0.82	2.87
		Grammar/Writing	109	9	38	23.90	6.29	0.83	2.59
		Quantitative Reasoning	109	6	38	20.96	7.81	0.88	2.75
	White	CLT	444	26	113	70.69	16.04	0.91	4.73
		Verbal Reasoning	444	7	38	23.48	6.06	0.78	2.82
		Grammar/Writing	444	6	39	25.88	5.53	0.79	2.51
		Quantitative Reasoning	444	6	38	21.34	7.17	0.85	2.77
1618	Non-White	CLT	100	6	104	54.54	18.72	0.93	4.81
		Verbal Reasoning	100	3	37	20.83	8.07	0.89	2.74
		Grammar/Writing	100	0	38	18.97	6.65	0.83	2.72
		Quantitative Reasoning	100	0	35	14.74	6.41	0.81	2.79
	White	CLT	107	37	106	75.45	17.35	0.93	4.49
		Verbal Reasoning	107	12	40	28.77	5.99	0.82	2.52
		Grammar/Writing	107	11	38	26.64	6.04	0.83	2.50
		Quantitative Reasoning	107	4	37	20.04	7.36	0.86	2.71

Reliability and Standard Error of Measurement by Subgroup and Testing Mode

The reliability and SEM are computed by subgroup for the three multiple-choice based tests, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning, and the composite CLT scores. As indicated previously, small sample size may result in unstable reliability of test scores across occasions when the total group is broken down for analysis by subgroup.

The analysis results are summarized in Table 11.5 by test form, gender, and testing mode. In Form 1517, the reliability coefficients of the CLT scores are 0.91-0.92 by testing mode with the corresponding SEM of 4.68-4.74 for females and 0.92-0.93 for males by testing mode with the corresponding SEM of 4.75-4.78. In Form 1618, the reliability is 0.92 with the SEM of 4.59 for females online testing and 0.95-0.96 with the SEMs of 4.59-4.76 for males across testing modes. For the three subtests, the reliability is in the range of 0.78-0.86 with the SEMs of 2.45-2.82 across genders and testing modes for Form 1517; while the reliability is in the range of 0.79-0.87 with the SEMs of 2.53-2.84 for Form 1618.

The results of analyses support the similar level of reliability and accuracy of test scores for both male and female students between different testing modes within each test form.

Table 11.6 provides the reliability and SEMs by test form, racial/ethnicity, and testing mode. In the Form 1517 online version, the reliability coefficient of the CLT scores is 0.92 for White students with the SEM of 4.73 and 0.93 for Non-White students with the SEM of 4.77. The reliability coefficients for the three subtests are 0.78-0.85 with the SEM of 2.51-2.83, respectively, for Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning for White students; while the reliability coefficients are 0.83-0.88 with the SEMs of 2.61-2.84 for Non-White students. In the paper-and-pencil version, the reliability of CLT scores is 0.91 (SEM=4.69) for White students and 0.93 (SEM=4.83) for Non-White students. Similarly, the reliability coefficients of the three subtests are 0.79-0.86 (SEMs=2.50-2.81) for White students and 0.80-0.87 (SEMs=2.57-2.91) for Non-White students.

Similar results are found in Form 1618. For the online version, the reliability of CLT scores is 0.92 (SEM=4.49) for White students and 0.93 (SEM=4.82) for Non-White students. The reliability coefficients for the three subtests are in the range of 0.78-0.86 (SEMs=2.50-2.72) for White Students and 0.79-0.90 (SEMs=2.71-2.80) for Non-White students. The reliability coefficient for the paper-and-pencil version is 0.95 (SEM=4.64) for White students and 0.94 (SEM=4.72) for Non-White students. The reliability coefficients for the three subtests are in the range of 0.89-0.90 (SEMs=2.45-2.66) for White Students and 0.84-0.87 (SEMs=2.72-2.79) for Non-White students.

The results of analyses provide evidence to support the similar level of internal consistency and accuracy of test scores between White and Non-White students across testing mode by test form. Caution should be taken in the interpretation of the results in the comparison of reliability and SEM based on small sample size. For example, the reliability of CLT scores is 0.95 for White students (N=23) on the paper-and-pencil version. Four items are excluded from estimates due to the zero variance.

Table 11.5. Descriptive Statistics of Test Scores by Test Form, Gender, and Testing Mode

Test	Gender Mode	Test	N	Min.	Max.	Mean	SD	Reliability	SEM
1517	Female/Online	CLT	188	35	109	68.17	16.05	0.91	4.73
		Verbal Reasoning	188	7	38	22.37	6.10	0.79	2.82
		Grammar/Writing	188	10	37	25.63	5.37	0.78	2.52
		Quantitative Reasoning	188	6	37	20.16	7.03	0.84	2.79
	Female/Paper	CLT	123	27	106	71.99	16.44	0.92	4.68
		Verbal Reasoning	123	9	37	24.17	6.21	0.80	2.81
		Grammar/Writing	123	9	38	26.64	5.52	0.80	2.45
		Quantitative Reasoning	123	6	37	21.18	7.33	0.86	2.75
	Male/Online	CLT	176	26	113	69.43	17.45	0.93	4.75
		Verbal Reasoning	176	8	39	22.98	6.26	0.79	2.84
		Grammar/Writing	176	6	39	24.57	6.08	0.83	2.53
		Quantitative Reasoning	176	7	38	21.88	7.48	0.87	2.75
Male/Paper	CLT	94	35	104	70.30	16.91	0.92	4.78	
	Verbal Reasoning	94	10	38	22.66	6.56	0.81	2.83	
	Grammar/Writing	94	9	37	25.32	6.08	0.82	2.58	
	Quantitative Reasoning	94	7	36	22.32	7.28	0.86	2.74	
1618	Female/Online	CLT	79	30	104	66.46	16.85	0.924	4.65
		Verbal Reasoning	79	9	38	26.01	6.84	0.855	2.61
		Grammar/Writing	79	8	38	23.99	6.27	0.828	2.60
		Quantitative Reasoning	79	5	31	16.46	5.72	0.762	2.79
	Female/Paper	CLT	3	58	96	83.00	21.66	n/a	n/a
		Verbal Reasoning	3	22	35	30.67	7.51	n/a	n/a
		Grammar/Writing	3	25	35	31.67	5.77	n/a	n/a
		Quantitative Reasoning	3	11	26	20.67	8.39	n/a	n/a
	Male/Online	CLT	65	19	106	70.14	22.97	0.96	4.59
		Verbal Reasoning	65	3	38	25.83	8.90	0.916	2.58
		Grammar/Writing	65	6	37	24.43	7.35	0.875	2.60
		Quantitative Reasoning	65	3	37	19.88	8.39	0.895	2.72
Male/Paper	CLT	68	6	105	58.25	20.89	0.95	4.76	
	Verbal Reasoning	68	6	40	22.32	8.11	0.89	2.71	
	Grammar/Writing	68	0	38	19.69	7.61	0.88	2.68	
	Quantitative Reasoning	68	0	36	16.24	7.57	0.87	2.75	

Table 11.6. Descriptive Statistics of Test Scores by Test Form, Race/Ethnicity, and Testing Mode

Test	Race/Mode	Test	N	Min.	Max.	Mean	SD	Reliability	SEM
1517	Non-White/ Online	CLT	63	27	109	66.63	18.58	0.93	4.77
		Verbal Reasoning	63	8	39	21.13	6.95	0.83	2.84
		Grammar/Writing	63	9	38	24.11	6.42	0.83	2.61
	Non-White/ Paper	Quantitative Reasoning	63	6	38	21.40	7.93	0.88	2.71
		CLT	46	33	106	64.78	18.24	0.93	4.83
		Verbal Reasoning	46	10	34	20.80	6.48	0.80	2.91
	White/ Online	Grammar/Writing	46	14	37	23.61	6.18	0.83	2.57
		Quantitative Reasoning	46	6	37	20.37	7.69	0.87	2.79
		CLT	283	26	113	69.51	16.22	0.92	4.73
	White/ Paper	Verbal Reasoning	283	7	37	23.09	5.99	0.78	2.83
		Grammar/Writing	283	6	39	25.43	5.53	0.79	2.51
		Quantitative Reasoning	283	7	38	20.99	7.15	0.85	2.79
White/ Paper	CLT	161	27	106	72.78	15.56	0.91	4.69	
	Verbal Reasoning	161	9	38	24.16	6.15	0.79	2.81	
	Grammar/Writing	161	9	38	26.66	5.44	0.79	2.50	
1618	Non-White/ Online	Quantitative Reasoning	161	6	37	21.96	7.18	0.86	2.73
		CLT	59	19	100	56.00	18.36	0.93	4.82
		Verbal Reasoning	59	3	37	21.20	8.35	0.90	2.71
	Non-White/ Paper	Grammar/Writing	59	6	35	19.95	6.07	0.80	2.74
		Quantitative Reasoning	59	3	35	14.85	6.09	0.79	2.80
		CLT	41	6	104	52.44	19.25	0.94	4.72
	White/ Online	Verbal Reasoning	41	6	35	20.29	7.73	0.87	2.79
		Grammar/Writing	41	0	38	17.56	7.26	0.86	2.72
		Quantitative Reasoning	41	0	31	14.59	6.92	0.84	2.77
	White/ Paper	CLT	84	41	106	76.40	16.30	0.92	4.49
		Verbal Reasoning	84	16	38	29.15	5.40	0.78	2.52
		Grammar/Writing	84	13	38	27.08	5.58	0.80	2.50
White/ Paper	Quantitative Reasoning	84	7	37	20.17	7.22	0.86	2.72	
	CLT	23	37	105	71.96	20.76	0.95	4.64	
	Verbal Reasoning	23	12	40	27.35	7.74	0.90	2.45	
White/ Paper	Grammar/Writing	23	11	36	25.04	7.42	0.89	2.46	
	Quantitative Reasoning	23	4	36	19.57	8.01	0.89	2.66	

Reliability and Standard Error of Measurement by School Type and Family Income

The reliability and SEM, as well as descriptive statistics, are computed based on the raw scores by school type and family income of each test form for the three multiple-choice based tests, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning, and the composite CLT scores.

The results in Table 11.7 show that the reliability coefficients of CLT scores are 0.92 (SEM=4.69-4.74) for home and private schools in Form 1517 and 0.93-0.96 (SEM=4.28-4.64) for charter, home, and private schools in Form 1618. The reliability coefficients of the three subtests are in the similar range of 0.79-0.86 (SEM=2.52-2.84) in Form 1517 and 0.79-0.88 (SEM=2.10-2.79) in Form 1618 across school types.

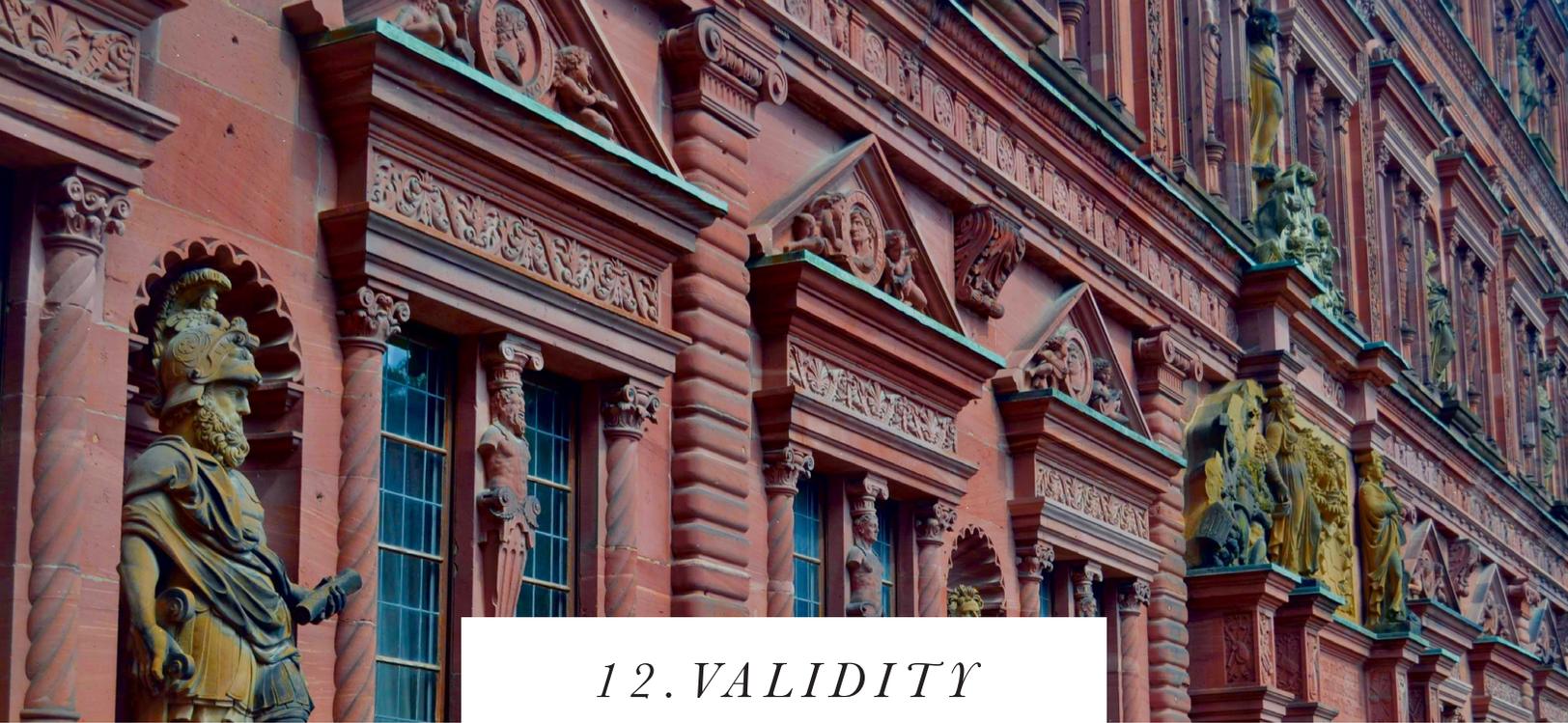
The results of analysis by self-reported family income are summarized in Table 11.8 for each test form. The reliability of CLT scores are 0.91-0.92 (SEM=4.58-4.87) across family income categories in Form 1517 and 0.92-0.95 (SEM=4.37-4.77) in Form 1618. The reliability coefficients for the three subtests are in the range of 0.74-0.87 (SEM=2.49-2.83) in Form 1517 and in the similar range of 0.75-0.88 (SEM=2.39-2.81) across the family income categories, respectively.

Table 11.7. Reliability and Standard Error of Measurement by School Type and Test Form

Form	Type	Test	N	Range	Min.	Max.	Mean	SD	Reliability	SEM
1517	Home School	CLT	86	70	34	104	69.78	16.57	0.92	4.69
		Verbal Reasoning	86	27	9	36	23.95	6.03	0.79	2.76
		Grammar/Writing	86	25	12	37	26.23	5.50	0.79	2.52
		Quantitative Reasoning	86	30	7	37	19.59	7.09	0.84	2.84
	Private School	CLT	641	87	26	113	68.72	16.74	0.92	4.74
		Verbal Reasoning	641	33	6	39	22.45	6.35	0.80	2.84
		Grammar/Writing	641	33	6	39	24.91	5.84	0.81	2.54
		Quantitative Reasoning	641	34	6	40	21.36	7.26	0.86	2.72
	Charter School	CLT	74	87	19	106	62.27	17.54	0.93	4.64
		Verbal Reasoning	74	35	3	38	23.30	7.75	0.88	2.68
		Grammar/Writing	74	27	9	36	22.15	5.86	0.79	2.69
		Quantitative Reasoning	74	34	3	37	16.82	6.40	0.81	2.79
1618	Home School	CLT	30	58	47	105	81.77	17.47	0.94	4.28
		Verbal Reasoning	30	24	16	40	31.80	6.07	0.88	2.10
		Grammar/Writing	30	22	16	38	29.70	5.68	0.83	2.34
	Private School	Quantitative Reasoning	30	28	8	36	20.27	7.77	0.88	2.69
		CLT	164	100	6	106	64.29	20.18	0.95	4.51
		Verbal Reasoning	164	32	6	38	24.45	7.71	0.88	2.67
Private School	Grammar/Writing	164	38	0	38	22.12	7.54	0.88	2.61	
	Quantitative Reasoning	164	35	0	35	17.72	7.13	0.85	2.76	

Table 11.8. Reliability and Standard Error of Measurement by Family Income and Test Form

Form	Test	Family Income of \$50,000 or Lower							
		N	Range	Min.	Max.	Mean	SD	Reliability	SEM
1517	CLT	58	76	27	103	65.02	17.21	0.92	4.87
	Verbal Reasoning	58	31	7	38	21.57	7.41	0.86	2.77
	Grammar/Writing	58	26	9	35	24.14	5.65	0.79	2.59
	Quantitative Reasoning	58	29	6	35	19.31	6.99	0.84	2.79
1618	CLT	39	85	19	104	59.36	16.85	0.92	4.77
	Verbal Reasoning	39	30	8	38	23.03	7.20	0.86	2.69
	Grammar/Writing	39	32	6	38	21.41	6.29	0.82	2.67
	Quantitative Reasoning	39	23	5	28	14.92	5.61	0.75	2.81
Family Income of \$50,001 - \$125,000									
1517	CLT	181	76	33	109	71.22	15.74	0.91	4.72
	Verbal Reasoning	181	30	9	39	23.87	5.55	0.74	2.83
	Grammar/Writing	181	27	11	38	25.72	5.70	0.81	2.48
	Quantitative Reasoning	181	31	6	37	21.62	7.12	0.85	2.76
1618	CLT	53	74	31	105	74.57	19.53	0.95	4.37
	Verbal Reasoning	53	33	7	40	28.17	7.32	0.88	2.54
	Grammar/Writing	53	23	12	35	25.66	6.57	0.85	2.55
	Quantitative Reasoning	53	28	8	36	20.74	7.72	0.88	2.67
Family Income of \$125,001 - \$225,000 or Higher									
1517	CLT	109	66	40	106	74.01	15.27	0.91	4.58
	Verbal Reasoning	109	27	10	37	23.90	6.11	0.79	2.80
	Grammar/Writing	109	21	14	35	26.56	4.89	0.74	2.49
	Quantitative Reasoning	109	31	7	38	23.55	7.55	0.87	2.72
1618	CLT	22	66	40	106	75.36	17.42	0.93	4.61
	Verbal Reasoning	22	20	18	38	29.27	6.40	0.84	2.56
	Grammar/Writing	22	27	9	36	26.73	6.17	0.85	2.39
	Quantitative Reasoning	22	24	11	35	19.36	6.59	0.83	2.72



12. VALIDITY

Overview

The ultimate use of test scores is to draw inferences about students' ability, competence, or behaviors in situations beyond that observed in the testing scenario. The high reliability of test scores implies consistency in test scores but cannot assure that the inferences drawn from test scores are valid and defensible.

Validity is another critical aspect that needs to be addressed in test development and evaluation according to the *Standards for Educational and Psychological Testing*,¹ and is related to fairness. While reliability addresses the consistency in test scores obtained from different forms, administrations, and time, validity addresses whether a test measures what it intends to measure. Validity refers to the degree to which evidence collected in the test scores and in the process of test development and test administration supports the inferences based on test scores as intended.²

According to the *Standards*, validity evidence is collected from the following aspects: content, response process, internal structure, relations with other variables, and consequences of testing. Validity evidence related to test content can be collected based on test specifications, alignment of test content with curriculum, and instruction if relevant for the purpose of the test. Further, test administration and scoring reflect more dimensions for content-related evidence of validity. Chapters 2, 3, 4, and 9 in this technical report provide such content-related evidence of validity.

Response process related evidence of validity can be collected in multiple ways. For example, students taking the test can be interviewed about how they respond to the items. Some think-aloud procedure can help item developers better understand test-takers' thinking and evaluate whether test-takers' thinking is consistent with what the item was intended to be. Further, students' problem-solving strategies could be investigated by observing students' responding behaviors, analyzing process data such as item response time and log files, and the relationship between responses and response process data.

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

2 Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13 - 103). New York: American Council on Education and Macmillan.

Evidence Based on Internal Structure

In addition, the investigation of the internal structure of a test can provide important validity evidence. The internal structure can be evaluated in terms of dimensionality, construct equivalence, measurement precision (in terms of reliability, standard error of measurement, and test information) and differential item functioning. In general, construct equivalence across the subgroups of the test-taker populations and differential item functioning are related to test fairness. Test fairness, as part of validity evidence, means that comparable opportunities have been provided to test-takers to demonstrate their abilities on the constructs a test intends to measure.³ In evaluating test fairness, such questions as whether the test measures the same construct in all relevant populations should be addressed. An investigation of the factor structure of a test and the invariance of the factor structure across subgroups of the student population can provide evidence of construct-related evidence of validity.⁴

To collect another source of evidence of validity, an investigation may be conducted of the relationship among the test scores and other variables such as SAT/ACT scores, high school and college GPA, and graduation. A multitrait-multimethod study will serve this purpose. Further, the influence of the CLT on instruction and school dropout rates can be examined to evaluate the intended and unintended consequences of testing.

The collection of validity evidence is an ongoing process. This technical report provides evidence from different sources in the test development and administration process. This chapter focuses on collecting evidence related to the internal structure of the CLT. It evaluates the internal structure of the three subjects: Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning in two test administrations, Form 1517 and Form 1618. It further investigates the influences of gender and ethnicity on the internal structure of these three subjects in each of the two forms. Both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are conducted using the domain scores.

EVALUATION OF THE INTERNAL STRUCTURE OF CLT USING EFA AND CFA

The evaluation of the internal structure of the CLT is conducted using the domain scores. Table 12.1 summarizes the number of items measuring each domain within each CLT subject test. The number of items assessing each domain across subjects ranged from 10 to 27. The number of items assessing each domain remained the same across the two forms. The descriptive statistics for the domain scores are summarized in Tables 12.2 and 12.3 respectively for Form 1517 and Form 1618. The distributions of the scores for each domain were similar across the two forms.

Table 12.1 The Number of Items Measuring Each Domain

Subject	Domain	Number of Items
Verbal Reasoning	Analysis	13
	Comprehension	27
Grammar/Writing	Grammar	20
	Writing	20
Quantitative Reasoning	Algebra	10
	Geometry	14
	Mathematical Reasoning	16

³ American Educational Research Association, et al., 2014.

⁴ Messick, S. (1995). Standards-based score interpretation: Establishing valid grounds for valid inferences. In *Proceedings of the joint conference on standard setting for large-scale assessments of the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES), Vol. II* (pp. 291-305). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Table 12.2 Descriptive Statistics for the Domain Scores for Form 1517

Domain	N	Minimum	Maximum	Mean	Std. Deviation
verbal_Analysis	780	.00	13.00	6.74	2.37
verbal_Compreh	780	4.00	26.00	15.84	4.54
writing_Grammar	780	3.00	20.00	14.82	3.17
writing_Writing	780	2.00	20.00	10.17	3.20
quan_Algebra	780	.00	10.00	5.45	2.08
quan_Geometry	780	.00	14.00	7.26	3.18
quan_Reasoning	780	.00	16.00	8.23	3.13

Table 12.3 Descriptive Statistics for the Domain Scores for Form 1618

Domain	N	Minimum	Maximum	Mean	Std. Deviation
verbal_Analysis	276	0.00	13.00	7.89	2.76
verbal_Compreh	276	2.00	27.00	17.21	5.53
writing_Grammar	276	0.00	20.00	13.63	3.82
writing_Writing	276	0.00	20.00	9.42	4.04
quan_Algebra	276	0.00	10.00	4.63	2.45
quan_Geometry	276	0.00	13.00	6.38	2.73
quan_Reasoning	276	0.00	15.00	6.73	2.97

CORRELATIONS BETWEEN THE DOMAIN SCORES AND THE SUBJECT TEST SCORES

The correlations between the domain scores and the subject scores for each test form were computed and summarized in Tables 12.4 and 12.5 for Forms 1517 and 1618, respectively. Across forms, similar patterns were observed. In general, the correlation between the Verbal Reasoning scores and the Grammar/Writing scores was higher than those between the Quantitative Reasoning score and the Verbal Reasoning or Grammar/Writing scores. The correlations between the Verbal Reasoning subject scores and the Grammar/Writing domain scores were higher than those between the Verbal Reasoning subject scores and the Quantitative Reasoning domain scores. The correlations between the Grammar/Writing subject scores and the Verbal Reasoning domain scores were higher than those between the Grammar/Writing subject scores and the Quantitative Reasoning domain scores. On the other hand, the correlations between the Quantitative Reasoning subject scores and the Grammar/Writing domain scores and those between the Quantitative Reasoning subject scores and the Verbal Reasoning domain scores were all comparatively lower.

The correlations reported in Tables 12.4 and 12.5 in the box at the right corner are all related to the domain scores. In general, the domain scores from the same subject test tended to be more highly correlated. The same is true for the correlations between the domain score between the Verbal Reasoning and Grammar/Writing subjects. On the other hand, the correlations between the domain score from the Quantitative Reasoning test and those from either the Verbal Reasoning test or the Grammar/Writing test were all lower. The patterns observed across multi-traits fall within expectations and provided both convergent and divergent validity evidence.

Table 12.4 Correlations among the Domain Scores and the Subject Scores for Form 1517

	V_ Orig	W_ Orig	Q_ Orig	V_ Analysis	V_ Compreh	W_ Grammar	W_ Writing	Q_ Algebra	Q_ Geometry	Q_ Reasoning
V_Orig	1									
W_Orig	.746**	1								
Q_Orig	.538**	.583**	1							
V_Analysis	.837**	.610**	.442**	1						
V_Compreh	.958**	.722**	.520**	.646**	1					
W_Grammar	.661**	.904**	.520**	.541**	.640**	1				
W_Writing	.688**	.906**	.535**	.562**	.667**	.638**	1			
Q_Algebra	.443**	.479**	.812**	.353**	.434**	.432**	.436**	1		
Q_Geometry	.451**	.510**	.892**	.364**	.439**	.462**	.461**	.611**	1	
Q_Reasoning	.504**	.524**	.890**	.427**	.480**	.458**	.491**	.611**	.661**	1

** Correlation is significant at the 0.01 level (2-tailed).

Table 12.5 Correlations among the Domain Scores and the Subject Scores for Form 1618

	V_ Orig	W_ Orig	Q_ Orig	V_ Analysis	V_ Compreh	W_ Grammar	W_ Writing	Q_ Algebra	Q_ Geometry	Q_ Reasoning
V_Orig	1									
W_Orig	.805**	1								
Q_Orig	.635**	.663**	1							
V_Analysis	.911**	.726**	.571**	1						
V_Compreh	.979**	.792**	.625**	.806**	1					
W_Grammar	.729**	.917**	.596**	.655**	.718**	1				
W_Writing	.755**	.926**	.626**	.682**	.742**	.700**	1			
Q_Algebra	.533**	.562**	.861**	.470**	.529**	.516**	.520**	1		
Q_Geometry	.559**	.560**	.870**	.510**	.546**	.491**	.541**	.657**	1	
Q_Reasoning	.559**	.601**	.871**	.503**	.550**	.543**	.565**	.622**	.611**	1

EXPLORATORY FACTOR ANALYSIS

Exploratory factor analyses were conducted based on the seven domain scores from the three subject tests for both test forms using the software Statistical Package for the Social Sciences (SPSS). Eigenvalues, eigenvalue differences and the percentage of variance explained by each factor were examined.

Form 1517, Eigenvalues and Eigenvalue Differences

The eigenvalues and the eigenvalue differences between factors for Form 1517 are summarized in Table 12.6. The eigenvalues for the first two factors were larger than 1. Kaiser (1960) recommends extracting the component based on the eigenvalue that is larger than 1.⁵ According to Kaiser's rule, two components were extracted.

Hattie (1985) suggests using the ratio of the difference between the first factor and the second factor to the difference between the second and the third factor to examine the relative strength of the first factor.⁶ This ratio was dubbed as the Factor Difference Ratio Index (FDRI) in Johnson, Yamashiro, and Yu (2003).⁷ If this ratio is larger than 3, the first factor is relatively strong. The eigenvalues for the first factor were larger than 4, and the difference between the first two factors was around 3. Based on this criterion, the first factor was relatively strong. The scree plot presented in Figure 12.1 for Form 1517 in general supports one dominant factor.

Table 12.6 Variance Explained for Form 1517

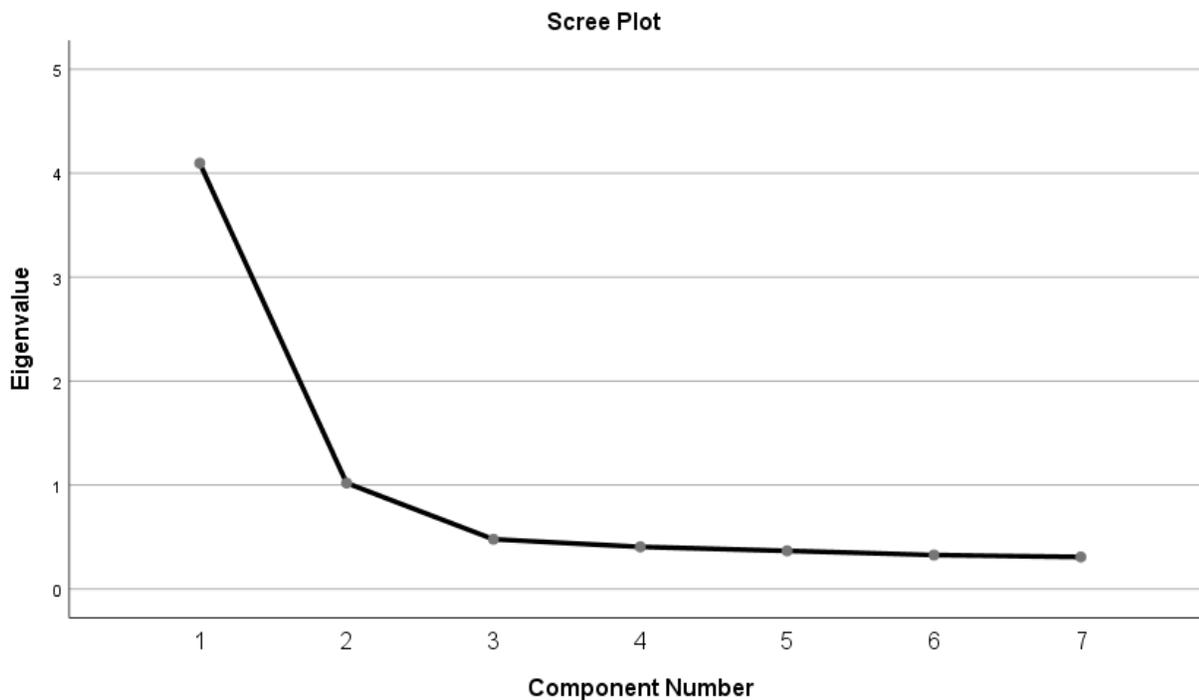
Component	Total	% of Variance	Cumulative %	Differences between Eigenvalues	Factor Difference Ratio Index
1	4.096	58.518	58.518	3.077	5.688
2	1.019	14.558	73.076	0.541	
3	.478	6.826	79.903		
4	.406	5.794	85.696		
5	.367	5.243	90.939		
6	.326	4.663	95.602		
7	.308	4.398	100.000		

5 Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

6 Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164.

7 Johnson, J. S., Yamashiro, A., & Yu, J. (2003). *ECPE annual report: 2002*. Ann Arbor, MI: English Language Institute, University of Michigan.

Figure 12.1. The scree plot for Form 1517 based on domain scores.



Percentage of Variance Explained

The percentage of variance explained by each factor is also presented in Table 12.6. Reckase (1979) suggested that a test is unidimensional if the first factor accounts for 20% or more of the total variance.⁸ For Form 1517, the first component explained over 58.5% variance while the second component explained about 14.6% of the total variance. In general, one dominant component was supported.

Table 12.7 presents the loading of each component. The domain scores in the Verbal Reasoning and Grammar/Writing subjects load positively on component 1 but negatively on component 2. This indicates that component 1 is related to Verbal Reasoning and Grammar/Writing skills. On the other hand, quantitative domain scores load on both component 1 and 2, but with a larger weight on component 1 as well. This indicates that the Quantitative Reasoning domain scores are highly related to both the Verbal Reasoning and the Grammar/Writing components as well.

Table 12.7 Component Matrix

	Component	
	1	2
verbal_Analysis	.727	-.401
verbal_Compreh	.811	-.341
writing_Grammar	.784	-.265
writing_Writing	.801	-.269
quan_Algebra	.716	.470
quan_Geometry	.742	.468
quan_Reasoning	.769	.400

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

⁸ Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4(3), 207-230.

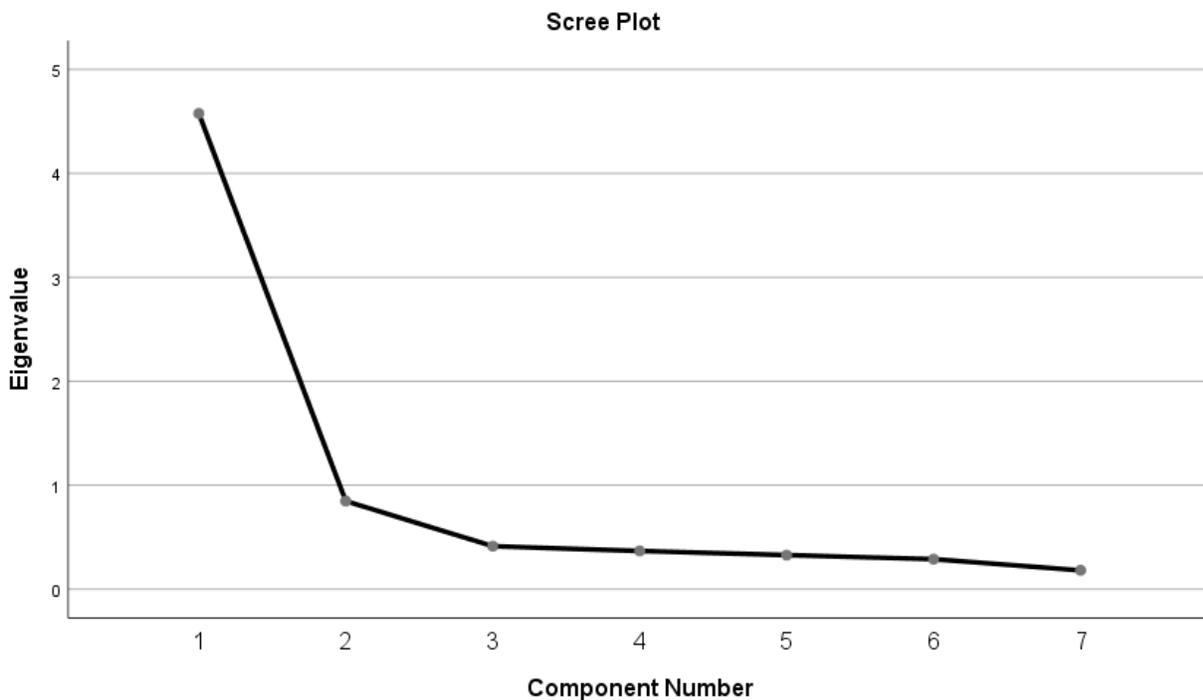
Form 1618, Eigenvalues and Eigenvalue Differences

The eigenvalues and the eigenvalue differences between factors for Form 1618 are summarized in Table 12.8. Only the eigenvalue for the first factor was larger than 1. The eigenvalue for the first factor were larger than 4, and the difference between the first two factors was above 3. According to Hattie (1985), the Factor Difference Ratio Index (FDRI) is larger than 3. The first factor is relatively dominant. Kaiser's rule (Kaiser, 1960) also suggested one component. The scree plot presented in Figure 12.2 for Form 1618 supports one factor.

Table 12.8 Variance Explained for Form 1618

Component	Total	% of Variance	Cumulative %	Differences between Eigenvalues	Factor Difference Ratio Index
1	4.576	65.367	65.367	3.729	8.592
2	.847	12.098	77.465	0.434	
3	.413	5.900	83.365		
4	.368	5.252	88.616		
5	.328	4.681	93.297		
6	.288	4.118	97.415		
7	.181	2.585	100.000		

Figure 12.2. The scree plot for Form 1618 based on domain scores.



Percentage of Variance Explained

The percentage of variance explained by each component for Form 1618 is presented in Table 12.7. The first component accounts for 65.4% of the total variance while the second component explained about 12% of the total variance. The results from EFA supported unidimensionality for Form 1618.

CONFIRMATORY FACTOR ANALYSIS

In general, the results from EFA conducted for Forms 1517 and 1618 were not consistent in identifying the number of factors extracted. In general, the empirical data from Form 1517 supported a two-factor model while that for Form 1618 supported a one-factor model. In theory, CLT was developed to assess students' ability in three content areas: Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning. Thus, three models: a one-factor model, a two-factor model, and a three-factor model were fitted to the domain scores for each form respectively. The three models are presented in Figures 12.3, 12.4, and 12.5, respectively.

Figure 12.3. The One-Factor Model.

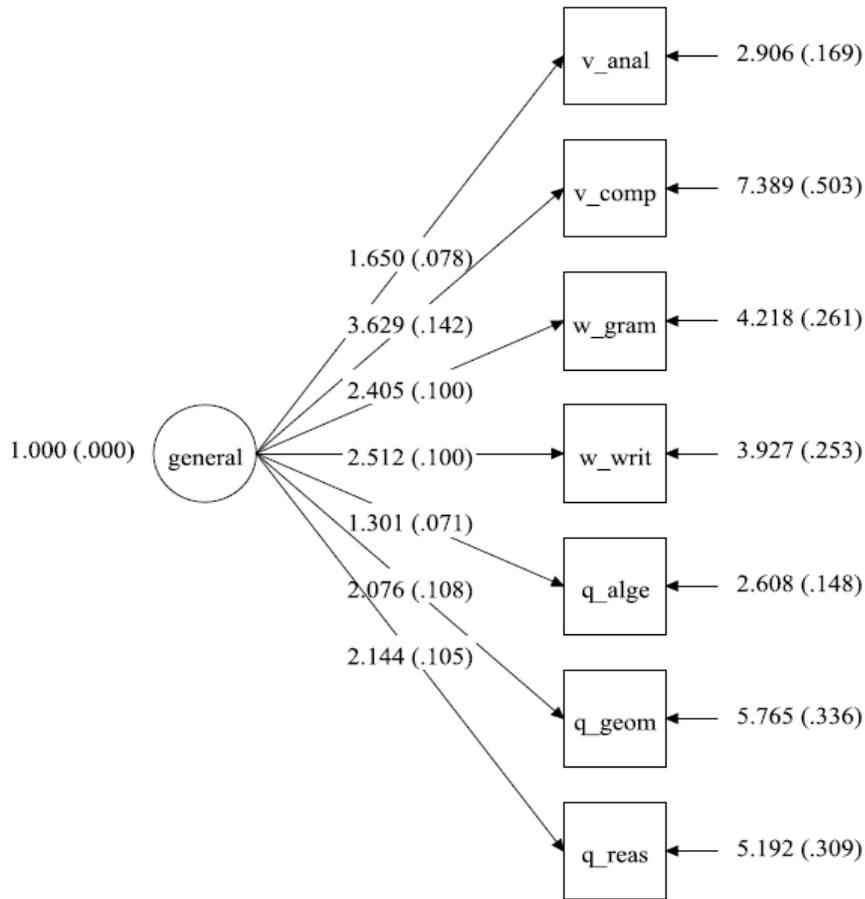


Figure 12.4. The Two-Factor Model.

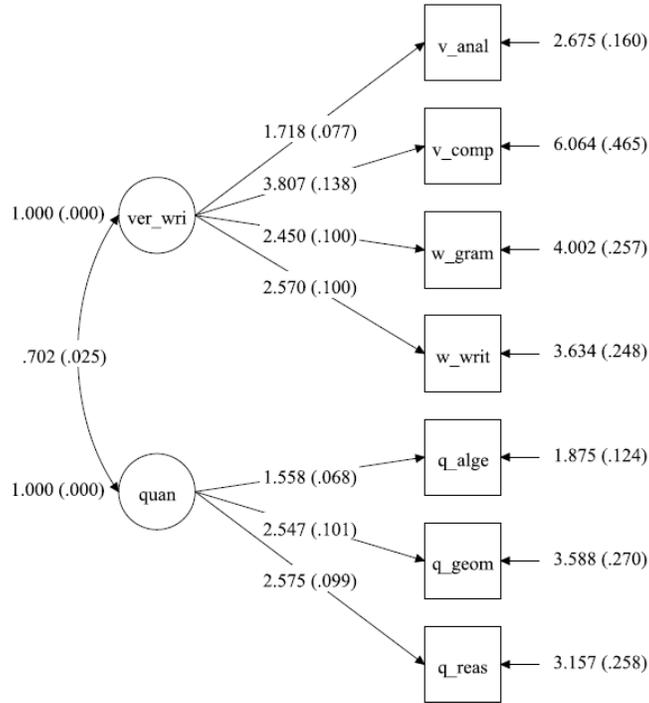
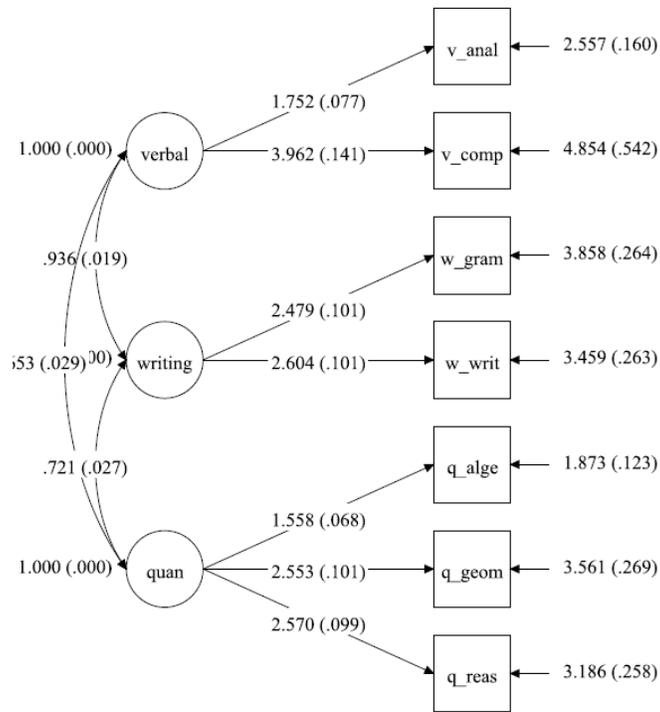


Figure 12.5. The Three-Factor Model.



Several goodness-of-fit indices (GFI) were used to evaluate the model-data fit. These include chi-square (χ^2), the comparative fit index (CFI), the Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA) and the standardized root mean square error residual (SRMR). Based on Hu and Bentler (1999), the cut values for a model with good fit are CFI > 0.95, TLI > 0.95, RMSEA < 0.06, and SRMR < 0.08.⁹ In practice, a rough guideline is that for absolute fit indices and incremental fit indices (such as CFI and TLI), cutoff values could be just above 0.90 and for fit indices based on residual matrices (such as RMSEA and SRMR), values below 0.10 or 0.05 are usually considered adequate. In general, these fit indexes supported the conclusion that both the two-factor and the three-factor models were the best fitting models.

Further, information-based relative fit indexes were also examined. They are Akaike's information criterion (AIC; Akaike, 1974)¹⁰ and Bayesian information criterion (BIC; Schwarz, 1978).¹¹ The best fitting model is the one that minimizes AIC or BIC. The fit indexes are computed as follows.

$$AIC = \overline{D(\xi)} + 2p$$

$$BIC = \overline{D(\xi)} + p \ln$$

where $\overline{D(\xi)}$ is the posterior mean of the deviance, a measure of fit; p is the number of model parameters to be estimated; N is the sample size; and $p_D = \overline{D(\xi)} - D(\hat{\xi})$ which is the difference between the posterior mean of the deviance ($\overline{D(\xi)}$) and the deviance of the posterior model ($D(\hat{\xi})$) based on the posterior estimates of the parameters. Both AIC and BIC as reported in Table 12.9 identified the three-factor model as the best fitting model. Also, the chi-square test as reported in Table 12.9 supported the three-factor model as the best fitting model. This is consistent with the theoretical design of CLT with content coverage in Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning.

Table 12.9 Model Comparison

Form	Model	AIC	BIC	χ^2 (df)	p-value	RMSEA	CFI	TLI	SRMR
1517	1-factor	25229.349	25327.194	396.356 (14)	<.001	0.187	0.859	0.789	0.072
	2-factor	24862.271	24964.776	27.279 (13)	0.011	0.038	0.995	0.992	0.016
	3-factor	24848.073	24959.896	9.081 (11)	0.614	<.001	1.000	1.001	0.009
1618	1-factor	9044.060	9120.088	131.844 (14)	<.001	0.175	0.906	0.858	0.062
	2-factor	8939.704	9019.353	25.488 (13)	0.020	0.059	0.990	0.984	0.023
	3-factor	8925.044	9011.933	6.828 (11)	0.813	<.001	1.000	1.006	0.013

In summary, this section assessed the internal structure of two test forms of CLT, Form 1517 and 1618. It intends to provide validity evidence related to the internal structure of the CLT forms. EFA supported different numbers of components to be extracted. CFA empirically identified the best fitting model, the three-factor model. This is consistent with the theoretical framework and the theoretical content model in the CLT design and development.

9 Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

10 Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

11 Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

Evidence Based on Content

In addition to being technically valid, the content of the CLT also passes a reasonableness test. According to the Standards for Educational and Psychological Testing (2014), “Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.”¹² Parameters for the CLT are designed to ensure that test results yield appropriate indicators of individuals’ capacity for higher-level thinking as well as preparation for college. The range of question types in each of the three test sections provide a reasonable assessment of the kind of knowledge and skills that colleges value.

On the Verbal Reasoning section, questions are broken down into two types: Comprehension questions, which include the subdomains “Passage as a Whole,” “Passage Details,” and “Passage Relationships”; and Analysis questions, which include the subdomains “Textual Analysis” and “Interpretation of Evidence.”

As a result, students are asked to engage with a text on two essential levels: first, their understanding of the text’s meaning, the author’s intent, and the information conveyed by the passage, and second, their ability to analyze and synthesize information in the text to draw valid conclusions. This reflects the multi-level analysis that students are required to engage in during high school, college, and beyond: students are asked not only to assess and comprehend a text, but to draw new ideas and conclusions from it.

On the Grammar/Writing section, questions are also broken down into two types: Grammar questions, which include the subdomains “Agreement” and “Punctuation and Sentence Structure”; and Writing questions, which include the subdomains “Structure,” “Style,” and “Word Choice.”

Here, Grammar questions serve to evaluate a student’s ability to use English standards and conventions properly, so as to clearly convey ideas and information. Writing questions serve to evaluate a student’s ability to use language and style to accurately and appropriately convey the tone, argument, and intent of the text. Both skills are essential for high-level writing: to succeed at the college level, students must not only have a grasp of the conventions required to convey their arguments properly, but also the ability to clearly and concisely communicate their ideas.

On the Quantitative Reasoning section, questions are broken down into three types: Algebra, Geometry, and Mathematical Reasoning. Algebra questions include the subdomains “Arithmetic and Operations” and “Algebraic Expressions and Equations.” Geometry questions include the subdomains “Coordinate Geometry,” “Properties of Shapes,” and “Trigonometry.” Mathematical Reasoning questions include the subdomains “Logic” and “Word Problems.”

The breakdown of Quantitative Reasoning questions into three types mirrors the types of logical reasoning and analysis skills that will serve students well in college and beyond. Algebra questions test students’ ability to understand and work with symbols. Geometry questions test students’ spatial abilities and understanding of shapes such as lines, triangles, squares, and other 2-D and 3-D shapes. Mathematical Reasoning questions test students’ logical abilities. These skills are not only necessary for students interested in pursuing higher-level mathematics or science coursework in college, but are important indications of a student’s ability to think clearly and logically, a crucial skill regardless of academic discipline.

Summary

In terms of both content and internal structure, the CLT exam demonstrates a high level of validity. Analysis of the test’s structure found that a three-factor model is a good fit for evaluating its domain scores, as the exam is composed of three equally weighted subject tests. The types of questions in each subject test correspond to key skills in reading, writing, and mathematics. CLT test scores are thus a legitimate measure of students’ aptitude and preparation for academic work at the college level and beyond.

12 AERA, 2014.

QUALITY CONTROL OF PSYCHOMETRIC ANALYSIS

Previous chapters in this technical report have presented the results of our psychometric analysis of the CLT exam at the item level, the subscore level, the subject level, and the overall test level. We performed all these analyses to collect validity evidence, assess the reliability of the test scores, evaluate and assure fairness, and ultimately to confirm that the CLT test measures what it is intended to measure, and with adequate precision.

Psychometrics is important in the whole process of test development. It plays an important role in the process of test development, beginning with creating test specifications and designing field tests, then in reviewing and analyzing test data. Further, test form construction needs input from psychometrics about the overall psychometric quality of the pulled new test forms. Finally, psychometrics contributes greatly to item analysis, differential item functioning analysis, scoring, calibration, equating, scaling, standard setting, and technical report writing. High quality psychometric work ensures high quality tests. Thus, quality control should be addressed in all these aspects of psychometric work.

The first step in quality control for psychometric work is to develop a process flowchart to standardize the procedure and the steps to follow. The second key element in quality control is to replicate the analysis with different teams. The third key step is to document what has been done, what has been found, and what action has been taken to ensure quality and take correction steps. The CLT has faithfully taken all three measures in the development of the CLT exam.

First, CLT identified all psychometric steps, procedures, and analyses at the very initial stage and during the entire process of CLT test development. This aspect was further reinforced when planning the contents of this technical report. We have had multiple rounds of discussions with the CLT staff to ensure that the right and defensible psychometric steps and procedures would be followed in developing the CLT. Further, through these discussions, we reached agreement as to the type of psychometric analyses that need to be conducted, how a specific psychometric analysis should be carried out, who is responsible for the task, how quality control should be performed on the output from each psychometric analysis, and who is responsible for quality control.

Second, to ensure that the outputs from psychometric analysis for CLT were error-free, our team of psychometricians conducted each analysis independently and compared the results afterwards. We discussed any discrepancy identified in the comparison of our results and explored further to identify the sources of any discrepancy.

This section documents what we did to perform quality control on the psychometric analysis results reported in this document, what we found, and what actions and/or correction steps we and CLT have taken to ensure quality.

Data cleaning is an important step in all psychometric analysis. When CLT delivered data for different CLT test forms to us, we ran basic analyses, such as descriptive statistics, on all variables, including the frequency of values or symbols a variable may take. We ran distributions of all variables to identify any outliers in each variable and in each student data record. The frequency table helped to identify any values or symbols which did not look reasonable. This reasonableness check of each variable and case laid a fundamental base for quality psychometric analysis. We further cleaned the data when needed. We included multiple sources of information in our quality control to ensure the cleanness

and the integrity of the data to be used for psychometric analyses. These include previous analysis data and results, previous technical reports, and other reports for special studies. We identified any potential issues and reported them to the CLT staff. Before beginning our analysis, we as a psychometric team held multiple rounds of discussions to ensure the quality of the data.

During the psychometric analysis stage, one of us led the psychometric task and ran the analysis, summarized the results, put them in the written document, and then shared it with the other, who was responsible for quality control. Two layers of checking were implemented, a within-person check and a between-person level of quality control. For the former layer of defense, each of us was responsible for running quality control of our own work before the other psychometric team member ran a quality control check.

Further, the quality control party independently conducted each analysis. We checked the reported values, and if the results matched, they were used in the final reporting. If a discrepancy was found, the two of us met and discussed potential sources for the discrepancy. We each did one or multiple rounds of screening of what was implemented in our own analysis. We first independently verified that the same standardized steps and procedures had been followed in the analysis. When we had followed standard procedures, we further explored other potential sources for discrepancy. These included different approaches to the exclusion rules applied in data cleaning, variable recoding, ways of dealing with missing values, changes that needed to be made when the analyses were conducted on different data sets, and any other special issues specific to an individual form. This process continued until the exact match was obtained. This within-person and between-person quality control policy established two layers of quality assurance.

As elaborated above, we communicated frequently and followed closely the standardization of the psychometric analysis procedure. We carefully checked and evaluated the reasonableness of the analysis results. Further, some psychometric analyses were automated to avoid human errors. Some analyses were conducted for the total and subgroups of the student population across different parallel forms. Automation helped to remove the errors in manual implementation of such analyses.

In general, to ensure quality control and to increase the efficiency and accuracy in the process of quality control, each step for conducting each psychometric analysis should be standardized and well documented so that replication by independent parties is feasible. Further, the action flowchart for the psychometric analysis, the steps to follow to ensure quality control of the analysis results, and the specifics that need to be controlled should all be streamlined, standardized, and documented as well.

Following these guidelines, we documented quality control results in writing. This helped track potential sources of discrepancy. Such documents are available upon request for relevant stakeholders of CLT who sign a confidentiality agreement. The psychometric analysis in Chapters 9-12 of this report is accompanied by tables with empirical data about the test items and reported scores, and the appendix presents further detailed information.

Hong Jiao, *Ph.D., University of Maryland, College Park*

Liru Zhang, *Ph.D., Independent Consultant*

Appendix

Table A1 Item P-Values

	Form 1517			Form 1618		
	Verbal Reasoning	Grammar/Writing	Quantitative Reasoning	Verbal Reasoning	Grammar/Writing	Quantitative Reasoning
	0.55	0.83	0.83	0.40	0.16	0.75
	0.70	0.82	0.73	0.84	0.80	0.65
	0.41	0.74	0.59	0.58	0.45	0.72
	0.64	0.95	0.83	0.82	0.85	0.86
	0.40	0.31	0.65	0.44	0.94	0.59
	0.86	0.88	0.54	0.77	0.74	0.78
	0.41	0.87	0.61	0.84	0.63	0.63
	0.55	0.83	0.85	0.77	0.51	0.54
	0.78	0.31	0.71	0.75	0.29	0.35
	0.47	0.83	0.68	0.48	0.87	0.57
	0.62	0.78	0.55	0.61	0.87	0.58
	0.68	0.83	0.78	0.67	0.34	0.31
	0.22	0.50	0.65	0.66	0.68	0.16
	0.80	0.41	0.70	0.87	0.39	0.55
	0.72	0.36	0.65	0.67	0.47	0.39
	0.68	0.75	0.71	0.41	0.86	0.54
	0.49	0.64	0.48	0.78	0.38	0.57
	0.43	0.65	0.55	0.42	0.63	0.46
	0.52	0.92	0.60	0.39	0.36	0.47
	0.59	0.21	0.41	0.56	0.80	0.52
	0.49	0.37	0.42	0.81	0.50	0.35
	0.64	0.13	0.53	0.79	0.87	0.46
	0.61	0.81	0.49	0.83	0.42	0.58
	0.24	0.71	0.46	0.50	0.49	0.43
	0.85	0.34	0.45	0.70	0.68	0.35
	0.81	0.41	0.33	0.80	0.56	0.50
	0.38	0.54	0.56	0.78	0.64	0.39
	0.60	0.81	0.41	0.74	0.80	0.39
	0.75	0.95	0.43	0.30	0.47	0.15
	0.39	0.68	0.48	0.49	0.72	0.29
	0.50	0.93	0.52	0.64	0.74	0.36
	0.55	0.47	0.24	0.76	0.41	0.38

Table A1 Item P-Values (Continued)

0.61	0.82	0.40	0.59	0.67	0.37
0.42	0.85	0.23	0.56	0.42	0.17
0.48	0.31	0.40	0.44	0.29	0.16
0.72	0.29	0.27	0.37	0.33	0.22
0.62	0.60	0.28	0.64	0.73	0.30
0.51	0.47	0.43	0.43	0.51	0.39
0.59	0.55	0.25	0.49	0.18	0.22
0.33	0.53	0.27	0.71	0.62	0.26

Table A2 Item Point-Biserial Correlation

Form 1517						Form 1618					
Verbal Reasoning		Grammar/ Writing		Quantitative Reasoning		Verbal Reasoning		Grammar/ Writing		Quantitative Reasoning	
With item	W/o item	With item	W/o item	With item	W/o item	With item	W/o item	With item	W/o item	With item	W/o item
0.42	0.35	0.11	0.05	0.41	0.36	0.22	0.15	0.19	0.14	0.43	0.38
0.38	0.31	0.30	0.24	0.50	0.46	0.33	0.29	0.47	0.43	0.57	0.52
0.37	0.30	0.43	0.36	0.24	0.18	0.35	0.29	0.49	0.43	0.53	0.48
0.30	0.23	0.32	0.29	0.38	0.33	0.46	0.42	0.43	0.38	0.35	0.30
0.21	0.13	0.35	0.27	0.35	0.29	0.34	0.28	0.42	0.39	0.42	0.36
0.34	0.29	0.31	0.26	0.52	0.47	0.49	0.45	0.34	0.28	0.37	0.32
0.38	0.31	0.33	0.27	0.52	0.47	0.46	0.43	0.34	0.28	0.41	0.35
0.29	0.21	0.32	0.26	0.38	0.34	0.51	0.47	0.33	0.27	0.56	0.51
0.32	0.26	0.19	0.11	0.44	0.38	0.43	0.38	0.21	0.15	0.47	0.41
0.09	0.01	0.42	0.37	0.47	0.42	0.52	0.47	0.47	0.43	0.55	0.49
0.27	0.19	0.40	0.34	0.47	0.41	0.38	0.32	0.44	0.41	0.44	0.38
0.43	0.37	0.41	0.35	0.38	0.33	0.30	0.25	0.52	0.47	0.26	0.20
0.05	-0.02	0.45	0.38	0.50	0.44	0.46	0.41	0.62	0.58	0.23	0.18
0.35	0.30	0.39	0.31	0.39	0.33	0.50	0.46	0.45	0.39	0.47	0.41
0.35	0.29	0.25	0.17	0.53	0.48	0.20	0.15	0.36	0.29	0.43	0.38
0.40	0.33	0.41	0.35	0.39	0.33	0.26	0.20	0.51	0.47	0.55	0.49
0.34	0.27	0.58	0.52	0.45	0.40	0.48	0.44	0.29	0.22	0.42	0.36
0.33	0.26	0.39	0.31	0.35	0.29	0.24	0.18	0.44	0.38	0.47	0.41
0.47	0.40	0.39	0.35	0.44	0.38	0.29	0.23	0.45	0.40	0.43	0.37
0.36	0.29	0.14	0.07	0.41	0.35	0.42	0.37	0.54	0.50	0.41	0.35
0.29	0.22	0.33	0.25	0.45	0.40	0.56	0.52	0.46	0.40	0.39	0.33
0.40	0.34	0.34	0.28	0.31	0.25	0.56	0.52	0.52	0.49	0.45	0.39
0.29	0.22	0.31	0.24	0.47	0.42	0.58	0.55	0.34	0.28	0.36	0.30
0.20	0.14	0.09	0.01	0.32	0.26	0.47	0.41	0.36	0.30	0.43	0.37

Table A2 tem Point-Biserial Correlation (Continued)

0.37	0.32	0.27	0.19	0.25	0.19	0.59	0.55	0.48	0.43	0.51	0.46
0.44	0.38	0.35	0.27	0.37	0.32	0.45	0.41	0.42	0.36	0.20	0.13
0.41	0.35	0.35	0.27	0.52	0.46	0.53	0.49	0.45	0.40	0.13	0.06
0.23	0.16	0.36	0.30	0.49	0.44	0.59	0.55	0.49	0.45	0.37	0.31
0.46	0.41	0.29	0.25	0.39	0.33	0.24	0.19	0.43	0.37	0.30	0.25
0.29	0.22	0.38	0.31	0.38	0.32	0.43	0.38	0.56	0.52	0.44	0.38
0.31	0.24	0.34	0.30	0.40	0.34	0.58	0.53	0.46	0.41	0.31	0.24
0.48	0.42	0.53	0.46	0.39	0.34	0.55	0.51	0.38	0.32	0.46	0.40
0.44	0.38	0.39	0.33	0.25	0.19	0.38	0.32	0.33	0.28	0.45	0.39
0.19	0.12	0.40	0.35	0.33	0.28	0.46	0.40	0.37	0.31	0.17	0.12
0.24	0.16	0.24	0.17	0.34	0.28	0.50	0.45	0.19	0.13	0.01	-0.04
0.43	0.37	0.30	0.23	0.41	0.36	0.48	0.43	0.27	0.21	0.14	0.09
0.32	0.25	0.39	0.31	0.29	0.23	0.54	0.49	0.54	0.50	0.28	0.22
0.45	0.39	0.24	0.16	0.11	0.04	0.39	0.34	0.22	0.15	0.45	0.39
0.38	0.31	0.46	0.39	0.33	0.27	0.55	0.50	0.27	0.22	0.12	0.06
0.40	0.34	0.42	0.35	0.26	0.20	0.51	0.47	0.50	0.44	0.22	0.16

Table A3 DIF Analysis Results for Form 1517 Verbal Reasoning Section

Gender			Race		
alphaMH	deltaMH	ETS Category	alphaMH	deltaMH	ETS Category
1.4305	-0.8414	A	1.1389	-0.3057	A
0.9892	0.0255	A	0.9213	0.1926	A
1.215	-0.4576	A	1.0086	-0.0202	A
1.1934	-0.4156	A	0.7572	0.6537	A
1.1698	-0.3686	A	0.9856	0.0342	A
1.1322	-0.2917	A	1.9771	-1.6018	C
1.595	-1.0972	B	0.8283	0.4426	A
1.059	-0.1347	A	0.8816	0.2961	A
1.0809	-0.1828	A	1.5308	-1.0006	B
0.9629	0.0888	A	0.9017	0.2433	A
0.852	0.3764	A	1.253	-0.5301	A
0.7878	0.5606	A	1.1434	-0.3149	A
0.7489	0.6796	A	1.1613	-0.3514	A
1.0412	-0.0948	A	1.2483	-0.5212	A
1.3593	-0.7214	A	1.0061	-0.0142	A
0.7054	0.8202	A	0.9211	0.1932	A
1.2677	-0.5575	A	0.9385	0.1493	A
1.9861	-1.6125	C	1.042	-0.0966	A
1.2537	-0.5313	A	0.7747	0.6	A
1.1798	-0.3885	A	1.1692	-0.3673	A
0.7658	0.627	A	1.1017	-0.2276	A
0.8293	0.44	A	0.6192	1.1264	B
1.0072	-0.0169	A	0.9576	0.1018	A
0.818	0.4722	A	1.934	-1.5501	C
0.4669	1.7898	C	0.9634	0.0876	A
0.4997	1.6305	C	1.2571	-0.5376	A
0.8349	0.424	A	0.9971	0.0068	A
0.9458	0.1308	A	1.1923	-0.4134	A
1.3185	-0.6498	A	0.6595	0.9783	A
0.8579	0.3601	A	0.666	0.9552	A
0.7305	0.738	A	1.339	-0.686	A
1.1212	-0.2688	A	0.656	0.9909	A
0.7476	0.6837	A	1.2218	-0.4708	A
1.0243	-0.0564	A	0.6081	1.1691	B
0.8389	0.4129	A	1.2569	-0.5374	A

Table A3 DIF Analysis Results for Form 1517 Verbal Reasoning Section (Continued)

0.634	1.0709	B	1.044	-0.1011	A
0.9366	0.1538	A	0.9363	0.1546	A
1.5696	-1.0595	B	0.893	0.2659	A
1.3334	-0.6762	A	1.0805	-0.1818	A
1.2597	-0.5426	A	1.4491	-0.8717	A

Table A4 DIF Analysis Results for Form 1517 Grammar/Writing Section

Gender			Race		
alphaMH	deltaMH	ETS Category	alphaMH	deltaMH	ETS Category
1.1389	-0.3057	A	1.3176	-0.6481	A
0.9213	0.1926	A	0.6775	0.915	A
1.0086	-0.0202	A	0.7246	0.757	A
0.7572	0.6537	A	1.1762	-0.3814	A
0.9856	0.0342	A	0.6046	1.1824	B
1.9771	-1.6018	C	1.0526	-0.1204	A
0.8283	0.4426	A	1.1105	-0.2463	A
0.8816	0.2961	A	0.6503	1.0113	B
1.5308	-1.0006	B	1.0557	-0.1273	A
0.9017	0.2433	A	0.8838	0.2904	A
1.253	-0.5301	A	0.8318	0.4328	A
1.1434	-0.3149	A	1.2119	-0.4517	A
1.1613	-0.3514	A	1.0665	-0.1513	A
1.2483	-0.5212	A	1.1641	-0.3571	A
1.0061	-0.0142	A	1.7181	-1.2718	B
0.9211	0.1932	A	1.1465	-0.3213	A
0.9385	0.1493	A	1.6481	-1.1741	B
1.042	-0.0966	A	0.5201	1.5361	C
0.7747	0.6	A	1.5776	-1.0713	B
1.1692	-0.3673	A	1.1625	-0.3538	A
1.1017	-0.2276	A	0.9487	0.1237	A
0.6192	1.1264	B	0.8702	0.3268	A
0.9576	0.1018	A	1.3163	-0.6459	A
1.934	-1.5501	C	1.9295	-1.5446	C
0.9634	0.0876	A	1.1336	-0.2948	A
1.2571	-0.5376	A	1.2902	-0.5987	A
0.9971	0.0068	A	0.7752	0.5985	A
1.1923	-0.4134	A	1.52	-0.984	A
0.6595	0.9783	A	1.597	-1.1001	B
0.666	0.9552	A	1.1231	-0.2729	A

Table A4 DIF Analysis Results for Form 1517 Grammar/Writing Section (Continued)

1.339	-0.686	A	1.4328	-0.845	A
0.656	0.9909	A	0.7711	0.6107	A
1.2218	-0.4708	A	1.0069	-0.0162	A
0.6081	1.1691	B	0.6252	1.1037	B
1.2569	-0.5374	A	0.8579	0.3601	A
1.044	-0.1011	A	0.9204	0.195	A
0.9363	0.1546	A	0.7715	0.6097	A
0.893	0.2659	A	0.4956	1.6496	C
1.0805	-0.1818	A	0.62	1.1234	B
1.4491	-0.8717	A	1.4642	-0.8961	A

Table A5 DIF Analysis Results for Form 1517 Quantitative Reasoning Section

Gender			Race		
alphaMH	deltaMH	ETS Category	alphaMH	deltaMH	ETS Category
1.0027	-0.0063	A	0.8922	0.268	A
1.0293	-0.0679	A	0.5632	1.3494	B
0.8135	0.485	A	1.4351	-0.8489	A
1.0728	-0.1652	A	1.9717	-1.5954	C
0.5559	1.3799	B	2.8763	-2.4827	C
1.0701	-0.1592	A	1.5547	-1.0371	B
0.7595	0.6464	A	0.8491	0.3845	A
0.9161	0.2059	A	1.3925	-0.7782	A
0.7792	0.5862	A	0.7587	0.6491	A
1.1958	-0.4201	A	1.734	-1.2935	B
1.2729	-0.567	A	0.5783	1.2871	B
1.5761	-1.0692	B	0.6952	0.8543	A
0.9324	0.1644	A	1.074	-0.1678	A
1.3201	-0.6526	A	1.5601	-1.0452	B
1.3791	-0.7553	A	1.9464	-1.5651	C
1.3395	-0.687	A	0.7444	0.6936	A
0.8502	0.3815	A	0.9762	0.0565	A
0.7524	0.6685	A	1.0557	-0.1274	A
0.7443	0.6941	A	1.2312	-0.4888	A
0.7519	0.67	A	0.74	0.7076	A
1.4137	-0.8135	A	1.3779	-0.7534	A
0.9289	0.1733	A	0.6881	0.8784	A
1.1229	-0.2724	A	1.3372	-0.6828	A
0.9659	0.0815	A	0.9666	0.0798	A
0.8455	0.3944	A	0.7984	0.5292	A

Table A5 DIF Analysis Results for Form 1517 Quantitative Reasoning Section (Continued)

0.8596	0.3556	A	1.0046	-0.0107	A
1.0938	-0.2107	A	0.7123	0.7972	A
0.871	0.3245	A	2.0565	-1.6944	C
1.2835	-0.5865	A	0.7696	0.6154	A
1.5732	-1.0649	B	1.2317	-0.4898	A
1.3709	-0.7414	A	2.6321	-2.2743	C
1.2308	-0.4881	A	1.142	-0.3121	A
0.8966	0.2565	A	0.6847	0.89	A
1.0847	-0.1911	A	1.0318	-0.0735	A
0.9453	0.1322	A	0.8932	0.2654	A
1.4825	-0.9252	A	1.7362	-1.2965	B
0.6528	1.0021	B	0.8438	0.3992	A
0.9514	0.1172	A	0.9955	0.0106	A
1.7006	-1.2477	B	1.5542	-1.0362	B
0.7782	0.5892	A	0.7605	0.6433	A

Table A6 DIF Analysis Results for Form 1618 Verbal Reasoning Section

alphaMH	Gender		alphaMH	Race	
	deltaMH	ETS Category		deltaMH	ETS Category
3.0271	-2.6029	C	0.5732	1.3077	B
1.0031	-0.0073	A	0.9014	0.2438	A
1.5475	-1.0261	B	1.4914	-0.9393	A
0.7213	0.7677	A	0.842	0.404	A
1.4026	-0.795	A	1.2444	-0.5139	A
0.5237	1.5199	C	0.6215	1.1177	B
2.5116	-2.1642	C	2.54	-2.1906	C
0.6276	1.0946	B	0.5499	1.4052	B
0.9509	0.1184	A	0.8065	0.5054	A
1.4906	-0.9381	A	1.6197	-1.1332	B
0.9445	0.1341	A	0.539	1.4525	B
1.0527	-0.1207	A	0.7076	0.8129	A
0.9299	0.1708	A	0.3023	2.8114	C
1.3556	-0.7149	A	0.1673	4.202	C
0.7366	0.7184	A	1.013	-0.0303	A
0.5936	1.2257	B	1.1856	-0.4	A
0.5731	1.3083	B	0.6757	0.9211	A
1.6476	-1.1735	B	1.8741	-1.4761	B
2.1152	-1.7605	C	1.0703	-0.1597	A
1.3057	-0.6268	A	0.5733	1.3074	B

Table A6 DIF Analysis Results for Form 1618 Verbal Reasoning Section (Continued)

0.2848	2.9515	C	0.8015	0.52	A
2.0925	-1.7351	C	1.7242	-1.2802	B
0.9624	0.09	A	0.531	1.4873	B
0.4946	1.6544	C	0.8378	0.4158	A
0.6859	0.8859	A	0.4568	1.8412	C
1.4005	-0.7915	A	1.6725	-1.2087	B
0.9287	0.1739	A	2.7435	-2.3717	C
2.4245	-2.0812	C	1.2503	-0.525	A
1.0677	-0.1539	A	1.4931	-0.942	A
1.5514	-1.0321	B	1.5252	-0.9921	A
1.2059	-0.44	A	0.8429	0.4018	A
0.8722	0.3214	A	0.6921	0.8647	A
0.4491	1.8813	C	0.879	0.3032	A
0.9935	0.0153	A	1.0909	-0.2044	A
1.0858	-0.1935	A	1.2056	-0.4394	A
0.6944	0.857	A	1.021	-0.0489	A
0.7182	0.7779	A	1.1004	-0.2248	A
1.277	-0.5745	A	0.9167	0.2045	A
0.702	0.8316	A	1.4702	-0.9057	A
0.4426	1.9156	C	1.4779	-0.918	A

Table A7 DIF Analysis Results for Form 1618 Grammar/Writing Section

Gender			Race		
alphaMH	deltaMH	ETS Category	alphaMH	deltaMH	ETS Category
0.831	0.4349	A	0.5765	1.2944	B
1.6442	-1.1685	B	1.7591	-1.3273	B
0.8005	0.5228	A	0.8485	0.386	A
0.9566	0.1043	A	0.9036	0.2383	A
0	Inf	C	1.4882	-0.9342	A
0.8354	0.4227	A	0.9645	0.085	A
1.0382	-0.0881	A	1.0469	-0.1077	A
1.7432	-1.3059	B	1.3079	-0.6308	A
0.8858	0.2848	A	0.9669	0.0791	A
0.7589	0.6484	A	2.4539	-2.1096	C
0.8208	0.464	A	0.9102	0.2211	A
1.3712	-0.7419	A	1.7351	-1.295	B
1.3675	-0.7355	A	1.5467	-1.0248	B
0.8645	0.3423	A	2.7824	-2.4048	C

Table A7 DIF Analysis Results for Form 1618 Grammar/Writing Section (Continued)

1.6402	-1.1628	B	0.5795	1.2819	B
0.7237	0.76	A	1.6687	-1.2032	B
0.529	1.4964	B	0.4888	1.682	C
1.1391	-0.3061	A	0.9407	0.1435	A
0.6979	0.8454	A	1.9592	-1.5804	C
1.3211	-0.6544	A	0.6823	0.8985	A
1.0115	-0.0269	A	0.7114	0.8002	A
0.7435	0.6964	A	1.0877	-0.1976	A
0.8802	0.2999	A	1.1789	-0.3869	A
1.538	-1.0116	B	0.3911	2.2062	C
1.3516	-0.708	A	0.8205	0.4649	A
0.4832	1.709	C	3.7089	-3.0802	C
0.8842	0.2891	A	0.616	1.1388	B
0.7396	0.709	A	1.371	-0.7416	A
1.8932	-1.4999	B	0.5829	1.2684	B
1.4484	-0.8706	A	1.7375	-1.2982	B
0.8521	0.376	A	2.6104	-2.2548	C
0.8271	0.4462	A	1.4177	-0.8202	A
1.2819	-0.5837	A	1.0454	-0.1044	A
1.1362	-0.3	A	0.7952	0.5384	A
0.9601	0.0958	A	0.4658	1.7952	C
0.858	0.3599	A	1.552	-1.0329	B
0.5618	1.355	B	4.6982	-3.6359	C
1.8096	-1.3938	B	1.5255	-0.9925	A
0.8334	0.4284	A	1.114	-0.2536	A
0.5883	1.2469	B	0.9305	0.1694	A

Table A8 DIF Analysis Results for Form 1618 Quantitative Reasoning Section

Gender			Race		
alphaMH	deltaMH	ETS Category	alphaMH	deltaMH	ETS Category
0.7508	0.6736	A	2.009	-1.6394	C
2.005	-1.6348	C	0.7488	0.6798	A
0.887	0.2818	A	0.8757	0.3119	A
1.3673	-0.7351	A	1.588	-1.0868	B
0.9254	0.1821	A	0.6411	1.0446	B
0.6681	0.9478	A	3.3416	-2.8352	C
0.3066	2.778	C	1.7254	-1.2818	B
0.9055	0.2332	A	0.6961	0.8514	A

Table A8 DIF Analysis Results for Form 1618 Quantitative Reasoning Section (Continued)

1.6295	-1.1474	B	0.6394	1.0509	B
0.8452	0.3953	A	0.7345	0.7252	A
1.8205	-1.4079	B	0.9635	0.0874	A
1.578	-1.0719	B	0.5439	1.4311	B
1.2447	-0.5145	A	1.1529	-0.3344	A
1.1559	-0.3405	A	1.0064	-0.015	A
2.0012	-1.6303	C	1.3099	-0.6343	A
1.1408	-0.3096	A	1.8889	-1.4946	B
0.3876	2.227	C	1.2276	-0.4819	A
1.9803	-1.6057	C	0.312	2.7373	C
0.6777	0.9142	A	0.7348	0.7242	A
1.1809	-0.3907	A	1.5522	-1.0332	B
0.5706	1.3185	B	1.0215	-0.05	A
0.811	0.4922	A	1.3854	-0.766	A
0.597	1.2122	B	1.9794	-1.6046	C
0.727	0.7492	A	1.2425	-0.5102	A
2.1021	-1.7459	C	0.9616	0.0919	A
0.786	0.566	A	0.8968	0.2561	A
0.7879	0.5601	A	0.6062	1.1761	B
0.9008	0.2456	A	1.1905	-0.4099	A
1.1463	-0.3209	A	0.2171	3.5889	C
1.0946	-0.2125	A	0.6594	0.9785	A
0.6938	0.8591	A	1.1114	-0.2482	A
0.5786	1.2858	B	0.7625	0.6373	A
1.4837	-0.9271	A	0.8467	0.3911	A
0.793	0.5449	A	1.2692	-0.5602	A
1.5788	-1.0731	B	0.7123	0.7973	A
2.2032	-1.8563	C	1.0613	-0.1398	A
0.9395	0.1466	A	0.452	1.8659	C
0.7317	0.7343	A	1.1084	-0.2418	A
0.6753	0.9227	A	0.3558	2.4283	C
1.5584	-1.0426	B	1.06	-0.1369	A