



CLT

2021 TECHNICAL REPORT

The Classic Learning Test

SEVENTH AND EIGHTH GRADE



2021 TECHNICAL REPORT
The Classic Learning Test
SEVENTH AND EIGHTH GRADE

CONTRIBUTORS

The following individuals helped to create this technical report. We greatly appreciate all their good and thoughtful work to provide this reference on the Classic Learning Test.

<i>Technical Manual Lead</i>	Tracy Gardner, Ph.D.	Stephanie Beattie
<i>Chapter Leads</i>	Stephanie Beattie	Natasha P. Wilson, Ph.D.
	Kimberly Farley	<i>Leadership Reviewers</i>
	Tracy Gardner, Ph.D.	Tracy Gardner, Ph.D.
	Joseph Loomis	Noah Tyler
	Mary Trent	<i>Graphic Designer</i>
	Natasha P. Wilson, Ph.D.	Meg Prom
<i>Writers/Reviewers</i>	Gabriel Blanchard	

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

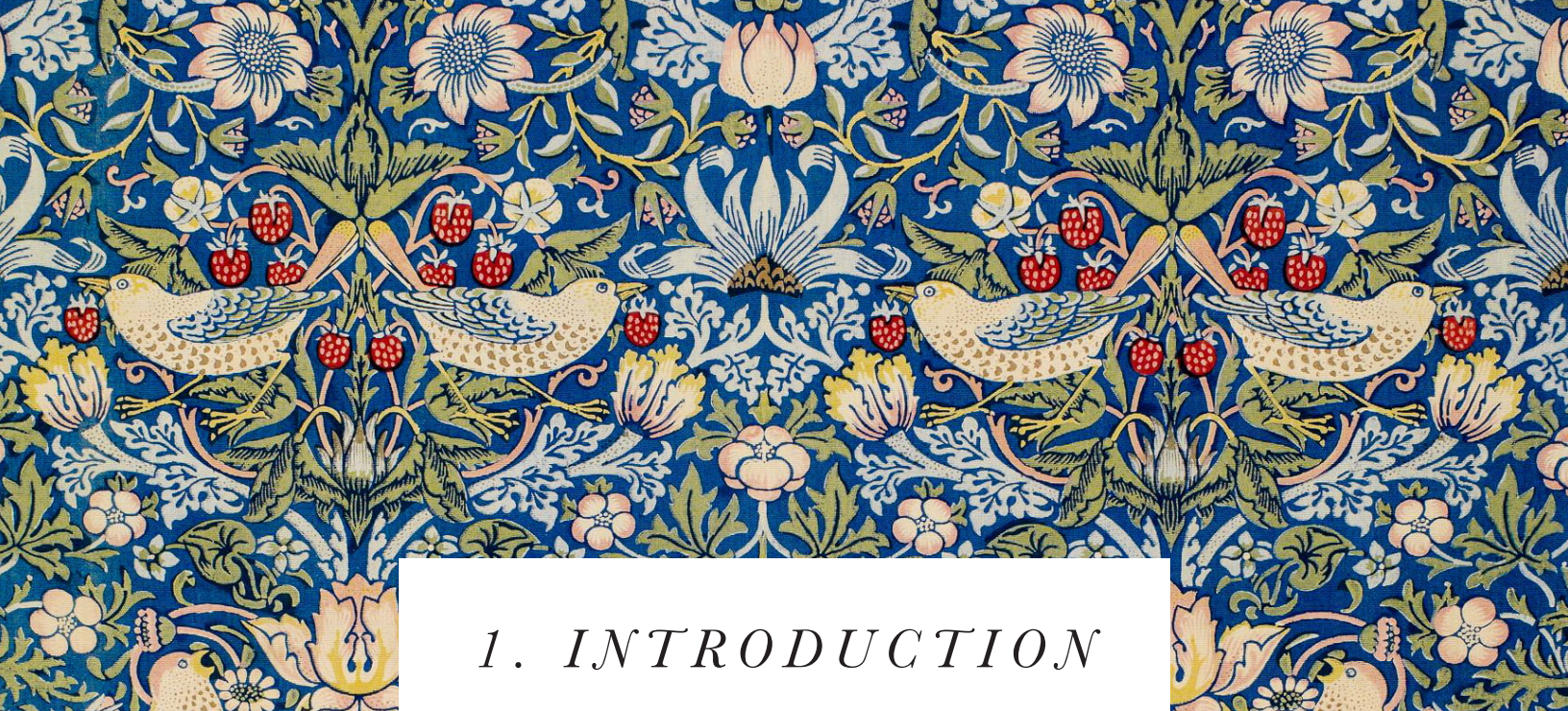
Jeremy Tate,
CLT President



TABLE OF CONTENTS

1. Introduction	10
Overview	
Improving Students' Test-Taking Experience	
Motivating Positive Change in Assessment and Education	
Contents of the Technical Report	
2. Standards and Content Coverage.....	14
Overview	
Verbal Reasoning Test	
Grammar/Writing Test	
Author Bank	
Quantitative Reasoning Test	
Calculator Policy	
Difficulty Levels	
3. Test Development	24
Overview	
Selecting and Training Item Writers	
Design, Review, and Uploading	
Item Review	
4. Test Administration.....	26
At-Home vs. In-School Testing	
Proctors	
Test Day Schedules	
Test Scores	
Testing Formats	
Test Day CLT8 Support	
5. Test Security	30
Data Security	
Data Privacy and Acceptable Use	
Access Control	
Monitoring and Auditing	
Incident Management and Response	
Test Day Processes and Procedures	
Summary	

6. Fairness and Accessibility	34
Overview	
Fairness During the Testing Process	
Fairness in Score Interpretations for Intended Uses	
Fairness in Test Accessibility	
Accommodations and Requests	
7. Item Analysis.....	48
Introduction	
Item Difficulty	
Item Discrimination	
Option/Distractor Analysis	
Differential Item Functioning	
Summary of Item Analysis Results for CLT8 Tests	
References	
8. CLT8 Report Scores, Reliability, and Standard Error of Measurement	64
Overview	
Summary Statistics, Reliability, and Standard Error of Measurement	
9. Validity.....	74
Overview	
Evidence Based on Internal Structure	
Evaluation of the Internal Structure of CLT8 using Exploratory Factor Analysis	
Exploratory Factor Analysis	
Evidence Based on Content	
Summary	
10. Quality Control (QC) of Psychometric Analysis	86
Appendices.....	88



1. INTRODUCTION

Overview

Classic Learning Initiatives (CLI) launched in December 2015, as a long-awaited alternative to the SAT and ACT. As of July 2021, more than fifty thousand students at schools across the United States have taken an exam from the CLT suite of assessments,¹ and over two hundred colleges and universities have adopted the CLT 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.

CLI launched the CLT8 in 2018. Modelled after the CLT, the CLT8 is a formative and summative assessment designed for 7th and 8th grade students. It is used by homeschool families, private schools, and charter schools to evaluate student performance at the beginning and end of the school year. In-depth analytics allow educators to identify student abilities and support curriculum decisions. The CLT8 is offered as part of a suite, including the CLT10 for 9th and 10th grade students and the CLT college admission exam.

Improving Students' Test-Taking Experience

For students, the CLT8 is refreshingly user-friendly and modern. It was designed with the goal of providing the best possible test-taking experience, and includes the following features:

- » Online platform accessible via students' own desktops, laptops, or tablets
- » Predictable format
- » Straightforward scoring: 120 questions, 120 possible points
- » Shorter test-taking time (135 minutes)
- » Scores released within two business days
- » In-depth Student Analytics

¹ The CLT suite of assessments is comprised by: 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 to enter high school.

² The full list of colleges which have adopted the CLT as an admissions exam is provided at <https://www.cltexam.com/colleges>.

ONLINE PLATFORM; STUDENT DEVICES

The CLT8 is primarily administered online, 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 reduces the risk 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 (desktops, laptops, or tablets). Using an unfamiliar device for a high-stakes test can be frustrating, as every device 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 ability to perform.

PREDICTABLE FORMAT

The CLT8 is designed for simplicity and balance. Each of the three sections has forty (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 anxieties that can come from a less regular test design.

Each section loads into a single browser window, so students can scroll to any part of that section without changing pages. A percentage progress bar is provided at the top of the page, giving students a visual sense of their progress on the exam.

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.

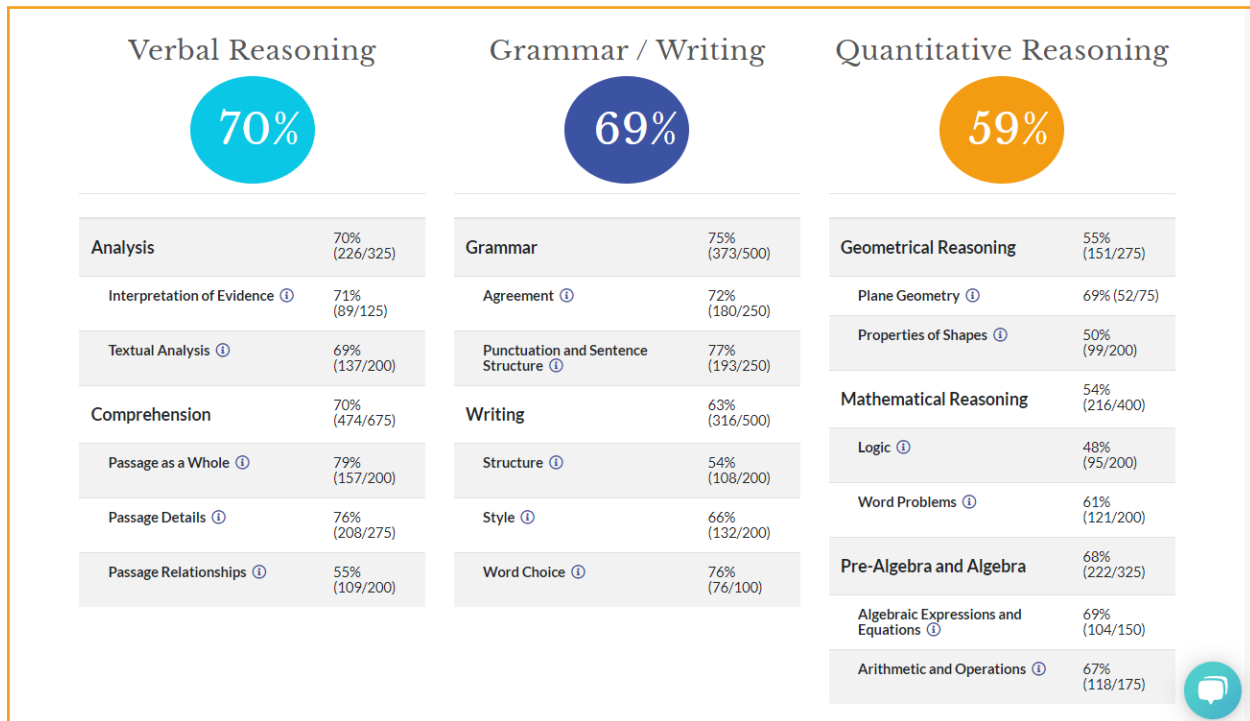
SHORTER TEST; FAST RESULTS

The CLT8 is 135 minutes long, or 2 hours and 15 minutes. The CLT8 was designed to be shorter than comparable tests in order to take as little as possible away from instruction time. The added information gathered by day-long or multi-day assessment regimes is of questionable value, due to evidence that the scores for many students can be negatively affected by exhaustion.

Students that test on a computer can access their scores within two business days of taking the test. Students that use a paper-based test will receive scores once the tests are scanned and processed after receipt.

IN-DEPTH ANALYTICS

For students that take the CLT8 in-school, student scores and analytics are accessible to school administrators, teachers, and parents. Because the CLT8 is an assessment tool, the score and analytics can be used to assess students' strengths and make plans for further improvement. From the school administration and instruction point of view, this data can help inform the material students are taught. The CLT8 is not based on a specific curriculum, but rather upon time-tested, traditional sources, as well as basic grammar, mathematics, and logic; the analytics data we provide serves as a tool to assess how a school's curriculum promotes both student achievement and aptitude. The analytics data are also available to students taking the test from home, and can be used in a similar fashion by homeschoolers.



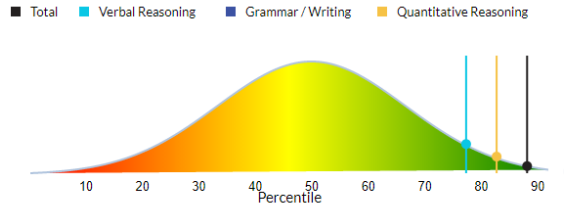
CLT analytics reports are easy to interpret and straightforward. They indicate measures essential to learning and education, and are not weighed down with excessive detail. Teachers can identify the areas of both mastery and the need for improvement for each student and class. All three sections of each exam contain domains and subdomains that pertain to the mastery of each section. For each subdomain, there are corresponding documents that define the subdomain, list the main skills being assessed, and provide sample questions within that subdomain. Teachers and administrators can utilize the analytics and corresponding documents to determine best practices and instruction. This information can be used to understand individual student performance and aptitude. Some of the notable areas of aptitude and achievement include:

- » Making high-level inferences about passage connections and information in the text
- » Identifying evidence supporting an argument or inference from the passage
- » Using evidence in a passage to come to a logical conclusion
- » Understanding how an author uses rhetorical devices in a passage
- » Mastery of grammatical agreement, punctuation, and sentence structure
- » Logically grasping an argument or statement and building evidence in an argument
- » Understanding the nuances of word choice in a sentence
- » Identifying false and true statements based on given information
- » Drawing conclusions about an unknown integer from given information
- » Mastery of algebraic expressions and equations
- » Mastery of geometry and trigonometry
- » Mastery of arithmetic and operations

Percentile Report

This report displays how your scores compare to all other examinees on this test type (CLT, CLT10, or CLT8). The percentiles indicate the percentage of students that scored lower than your score, in each category. The chart displays where each of your scores fall among all examinees. The area in the middle contains the majority of students. Scores right-of-middle indicate above-average performance, whereas scores left-of-middle indicate below-average performance.

Section	Student Percentile
Verbal Reasoning	77th
Grammar / Writing	95th
Quantitative Reasoning	82nd
Total Score	88th



Academic Domains and Subdomains Report

Below are analytics on how you performed by question type. These question types are broken out by academic domains and academic subdomains. The metric, % correct, indicates what percentage of questions you got correct within each category

Top Question Types

Question Type by Academic Domain-Subdomain Pairing	% Correct
Writing - Style ⓘ	100% (8/8)
Geometrical Reasoning - Plane Geometry ⓘ	100% (3/3)
Pre-Algebra and Algebra - Algebraic Expressions and Equations ⓘ	100% (6/6)
Grammar - Agreement ⓘ	100% (10/10)

Areas for Improvement

Question Type by Academic Domain-Subdomain Pairing	% Correct
Geometrical Reasoning - Properties of Shapes ⓘ	38% (3/8)
Pre-Algebra and Algebra - Arithmetic and Operations ⓘ	43% (3/7)
Comprehension - Passage Relationships ⓘ	63% (5/8)
Writing - Structure ⓘ	75% (6/8)

MOTIVATING POSITIVE CHANGE IN ASSESSMENT AND EDUCATION

Besides offering a logistically better and more positive test experience, the CLT8 sets itself apart by its content. The CLT8 aims to change the landscape of assessment, and education generally, by providing a standard that is intellectually richer and more rigorous than other end-of-grade assessments.

Hallmarks of the CLT8’s content are that it:

- » Assesses both aptitude and achievement
- » Is more rigorous
- » Features rich reading passages
- » Supports strong educational choices

APTITUDE AND ACHIEVEMENT

The CLT8 aims to assess not only students’ achievement, but also their aptitude. This is important because a test including an aptitude component makes provision for the diversity of curricula within American education. An achievement test, by definition, has to assume a particular body of content or set of academic standards to master; the CLT8 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 any one-size-fits-all curriculum.

Of course, students must draw upon the teaching they have received in order to demonstrate what they have learned. “Achievement” within a domain of knowledge is a vital and necessary aspect of

assessment, and it is relevant for the CLT8. Students preparing for the rigors of high school education, as well as school administrators, teachers, and parents who prepare these students, want to know that they are progressing successfully through the curriculum. Students at this stage in their education want to be challenged and are discovering their innate intellectual potential. Because the CLT8 is both an achievement and aptitude test, students are provided a window into their own unique set of intellectual strengths, while also given the tools to make incremental improvements in their less developed areas through the analytics provided along with the test.

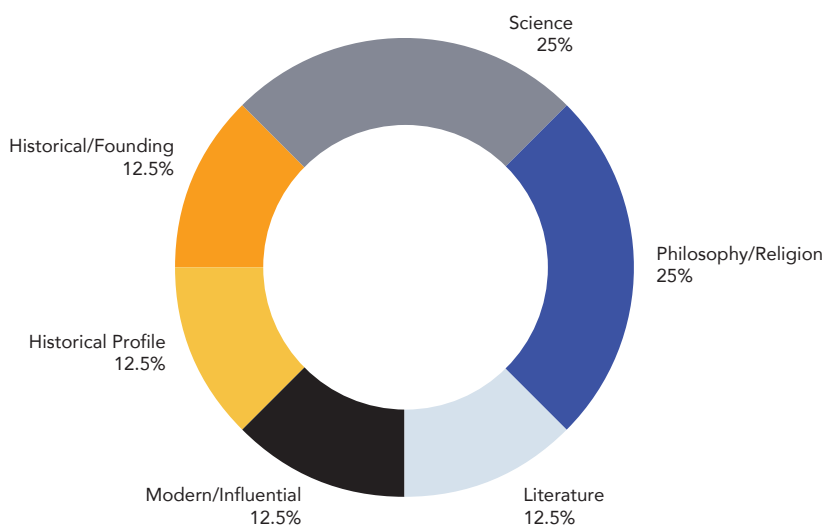
RICH READING PASSAGES

For two-thirds of its reading passages, the CLT8 uses an author bank of more than one hundred 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 CLT8 intentionally acknowledges the particular significance of certain authors. (The full author bank is listed in Chapter 2.)

The CLT8 uses older sources because these texts have stood the test of time. Shakespeare was relevant three hundred years ago, fifty years ago, and today, and it is likely he will remain relevant fifty and three hundred years from now. The passages used on the CLT8 are intended to respect students' dignity and capability by being truly worthy of their time and attention. Using older texts also gives the CLT8 a healthy neutrality on contemporary political matters. Rather than favoring either side of the political spectrum, the CLT8 draws from works that are grounded and established in tradition.

The CLT8 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.

CLT8 Reading Passages



SUPPORTS EDUCATOR CHOICES

Ultimately, the CLT8 seeks to enhance the way young people are educated. Woven into the assessment is the message that literature, logic, tradition, and the virtues are central to the meaning of education. Whether a CLT8 passage reinforces an existing familiarity with an author, or makes the first introduction to a thinker who will become a student’s favorite, a deep synergy is able to develop between the assessment and the schoolroom. With the introduction of the CLT8, schools and homeschooling families have a fresh incentive to focus on enduring ideas.

CONTENTS OF THE TECHNICAL REPORT

This technical report is a guide explaining the details of how the exam works. Chapters 2-6 describe the design and administration of the CLT8. Specifically, 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 CLI takes to develop, edit, and prepare each test for administration. Chapters 4, 5, and 6 explain how the CLT8 is administered and describe the measures taken to ensure the test’s security and fairness. Chapters 7-10 explain and analyze the test’s metrics. Specifically, Chapter 7 presents an item analysis and demographic statistics on CLT8 test questions and reported scores and Chapters 8 and 9, respectively, quantify the test’s reliability and validity. Chapter 10 summarizes the quality control procedure of the data presented in Chapters 7-9.



2.

STANDARDS AND CONTENT COVERAGE

Overview

The CLT8, an end-of-grade assessment for 7th and 8th graders, 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. The CLT8 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 “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 CLT8 focuses on ideas that matter on a grander scale. These include perennial questions about human nature, the physical world, lessons from history, and universal mathematical concepts.

Each CLT8 exam consists of three sections: Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning. These sections are consistent across the suite of CLT assessments (CLT8, CLT10, and CLT), providing a pathway of assessment through to the college entrance exam.

The content of the CLT8 test is distinct from many other assessments at this level. First, instead of the majority of reading passages coming from contemporary sources, the CLT8’s two English sections primarily use selections from classic authors who have shaped history, literature, and philosophy in foundational ways through the centuries. The CLT8 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 focuses more on assessing mathematical reasoning capacity than on testing specific mathematical skills or knowledge.

The hallmark of the CLT8 is that it is based on enduring concepts accessible to students from a variety of educational backgrounds. This ensures that a greater diversity of students have an opportunity to access high-quality testing.

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 passages on different subjects, and are tested on their ability to comprehend the text and synthesize ideas within it. 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 SECTION BLUEPRINTS

In 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.”

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

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 second passage of the four, which is always the Science passage.

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 CLT8 includes analogy questions to assess higher order logical reasoning and synthesis. The content of the CLT8’s analogies refer to concepts within the accompanying 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 COMPLEXITY AND TYPE

Each Verbal Reasoning section consists of four passages: three longer passages and one passage composed of two shorter excerpts presented together. They are arranged in this order:

- » Literature
- » Science
- » Philosophy/Religion
- » Historical/Founding Documents (paired excerpts)

Tests are calibrated so that each Verbal Reasoning section has a total word count between 1,700 -1,750, for an average of 1,725 words total.

VERBAL REASONING SAMPLE QUESTIONS

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

Passage as a Whole

1. Which of the following best describes Kotick's main mission in the passage?

- A) To find the white seal
- B) To return home after his long adventures
- C) To find a place where seals can live in safety
- D) To fulfill the prophecy of the white seal

Sample Question #1 based on Rudyard Kipling's "The White Seal" in The Jungle Book

Passage Details

2. In the final sentence of Passage 2, Paragraph 2, the author of Passage 2 uses "shade" and "sunshine" to refer to which of the following?

- A) Darkness and sunlight
- B) Hard times and happy times
- C) Rest and work
- D) Evil and good

Sample Question #2 based on Thomas Jefferson's Letter to Mrs. Conway

Passage Relationships

3. Venomous snakebite : neglected tropical disease ::

- A) venom : snake
- B) horse : animal
- C) toxin : protein
- D) rashes : nausea

Sample Question #3 based on Léa Surugue's Why is it so hard to stop people from dying from a snakebite?

Textual Analysis

4. The author suggests which of the following in the passage?

- A) When it comes to declarations of injustice, men can often act hypocritically.
- B) Most people are just, no matter what temptations they face in life.
- C) Only those who act justly will acquire power and riches.
- D) Fairness is in the eye of the beholder and men must use every tool to their advantage.

Sample Question #4 based on Plato's The Republic

Interpretation of Evidence

5. Which lines in the passage best support the answer to the previous question?

- A) Paragraph 1, Sentence 2 ("There he . . . seals")
- B) Paragraph 2, Sentence 2 ("In another . . . yourself")
- C) Paragraph 3, Sentence 1 ("Even gentle . . . killing")
- D) Paragraph 3, Sentence 2 ("Go and . . . Kotick")

Sample Question #5 based on Rudyard Kipling's "The White Seal" in The Jungle Book

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 on different subjects, and are tested on their ability to correct errors within each 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 correct spelling, 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 of Choice

GRAMMAR/WRITING SECTION BLUEPRINTS

In 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 COMPLEXITY AND TYPE

Each Grammar/Writing section consists of four passages. They are arranged in this 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 345-445 words. The total must be between 1,550-1,600 words, for an average of 1,575 words total.

GRAMMAR/WRITING SAMPLE QUESTIONS

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

Agreement

1. **They became vain in their own conceits because they will choose to be great rather than humble.**
 - A) NO CHANGE
 - B) chose
 - C) are chosen
 - D) will have been chosen

Sample Question #1 based on Thomas a Kempis’s “The Doctrine of Truth” in The Imitation of Christ.

St. Thomas à Kempis	Winston Churchill	David Hume	Louis Pasteur
Æschylus	Cicero	Edmund Husserl	Max Planck
Æsop	Nicolaus Copernicus	Henrik Ibsen	Plato
Louisa May Alcott	Confucius	William James	Plutarch
Dante Alighieri	Charles Darwin	Thomas Jefferson	Edgar Allen Poe
St. Anselm of Canterbury	René Descartes	Pope St. John Paul II	Marcel Proust
Susan B. Anthony	Charles Dickens	Carl Jung	Pseudo-Dionysius
Archimedes	Fyodor Dostoevsky	Franz Kafka	Claudius Ptolemy
Hannah Arendt	Frederick Douglass	Immanuel Kant	John Rawls
St. Athanasius	W. E. B. Du Bois	Johannes Kepler	Jean-Jacques Rousseau
Jane Austen	Jonathon Edwards	Søren Kierkegaard	Jean-Paul Sartre
Averroës	Albert Einstein	Martin Luther King, Jr.	Dorothy L. Sayers
Avicenna	George Eliot	Antoine Lavoisier	William Shakespeare
Francis Bacon	Epictetus	Harper Lee	Mary Shelley
James Baldwin	Desiderius Erasmus	Gottfried Leibniz	Adam Smith
Simone de Beauvoir	Euclid	C. S. Lewis	Aleksandr Solzhenitsyn
Jeremy Bentham	Euripides	John Locke	Sophocles
<i>Beowulf</i>	F. Scott Fitzgerald	Lucretius	John Steinbeck
Mary McLeod Bethune	Benjamin Franklin	Martin Luther	Harriet Beecher Stowe
The <i>Bhagavad Gītā</i>	Sigmund Freud	Niccolò Machiavelli	Tacitus
Boëthius	Galileo Galilei	James Madison	St. Teresa of Ávila
Niels Bohr	Mahatma Gandhi	Maimonides	Tertullian
Dietrich Bonhoeffer	Gabriel García Márquez	Karl Marx	Thucydides
Jorge Luis Borges	Edward Gibbon	Herman Melville	Alexis de Tocqueville
Charlotte Brontë	Edward Gibbon	Gregor Mendel	J. R. R. Tolkien
Emily Brontë	Johann Wolfgang von Goethe	John Milton	Leo Tolstoy
Martin Buber	William Harvey	Michel de Montaigne	Harriet Tubman
John Bunyan	Friedrich Hayek	St. Thomas More	Mark Twain
Julius Cæsar	Georg Wilhelm Friedrich Hegel	St. John Henry Newman	Virgil
John Calvin	Ernest Hemingway	Isaac Newton	Voltaire
Willa Cather	Herodotus	Friedrich Nietzsche	Booker T. Washington
St. Catherine of Siena	Hippocrates	Flannery O'Connor	Alfred North Whitehead
Geoffrey Chaucer	Thomas Hobbes	George Orwell	Ludwig Wittgenstein
G. K. Chesterton	Homer	Ovid	Mary Wollstonecraft
Anton Chekhov		Blaise Pascal	

Punctuation and Sentence Structure

2. ...so far as his own sculptures are ...so far as his own sculptures are concerned except the Cacciatore, which he sent to the Great Exhibition...

- A) NO CHANGE
- B) concerned; and except the
- C) concerned except, the
- D) concerned; except the

Sample Question #2 based on George Eliot's Letter About Rome Artists

Structure

3. **Weapons of such a devastating nature naturally raise questions of ethics in war.**

- A) NO CHANGE
- B) Fortunately, the United States is the only country yet to have harnessed this power.
- C) It took many years of arduous research and development to produce such weapons.
- D) But the dread secret and the fearful engines of atomic might are not ours alone.

Sample Question #3 based on Dwight D. Eisenhower's Atoms For Peace speech

Style

4. Which of the following choices is the clearest and most concise way to convey all of the information in the sentence?

This year, I joined the European Society of Cardiology's (ESC) annual congress to learn how the field of cardiology in sunny Paris is evolving within the rapidly changing landscape of medicine.

This year, I joined the European Society of Cardiology's (ESC) annual congress to learn how the field of cardiology in sunny Paris is evolving within the rapidly changing landscape of medicine.

- A) NO CHANGE
- B) This year, I joined in sunny Paris to learn how the European Society of Cardiology's (ESC) annual congress, the field of cardiology, is evolving within the rapidly changing landscape of medicine.
- C) This year I joined the field of cardiology in sunny Paris to learn how the European Society of Cardiology's (ESC) annual congress within the rapidly changing landscape of medicine is evolving.
- D) This year, I joined the European Society of Cardiology's (ESC) annual congress in sunny Paris to learn how the field of cardiology is evolving within the rapidly changing landscape of medicine.

Sample Question #4 based on Adya Misra's The Changing Face of Cardiovascular Medicine: Insights from the European Society of Cardiology Annual Congress."

Word Choice

5. **A good and devout man arranges in his mind the things he has to do not according to the whims of evil information but according to the dictates of right reason. but according to the dictates of right reason.**

- A) NO CHANGE
- B) installation
- C) inclination
- D) indisposition

Sample Question #5 based on Thomas a Kempis's "The Doctrine of Truth" in The Imitation of Christ.

Author Bank

The CLT8 draws on sources that have helped shape the course of Western intellectual thought, including authors of diverse backgrounds and philosophies. As of August 2021, two-thirds of all CLT-suite exam 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. This list is periodically updated on our website (www.cltxam.com/authors) on our website (www.cltxam.com/authors).

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 reasoning ability across different domains.

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

QUANTITATIVE REASONING SECTION SECTION BLUEPRINTS

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

Of the 40 Quantitative Reasoning questions, there are 14 Pre-Algebra and Algebra questions, 10 Geometrical Reasoning questions, and 16 Mathematical Reasoning questions, as of 2021. There also are 5 figures and 4 science-based question and 4 science-based questions 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

1. A set of four positive consecutive integers includes three prime numbers. What is the sum of this set?
- A. 6
 - B. 10
 - C. 14
 - D. 26

Algebraic Expressions and Equations

2. If the equation $x^2 + bx + 4 = 0$ has only one solution, and if b is an integer, which of the following could be the value of b ?
- A. 2
 - B. 4
 - C. 6
 - D. 8

Plane Geometry

3. Which transformation of point $A(-2,3)$ results in the image $A'(2,3)$?
- A. A horizontal shift of +2
 - B. A reflection across the y -axis
 - C. A rotation of 90° 90 degrees clockwise about the origin
 - D. A reflection across the line $y=x$

Properties of Shapes

4. A poster is delivered in a right cylindrical container made of cardboard, with a plastic top and a plastic bottom. If the radius of the top and bottom of the container is 10 centimeters and the height of the container is centimeters, what is the surface area of the cardboard portion of the container?
- A. 1000 square centimeters
 - B. 2000 square centimeters
 - C. 1000π square centimeters
 - D. 2000π square centimeters

Logic

5. How many integers between 100 and 120 (inclusive) meet both of the conditions given in the statements below?
1. The sum of the digits of the integer is greater than 5.
 2. The integer is prime.
- A. 0
 - B. 2
 - C. 3
 - D. 5

Word Problems

6. Annabelle’s English class has 20% more students than her social studies class. If her social studies class has 20 students, how many students are in her English class?
- A. 18 students
 - B. 22 students
 - C. 24 students
 - D. 42 students

Calculator Policy

Calculators are not allowed on the CLT8, including on the Quantitative Reasoning section. Questions are designed to be solvable without using a calculator.

The CLT8 is meant to test students’ logical reasoning skills and ability to understand and simplify complex topics, rather than testing their ability to perform complicated calculations. This policy also secures test integrity and simplifies the test by avoiding the need to specify 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 Text Evaluator(TM), a passage calibration software with grade level ratings, to analyze the difficulty level of each passage and ensure it falls within an appropriate range.

Difficulty levels of questions are scored on a scale of 1 through 5: each section of the test contains eight questions at each difficulty level, for a total of twenty-four questions at each difficulty level. In the Verbal Reasoning and Grammar/Writing section, difficulty levels are distributed evenly throughout each passage. Each passage, for which there are ten questions, has two questions of each difficulty level. In 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, higher-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.



3. TEST DEVELOPMENT

Overview

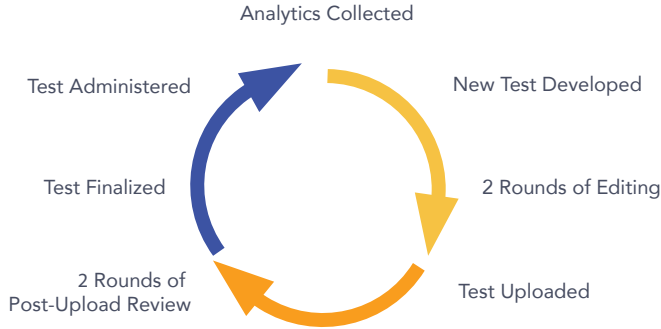
The CLI Test Development Team of Classic Learning Initiatives (CLI) write and edit each test according to a specific set of parameters. The Test Development, Editorial, and Operations Teams work together in the test preparation process. They follow a schedule of development, review, and uploading, so that every test undergoes quality control and is ready on time. The CLI Team analyzes the results of each previous exam, and uses that data in the creation of future tests.

Selecting and Training Item Writers

The CLI Test Development Team 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 suite of exams (CLT8, CLT10, and CLT). Their work then goes through multiple rounds of revision and editing to ensure that each section maintains the high standards of the CLT8, and is consistent, clear, and accurate.

Design, Review, and Uploading

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



DEVELOPMENT AND EDITORIAL PROCESS

After a test is administered, once the analytics from it become available, the CLI Test Development Team meets to discuss the performance of the test form, its sub-sections and items, and to re-calibrate test development plans as necessary. The CLI Test Development Team looks at question performance within each specific difficulty level, focusing special attention on any questions that perform outside of expectations. At the end of each school year, the CLI 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 CLI 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 CLI Test Development Team confirms that question categories are accurate, difficulty levels are well-calibrated, and questions meet CLI quality standards. In-house editors help guide the test through the editorial process. The CLI Test Development 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, overseen by the CLI Test Development Team.

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

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

UPLOADING PROCESS

Before the initial rounds of editing and proofreading, CLI Test Development Team uploads the test. The CLI Test Development 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”).

POST-UPLOAD REVIEW PROCESS

Once all of the passages, images, and test questions are replicated in the website, the CLI Test Development 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 CLI Test Development Team performs two rounds of post-upload reviews of the test. The final reviewer is completely new to the test. The two rounds of post-upload review include the CLI Test Development Team doing 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 CLI Test Development Team creates and formats the paper document using the final version of the uploaded test. The paper test is then 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, the CLI Team completes a thorough item analysis of the test. The CLI Team examines p-values for each item. The CLI Team 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 for the CLT8 administered on May 5, 2021 is provided in Chapter 7.



4. TEST ADMINISTRATION

At-Home vs. In-School Testing

The CLT8 is offered multiple times per year as an at-home test, open to all students, and as an in-school test for students at particular institutions which contract with CLI, typically on the same dates. As of the 2021-2022 academic year, the CLT8 is offered on four at-home and in-school test dates, twice in the fall and twice in the spring.

TESTING FROM HOME

Students interested in taking the CLT8 at home can do so on their own desktops, laptops, or tablets. In these cases, a parent typically acts as proctor and is asked to follow the at-home proctor guidelines and ensure that their child follows the honor code.

Students testing from home must have a proctor at home with them at all times. If necessary, students can take the test from another location such as a library, church, or a friend or relative's home, provided the proctor is on location; this may be necessary due to internet connectivity constraints or availability of a computer.

In order to register for at-home testing, students must create a profile on the CLT website and sign up for the specified exam date. Once registration and payment are completed, the student receives test day instructions and while the proctor receives the proctor manual. On test day, students sign into their profile to access the test.

IN-SCHOOL TESTS

For in-school testing, CLI creates a school administrator account in which a school can register students through a bulk import process. This process is outlined under the school's administrator profile and the CLI Operation Team also emails these instructions to administrators. The process consists of uploading student names, usernames, and passwords into CLI's system to a unique profile for them.

In-school tests at contracting U.S. secondary schools may be offered online or on paper. The administration process is identical to the at-home CLT8. Paper tests must be requested by administrators at least 3 weeks in advance of the test date, and paper test scoring requires up to 30 days after the last day of the test to complete.

Proctors

All CLT8 administrations are proctored, whether the test is being taken in-school or at-home. This is to ensure the integrity and security of the test-taking experience. Proctors are responsible for monitoring the timing for each section, ensuring that students have the proper instructions for the test and that there are no unacceptable items in the room where the test will be taken (among other duties).

In-school proctors are generally (but do not have to be) connected with the institution that is serving as a testing site. CLI communicates with school administration to coordinate proctor selection. The selection is made by the school administration according to guidelines found in the Proctor Manual. CLT8 at-home proctors must be twenty-one years of age or older, and may be related to one or more of the students they are proctoring. In-school CLT8 proctors follow the same proctor guidelines; the proctor may not be related to any of the the students. 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.

CLI staff members coordinate with school administration and proctors ahead of time to provide them with proctor training in the form of the manual. Proctors must familiarize themselves with the proctor manual, which carefully outlines the procedures for proctoring the CLT8.¹

For students testing at home, proctors are typically the parents of the student. Guidelines similar to those provided to in-school proctors are provided to the parent-proctors and are expected to be followed.²

Test Day Schedules

The CLT8 testing schedule is dependent on the time zone of the test location and whether or not the student has been granted extra time due to accommodations.

Students are required to arrive ten minutes before the scheduled test time. Starting at the scheduled test time, ten minutes are allotted for the proctor to provide instructions.

Students have 45 minutes for the Verbal Reasoning section, 40 minutes for Grammar/Writing, and 50 minutes for Quantitative Reasoning, for a total test-taking time of 135 minutes. There is a 10-minute restroom break between the Grammar/Writing and Quantitative Reasoning sections.

Test Scores

CLT8 scores are released the Tuesday following the last day of test administration for students taking the test online (whether at home or in-school). Scores are released within 30 days for paper test students. To see their scores, all students, whether in-school online or paper, or at home online, simply log into their student accounts on cltexam.com.

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 CLT8 school administrator account. Only persons known by CLI to work at the school in question (usually the school administrator role) are granted access to this account. Support from CLI is available for school administrators, teachers, and parents for analyzing the data provided through the CLT8 Analytics.

1 Proctor Manual for In-school CLT8 -2021-2022 (available upon request)

2 Proctor Manual for At-Home CLT8s - 2020-2021 (available upon request)

Testing Formats

The CLT8 is administered in two modes: online, and with paper and pencil. The test is primarily administered online through <http://cltexam.com>, and occasionally with paper and pencil. The latter is only available for in-school testing, and may be obtained by special request of the school.

ONLINE ADMINISTRATION

Students who take the CLT8 in-school do so online on their own devices, which they bring to the testing site. This device may be a laptop computer (including a Chromebook) or a tablet. The test works on most modern devices, and is compatible with most browsers (Chrome and Firefox are the recommended browsers), although it is not compatible with Safari or Microsoft Edge. It requires a reliable internet connection with Javascript enabled. Schools which are furnished with desktop computers may choose to administer the test on these instead of on students' own devices.

Questions in the Quantitative Reasoning portion of the exam may include mathematical notation. Mathematical notation is scripted in HTML (MathML), 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 exceptional, and is discussed individually with the testing site contact to ensure standardization.

At-home test takers must provide their own devices, which may be desktops, laptops, or tablets. The same browser and Javascript requirements apply.

PAPER OPTION

The CLT8 may also be administered on paper, upon special request from the schools. Schools testing on paper pay a paper test surcharge to cover the additional overhead costs. This option is available for in-school test takers only, not for at-home test takers. 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 addressed to the attention of the school's primary point of contact. As with the online CLT8, proctors are expected to follow a strict process, outlined in the paper test manual.
- » School administrators must register their students two (2) weeks before test day to allow CLI time to ensure that students are registered and have access to their online accounts which, for online test-takers, will allow them to take the test. School administrators do this by uploading a .csv of students' usernames and passwords into the CLT system.
- » Students complete a paper answer sheet. School administrators are instructed to scan these and to send a scan to the CLI Operations Team.
- » Students receive their scores within 30 days of the test date. CLI converts the scans of their answer sheets into a .csv spreadsheet using [InspiroScan](#). The CLI Operations Team then cross-references the spreadsheet with the original answer sheets, to ensure that each student's answers are faithfully represented therein. CLI then uploads these documents into the students' CLT accounts.
- » Once scores are imported and executed in the CLI system, administrators and students are able to access these scores, as well as their analytics, using their cltexam.com account.

Test Day CLT8 Support

For each test day, CLI 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 CLI Operations Team and members of the CLI Technology Team, as well as the regular CLI Customer Service Team. For this purpose, CLI uses the Hubspot Chat window on the cltexam.com website to field all questions submitted online. The CLI Customer Service Team is also able to field questions via telephone. For both in-school and at-home CLT8 students, 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*

Classic Learning Initiatives (CLI) test security is designed to ensure privacy and honesty among test-takers, and is broken down into general data security and test day processes and procedures.

Data Security

CLI trains all its employees on the high sensitivity levels of CLT8 data, including the access and use of confidential material such as personally identifiable information (PII). CLI requires each employee to acknowledge and sign internal policies regarding the acceptable use of CLT8 data.

Data Privacy and Acceptable Use

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

STUDENTS

Students who take the CLT8 at school will have access to their scores and analytics. Their scores and analytics will also be available to school administrators, teachers, and parents.

PROCTORS

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

CLT8 data may be accessed either through the web application or through the database directly. All users must be authenticated to access CLT8 data; 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 most minimal access level (comparable to students' own access level).
 - » **Support Access :** CLI employees are granted a support role in order to access necessary information to serve customers. Users in a support role can view test registrations and student data, but they cannot access the test management section of the application.
 - » **Privileged Access :** A limited number of CLI employees have privileged access, which 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 select members of 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. The read-only account is used unless a critical change is required.
- **Data Access:** All CLT8 data is stored in a secure cloud environment that is not accessible to CLI employees in general, only to authorized members of the technical and operation teams. 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, or 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 CLI 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 CLT8 only under secure, supervised conditions. These conditions are as follows:

- **In-school or at-home:** There are two ways that students can take the CLT8, at a CLI-approved partner school or at home with an approved proctor (typically a parent or legal guardian).
- **Proctors:** CLT8 proctors must be twenty-one years of age or older, and may be related to one or more 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. CLI creates CLT accounts for all proctors and provides them with the CLT8 proctor manual.

- **Admitting Students into the Testing Room (in-school and at-home):** On test day, proctors have the final list of CLT8 students for their specific school 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 each 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

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 either stowed in a backpack or purse or left with the proctor)
- » Watches or other such devices that do have internet availability, the ability to communicate with other students, or a calculator

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.

ANOMALIES

Proctors must submit the Administration and Anomaly Report to CLI after the exam, 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 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

CLI 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, CLI provides an exam that is fair, reliable, and accurate.

¹ CLT Administration and Anomaly Report: <https://info.cltxam.com/clt-administration-and-anomaly-report>.



6. *FAIRNESS & ACCESSIBILITY*

Overview

Classic Learning Initiatives (CLI) is committed to providing every student a fair test-taking experience by ensuring the security, integrity, and validity of its examinations. CLI is committed to providing access to its programs and services to students with documented disabilities. A disability is a physical or mental impairment that substantially limits a major life activity.

CLI therefore offers a range of accommodations for students with documented learning or physical disabilities, in accordance with 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 provide equality of access to the test the CLI seeks to promote cognitive diversity and minimize bias.

Test accommodations are adaptations to the exam that can help ensure that the test measures what it is designed to measure, rather than the negative effects of a person's disability. The purpose of test accommodations is to provide candidates with full access to the test – not to guarantee improved performance, a passing score, test completion, or any other specific outcome.

Fairness During the Testing Process

All CLT8 testing takes into account learning differences and disabilities in accordance with the Standards for Educational and Psychological Testing. CLI also considers fairness in testing a top concern, and persistently works to minimize bias and ensure a universally accessible design.

Fairness in Score Interpretations for Intended Uses

The purpose of the CLT8 exam is to assess students' abilities in the areas of Verbal Reasoning, Grammar and Writing, and Quantitative Reasoning at the end of 7th and 8th grade.

Fairness in Test Accessibility

CLI provides testing accommodations to students with documented disabilities to make testing equally accessible to all. Test accommodations are individualized and considered on a case-by-case basis.

Regardless of diagnosis, all individuals seeking disability-related accommodations must provide evidence that their condition rises to the level of a disability and provide information about functional limitations in areas central to daily life. Simply demonstrating that an individual meets diagnostic criteria for a particular disorder does not automatically mean that the person is disabled or entitled to test accommodations. Accommodations must be appropriate to the particular task and setting involved.

Accommodations and Requests

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 previously granted testing accommodations on the test and who are interested in testing accommodations for additional exams should contact our Accommodations team directly, via the email provided in the Accommodations Request Form.

All accommodations request forms must be submitted on behalf of individual students at least four weeks in advance of the testing date. An Accommodations Request Form submitted for more than one student will not be considered.

When accommodations requests are submitted by school administrators on behalf of individual students, parents must also submit a Consent Form for Releasing Accommodations Documentation which authorizes the student's school to release accommodations-related documentation to CLI.

Approved accommodations on the exam may include:

- » Large font exam
- » Additional test time
- » Separate testing room
- » Additional breaks
- » Allowance for use of zoom feature for students with low vision
- » Allowance for use of screen readers for students with low-to-no vision

Review Timeline

To ensure the timely fulfillment of accommodations requests, requests for extended time accommodations must be submitted, with supporting documentation, at least four weeks before the test date.²

CLI reviews accommodations requests and submitted documentation and will contact the submitter by email about any matters requiring clarification. CLT will notify students regarding their approval status within two (2) weeks of receiving the pertinent accommodations request form.

CLI staff will make every effort to review and approve requests; however, CLI cannot guarantee a full review for requests received after the accommodations deadline. In order to be fair to all candidates, accommodations requests are reviewed in the order they are received; requests may not be expedited.

Testers may appeal an accommodation decision if their requested accommodations were not approved. Successful appeals should include a specific reason for appeal, as well as additional documentation beyond what was included in the original request.³

1 CLT Testing Accommodations Request Form, <https://info.cltexam.com/clt-accommodations-request-form>.

2 CLT Testing Accommodations Guidelines (available upon request)

3 CLT Testing Accommodations Appeal Guidelines (available upon request)

7. ITEM ANALYSIS

INTRODUCTION

Item analyses are conducted to evaluate the psychometric properties of the individual items on a test form based on the Classical Test Theory (CTT) framework. The items on the assessment are dichotomously scored multiple choice items. Thus, items are evaluated in terms of item difficulty, item discrimination, and option/distractor analysis (Hambleton & Jones, 1993). Because students taking the assessment differ demographically (e.g., Gender, School Type, Total Household Income), Differential Item Functioning (DIF) analysis is also used to evaluate the items. The following sections provide details on each of the item analyses.

ITEM DIFFICULTY

In CTT, the item difficulty of dichotomously scored multiple choice items is defined as the proportion of examinees who obtained a correct response (McDonald, 1999). The item difficulty is equivalent to the mean item score (see Equation 7.1) and is also known as the item's p -value. It can be computed using the following equation:

$$p - value_i = \underline{x}_i = \frac{\sum_{j=1}^I x_{ij}}{N_i} \quad (7.1)$$

where x_{ij} is the score on item i ($i = 1, 2, 3, \dots, I$) for student j ($j = 1, 2, 3, \dots, N$), and N_i is the total number of students who were administered the i th item.

The item difficulty ranges from 0 to 1. A high item difficulty statistic means that a large proportion of examinees selected the correct response which suggests that the item is relatively easy; conversely, a low item difficulty statistic means that a small proportion of examinees selected the correct response, which suggests that the item is relatively difficult. It is desired that the item difficulties on the CLT assessment fall within the range of 0.3 to 0.9. Items that fall outside of this range are flagged for further review.

A limitation of CTT is that the item statistics are sample-dependent, and thus vary depending upon the specific group of examinees who were administered the item on that occasion (Hambleton & Jones, 1993; McDonald, 1999). That is, given a specific item administered on two occasions, the item difficulty may be higher for a high ability group than it would be for a low ability group. Consequently, the item difficulty may not be comparable across test administrations.

ITEM DISCRIMINATION

Item discrimination is another measure that is often used to evaluate the psychometric properties of an item within the CTT framework (Attali & Fraenkel, 2000; Hambleton & Jones, 1993). This index describes the relationship between students on an item and their performance on the overall test. Item discrimination can be quantified in terms of the item-total correlation. Because the items on the CLT8 are all dichotomously scored, the point-biserial correlation is an appropriate type of correlation (Attali & Fraenkel, 2000). The point-biserial correlation can be computed as follows:

$$pbis_i = \frac{\bar{X}_c - \bar{X}}{S} \sqrt{\frac{p\text{-value}_i}{1-(p\text{-value}_i)}} \quad (7.2)$$

where \bar{X}_c is the mean total test score with the i th item score deleted for examinees who selected a correct response, \bar{X} is the mean total test score with the i th item score deleted for all examinees, S is the standard deviation of the total test scores with the i th item score deleted for all examinees, and $p\text{-value}_i$ of the i th item.

Point-biserial correlations values range from -1 to +1. A large point-biserial correlation coefficient indicates that the item can distinguish between examinees with low and high total test scores. It is desired that the point-biserial correlations of items on the CLT8 assessment are greater than or equal to 0.15. Items with a point-biserial correlation less than 0.15 are flagged for further review.

As previously discussed, a limitation of CTT is that the item statistics are sample-dependent, and thus vary depending upon the specific group of examinees who were administered the item on that occasion (Hambleton & Jones, 1993; McDonald, 1999). That is, given a specific item administered on two occasions, the item discrimination may differ. Consequently, the item discrimination may not be comparable across test administrations.

OPTION/DISTRACTOR ANALYSIS

The items on the CLT assessment are all multiple-choice items with four response options. The option/distractor analysis shows the proportion of students choosing each of the response options and examinees who chose not to select a response option (i.e., omit). The proportion can be calculated using the following formula:

$$proportion_{oi} = \frac{N_{oi}}{N_i} \quad (7.3)$$

where N_{oi} represents the number of examinees select the response option/distractor or omitting the item. The proportion of examinees selecting the correct response option is equivalent to the item difficulty (i.e., $p\text{-value}$). Items where the proportion of examinees choosing to omit a response is greater than .05 are flagged for further review.

In addition to the proportion of examinees selecting each response and omitting the item, the point-biserial correlation (i.e., option/distractor point-biserial correlation) may be used to evaluate the quality of an item (Attali & Fraenkel, 2000). The option/distractor point biserial can be calculated as follows:

$$pbis_{oi} = \frac{\bar{X}_o - \bar{X}}{S} \sqrt{\frac{proportion_{oi}}{1-proportion_{oi}}} \quad (7.4)$$

where \bar{X}_o is the mean total test score with the i th item score deleted for examinees who selected response option o or omit, and $proportion_{oi}$ is the proportion of examinees selecting the option of the i th item.

For an item with good discriminating power (i.e., a large point biserial), it is expected that more high ability students will choose the correct option, while the low ability students would be attracted to the other response options/distractors, which often represent different types of examinee misconceptions. Thus,

the correct response option should have a positive option point-biserial correlation, which is equivalent to the item discrimination, and the other options/distractors or omits will have a negative option point-biserial correlations. A positive option point biserial for a response option that is not the key or the omit category warrants further item review.

DIFFERENTIAL ITEM FUNCTIONING

To further investigate the item psychometric properties, differential item functioning (DIF) is conducted. DIF analyses investigate differences in item-level performance between groups of examinees that are matched based upon scores derived from the measurement instrument (Dorans & Holland, 1993; Magis, Béland, Tuerlinckx, & De Boeck, 2010). The results from the DIF analyses provide evidence that the item scores derived from the measurement instrument (i.e., CLT) have the same meaning across subgroups, and that the item is not potentially biased. Consequently, DIF analyses are an integral part of validity evidence.

The items on the CLT8 are all dichotomously scored multiple choice items. Furthermore, the DIF analyses will be based upon data from two (i.e., reference and focal) groups. Thus, the Mantel-Haenszel (MH) method (Mantel & Haenszel, 1959) is used to detect DIF (Dorans & Holland, 1993; Magis, et al., 2010).

The MH method is based upon the analysis of 2×2 contingency tables created using the data from M ($m = 1, 2, 3, \dots, M$) strata (Dorans & Holland, 1993). The strata are derived using the CLT8 total scores, and matched samples are created using the N_{Rm} and N_{Fm} examinees in the reference and focal groups, respectively. Contingency tables for each stratum (m) are constructed using the counts of scores on the i th item from each group. An example of a 2 (groups) \times 2 (item scores) contingency table is illustrated in Table 7.1 where A_m and B_m represent the count of examinees from the reference group with a score of 1 and 0, respectively, C_m and D_m represents the count of examinees from the focal group with a score of 1 and 0, respectively, N_{1m} and N_{0m} represent the number of examinees with a score point of 1 and 0, respectively, and T_m represents the total number of examinees in the m th stratum (i.e., $T_m = A_m + B_m + C_m + D_m$).

Table 7.1– Crosstabulation Table for Group Membership and Item Scores

		Score on item i		
		0	Total	
Group	Reference	A_m	B_m	N_{Rm}
	Focal	C_m	D_m	N_{Fm}
Total		N_{1m}	N_{0m}	T_m

The Mantel-Haenszel common odds ratio (α_{MH}) can be estimated using the following equation:

$$\alpha_{MH} = \frac{\sum_{m=1}^M \frac{A_m D_m}{T_m}}{\sum_{m=1}^M \frac{B_m C_m}{T_m}}$$

Lastly, as proposed by Dorans and Holland (1993), the DIF effect size can be quantified using the MH Delta difference statistic:

$$D - DIF_{MH} = -2.35 \ln(\alpha_{MH}).$$

Based upon the ETS DIF classification rules (Dorans & Holland, 1993), items are classed into three categories based upon the absolute value of the MH Delta difference statistic (i.e., $|D - DIF_{MH}|$). Items with an absolute value of the MH Delta difference statistic that is less than 1.0 are in the negligible or A category, items with

an absolute value of the MH Delta difference statistic that is greater than 1.0 but less than 1.5 are in the intermediate or B category, and items with an absolute value of the MH Delta difference statistic that is greater than or equal to 1.5 are in the large or C category. Items in category B and C are flagged for further review.

It is important to note that DIF does not necessarily mean that an item is biased. The presence of DIF indicates only that the students with equal ability from different subgroups have an unequal probability of correctly responding to an item. An item is biased if it measures an attribute(s) irrelevant to the intended construct (i.e., construct irrelevant variance); this should only be determined by expert review of item content.

Summary of Item Analysis Results for CLT Tests

All analyses introduced in the above section were conducted for the May 2021 CLT administrations (i.e., Form 1023).

ITEM DIFFICULTY

The item difficulties were calculated using the `itemAnalysis` function from the CTT package (Willse, 2018) in R. Table 7.2 presents the descriptive statistics of p-values for all items within the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections of the CLT assessment. Of the three sections, the Quantitative Reasoning section has the lowest mean item difficulty ($\bar{x}_i = 0.48$) and the most (7) items flagged as difficult (i.e., p-value < .3). The p-value for individual items can be found in Appendix A1.

Table 7.2 – Descriptive Statistics of Item Difficulty (P-values) Item Discrimination

SUBJECT	N	MEAN	SD	MIN	MAX	P-VALUE < .3	P-VALUE > .9
Verbal Reasoning	40	0.64	0.16	0.24	0.94	1	2
Grammar/Writing	40	0.64	0.16	0.21	0.92	1	1
Quantitative Reasoning	40	0.48	0.19	0.16	0.92	7	1

ITEM DISCRIMINATION

The item discriminations were calculated using the `itemAnalysis` function from the CTT package (Willse, 2018) in R. Table 7.3 presents the descriptive statistics of point biserials for all items within the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections of the CLT assessment. The mean point-biserial correlations range from 0.29 to 0.35. The Quantitative Reasoning section has the most items flagged for low item discrimination (8). The point-biserial for individual items can be found in Appendix A1.

Table 7.3 – Descriptive Statistics of Item Discrimination (Point-Biserial Correlations)

SUBJECT	N	MEAN	SD	MIN	MAX	PBIS < .15
Verbal Reasoning	40	0.33	0.10	0.11	0.48	4
Grammar/Writing	40	0.35	0.08	0.16	0.46	0
Quantitative Reasoning	40	0.29	0.12	0.04	0.52	8

Option/Distractor Analysis

The analyses were conducted using the `distractorAnalysis` function from the CTT package (Willse, 2018) in R. 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 and the female group as the focal group. For the race DIF analysis, the White student group was treated as the reference group and the Non-White student group was treated as the focal group. Students with missing group indicators were excluded from the DIF analyses. The `difMH` function from the `difR` package (Magis, Beland, Tuerlinckx, & De Boeck, 2010) in R was used with the default setting.

GENDER

Gender identifications are based on the available self-reported information from examinees. The frequency distributions of the self-reported responses for gender identification are presented in Table 7.4. Gender identification responses that are “Empty string” and “Not Provided” are excluded from the analysis. Consequently, the subsequent analyses are based upon 1,565 Males and 1,546 Females.

Table 7.4 – Frequency Distribution of Self-Reported Gender Identification

GENDER	FREQUENCY	% TOTAL	CUMULATIVE % TOTAL
Female	1,546	44.48	44.48
Male	1,565	45.02	89.50
Not Provided	365	10.50	100.00
Total	3,476	100.00	100.00

The number of items flagged with gender DIF for each section of the CLT8 is summarized in Table 7.5. In general, most of the items in each section were classified with Category A DIF, which is negligible, across all of the sections of the CLT. The detailed information about the DIF analysis results for individual items can be found in Appendix A1.

Table 7.5 – Differential Item Functioning Based using Gender

SUBJECT	NN	A	B	C
Verbal Reasoning	40	39	1	0
Grammar/Writing	40	36	4	0
Quantitative Reasoning	40	35	4	1

RACE/ETHNICITY

Racial/Ethnicity identifications are based on the available self-reported information from examinees. The frequency distributions of the self-reported responses for race/ethnicity identification are presented in Table 7.6. Records where the identification is Not Provided are excluded from the analysis. Due to the small number of examinees in some racial groups (see Table 7.6), such as Hispanic or Latino ($N = 313$), Black or African American ($N = 86$), and American Indian or Alaska Native ($N = 25$), the analyses are based on two general categories, White and Non-White. Examinees who self-identified as White are in the White group ($N = 1,984$) and examinees who self-identified as American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and Other are in the Non-White group ($N = 796$).

Table 7.6 – Frequency Distribution of Self-Reported Race/Ethnicity Identification

RACE/ETHNICITY	FREQUENCY	% TOTAL	CUMULATIVE % TOTAL
American Indian or Alaska Native	25	0.72	0.72
Asian	166	4.78	5.49
Black or African American	86	2.47	7.97
Hispanic or Latino	313	9.00	16.97
Native Hawaiian or Other Pacific Islander	57	1.64	18.61
White	1,984	57.08	75.69
Other	149	4.29	79.98
Not Provided	696	20.02	100.00
Total	3,476	100.00	100.00

The number of items flagged with race/ethnicity DIF for each section of the CLT is summarized in Table 7.7. No items were flagged for C DIF in any of the sections of the CLT assessment. The detailed information about the DIF analysis results for individual items can be found in Appendix A1.

Table 7.7 – Differential Item Functioning Based using Race/Ethnicity

SUBJECT	N	A	B	C
Verbal Reasoning	40	39	1	0
Grammar/Writing	40	40	0	0
Quantitative Reasoning	40	40	0	0

References

- Attali, Y., & Fraenkel, T. (2000). The point-biserial as a discrimination index for distractors in multiple-choice items: Deficiencies in usage and an alternative. *Journal of Educational Measurement*, 37(1), 77-86.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Lawrence Erlbaum Associates, Inc.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their application to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Holland, P. W., & Thayer, D. T. (1985). *An alternate definition of the ETS delta scale of item difficulty* (Research Report RR-85-43). Princeton, NJ: Educational Testing Service.
- Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847–862. <https://doi.org/10.3758/BRM.42.3.847>
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719–748.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Lawrence Erlbaum Associates Publishers.
- Willse, J. T. (2018). *CTT: Classical Test Theory Functions*. (R package version 2.3.3) [Computer software]. <https://mirror.rcg.sfu.ca/mirror/CRAN/web/packages/CTT/CTT.pdf>



8.

*DESCRIPTIVE STATISTICS,
RELIABILITY, AND STANDARD
ERROR OF MEASUREMENT*

OVERVIEW

The CLT8 contains three multiple-choice tests: Verbal Reasoning (VR), Grammar/Writing (GW), and Quantitative Reasoning (QR). 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.

CLT8 scores are scaled from a raw score to a scaled score. The scaled scores are reported to examinees on a 0-120 scale and are determined with a method which uses repeat-person scores, item analysis data, and score distribution data. CLT8 scales each subject section by examining the current test administration item and person scores. These raw scores are then compared with repeat-person scores, taking into account the distributional shapes of the raw scores as compared to previous test administrations. From this information, a scaled score is produced which places the raw CLT8 score (section scores and total score) onto a scale that is consistent across test administrations and scaled for test difficulty. The descriptive statistics provided in this chapter are based upon the total raw scores.

In addition to the composite scores, reliability estimates are reported. Reliability refers to the consistency in test scores (Traub & Rowley, 1991). 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.

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 (Cronbach, 1951; McDonald, 1999). Reliability coefficients are usually estimated using a single test administration by calculating the inter-item covariances or correlations. Cronbach's alpha (1951), which is a measure of internal consistency, is one of the most widely used estimates of test reliability. Cronbach's alpha can be computed using the formula:

$$\alpha_I = \frac{I}{I-1} \left(1 - \frac{\sum_{i=1}^I \text{var}(x_i)}{\text{var}(X)} \right) \quad (8.1)$$

where a_i is the reliability estimate of the i ($i = 1, 2, 3, \dots, I$) items, $\text{var}(x_i)$ is the variance of item i , and $\text{var}(X)$ is the variance of the total scores. 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.

Finally, the Standard Error of Measurement (SEM) for each total score is reported. The SEM provides another indicator of the accuracy of test scores by quantifying the amount of error or inconsistency in test scores. The SEM can be computed using the following formula:

$$SEM = SD(X)\sqrt{1 - \alpha_I} \quad (8.2)$$

where $SD(X)$ is the standard deviation of the total scores.

In the following sections, descriptive statistics, reliability, and SEM are presented using unadjusted scores by subgroup (i.e., gender and race), school type, and student family income on the administration on May 5, 2021, form pclt/1023. Identifications of gender, race/ethnicity, and family income are based on available self-reported information from examinees.

SUMMARY STATISTICS, RELIABILITY, AND STANDARD ERROR OF MEASUREMENT

The descriptive statistics of minimum and maximum scores, mean, and standard deviation (SD) derived from raw scores, estimated reliability, and standard error of measurement (SEM) of the CLT8 scores and the Verbal Reasoning (VR), Grammar/Writing (GW), and Quantitative Reasoning (QR) section scores across all examinees are reported in this section.

Descriptive statistics presented in Table 8.1 suggest that the mean and standard deviation of the CLT8 scores are 70.38 and 18.77, respectively. The average scores for the VR, GW, and QR subtests are 25.76, 25.49, and 19.14, respectively. The standard deviations for the VR, GW and QR subtests are 6.96, 7.36, and 6.89, respectively. The reliability of the CLT8 scores is 0.94 with a SEM of 4.71. The reliability coefficients for the three subtests range from 0.84 to 0.87, and the SEMs range from 2.65 to 2.67. The results suggest that form 1023 provides highly reliable CLT8 scores and moderately highly reliable Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning subtest scores.

Table 8.1 – Descriptive Statistics of Test Scores

TEST	N	MIN	MAX	MEAN	SD	RELIABILITY	SEM
CLT8	3,476	14	116	70.38	18.77	0.94	4.71
Verbal Reasoning	3,476	3	40	25.76	6.96	0.85	2.67
Grammar/Writing	3,476	0	40	25.49	7.36	0.87	2.65
Quantitative Reasoning	3,476	0	40	19.14	6.89	0.84	2.76

SUMMARY STATISTICS, RELIABILITY, AND STANDARD ERROR OF MEASUREMENT BY SUBGROUP

The estimated reliability and SEM are computed by subgroup for composite CLT8 scores and the three multiple-choice based tests, Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning. It is important to note that when the variance of an item is zero, it is removed from the analysis.

GENDER

Gender identifications are based on the available self-reported information from examinees. The frequency distributions of the self-reported responses for gender identification are presented in Table 8.2. Gender identification responses that are “Empty string” and “Not Provided” are excluded from the analysis. Consequently, the subsequent analyses are based upon 1,565 Males and 1,546 Females.

Table 8.2 – Frequency Distribution of Self-Reported Gender Identification

GENDER	FREQUENCY	% TOTAL	CUMULATIVE % TOTAL
Female	1,546	44.48	44.48
Male	1,565	45.02	89.50
Not Provided	365	10.50	100.00
Total	3,476	100.00	100.00

Table 8.3 provides the analysis results by gender. The results suggest similar performance between males ($M = 69.13$, $SD = 19.38$) and females ($M = 70.97$, $SD = 17.94$) on the overall assessment (see Table 8.3). On average, female examinees performed slightly better than male students on the Verbal Reasoning and Grammar/Writing sections and male students performed better on the Quantitative Reasoning section (see Table 8.3).

Table 8.3 – Descriptive Statistics, Reliability, and SEM of Test Scores by Gender

GENDER	TEST	N	MIN	MAX	MEAN	SD	RELIABILITY	SEM
Male	CLT8	1,565	14	116	69.13	19.38	0.94	4.74
	Verbal Reasoning		3	39	25.10	7.03	0.85	2.69
	Grammar/Writing		0	40	24.42	7.60	0.87	2.69
	Quantitative Reasoning		0	40	19.62	7.10	0.85	2.75
Female	CLT8	1,546	18	116	70.97	17.94	0.93	4.68
	Verbal Reasoning		6	40	26.19	6.80	0.85	2.65
	Grammar/Writing		0	40	26.40	6.97	0.86	2.62
	Quantitative Reasoning		0	40	18.39	6.54	0.82	2.76

The reliability coefficients given the CLT8 scores for male and female examinees are 0.94 and 0.93, respectively. The SEMs of the CLT8 scores for male and female examinees are 4.74 and 4.68, respectively. The reliability coefficients of the VR, GW, and QR subtests range from 0.85 to 0.87 for male examinees, and the SEMs range from 2.69 to 2.75. The reliability coefficients of the VR, GW, and QR subtests range from 0.82 to 0.86 for female examinees, and the SEMs range from 2.62 to 2.76. The results of analyses show highly reliable CLT8 scores and moderately highly reliable subtest scores for both male and female examinees.

RACE/ETHNICITY

Racial/Ethnicity identifications are based on the available self-reported information from examinees. The frequency distributions of the self-reported responses for race/ethnicity identification are presented in Table 8.4. Records where the identification is Not Provided are excluded from the analysis. Due to the small number of examinees in some racial groups (see Table 8.4), such as Hispanic or Latino ($N = 313$), Black or African American ($N = 86$), and American Indian or Alaska Native ($N = 25$), the analyses are based on two general categories, White and Non-White. Examinees who self-identified as White are in the White group and examinees who self-identified as American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and Other are in the Non-White group ($N = 796$).

Table 8.4 – Frequency Distribution of Self-Reported Race/Ethnicity Identification

RACE/ETHNICITY	FREQUENCY	% TOTAL	CUMULATIVE % TOTAL
American Indian or Alaska Native	25	0.72	0.72
Asian	166	4.78	5.49
Black or African American	86	2.47	7.97
Hispanic or Latino	313	9.00	16.97
Native Hawaiian or Other Pacific Islander	57	1.64	18.61
White	1,984	57.08	75.69
Other	149	4.29	79.98
Not Provided	696	20.02	100.00
Total	3,476	100.00	100.00

Table 8.5 summarizes the analysis results by White and Non-White examinee groups. The average CLT8 score for White examinees ($M = 72.48$, $SD = 17.66$) is higher than the Non-White examinee group ($M = 63.18$, $SD = 19.33$) group. On the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections, the mean scores for White Students were 26.50, 26.31, and 19.67, respectively. For the Non-White examinee group, the mean scores on the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections were 23.19, 22.84, and 17.14, respectively.

Table 8.5 – Descriptive Statistics, Reliability, and SEM of Test Scores by White vs. Non-White Groups

GROUP	TEST	N	MIN	MAX	MEAN	SD	RELIABILITY	SEM
Non-White	CLT	796	18	113	63.18	19.33	0.94	4.80
	Verbal Reasoning		4	40	23.19	7.29	0.86	2.74
	Grammar/Writing		1	40	22.84	7.65	0.87	2.74
	Quantitative Reasoning		0	37	17.14	6.80	0.84	2.75
White	CLT	1,984	18	116	72.48	17.66	0.93	4.69
	Verbal Reasoning		6	40	26.50	6.52	0.84	2.65
	Grammar/Writing		0	40	26.31	6.97	0.86	2.63
	Quantitative Reasoning		0	40	19.67	6.72	0.83	2.76

Table 8.5 also provides reliability coefficients and SEMs by White and Non-White examinee groups. The reliability and SEM of the CLT8 scores are 0.93 and 4.69 for the White examinee group, respectively. The reliability and SEM of the CLT8 scores are 0.94 and 4.80 for the Non-White examinee group, respectively. For the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections, the reliability coefficients and SEMs range from 0.83 to 0.86 and 2.63 to 2.76, respectively, for White students and 0.84 to 0.87 and 2.74 to 2.75, respectively, for Non-White students. The results of analyses provide evidence to support highly reliable CLT8 scores and moderately highly reliable subtest scores for both White and Non-White examinee groups. The results also indicate the comparability of test scores for internal consistency and accuracy of test scores.

RELIABILITY AND STANDARD ERROR OF MEASUREMENT BY SCHOOL TYPE AND FAMILY INCOME

The descriptive statistics, reliability and SEM are computed based on the raw scores by school type and family income for the composite CLT8 score and the Verbal Reasoning, Grammar/ Writing, and Quantitative Reasoning section scores. It is important to note that when the variance of an item is zero, it is removed from the analysis.

SCHOOL TYPE

Table 8.6 presents the frequency distribution of school type. It is important to note that a small sample size may result in statistics with large sampling error, so caution should be taken in the interpretations of those statistics, particularly in comparison between or across subgroups. As shown in Table 8.6, there are fewer than 25 examinees who attend a Magnet, Privado, or Public school. Consequently, statistics are reported only for examinees who attend a Home School, Private School, or Charter School.

Table 8.6 – Frequency Distribution of School Type

SCHOOL TYPE	FREQUENCY	% VALID	CUMULATIVE % VALID
Charter	233	6.70	11.42
Homeschool	983	28.28	39.70
Magnet	1	0.03	39.73
Privado	3	0.09	41.57
Private	2,011	57.85	99.42
Public	20	0.58	100.00
Other	61	1.75	41.48
Not Provided	164	4.72	4.72
Total	3,476	100.00	100.00

Table 8.7 presents the descriptive statistics of CLT8 scores and scores on the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections by school type. The average CLT8 score is higher for Home School ($M = 75.66$, $SD = 18.12$) than that for Private School ($M = 68.44$, $SD = 18.20$), and Charter School ($M = 64.62$, $SD = 19.42$) examinees. This also holds true for the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning sections. The results in Table 8.7 show that the reliability coefficients and SEM of the CLT8 scores range from 0.93 to 0.94 and from 4.60 to 4.80, respectively. The reliability coefficients of the three subtests range from 0.82 to 0.86 and the SEMs range from 2.56 to 2.76 across the school types. This suggests that the reliability of the scores from the CLT8 assessment are similar across school types.

Table 8.7 – Descriptive Statistics, Reliability, and SEM of Test Scores by School Type

TYPE	TEST	N	MIN	MAX	MEAN	SD	RELIABILITY	SEM
Home School	CLT	983	18	116	75.66	18.12	0.94	4.60
	Verbal Reasoning		6	40	27.50	6.54	0.84	2.58
	Grammar/Writing		1	40	27.47	7.03	0.87	2.56
	Quantitative Reasoning		3	40	20.69	7.25	0.86	2.72
Private School	CLT	2,011	14	116	68.44	18.20	0.93	4.74
	Verbal Reasoning		3	40	25.14	6.84	0.85	2.69
	Grammar/Writing		0	40	24.78	7.20	0.86	2.68
	Quantitative Reasoning		0	39	18.51	6.52	0.82	2.76
Charter School	CLT	233	18	110	64.62	19.42	0.94	4.80
	Verbal Reasoning		4	37	23.55	7.35	0.86	2.74
	Grammar/Writing		1	39	23.10	7.78	0.88	2.73
	Quantitative Reasoning		0	37	17.97	6.81	0.84	2.76

FAMILY INCOME

The frequency distribution of Family Income is presented in Table 8.8. Another noteworthy observation is that approximately 70% of examinees either did not provide a response (i.e., Not Provided) or selected the Unsure/Prefer Not to Respond response option. Caution should be taken in the interpretations of statistics based upon small sample sizes due to large sampling error. For this reason, the self-reported family incomes are collapsed into three general categories: \$50,000 or lower, \$50,001-\$125,000, and \$125,001-\$225,000 or higher.

Table 8.8 – Frequency Distribution of Family Income

HOUSEHOLD INCOME	FREQUENCY	% TOTAL	CUMULATIVE % TOTAL
\$0 - \$25,000	41	1.18	1.18
\$25,000 - \$50,000	122	3.51	4.69
\$50,000 - \$75,000	199	5.72	10.41
\$75,000 - \$125,000	335	9.64	20.05
\$125,000 - \$225,000	219	6.30	26.35
More than \$225,000	124	3.57	29.92
Unsure/Prefer Not to Respond	1,107	31.85	61.77
Not Provided	1,329	38.23	100.00
Total	3,476	100.00	100.00

Descriptive statistics of the overall CLT8 scores and the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning section scores are presented in Table 8.9 by family income category. The average CLT8 score for the Family Income of \$50,000 or Lower ($M = 59.89$, $SD = 17.69$) group was more than 10 points lower than the Family Income of \$50,001 – \$125,000 ($M = 71.61$, $SD = 18.15$) and Family Income of \$125,001 or Higher ($M = 72.91$, $SD = 19.16$) groups. Similarly, the average Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning section scores were lower for the Family Income of \$50,000 or Lower group than the Family Income of \$50,001–\$125,000 and Family Income of \$125,001 or Higher groups (see Table 8.9).

Table 8.9 also reports the reliability and SEMs of the CLT8 scores and Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning section scores. The reliability of CLT8 scores range from 0.92 to 0.94, and the SEMs range from 4.67 to 4.85 across family income categories (see Table 8.9). Except for the reliability of the Quantitative Reasoning section scores for the Family Income of \$50,000 or Lower group ($\alpha_{QR} = 0.79$), the reliability of the section scores were above 0.84 for all groups. Lastly, the SEMs of the section scores ranged from 2.61 to 2.78 across the three family income groups.

Table 8.9 – Descriptive Statistics, Reliability, and SEM of Test Scores by Family Income

TEST	N	MIN	MAX	MEAN	SD	RELIABILITY	SEM
Family Income of \$50,000 or Lower							
CLT	163	18	104	59.89	17.69	0.92	4.85
Verbal Reasoning		6	38	22.36	7.04	0.84	2.78
Grammar/Writing		4	38	21.90	7.48	0.86	2.76
Quantitative Reasoning		5	36	15.63	6.03	0.79	2.76
Family Income of \$50,001 - \$125,000							
CLT	534	22	116	71.61	18.15	0.93	4.70
Verbal Reasoning		6	39	26.30	6.75	0.85	2.65
Grammar/Writing		3	39	26.09	6.98	0.86	2.64
Quantitative Reasoning		1	38	19.22	7.02	0.85	2.76
Family Income of \$125,001 or Higher							
CLT	343	21	111	72.91	19.16	0.94	4.67
Verbal Reasoning		4	39	26.40	6.91	0.85	2.64
Grammar/Writing		0	39	26.28	7.39	0.87	2.61
Quantitative Reasoning		0	39	20.23	7.16	0.85	2.75

References

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Lawrence Erlbaum Associates Publishers.
- Traub, R. E, & Rowley, G. L. (1991). An NCME instructional module on understanding reliability. *Educational Measurement: Issues and Practice*, *10*(1),37-45. <https://doi.org/10.1111/j.1745-3992.1991.tb00183.x>



9. *VALIDITY*

The ultimate use of test scores is to make inferences about students' ability, competence, or behaviors in situations beyond that observed in the testing scenario. The previous chapter detailed the evidence related to the reliability of the CLT8 scores, and the evidence suggests that the CLT8 test scores are highly reliable (see Chapter 8). Nonetheless, additional evidence is needed to assure that the inferences drawn from the CLT 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 (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), and it is related to fairness. While reliability addresses the consistency in test scores obtained from different forms and administrations, 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 (Messick, 1987).

According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), 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.

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 as intended. 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.

The collection of validity evidence is an ongoing process. This technical report provides evidence from different sources in the test development and administration process. Chapters 2, 3, 4, and 9 in this technical report provide such content-related evidence of validity. This chapter focuses on collecting evidence related to the internal structure of the CLT8.

Evidence Based on Internal Structure

The investigation of the internal structure of a test can provide important validity evidence (AERA, APA, & NCME, 2014). The internal structure can be evaluated in terms of dimensionality, construct equivalence, measurement precision (in terms of reliability, standard error of measurement, and/or test information) and differential item functioning. In general, construct equivalence across the subgroups of the examinee 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 examinees to demonstrate their abilities on the constructs a test intends to measure (pp. 51-53). Furthermore, evaluations of test fairness question whether the test measures the same construct in all relevant populations. For instance, 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.

EVALUATION OF THE INTERNAL STRUCTURE OF CLT USING EXPLORATORY FACTOR ANALYSIS

The evaluation of the internal structure of the CLT8 is conducted using the domain scores. Theoretically, the items on the CLT8 are measures of a dominant underlying construct. An Exploratory Factor Analysis (EFA) is an appropriate method when the goal of the analysis is to “arrive at a parsimonious representation of the associations among measured variables” (Fabrigar, Wegener, MacCallum, & Strahan, 1999, p. 275). Thus, by conducting an EFA, empirical evidence is collected to assess the assumption that there is a dominant construct underlying the data. Table 9.1 summarizes the number of items measuring each domain within each CLT8 subject test. The number of items assessing each domain across subjects ranged from 11 to 27. The descriptive statistics for the domain scores are summarized in Table 9.2.

Table 9.1 – 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	Geometrical Reasoning	11
	Mathematical Reasoning	16
	Pre-Algebra and Algebra	13

Table 9.2 – Descriptive Statistics for the Domain Scores

SUBJECT	DOMAIN	N	MINIMUM	MAXIMUM	MEAN	STANDARD DEVIATION
Verbal Reasoning	Analysis	3,476	0	13	8.19	2.40
	Comprehension	3,476	3	27	17.56	5.04
Grammar/Writing	Grammar	3,476	0	20	13.54	3.92
	Writing	3,476	0	20	11.95	3.98
Quantitative Reasoning	Geometrical Reasoning	3,476	0	11	4.58	2.32
	Mathematical Reasoning	3,476	0	16	7.20	3.11
	Pre-Algebra and Algebra	3,476	0	13	7.36	2.59

CORRELATIONS BETWEEN THE DOMAIN SCORES AND THE SUBJECT TEST SCORES

The correlations between the domain scores are computed and summarized in Table 9.3. In general, the domain scores from the same subject section tend to be more highly correlated (see Table 9.3). The same is true for the correlations between the domain scores between the Verbal Reasoning and Grammar/Writing subject sections, as the correlations ranged from 0.56 to 0.72. In contrast, the correlations between the domain scores from the Quantitative Reasoning section and those from either the Verbal Reasoning or the Grammar/Writing sections were all relatively lower, as the correlations ranged from 0.40 to 0.58 and 0.46 to 0.56, respectively. The patterns observed across the domain scores fall within expectations and provide both convergent and divergent validity evidence.

Table 9.3 – Correlations among the Domain Scores

SUBJECT	DOMAIN	ANALYSIS	COMPREHENSION	GRAMMAR	WRITING	GEOMETRICAL REASONING	PRE-ALGEBRA & ALGEBRA	QUANTITATIVE REASONING
Verbal Reasoning	Analysis	1.00						
	Comprehension	0.71	1.00					
Grammar/Writing	Grammar	0.56	0.68	1.00				
	Writing	0.60	0.72	0.73	1.00			
Quantitative Reasoning	Geometrical Reasoning	0.40	0.48	0.46	0.46	1.00		
	Pre-Algebra and Algebra	0.42	0.51	0.53	0.52	0.57	1.00	
	Mathematical Reasoning	0.49	0.58	0.52	0.56	0.63	0.61	1.00

Note. All correlations are significant at the 0.01 level (2-tailed). Domain scores within the same subject are in bold. Exploratory Factor Analysis

Exploratory factor analyses of the domain scores from the CLT were conducted using the psych package in R (Revelle, 2021). Specifically, a (uniterated) principal factor analysis where the squared multiple correlations are set as the initial communalities using the “fa” function. A three-factor orthogonal model was the largest model to lead to a solution. The results that follow are based upon this solution. The eigenvalues, the percentage of explained common variance by factor, and the proportion of eigenvalues were all examined.

The eigenvalues, the percentage of explained common variance, and the ratio of the eigenvalues are summarized in Table 9.4. The eigenvalue for the first factor is approaching 4, and the difference between the first two factors is around 3.5 (see Table 9.4). Moreover, the first factor accounts for more than 85% of the common variance. Lastly, the eigenvalue of the first factor is 9 times larger than the eigenvalue of the second factor. Based on these criteria, the first factor was relatively strong, and it is plausible that a dominant underlying dimension is influencing the domain scores.

Table 9.4 – Exploratory Factor Analysis Results: Eigenvalues and Variance Explained

FACTOR	EIGENVALUE	EXPLAINED COMMON VARIANCE	CUMULATIVE EXPLAINED COMMON VARIANCE	RATIO OF EIGENVALUES
1	3.944	89.45%	89.45%	9.91
2	0.398	9.02%	98.47%	5.94
3	0.067	1.53%	100.00%	

Table 9.5 presents the factor loadings from the EFA of the domain scores. Factor loadings are in bold if the absolute value is greater than 0.3. A value of 0.3 is chosen because the factor loading squared is an estimate of the amount of variance explained by the factor. Thus, a value of 0.3 suggests that 9% of the domain score’s variance is explained by the factor. All of the domain scores have a factor loading greater than 0.65 on the first extracted factor (see Table 9.5). Only the Geometrical Reasoning domain score has an absolute factor loading greater than 0.3 on the second factor. Lastly, none of the domain scores have an absolute factor loading greater than 0.3 on the third factor.

Table 9.5 – Exploratory Factor Analysis Results: Factor Loadings

SUBJECT	DOMAIN	FACTOR 1	FACTOR 2	FACTOR 3
Verbal Reasoning	Analysis	0.714	-0.219	0.147
	Comprehension	0.834	-0.208	0.093
Grammar/Writing	Grammar	0.784	-0.153	-0.145
	Writing	0.812	-0.182	-0.111
Quantitative Reasoning	Geometrical Reasoning	0.658	0.334	0.024
	Mathematical Reasoning	0.745	0.269	0.041
	Pre-Algebra and Algebra	0.689	0.258	-0.037

Note. Factor loadings with an absolute value greater than .3 are in bold.

It is important to note, however, that even though a rationale for a factor loading cutoff criteria 0.3 is provided, the choice of the value can still be viewed as arbitrary (DiStefano, Zhu, & Míndrilă, 2009). For instance, it can easily be argued that a factor loading cutoff criteria value of 0.5 is more appropriate, as that suggests that 25% of the domain score's variance is explained by the factor. In this instance, a factor loading cutoff criteria of 0.5 allows for clearer interpretations of the EFA results being in support of a unidimensional model.

In summary, this section assessed the internal structure of the CLT. The results from the EFA provide validity evidence related to the internal structure of the CLT forms. Fabrigar, Wegener, MacCallum, and Strahan (1999) argue that researchers should “balance the need for parsimony (i.e., a model with relatively few common factors) against the need for plausibility (i.e., a model with a sufficient number of common factors to adequately account for the correlations among measured variables)” when selecting the number of factors in the model (p. 277). There is sufficient empirical evidence to suggest that there is a single dominant factor underlying the measures on the CLT. This is consistent with the theoretical framework and the theoretical content model in the CLT design and development. Consequently, the assumption of a unidimensional model underlying the CLT is both parsimonious and plausible.

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” (p. 11). The test specification parameters for the CLT8 are designed to ensure that the test results yield appropriate indicators of individuals' capacity for higher-level thinking and preparation for college. The range of question types in the Verbal Reasoning, Grammar/Writing, and Quantitative Reasoning 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: (1) Comprehension questions, which include the subdomains “Passage as a Whole”, “Passage Details”, and “Passage Relationships,” and (2) 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: (1) their understanding of the text's meaning, the author's intent, and the information conveyed by the passage and (2) 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; that is, 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: (1) Grammar questions, which include the subdomains “Agreement” and “Punctuation and Sentence Structure,” and (2) Writing questions, which include the subdomains “Structure,” “Style,” and “Word Choice.” 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.

On the Quantitative Reasoning section, questions are broken down into three types: (1) Pre-Algebra and Algebra, (2) Geometrical Reasoning, and (3) Mathematical Reasoning. Pre-Algebra and Algebra questions include the subdomains “Arithmetic and Operations” and “Algebraic Expressions and Equations.” Geometrical Reasoning questions include the subdomains “Plane Geometry” and

“Properties of Shapes.” Mathematical Reasoning questions include the subdomains of “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 high school and beyond. Pre-Algebra and 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 building blocks for students taking more advanced math courses in high school and beyond, but are also important indicators of a student’s ability to think clearly and logically, which are crucial skills needed in all academic disciplines.

Summary

In terms of both content and internal structure, the CLT8 exam demonstrates a high level of validity. Analysis of the test’s structure suggest that a dominant dimension is underlying the measured from the CLT8 assessment. Furthermore, the types of questions in each subject test correspond to key skills in reading, writing, and mathematics. CLT8 test scores are thus a legitimate measure of students’ aptitude and preparation for academic work at the college level and beyond.

References

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

DiStefano, C., Zhu, M., & Mindriță, D. (2009). - Understanding and using factor scores. Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, 14(20). <https://doi.org/10.7275/da8t-4g52>

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>

Messick, S. (1987). *Validity*. (ETS Research Report Series RR-87-40). Educational Testing Services. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2330-8516.1987.tb00244.x>

Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research* (R package version 2.1.6) [Computer software]. Northwestern University, Evanston, Illinois. <https://CRAN.R-project.org/package=psych>.



10.

QUALITY CONTROL OF PSYCHOMETRIC ANALYSES

Chapters 7-9 of the technical report present the results of the psychometric analyses of the CLT8 exam at the item, subscore, subject, and overall test levels. These analyses are intended to collect validity evidence, assess the reliability of the test scores, evaluate and assure fairness, and confirm that the CLT8 test measures what it is intended to measure with adequate precision. Quality control should be ensured in all aspects of the psychometric analyses. The first step in quality control for psychometric work is to develop a process flowchart to standardize the procedure and the steps to follow. CLT8 identified all psychometric steps, procedures, and analyses when planning the contents of this technical report to ensure that the right and defensible psychometric steps and procedures would be followed in developing the CLT8 technical report. The specifications for the psychometric analyses and the steps to follow to ensure quality control of the analysis results were all streamlined, standardized, and documented to ensure that replication by independent parties was feasible.

The second key element in quality control is to independently replicate the analysis with different analysts. Prior to conducting the independent psychometric analyses, meetings were held to ensure that all parties understood the expectations when conducting the psychometric analyses. One of the analysts led the psychometric task by running the analysis, summarizing the results in a written document, and then sharing it with the other analyst who was responsible for quality control. The second analysts independently conducted each analysis and compared their results with the results from the lead analyst. If the results from the independent runs matched, they were used in the published version of the technical report. If a discrepancy was found, the analysts met and explored potential sources for the discrepancy, such as different approaches to the exclusion rules applied in data cleaning, variable recoding, and ways of dealing with missing values. The analyses were re-ran and compared again. This process continued until an exact match was obtained. Once an exact match was obtained, the results were used in the published version of the technical report.

Table A1 Item Statistics for Form 1023

ITEM ID	P-VALUE	POINT BISERIAL	GENDER			ETHNIC		
			α_{MH}	$D - DIF_{MH}$	DIF	α_{MH}	$D - DIF_{MH}$	DIF
1123	0.5946	0.2915	0.8013	0.5206	A	1.1064	-0.2376	A
1124	0.7759	0.1671	0.9441	0.1351	A	1.0516	-0.1181	A
1125	0.5152	0.3144	0.7509	0.6733	A	1.3481	-0.7019	A
1127	0.6643	0.2588	0.8888	0.2770	A	0.9689	0.0743	A
1126	0.8705	0.3308	0.7069	0.8151	A	1.0669	-0.1523	A
1128	0.9396	0.2829	0.9858	0.0335	A	0.8933	0.2650	A
1129	0.6453	0.4215	1.0489	-0.1122	A	1.6999	-1.2468	B
1130	0.7080	0.1129	1.0678	-0.1541	A	1.0276	-0.0641	A
1131	0.6217	0.4104	0.6439	1.0347	B	0.8845	0.2884	A
1133	0.4543	0.4224	1.0985	-0.2208	A	0.9818	0.0432	A
1135	0.8340	0.3161	0.8979	0.2532	A	0.7295	0.7413	A
1134	0.3789	0.4800	0.7212	0.7681	A	0.9279	0.1758	A
1132	0.7589	0.2763	0.9512	0.1177	A	1.0675	-0.1535	A
1136	0.7782	0.4371	1.0416	-0.0958	A	0.8648	0.3413	A
1137	0.7434	0.1114	0.9369	0.1531	A	1.1330	-0.2935	A
1138	0.6260	0.4446	1.3379	-0.6841	A	1.1507	-0.3298	A
1140	0.5101	0.4029	0.8213	0.4627	A	0.7448	0.6923	A
1139	0.9396	0.2735	0.8141	0.4833	A	0.9248	0.1837	A
1141	0.6263	0.4430	1.0456	-0.1049	A	0.9977	0.0054	A
1142	0.5409	0.1985	0.8620	0.3489	A	1.0095	-0.0223	A
1143	0.8202	0.2756	0.7755	0.5975	A	0.8632	0.3456	A
3109	0.4275	0.3899	0.9503	0.1198	A	0.8986	0.2513	A
1145	0.5935	0.2633	1.2274	-0.4815	A	0.9703	0.0708	A
1146	0.3743	0.1089	0.9987	0.0030	A	1.0029	-0.0068	A
1147	0.5432	0.4276	1.0700	-0.1591	A	1.0925	-0.2079	A
1148	0.5564	0.3591	1.2079	-0.4439	A	1.1674	-0.3637	A
1149	0.7980	0.3684	0.8416	0.4052	A	0.8103	0.4945	A
1150	0.6551	0.4002	0.8490	0.3847	A	1.2166	-0.4608	A

1151	0.6513	0.2898	1.2684	-0.5587	A	1.0995	-0.2229	A
3110	0.2445	0.1356	0.9088	0.2247	A	1.1193	-0.2648	A
1154	0.7158	0.4225	1.1452	-0.3186	A	1.1185	-0.2632	A
1153	0.7532	0.2775	0.7437	0.6960	A	0.8548	0.3688	A
1155	0.7189	0.3439	0.8533	0.3727	A	0.7249	0.7561	A
1156	0.8147	0.3587	1.0308	-0.0713	A	1.0235	-0.0547	A
1157	0.7756	0.4540	0.9481	0.1252	A	0.9770	0.0548	A
1158	0.7440	0.3620	0.8824	0.2941	A	1.2236	-0.4743	A
1159	0.3518	0.2512	0.9233	0.1875	A	0.9272	0.1775	A
1160	0.6453	0.3322	0.9852	0.0349	A	1.1284	-0.2838	A
1162	0.5529	0.4283	0.8976	0.2539	A	0.9905	0.0224	A
1161	0.4963	0.3582	0.9922	0.0185	A	0.8291	0.4406	A
441	0.4963	0.3294	0.9810	0.0451	A	0.8123	0.4885	A
442	0.3159	0.2845	0.6792	0.9092	A	0.8453	0.3950	A
443	0.4819	0.3685	0.6983	0.8440	A	0.8092	0.4976	A
444	0.5521	0.3109	0.9816	0.0437	A	1.0576	-0.1316	A
445	0.7250	0.3809	0.7861	0.5655	A	0.8652	0.3402	A
446	0.8124	0.3420	0.8301	0.4376	A	0.9041	0.2370	A
3113	0.5360	0.3989	0.8935	0.2646	A	1.0995	-0.2230	A
448	0.4695	0.3272	1.1349	-0.2975	A	0.9892	0.0256	A
449	0.3130	0.1792	1.2431	-0.5114	A	1.0147	-0.0343	A
1292	0.8156	0.4612	0.9453	0.1321	A	0.9256	0.1816	A
1293	0.5409	0.3157	0.7842	0.5714	A	0.9604	0.0948	A
1294	0.3875	0.3592	0.7392	0.7103	A	1.2440	-0.5130	A
1295	0.7080	0.1612	0.8887	0.2774	A	1.1101	-0.2455	A
1296	0.7900	0.3169	0.6731	0.9304	A	0.9299	0.1709	A
1297	0.8075	0.4142	0.6063	1.1759	B	0.9169	0.2038	A
1298	0.5426	0.2896	0.7483	0.6813	A	1.1766	-0.3822	A
1299	0.9226	0.2809	0.7299	0.7400	A	0.7443	0.6939	A
1300	0.6027	0.4203	0.8217	0.4615	A	1.0198	-0.0461	A

1301	0.8047	0.4443	0.6105	1.1597	B	1.1835	-0.3959	A
1302	0.7523	0.4049	1.0428	-0.0984	A	1.1443	-0.3167	A
1183	0.5256	0.3794	0.7430	0.6981	A	0.9629	0.0889	A
1184	0.6315	0.4471	0.8444	0.3976	A	1.2082	-0.4445	A
1185	0.6683	0.4381	0.7415	0.7029	A	1.1730	-0.3750	A
1186	0.5480	0.3792	0.7398	0.7081	A	1.0798	-0.1803	A
1187	0.6487	0.1590	0.9501	0.1203	A	0.8013	0.5206	A
1188	0.7517	0.2682	0.7656	0.6277	A	1.1625	-0.3538	A
1189	0.5032	0.4100	0.7510	0.6729	A	0.9528	0.1137	A
1190	0.8055	0.4334	1.1461	-0.3204	A	0.8158	0.4784	A
1191	0.7365	0.3498	0.7695	0.6157	A	1.1608	-0.3504	A
1192	0.6257	0.3280	0.6354	1.0656	B	1.0898	-0.2021	A
1193	0.7345	0.4218	0.8193	0.4684	A	0.9339	0.1606	A
1194	0.6286	0.3043	1.1274	-0.2817	A	0.9879	0.0286	A
1195	0.8996	0.3679	0.8385	0.4140	A	0.7890	0.5570	A
1196	0.2074	0.2810	1.1200	-0.2664	A	1.1789	-0.3868	A
1197	0.6470	0.3955	0.8538	0.3713	A	1.0470	-0.1080	A
1198	0.6082	0.3225	0.8075	0.5025	A	0.8678	0.3331	A
1199	0.6985	0.4538	0.8755	0.3126	A	0.8294	0.4397	A
1200	0.7707	0.3274	0.8614	0.3506	A	1.2799	-0.5799	A
1201	0.8317	0.4167	0.6289	1.0900	B	0.8710	0.3246	A
1202	0.6413	0.3784	0.8697	0.3280	A	1.2066	-0.4414	A
1083	0.9177	0.3198	1.2356	-0.4972	A	1.1278	-0.2826	A
1084	0.3680	0.3756	1.4602	-0.8896	A	1.4619	-0.8923	A
1085	0.8498	0.1579	0.9218	0.1913	A	0.8391	0.4121	A
1086	0.5495	0.3290	1.2421	-0.5096	A	0.9824	0.0417	A
1087	0.2750	0.4208	1.4409	-0.8584	A	0.9266	0.1791	A
1088	0.4940	0.2153	0.9262	0.1800	A	0.8918	0.2690	A
1089	0.4856	0.3108	1.1988	-0.4262	A	1.0060	-0.0140	A
1090	0.8329	0.3860	1.6953	-1.2405	B	1.3998	-0.7903	A
1091	0.3061	0.3397	1.3090	-0.6327	A	0.9387	0.1486	A

1092	0.4077	0.3928	1.7670	-1.3378	B	1.3879	-0.7703	A
1775	0.7583	0.3769	1.1450	-0.3181	A	0.9577	0.1015	A
1777	0.8587	0.3261	1.8841	-1.4886	B	0.7412	0.7038	A
1095	0.4160	0.2788	1.2129	-0.4536	A	0.9272	0.1777	A
1096	0.7140	0.3898	0.9783	0.0515	A	0.9505	0.1194	A
1097	0.5647	0.4125	1.9263	-1.5407	C	1.0880	-0.1981	A
1784	0.6105	0.2882	1.0278	-0.0644	A	0.7433	0.6972	A
1539	0.5785	0.2974	1.2654	-0.5532	A	0.8530	0.3737	A
1100	0.5400	0.5075	1.4331	-0.8456	A	1.2363	-0.4984	A
1101	0.4586	0.3407	1.2641	-0.5508	A	1.0007	-0.0016	A
1544	0.4603	0.4043	1.4317	-0.8433	A	0.8941	0.2629	A
1103	0.6082	0.3888	1.3437	-0.6942	A	0.8383	0.4145	A
1104	0.5279	0.1354	1.1460	-0.3203	A	1.1058	-0.2362	A
3175	0.5719	0.3696	1.4814	-0.9235	A	1.1643	-0.3575	A
1106	0.4537	0.5185	1.3513	-0.7076	A	1.3559	-0.7155	A
1532	0.4635	0.2624	1.2678	-0.5576	A	1.1276	-0.2823	A
2041	0.4517	0.3462	1.1927	-0.4142	A	1.0187	-0.0434	A
1109	0.3979	0.0786	1.1169	-0.2598	A	0.8193	0.4682	A
1110	0.3777	0.1648	1.1433	-0.3146	A	0.9664	0.0803	A
1111	0.3654	0.3353	1.1790	-0.3869	A	0.8930	0.2659	A
3176	0.3780	0.1087	1.0311	-0.0720	A	0.8337	0.4273	A
1113	0.1746	0.1236	1.0185	-0.0430	A	1.0480	-0.1103	A
1114	0.5475	0.4136	1.7230	-1.2786	B	1.1097	-0.2447	A
3114	0.1577	0.1854	1.0414	-0.0952	A	1.1430	-0.3142	A
1551	0.4266	0.3945	1.1159	-0.2578	A	0.9344	0.1595	A
1117	0.2581	0.2192	1.2107	-0.4493	A	0.8669	0.3356	A
1118	0.3636	0.1435	1.0664	-0.1510	A	1.1540	-0.3366	A
1802	0.4404	0.3814	1.2034	-0.4350	A	0.9266	0.1791	A
1120	0.2687	0.0768	1.2452	-0.5154	A	0.8159	0.4782	A
1121	0.1631	0.1088	1.2808	-0.5815	A	0.7135	0.7934	A
3177	0.2957	0.0435	1.2322	-0.4906	A	0.9174	0.2026	A