CollegeBoard

The Cognitively Complex Thinking Required by Select SAT® Suite Questions

Evidence from English Learners (ELs)

College Board
August 2025

The Cognitively Complex Thinking Required by Select SAT Suite Questions:

Evidence from English Learners (ELs)

August 2025

Jim Patterson | **Lead author; Reading and Writing section analysis**

Michael Gosche | **Math section analysis**

Jay Happel | **Sample analysis**

Beth Oxler, Georgina Keenan, Nancy Burkholder | **Editorial services**

Vidlet, Inc. | **Cognitive interviews**

Suggested Citation:
College Board. 2025. *The Cognitively Complex Thinking Required by Select SAT Suite Questions: Evidence from English Learners (ELs).* College Board.

About College Board

College Board reaches more than 7 million students a year, helping them navigate the path from high school to college and career. Our not-for-profit membership organization was founded more than 120 years ago. We pioneered programs like the SAT® and AP® to expand opportunities for students and help them develop the skills they need. Our BigFuture® program helps students plan for college, pay for college, and explore careers. Learn more at **cb.org**.

© 2025 College Board. College Board, AP, BigFuture, SAT, and the acorn logo are registered trademarks of College Board. PSAT/NMSQT is a registered trademark of College Board and National Merit Scholarship Corporation. Bluebook and PSAT are trademarks of College Board. Desmos and related trademarks are property of Desmos Studio PBC.

Contents

Executive Summary v

Section 1: Introduction 1

Document Preview 2

Section 2: Literature Review 3

Verbal Protocols as Data in Social Science Research 3

Verbal Protocols as Data in Research on the Designs of Large-Scale Standardized Assessments 5

Threats to Verbal Protocol Validity and Reliability 6

Concurrent and Retrospective Verbalizations 8

Methodological Implications for the Present Study 9

Section 3: Methodology 10

Test Question Selection 10

Question Type–Level Construct Definition 13

Protocol Development 14

Test Question Delivery Method 14

Tools Available to Participants 14

Sample Definition, Selection Criteria, Recruitment, and Characteristics 15

- Sample Definition 15
- Sample Selection Criteria 16
- Sample Recruitment 16
- Sample Characteristics 17

Coding and Analysis 21

- Coding 21
- Analysis 22

Section 4: Results 26

Reading and Writing 26

- Participant and Question Performance 26
- Participant Performance Vignettes 30
- Participant Perceptions 61

Math 72

- Participant and Question Performance 72
- Participant Performance Vignettes 74
- Participant Perceptions 96

Section 5: Discussion 102

Reading and Writing 102

- Participant Performance 102
- Question Performance 103
- Participant Performance Vignettes 104
- Participant Perceptions 104

Math 105
Participant Performance 105
Question Performance..... 105
Participant Performance Vignettes..... 105
Participant Perceptions 105
General Discussion 106
Quantitative Results Summary: Participant and Question Differentials..... 106
Qualitative Results Summary: Participant Performance Vignettes and Participant Perceptions 106
Study Limitations 107
Section 6: Conclusion 109
References 110
Appendix..... 114
Exhibit 1: Recruitment Solicitation 114
Exhibit 2: Recruitment Screener (Survey) 115
Exhibit 3: Consent Form 120
Exhibit 4: Interview Session Training Questions..... 122
Reading and Writing 122
Math 124

Tables

Table 1. Characteristics of Reading and Writing (RW) and Math Questions Presented to Study Participants. 11
Table 2. Reading and Writing Participant Roster by Demographics and EL Status. 18
Table 3. Math Participant Roster by Demographics and EL Status. 20
Table 4. Participant Performance Level (PL) Definitions. 23
Table 5. Participant and Question Differentials, by Test Section..... 106

Figures

Figure 1. Reading and Writing Participant and Question Performance Summary Matrix. 27
Figure 2. Math Participant and Question Performance Summary Matrix. 73

Executive Summary

This report documents the findings of a think-aloud (cognitive lab) study conducted with students who are English learners (ELs) as they answered a set of either SAT® Suite Reading and Writing or Math questions. The research goals were, first, to ascertain, via qualitative and quantitative means, whether these EL students were able to demonstrate cognitively complex thinking in line with the question types' constructs and college and career readiness requirements and, second, to explore whether participants' performance on the questions or their postexperience reflections on the think-aloud activity would uncover any construct-irrelevant barriers to their success on such questions.

Twenty high school juniors and seniors who indicated being English learners and met other criteria were selected to participate in the Reading and Writing segment of the study, while an additional eighteen such students participated in the Math segment. Each participant was asked to think aloud (narrate their thoughts) to a moderator supplied by vendor Vidlet, Inc., as they answered up to fifteen Reading and Writing or Math questions (selected to be broadly representative of the sections' domains) and to answer a standardized series of postexperience interview questions. Participants engaged with the test questions via Bluebook™, the custom-built testing application developed by College Board to administer the SAT Suite tests in their digital-adaptive formats, and had access to the app's universal tools. Within the constraints of selection criteria, small sample sizes, and the self-selection methodology, the resulting Reading and Writing and Math participant pools were somewhat diverse in terms of gender, race/ethnicity, grade in school, and self-reported English language acquisition levels, but students with elevated high school GPAs (HSGPAs) were overrepresented.

The focal portions of the sessions, which were scheduled for roughly two hours and for which participants were compensated via gift card, were video recorded. The transcripts produced from these sessions were analyzed qualitatively and quantitatively by College Board subject matter experts relative to lists of predefined required (Reading and Writing) or expected (Math) behaviors, which operationally defined the questions' constructs by question type. The researchers performed coding in MAXQDA, a qualitative/mixed-methods research software package, and tabulated results in Microsoft Excel. Each participant-by-question

interaction was assigned one of up to five performance levels (PLs), with PL 1 representing the most successful performance (answering a given question correctly while also demonstrating all required behaviors [Reading and Writing] or at least one expected behavior [Math]) and PL 5 representing the least successful (answering a given question incorrectly and demonstrating no required or expected behaviors).

The College Board researchers analyzed the coded transcripts on three dimensions:

1. **Participant performance** was analyzed in terms of the number and proportion of correctly answered questions for which participants demonstrated appropriate cognitive behaviors. Vignettes (transcript excerpts) from select participants were used to illustrate demonstrations of the cognitively complex thinking elicited by the test questions.
2. **Question performance** was analyzed in terms of the number and proportion of correctly answering participants who also demonstrated appropriate cognitive behaviors.
3. **Participant perceptions** of the question-answering activity, in the form of responses to postexperience interview questions, were analyzed for both general themes and for any cases in which participants identified potential construct-irrelevant barriers to their success in the activity and to SAT Suite test taking more broadly.

The main metric used to assess participant performance was the *participant differential* (D_p). Mathematically, D_p represents the arithmetic difference between (1) the number of Reading and Writing or Math questions a given participant answered correctly and (2) the number of such questions for which the participant demonstrated all required behaviors (Reading and Writing) or at least one expected behavior (Math). Conceptually, D_p represents the “difference” between simply answering a given question correctly and doing so while also exhibiting appropriate behaviors. Because participants answered a variable number of test questions during the activity, the threshold for a “good” D_p was set at 70 percent, meaning that a given participant needed to demonstrate appropriate behaviors for at least 70 percent of the questions they answered correctly. Vignettes (transcript excerpts) from participants attaining PL 1 on each test question are provided and serve as a second source of evidence respecting participant performance on the questions.

The main metric used to assess question performance was the *question differential* (D_q). Similar to D_p, D_q represents, in mathematical terms, the arithmetic difference between (1) the number of participants answering a given question correctly and (2) the number of such participants who also demonstrated appropriate behaviors. Conceptually, D_q represents the “difference” between the number of participants who simply answered a given question correctly and the number who did so while also demonstrating appropriate behaviors. The threshold for a “good” D_q was again set at 70 percent, meaning, in this case, that for a given question, at least 70 percent of correctly answering participants also demonstrated appropriate behaviors.

Participant perceptions of the think-aloud activity were collected via a standardized set of postexperience interview questions. Responses to these questions were analyzed both for general themes and for indicators that participants had been affected by construct-irrelevant barriers and were thus impeded from demonstrating the full extent of their subject matter knowledge.

This report delineates three key findings:

- **Participant performance.** Fifteen of twenty Reading and Writing participants (75 percent) and sixteen of eighteen Math participants (89 percent) met or exceeded the threshold for a good D_p, providing evidence that EL students are able to demonstrate cognitively complex thinking in line with the question types' constructs. Additionally, vignettes exhibiting PL 1 were obtained for all fifteen Reading and Writing and all fifteen Math questions, providing additional support for the claim that EL students can demonstrate cognitively complex thinking via SAT Suite test questions.
- **Question performance.** Thirteen of fifteen Reading and Writing questions (87 percent) and all fifteen Math questions (100 percent) met or exceeded the threshold for a good D_q, providing evidence that, overall, the presented questions were capable of eliciting cognitively complex thinking from EL students.
- **Participant perceptions.** No clear evidence of construct-irrelevant barriers not already addressed by the provision of testing supports emerged from participant responses to the postexperience interview questions or observation of participant question-answering behavior during the think-aloud activity.

The generalizability of the results of this study is limited by several factors, including the study's small sample sizes, the artificiality of the think-aloud methodology itself, and the possibility (though, as it turned out, likely not the reality) that some participants may have previously encountered the studied SAT Suite test questions as part of their normal test preparation activities.

Section 1: Introduction

The following report presents the methodology, findings, and implications of a verbal protocol study conducted in 2024 by College Board, with support from vendor Vidlet, Inc., involving samples of high school juniors and seniors who identify as English learners (ELs) as they thought aloud through a series of either SAT Suite Reading and Writing or Math questions.

The research goals of this study were twofold:

- Does evidence gathered from qualitative and quantitative analysis of transcripts from samples of high school juniors and seniors who are English learners support the conclusion that select SAT Suite test questions are capable of eliciting cognitively complex thinking from English learners in line with college and career readiness expectations and the question types' constructs?
- Is evidence gathered from these transcripts and/or responses to postexperience interview questions suggestive of potential non-content-related (i.e., *construct-irrelevant*) impediments to EL students' ability to demonstrate the full extent of what they know and can do in the literacy and math domains of the SAT Suite tests? If so, have these impediments been addressed by the provision of testing supports, such as extra time?

In brief, this study, one of several verbal protocol studies of the SAT Suite conducted by College Board (College Board and HumRRO 2020; College Board 2024a, 2025a, 2025b), engaged samples of high school juniors and seniors in thinking aloud—verbalizing their thought processes—as they answered a series of either Reading and Writing or Math test questions selected from the official practice environment. Transcripts of these moderator-led sessions were produced and then analyzed for evidence of participants having exhibited cognitively complex behaviors associated with the various Reading and Writing and Math question types administered. Each participant-by-question interaction was evaluated for these behaviors as well as for whether the question was answered correctly or incorrectly, and then performance levels were assigned. Metrics called *differentials* were determined for each participant and for each Reading and Writing and Math test question, with the criteria for successful results being, respectively, that each participant demonstrated appropriate cognitive behaviors at least 70 percent of the time when answering questions correctly and that at least 70 percent of the time, participants demonstrated appropriate cognitive behaviors while answering a given question correctly. Transcript vignettes (excerpts) exemplifying participants correctly answering a given question and exhibiting appropriate behaviors were identified and served as a second source of evidence for this study. Responses to a standardized set of postexperience interview questions were also analyzed and served as an additional evidence source.

Document Preview

Section 2: Literature Review offers a brief overview of the research literature consensus on the validity of using a concurrent verbal protocol/think-aloud methodology as a means of gaining insight into cognitive processes that would otherwise be inaccessible or prone to retrospective or inferential bias. Section 3: Methodology details the method used to conduct the study and analyzes the enacted student samples along demographic lines. Section 4: Results presents the qualitative and quantitative findings obtained from the study, including summative metrics, question-by-question transcript vignettes, and analysis of postexperience interview question responses. Section 5: Discussion interprets the findings presented in the preceding section, draws conclusions and implications, and considers the study's limitations. Section 6: Conclusion briefly wraps up the body of the report. Following the references is an appendix containing the recruitment materials and excerpts of the verbal protocols used by College Board and Vidlet in carrying out the study's data collection.

Section 2: Literature Review

Verbal Protocols as Data in Social Science Research

The formal use of verbal protocols as a research tool to uncover otherwise unobservable cognitive processes extends back at least a century (Ericsson and Simon 1993). The scholarly consensus over the last half century has supported the use of verbal protocols as a data collection tool within a range of limitations and constraints, discussed more thoroughly below (Russo et al. 1989; Bainbridge and Sanderson 1995; Goos and Galbraith 1996; Branch 2013). Verbal protocol studies have illuminated participant thought processes in a wide range of areas, including business management (Isenberg 1986), marketing and consumer choice (Bolton 1993; Bettman and Park 1980), computer programming (Vessey 1986), engineering (Atman and Turns 2001), accounting (Biggs and Mock 1983), nursing (Haffer 1990), information systems (Nguyen and Shanks 2007), library science (Branch 2001), human geography (Lundberg 1984), and education (Suto and Greatorex 2008).

Education has, in fact, been one of the more fertile areas for verbal protocol studies in recent years. The appeal of the methodology to this field is intuitively obvious. Researchers, teachers, curriculum specialists, and other stakeholders are committed to developing and implementing instructional methods and materials that promote student learning, but such learning takes place, often silently and unobserved, in students' heads. Without some sense of how students themselves are engaging (or not engaging) with these methods and materials, we can't fully or fairly account for the success or failure of these interventions.

One foundational verbal protocol study in the education field was that of Pressley and Afflerbach (1995), who used and refined the approach in an effort to create a model of conscious mental processes enacted during reading. A particular area of focus for many literacy-related verbal protocol studies has been distinguishing the behaviors of more and less successful readers. For example, Kletzien (1991) employed verbal protocols to attempt to differentiate strategy use by high school–age students of higher and lower reading achievement levels as they engaged with successively more challenging expository passages. Kletzien found that both groups of participants used similar strategies but that those with better comprehension skills used more, and more varied, strategies as the texts became harder. Magliano and Millis (2003) used verbal protocol analysis to help develop a latent semantic analysis–based computerized reading comprehension measure. Drawing on prior work and their 2003 study, the researchers found that “good readers emphasize establishing coherence[,] and poor readers emphasize the contents of the current sentence” as they read (255). More recently, Cho et al. (2018) qualitatively and quantitatively analyzed the verbal responses of ten more and ten less successful online readers in an effort to determine how these two groups differed in their cognitive approaches to analyzing a controversial topic. The authors concluded that the more successful readers engaged in the work in ways “notably different” (215) from those of their less successful peers in terms of extent of source evaluation and application of metacognitive strategies related to successfully accomplishing the task.

Verbal protocol analysis has also been used successfully to explore participants’ thought processes as they engage in math tasks. For instance, Goos and Galbraith (1996) used the methodology to determine that two high school seniors collaborating on a series of problems in an applied math course exhibited “differing, but complementary, metacognitive strengths” (255), which typically aided in their joint problem-solving. Montague and Applegate (1993) analyzed the verbal protocols from eighty-one middle school students, roughly a third of whom were selected randomly from pools of learning disabled, average-achieving, and gifted students in a large southeastern metropolitan district. The researchers found that when presented with a range of problems in math, students identified as gifted were more strategic in their solving approaches than students in the other two achievement groups; that perceived difficulty of math problems seemed to affect students’ perseverance and cognition; and that “students with LD [learning disabilities] approach[ed] problem solving in a qualitatively different manner than their more proficient peers” (29). Özcan et al. (2017) also used verbal protocol analysis to examine math problem-solving approaches used by students, in this case sixty-nine sixth graders sampled across achievement levels. Among their findings, the researchers determined that those students who employed an incorrect process in solving a nonroutine math problem “mostly [did] operations aimlessly” and approached the word problem superficially (139–40).

As indicated above, the verbal protocol method has been employed successfully with students with learning disabilities. Özkubat and Özmen (2021) used think-aloud protocols as one tool to evaluate the math problem-solving skills of both sixth-grade students with learning disabilities and low- and average-ability students without such disabilities. Deshpande et al.’s (2021) small-scale examination of high school students’ problem-solving abilities in geometry used think-alouds to illuminate cognitive and metacognitive strategies employed by students with and without learning disabilities. Similarly, Botsas (2017) used think-aloud protocols to explore the cognitive and metacognitive strategy use of fifth- and sixth-grade students with and without learning disabilities as they read both narrative and expository science texts.

Verbal protocol studies have also frequently been used to study the cognitive (and metacognitive) processes of language learners as they acquire a second or subsequent language or perform other academic tasks. Yayli (2010) employed both think-aloud and retrospective methods to investigate the reading-related cognitive and metacognitive strategies of proficient and less proficient readers enrolled in a university-level English language teaching department in Turkey. Bowles and Gastañaga (2022) used a think-aloud method as one approach to assessing the impact of various forms of written corrective feedback given to heritage language, second-language, and third-language university-level learners of Spanish on their short essays. Al-Maani et al. (2024) used think-alouds to examine the language learning strategies used by intermediate and advanced Jordanian English as a foreign language (EFL) college seniors as they performed reading, writing, and listening tasks.

Though obviously not exhaustive, the above overview of verbal protocol studies in literacy and math education establishes that the methodology has been used to examine a broad range of cognitive and metacognitive activities in an array of fields. Moreover, in educational research, this approach has been used successfully in both literacy and math (as well as in other subject areas) with numerous categories of students, including younger and older students, higher- and lower-achieving students, native language speakers and language learners, and students who are neurodivergent as well as students who aren't.

Verbal Protocols as Data in Research on the Designs of Large-Scale Standardized Assessments

Of particular relevance to the present study is the use of the think-aloud methodology to analyze and evaluate elements of the design of large-scale standardized assessments. One such study is that of Johnstone et al. (2006), who concluded that the cognitive lab methodology elicited useful information about construct-irrelevant barriers in math test design from several student population subgroups of educational concern, including students with learning disabilities, students with hearing impairments, and English learners, as well as from English-proficient students without disabilities. By contrast, the researchers found students with cognitive impairments lacked the requisite verbalization capacities during problem-solving. Of further note, the authors found the methodology yielded little data on the hardest math test items studied because of the difficulties participants had in simultaneously solving these problems and verbalizing their approaches. A similar study, this time by Johnstone et al. (2007), explored a variety of ways of making grade 8 reading items more comprehensible. Using a think-aloud methodology with recently promoted eighth-grade students, the team determined that the use of "non-construct vocabulary"—that is, undefined specialized subject area terms—could pose (correctable) barriers to student performance, while such interventions as reducing passage word counts and boldfacing key words didn't seem to influence achievement.

Threats to Verbal Protocol Validity and Reliability

Although the preceding account clearly establishes that verbal protocol analysis has been extensively used in social science research, including in education, serious concerns about the validity of the method have been raised over the years that require and have received fair-minded consideration and response.

One of the earliest and most influential critiques of verbal protocols as data came from Nisbett and Wilson (1977). Drawing from then-burgeoning critiques of introspection-based research methods, the authors posited three major conclusions:

1. "The accuracy of subjective reports [of higher-order thinking involving inferences] is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports.
2. "When reporting on the effects of stimuli, people may not interrogate a memory of the cognitive processes that operated on the stimuli; instead, they may base their reports on implicit, a priori theories about the causal connection between stimulus and response. . . .
3. "Subjective reports about higher mental processes are sometimes correct, but even the instances of correct report are not due to direct introspective awareness. Instead, they are due to the incidentally correct employment of a priori causal theories" (233).

Rather than outright rejecting these concerns, Ericsson and Simon (1993) countered with a simple mental processing model that differentiates between information stored in a person's short-term memory (STM) and long-term memory (LTM). Specifically, the authors contended that "information recently acquired (attended to or heeded) by the central processor is kept in STM, and is directly accessible for further processing (e.g., for producing verbal reports), whereas information from LTM must first be retrieved (transferred to STM) before it can be reported" (11). In other words, participants in verbal protocol studies should be able to give accurate accounts of their cognition during or shortly after experiencing a stimulus, such as a novel task to be solved; by contrast, verbal accounts that depend on recall and interpretation of past stimuli (i.e., that require, in Ericsson and Simon's model, retrieval from LTM) are more prone to the kinds of validity errors that Nisbett and Wilson (1977) identified.

Subsequent researchers have further codified potential threats to the accuracy of verbal protocols as data sources. Bainbridge and Sanderson (1995), for example, identified several ways in which verbal reports can be distorted, with the aim of encouraging researchers to find ways to minimize or eliminate these risk factors. Potential distortion sources identified by Bainbridge and Sanderson include the following:

1. Altering the nature and performance of a task merely by asking for a verbalization
2. Placing participants under significant time pressure, which can lead to glossing over steps in cognition
3. Social and self-presentation biases leading participants to give what they think are expected or socially acceptable answers

4. Asking participants to verbally discuss processes (e.g., perceptual-motor skills) that are typically performed nonverbally and outside of conscious thought
5. Participants being unable to articulate everything they know about and can do with a given stimulus (e.g., a problem-solving task), meaning that "verbal protocol evidence may provide only a limited sample of the total knowledge available to the person being studied" (173)

Stratman and Hamp-Lyons (1994) conceptualized threats to the accuracy of verbal protocols as problems of *reactivity*, or the verbal protocol methodology itself altering the cognitive processes intended to be studied. Challenges identified by the authors include flawed verbalization directions given to participants; the difficulty participants often experience in simultaneously thinking and verbalizing; the impact on participants of hearing their own voices during verbalization; the impact of participants learning about themselves during the verbalization process (rather than simply reporting); and the possibility of experimenters inadvertently cueing expected or desired responses through their words or actions. Similarly, Kirk and Ashcraft (2001, 158–59) identified three sources of threat to verbal protocol accuracy: veridicality ("whether the verbal reports accurately reflected the underlying cognitive processes"), reactivity ("the possibility that the verbal report requirement may have altered the mental processing that normally occurs"), and demand-induced bias ("the possibility that aspects of the experimental procedures suggested to participants what kinds of verbal reports and solutions were expected").

The consensus among researchers has been to treat issues of (in Kirk and Ashcraft's formulation) veridicality, reactivity, and demand-induced bias seriously without abandoning the methodology. For instance, Leow and Morgan-Short (2004), echoing Ericsson and Simon and others, suggest that verbal protocol approaches be limited to eliciting "introspective, nonmetalinguistic verbalizations" (36)—that is, verbalizations made concurrent with task performance, rather than retrospectively after the task, and focused on description of behaviors rather than attempts at explanations about why certain behaviors were performed. The researchers' study specifically examined whether the act of thinking aloud altered performance on a reading task given to college-age students and found no such evidence when students in the think-aloud and control (non-think-aloud) conditions were compared statistically. By contrast, Kirk and Ashcraft (2001), in their study of adult use of strategies in the solving of simple arithmetic problems and who also employed a "silent" control group, found questionable veridicality and signs of reactivity. (We speculate, along the lines of Bainbridge and Sanderson's [1995] cautions quoted above, that this outcome may have resulted in part because the task—simple arithmetic with college-age participants—was too routine, and therefore too far out of conscious understanding, for meaningful verbal protocol analysis.) They advocate for a careful analysis of instructions given to participants to minimize potential bias in response and for the use of a nonverbalizing control group to serve as a baseline. Russo et al. (1989) similarly call for the use of "silent" control conditions, as they found it impossible to determine a priori using then-existing theory which tasks were likely to provoke reactivity in participants.

Concurrent and Retrospective Verbalizations

The preceding discussion and the general research consensus (e.g., Russo et al. 1989) suggest that concurrent verbal protocols are more trustworthy than are retrospective ones. This stands to reason, as it should be easier for participants to accurately verbalize in-the-moment cognition during task performance than re-create their thought processes sometime after the fact. In accordance, the present study relies on concurrent verbal protocols and emphasizes description of behaviors performed by participants rather than the motivations behind their behaviors.

Some researchers, however, have made a case for a hybridized approach, one that makes use of both concurrent and retrospective dimensions. Johnstone et al. (2006) advocated for such a blended approach, contending that it counterbalanced both the propensity of think-aloud verbalizations to be "incoherent" (2) and that of interviews to elicit potentially inaccurate retrospective explanations of behaviors already encoded into long-term memory.

While noting several concerns about the use of data requiring participants to retrieve information from long-term memory, Taylor and Dionne (2000) advocate for the value of retrospective debriefing (RD) in tandem with concurrent verbal protocols (CVP), which they found obtained "a richer account of problem-solving strategy than did either method used alone." Specifically:

> When problem solvers are requested to think aloud while solving a problem (CVP), and then to describe how they solved the problem (RD), CVP data can be used to provide data-based cues to guide the collection of RD data on a specific problem-solving event. . . . In turn, convergent information about the same event contained in the broader spectrum of RD data can be used by researchers to elaborate CVP data, which tend to focus on the control of the problem-solving process. . . . Equally important are instances in which CVP and RD data diverge. These divergent reports offer opportunities for critical examination and clarification of both the problem solver's knowledge and the CVP and RD methodologies. As a result of using the two methodologies as complementary data sources, the richness of data available on a particular event is enhanced. (417)

In addition to the precautions various authors already cited have offered to increase the validity and reliability of concurrent verbal protocols, Taylor and Dionne (2000) propose additional considerations for limiting threats to the accuracy of retrospective debriefings. These include keeping the focus of questions on neutral and complete reportage; conducting the interview as close as possible in time to the experience itself; stressing with participants the need for accuracy; limiting the number of tasks asked about; focusing when possible on specific, important moments in the verbal protocols; using probes carefully to flesh out detail and check researcher understanding without being leading; and keeping the focus on description rather than interpretation ("'what' and 'which' rather than 'why'"; 417).

Methodological Implications for the Present Study

In a number of ways, the present study closely attends to the critiques levied against and cautions raised concerning the use of verbal protocols as data. First, the study was designed primarily to elicit what Leow and Morgan-Short (2004, 36) referred to as "introspective, nonmetalinguistic verbalizations" by recording participants' concurrent reports of their behaviors while answering test questions. Second, the study was designed to gather retrospective debriefing data, in the form of standardized postexperience interviews with participants, as a secondary data source while paying heed to Taylor and Dionne's (2000) recommendations for limiting reactivity in questioning. Third, the initial instructions given to participants for the concurrent verbal protocols were kept as simple and nondirective (in Taylor and Dionne's words, as "infrequent and neutral"; 415) as possible, and interviewers were directed to prompt students only when they had lapsed into silence for a period of time or were clearly working without verbalizing. Fourth, the tasks posed by the SAT Suite test questions given to participants are sufficiently nonroutine to be likely to evoke conscious, accurate reports of inline processing as participants work through them. Finally, the present study was originally conceived as a follow-up to a previously published cognitive lab study involving a cross section of the SAT Suite test-taking population (College Board 2024a), which meant that the results of a "control" group of sorts would have been available for comparison to the results of this study. However, it proved logistically impossible to administer the same test questions by the same means to the participants in this study as it was to the participants of the prior study and impractical to add a new control group, so the present study has to stand on its own.

Section 3: Methodology

Test Question Selection

College Board subject matter experts began the research process for this study by identifying sets of SAT Suite Reading and Writing and Math test questions that would represent as many of the key skill/knowledge elements of the test sections' designs as possible. Because the designs of and specifications for all SAT Suite tests—the SAT, PSAT/NMSQT®, PSAT™ 10, and PSAT™ 8/9—are intentionally similar (College Board 2024b), the selected questions as sets could fairly be said to represent those encountered in the suite as a whole rather than in just one of the tests.

Consistent with the approach used in a prior cognitive lab study (College Board 2024a), the present study intentionally excluded questions from the Reading and Writing section's Standard English Conventions content domain. Although facility with the conventions of Standard English is highly valued in academic and career settings, the strongly rule-based nature of tasks in this domain makes these questions unlikely to elicit rich responses from students in a verbal protocol setting, and College Board makes no strong claim about the cognitive complexity of these questions. All other Reading and Writing content domains and all Math content domains were represented by multiple test questions in the question sample selected.

Fifteen Reading and Writing questions and fifteen Math questions were ultimately selected for study. These questions were drawn from actual SAT Suite item pools rather than developed specifically for this study and were therefore representative of questions students might encounter on test day. For logistical reasons, all questions used in the study were drawn from a linear (nonadaptive) version of an extant SAT practice test form that had recently been made available to students. This choice increased somewhat the risk that one or more participants would have encountered these questions previously as part of full-form test practice (a point returned to in this report's subsection on study limitations in Section 5: Discussion), but it also ensured that participants were presented with questions in combinations that could organically occur as part of authentic testing (or authentic practice, as the same procedures used to generate operational test forms are used to produce official full-length practice tests).

Collectively, the Reading and Writing and Math question samples represent a wide range of content domains, skill/knowledge testing points, subject areas, question difficulty levels, stimulus text complexities (Reading and Writing only), and question formats consistent with the tests' designs. All questions used in the study, like all those of the SAT Suite, are discrete, meaning that no set-based questions were used and that each question could be answered independently of all others.

Table 1 summarizes the most salient characteristics of the Reading and Writing (RW) and Math test questions presented to participants in this study. An explanation of the table's columns immediately follows.

Table 1. Characteristics of Reading and Writing (RW) and Math Questions Presented to Study Participants.

Test Section	Q#	Content Domain	Skill/Knowledge Testing Point	Subject Area	TC (*RW only*)	PSB	Question Format
Reading and Writing	1	Craft and Structure	Words in Context	SCI	PSR	7	MC
	2		Text Structure and Purpose	LIT	MID	3	MC
	3		Text Structure and Purpose	HSS	PSR	7	MC
	4	Information and Ideas	Command of Evidence: Quantitative	SCI	SCO	4	MC
	5		Command of Evidence: Textual	LIT	SCO	4	MC
	6	Expression of Ideas	Transitions	HSS	SCO	5	MC
	7		Rhetorical Synthesis	HUM	MID	4	MC
	8		Rhetorical Synthesis	SCI	PSR	5	MC
	9	Craft and Structure	Words in Context	SCI	PSR	4	MC
	10		Cross-Text Connections	HUM	SCO	4	MC
	11	Information and Ideas	Central Ideas and Details	LIT	SCO	3	MC
	12		Central Ideas and Details	HUM	PSR	6	MC
	13		Command of Evidence: Textual	SCI	SCO	4	MC
	14		Command of Evidence: Quantitative	SCI	PSR	7	MC
	15		Inferences	HSS	MID	4	MC
Math	1	Algebra	Linear Inequalities: Identify	SCI		4	MC
	2	Problem-Solving and Data Analysis	Ratios	RWT		5	MC
	3	Geometry and Trigonometry	Circles	None		6	MC
	4	Advanced Math	Nonlinear Functions: Rewrite	None		7	MC
	5	Problem-Solving and Data Analysis	Percentages	None		7	MC
	6	Advanced Math	Nonlinear Functions: Make Connections	None		7	MC
	7	Algebra	Linear Functions: Identify	SCI		2	MC
	8	Geometry and Trigonometry	Measures of Angles in a Triangle	None		3	MC
	9	Advanced Math	Nonlinear Functions: Interpret	SCI		4	MC
	10	Problem-Solving and Data Analysis	Scatterplot	None		4	MC
	11	Problem-Solving and Data Analysis	Probability	RWT		4	MC
	12	Advanced Math	Nonlinear Equations: Solve	None		5	SPR
	13	Algebra	Linear Equations in Two Variables: Make Connections	None		5	SPR
	14	Geometry and Trigonometry	Scale Factor and Area	None		6	MC
	15	Algebra	Systems of Two Linear Equations in Two Variables: Solve	None		6	SPR

Table 1 displays key traits of each of the SAT test questions used in this study.

- **Test section.** Reading and Writing or Math
- **Q#.** Question number (1–15), representing the order in which the questions were presented to participants
- **Content domain.** One of the major conceptual divisions within each of the two test sections: Information and Ideas, Craft and Structure, and Expression of Ideas in Reading and Writing; Algebra, Advanced Math, Problem-Solving and Data Analysis, and Geometry and Trigonometry in Math
- **Skill/knowledge testing point.** The skill/knowledge element targeted by the question (e.g., Words in Context in Reading and Writing; Probability in Math)
- **Subject area.** The content area, if any, sampled by the question: literature (LIT), history/social studies (HSS), the humanities (HUM), or science (SCI) in Reading and Writing; science (SCI) or real-world topics (RWT) in Math. (Social studies, a third content area sampled by SAT Suite Math questions, was not represented.) Math questions with a subject area of "None" test aspects of "pure" mathematics outside of context.
- **TC.** Stimulus text complexity. Reading and Writing test passages (only) are formally rated for text complexity by College Board subject matter experts using both quantitative and qualitative means. Passages developed for the section fall into one of three categories:
    - MID: Middle school/junior high school level (equivalent to grades 6–8)
    - SCO: Upper secondary level (grades 9–11)
    - PSR: Postsecondary readiness level (grades 12–14)
- **PSB.** Performance score band, a numerical rating of a question's statistical difficulty aligned to the test sections' scales. In SAT Suite terms, questions in PSBs 1 to 3 are considered easy and are associated with Reading and Writing section scores from 200 (the lowest possible) to 480 and with Math section scores from 200 to 460 (out of 800, in ten-point intervals). Questions in PSBs 4 and 5 are considered medium difficulty and are associated with Reading and Writing section scores from 490 to 600 and with Math section scores from 470 to 600. Questions in PSBs 6 and 7 are considered hard and are associated with Reading and Writing and Math section scores from 610 to 800. Each test section's question sample included questions typically ranging in PSB from 3 to 7; with one exception in Math, questions in PSBs 1 and 2 were excluded from selection, as the research literature (e.g., Bainbridge and Sanderson 1995) suggests that such relatively cognitively simple tasks are unlikely to elicit much conscious thought from test takers.
- **Question format.** All Reading and Writing questions, both in the study and on the actual SAT Suite tests, are in the four-option multiple-choice (MC) format, with each question having a single best answer (*key*). Math questions are either in this same MC format or in the student-produced response (SPR) format, for which students must generate and enter their own answers without the structure and support of provided answer choices.

As a group, the fifteen sampled Reading and Writing questions represented three of the section's four content domains (with Standard English Conventions being excluded, as previously noted), all major skill/knowledge testing points within those three domains, all four sampled subject areas, all three sampled stimulus text complexity levels, and all levels of difficulty from 3 (easy) to 7 (hard). As a group, the Math questions represented all four of the section's content domains, many skill/knowledge testing points within those domains, in-context questions representing two of three sampled subject areas as well as questions set outside of context, all levels of difficulty from 2 to 7, and both multiple-choice and student-produced response formats.

In addition to the fifteen Reading and Writing and fifteen Math questions formally presented to participants, three questions from each section were incorporated into participant training. Before a given participant did their own thinking aloud on the fifteen study questions in either Reading and Writing or Math, the session moderator, following a script, exemplified thinking aloud through a sample question from the same section, after which the participant would have one or (if deemed necessary by the moderator) two opportunities to practice thinking aloud themselves before beginning the actual question set. These training questions were drawn from the same practice test form from which all other questions were taken and can be found in the appendix. The practice portions of sessions were neither recorded nor analyzed.

Question Type–Level Construct Definition

The College Board subject matter experts who selected the questions for the study also identified constructs for the questions by skill/knowledge testing point. These *constructs*, in the form of lists of behaviors demonstrable by test takers, describe the kinds of cognitively complex thinking students are expected to exhibit if they approach answering the questions as intended by the test developers.

For each Reading and Writing testing point (e.g., Words in Context), staff developed a list of behaviors test takers were required to exhibit in order to answer each question as intended. Because many Math questions include, by design, multiple and often mutually exclusive pathways test takers may pursue in answering, these behaviors were defined as expected rather than required, and participants needed only to exhibit at least one of them to be considered as having enacted the construct. Answering correctly was always a required Reading and Writing behavior; for Math, participants' correct and incorrect answers for each question were tracked separately from the behavior list. Additionally, both Reading and Writing and Math staff identified generic sets of common behaviors that skillful test takers may or may not exhibit while answering questions; these optional behaviors were coded for but not analyzed in this report.

These construct definitions (lists of behaviors) can be found with their associated test questions in Section 4: Results.

The constructs (required/expected behaviors) used for this study are highly similar to the ones used in previous research (College Board 2024a), with some refinements made to better reflect learnings from the prior study.

Protocol Development

The lead author of this study, in collaboration with other College Board researchers and vendor Vidlet, Inc., developed closely parallel Reading and Writing and Math protocols for conducting the cognitive interviews in which students would participate. These protocols were designed as guides for the moderators conducting sessions with participants. The guides included general instructions for conducting the sessions, scripts for moderators to follow, and suggested probes and prompts that moderators could use during sessions should participants lapse into extended silence while working through the test questions. Consistent with best practices (as discussed in Section 2: Literature Review), moderators were directed to limit probes and prompts as much as possible and to make them as nondirective as possible (e.g., "Please keep thinking aloud") so as not to unduly influence participants' responses. Moderators were also advised against asking participants to clarify or explain their responses, as such would divert participants from direct, concurrent reporting of their thinking and actions in the moment to less reliable retrospective inferences. Vidlet moderators were briefed and trained on the protocol and given multiple opportunities to provide feedback and suggest refinements.

Test Question Delivery Method

SAT Suite test questions presented to participants during the think-aloud activity (including its training portion) were administered via Bluebook™, the custom-built test application College Board uses to give the SAT Suite tests in their standard digital-adaptive form. The use of Bluebook, which most students use to take SAT Suite tests and engage in full-form practice, enhanced the study's verisimilitude, gave participants ready and standardized access to the universal tools available in Bluebook (including a built-in version of the Desmos® graphing calculator for the Math section), and overall represented a methodological improvement relative to the prior cognitive lab study investigating students' interactions with the digital-adaptive SAT Suite (College Board 2024a), but it did come with its own limitation. In contrast to the prior study, in which only the focal test questions (and training questions) were presented via a third-party digital survey tool, participants in this study had to "skip around" to the specific focal questions, as directed by a moderator following the protocol script. On very rare occasions, this resulted in participants being misdirected to an "incorrect" question (i.e., one in the test form being used but not one of the focal questions); these few instances, as well as a small number of additional cases in which participants ran out of time to answer particular questions, are effectively discounted by the methodology, as the metrics calculated consider only numbers and proportions of correctly answered questions. To account for the fact that the digital-adaptive test sections are divided into two separately timed modules and that test takers can't return to the first module once they've moved on to the second, moderators were directed to inform participants they could review their responses (or lack of responses) to the focal questions in the first module before advancing to the second.

Tools Available to Participants

All participants in both the Reading and Writing and Math segments of this study had access to the full range of universal tools available in Bluebook (see

College Board 2024b, section 2.2.7.2). This suite of tools includes a graphing calculator built into the app and available for the Math section (only); alternatively, participants could make use of their own handheld calculators, provided those devices conformed to College Board's SAT Suite calculator policy. Participants in the Math segment also had access to a set of common formulas and could make use of scratch paper.

For this EL study, participants were allowed to bring a family member or friend to the sessions to act as a translator when interacting with the moderator; no participants in either the Reading and Writing or Math segment did so. Moderators were also allowed, per the protocol for this study, to employ extended prompting to provide structure and support as needed for participants, particularly those with low English language acquisition levels, as they thought aloud; no moderator for either segment recorded having done so.

Sample Definition, Selection Criteria, Recruitment, and Characteristics

SAMPLE DEFINITION

For its 2024 cognitive lab studies, College Board sought members of the SAT test-taking population who fit into one (or possibly more) of three categories: students with a specific learning disorder affecting reading (also known as dyslexia) (College Board 2025a), students with attention deficit hyperactivity disorder (ADHD) (College Board 2025b), and students who were English learners (ELs). The present study reports the results of the study involving students who identified as being ELs.

As part of the sample selection screener (see appendix), prospective participants were asked to indicate whether they were English learners (and/or had ADHD or a specific learning disorder affecting reading). If the answer was "yes," they were further asked to specify (1) how often they communicate in English in daily life (often, sometimes, rarely), (2) which language(s) they typically speak at home (only in English, only in a language other than English, in English and one or more other languages), (3) which language(s) other than English they know well, and (4) their current level of English language acquisition, rated on a six-point scale (derived from the Common European Framework of Reference for Languages [CEFR] version 3.3 [Council of Europe, n.d.]) ranging from "I can understand familiar everyday expressions and very basic phrases in English" to "I can easily understand nearly any text in English." ELs were also told that they could ask a family member or friend to act as a translator for all or part of the think-aloud activity if they so wished and to indicate whether they would or wouldn't likely do so.

This self-identification method did raise the possibility that one or more participants would self-identify as being an EL when they weren't merely to participate in the study and receive its incentive. To militate against this possibility, the screener didn't specify that ELs were being sought, and the initial query about being an EL was mingled with other possible conditions and statuses, including ones not expressly sought for this or other studies (e.g., students who are deaf or hard of hearing). As it turned out, observations of students in both the Reading and Writing and Math conditions gave no evidence of self-misidentification.

SAMPLE SELECTION CRITERIA

Prospective participants were deemed eligible for selection if they met the following criteria:

- They were students in either grade 11 or 12.
- They attended school in the United States.
- They answered "yes" when asked whether they were an EL.
- They provided other required demographic information, including gender, race/ethnicity, and self-reported high school GPA (HSGPA).[1]

  Note: Students were allowed to indicate that they preferred not to respond to the gender and/or race/ethnicity questions without being excluded from consideration.
- They were willing and able to productively participate in a virtual cognitive interview session of up to 120 minutes in length.

Self-reported HSGPA was used as the proxy for student academic achievement in this study. This was necessary because, as discussed immediately below, Vidlet operated as the primary student recruiter, and it was therefore not possible, for logistical and privacy reasons, to link prospective participants to any previous SAT Suite scores they may have had on file with College Board. Students were asked on the screener to report past SAT or PSAT-related test scores, but doing so was not a requirement, and as all students (with two exceptions in Reading and Writing) provided HSGPA information while not all students provided self-reported SAT/PSAT test scores, the latter weren't considered in this study. This is theoretically a limitation of the study, but evidence (e.g., Sanchez and Buddin 2016) suggests that self-reported HSGPAs are generally sufficiently accurate for research purposes.

SAMPLE RECRUITMENT

In June 2024, College Board contacted vendor Vidlet, Inc., an organization that had successfully aided in a prior cognitive lab study (College Board 2024a), to support a research initiative to learn more about how students from various subpopulations of interest—students with a specific learning disorder affecting reading (also known as dyslexia), students with attention deficit hyperactivity disorder (ADHD), and students who are English learners—experience SAT Suite testing.

Prior to recruitment, College Board and Vidlet jointly worked on a sample selection screener (survey) that would be given electronically to prospective participants to complete (see appendix). This screener was designed to collect eligibility information as well as a limited range of demographic detail (e.g., grade in school, gender, race/ethnicity) intended to ensure breadth in sample selection. Demographic survey items deemed potentially sensitive (e.g., gender, race/ethnicity) included a "prefer not to respond" option, and choosing this didn't disqualify the candidate from consideration.

Also prior to recruitment, College Board determined that an incentive of $150 per participant would fairly compensate students for their time and effort. This incentive would come in the form of a gift card, which could be used in a variety of ways.

[1] Two Reading and Writing participants (RW47 and RW61) failed to provide HSGPA and were inadvertently included in the study.

Vidlet recruited students primarily through its panel and email outreach processes; a small number of additional potential contacts were provided by College Board. The recruitment solicitation (see appendix) highlighted that participants would have an opportunity to provide feedback to influence SAT testing and that they'd receive an incentive of $150 on successful completion of the activity. After initial intake by the Vidlet team, participant information was de-identified and sent to College Board to ensure as diverse a selection as possible (given small sample sizes) by gender, race/ethnicity, geography, and self-reported HSGPA. Recruitment occurred on a rolling basis, meaning that some students were interviewed while others were still being identified.

Once students had confirmed their participation in the study, Vidlet collected consent forms (see appendix). These consent forms, which were either signed by students themselves (if they were age eighteen or over) or a parent/guardian (if not), described the nature of the activity, explained what participants would be asked to do, and made participants aware that they could opt out of some or all of the activity for any reason if they so chose (although successful completion of the activity was required to receive the incentive).

Participants were then assigned randomly by Vidlet to either the Reading and Writing or Math segment. Each segment consisted of two main elements: (1) a think-aloud portion, in which participants shared their thoughts concurrently as they worked through a set of SAT Suite test questions and (2) a postexperience interview using a standardized set of questions focused on participants' impression of the think-aloud activity as well as self-identified sources of challenge in answering particular questions or categories of questions. Collectively, these components were scheduled to take no more than 120 minutes.

Recruitment and interviewing for this EL-focused study took place concurrently with recruitment and interviewing for two other cognitive lab studies: those involving students with a specific learning disorder affecting reading (College Board 2025a) and students with ADHD (College Board 2025b). Over the course of approximately four months, the Vidlet research team led a total of about 120 students, divided roughly equally across the three subgroups of interest, through cognitive interview sessions structured according to protocol documents developed by College Board and vetted by Vidlet. No student was allowed to participate in more than one study.

SAMPLE CHARACTERISTICS

Reading and Writing

Table 2 displays the roster of Reading and Writing participants. For each participant, the table includes the participant identifier (a unique code used in place of a student's name); demographic information, including the participant's gender, race, ethnicity, home state, grade in school, and self-reported HSGPA; and information about the participant's EL status, including (1) how often they communicated in English in their daily lives, (2) which language(s) they typically speak at home, (3) which language(s) other than English they know well, and (4) their current level of English language acquisition. Definitions for "Language(s) Spoken at Home" and "English Language Acquisition Level" are discussed below.

Table 2. Reading and Writing Participant Roster by Demographics and EL Status.

	Demographics						EL Status				
Part. ID	Gender	Race	Ethnicity	Home State	Grade in School	Self-Reported HSGPA	EL?	Freq. of English Communication	Language(s) Spoken at Home	Known Language(s) Besides English	English Language Acquisition Level
RW5	Female	Asian	Not of Hispanic, Latino, or Spanish origin	TX	12	A– (90–92)	Yes	Often	Another	Vietnamese	5
RW25	Male	Black or African American	Not of Hispanic, Latino, or Spanish origin	GA	11	A+ (97–100)	Yes	Often	Another	Arabic	1
RW28	Male	Asian	Not of Hispanic, Latino, or Spanish origin	NY	11	A (93–96)	Yes	Often	Another	Bangla	2
RW29	Male	Asian	*NR*	TX	11	B– (80–82)	Yes	Often	Another	Pashto, Farsi, Urdu	3
RW31	Female	Black or African American	Not of Hispanic, Latino, or Spanish origin	TX	11	B+ (87–89)	Yes	Often	English and other	Tigrinya, Amharic	4
RW35	Female	White	Puerto Rican	PA	12	A+ (97–100)	Yes	*NR*	*NR*	Spanish	*NR*
RW46	Male	Asian	Not of Hispanic, Latino, or Spanish origin	IL	12	A+ (97–100)	Yes	*NR*	*NR*	Gujarati	*NR*
RW47	Female	White	Not of Hispanic, Latino, or Spanish origin	CA	11	*NR*	Yes	Often	English	French	6
RW48	Female	Asian	Not of Hispanic, Latino, or Spanish origin	CA	11	B (83–86)	Yes	Often	Another	Vietnamese	3
RW49	Male	Black or African American	Not of Hispanic, Latino, or Spanish origin	FL	12	A (93–96)	Yes	Sometimes	Another	Amharic	1
RW50	Male	Asian	Not of Hispanic, Latino, or Spanish origin	NJ	11	A (93–96)	Yes	Often	Another	Malayalam, Tamil	3
RW51	Female	Asian	Not of Hispanic, Latino, or Spanish origin	TX	11	A (93–96)	Yes	*NR*	*NR*	Telugu, Hindi	*NR*
RW52	Female	Black or African American	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	12	A+ (97–100)	Yes	*NR*	*NR*	Spanish	*NR*
RW53	Male	*NR*	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	TX	12	B+ (87–89)	Yes	Often	Another	Spanish	3
RW54	Female	White	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	12	A+ (97–100)	Yes	*NR*	*NR*	Spanish	*NR*
RW55	Female	Asian	Not of Hispanic, Latino, or Spanish origin	HI	12	A (93–96)	Yes	Often	Another	Mandarin, Cantonese	3
RW57	Female	Asian	Not of Hispanic, Latino, or Spanish origin	VA	11	A (93–96)	Yes	Often	Another	Mandarin, Cantonese	3
RW59	Male	Asian	Not of Hispanic, Latino, or Spanish origin	NJ	12	A+ (97–100)	Yes	Often	Another	Mandarin	3
RW60	Female	White	Not of Hispanic, Latino, or Spanish origin	NE	11	A+ (97–100)	Yes	Often	English and other	Arabic	5
RW61	Female	Native American or Alaska Native	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	TX	11	*NR*	Yes	Often	English and other	Spanish	4

NR: No response

Definitions for Language(s) Spoken at Home (“In which language[s] do you typically speak at home?”):

- English = Only in English
- Another = Only in a language other than English
- English and other = In English and one or more other languages

Definitions for English Language Acquisition Level (“Which of the following best describes your current level of English language acquisition?”):

- 1 = I can understand familiar everyday expressions and very basic phrases in English.
- 2 = I can understand sentences and frequently used expressions in English.
- 3 = I can understand the main points of clear texts on familiar subjects in English.
- 4 = I can understand the main ideas of complex texts in English.
- 5 = I can understand a wide range of demanding, longer texts in English.
- 6 = I can easily understand nearly any text in English.

Table 2 suggests that within the strictures of the small sample size (*n* = 20) and self-selection methodology used for this study, Reading and Writing participants represented a relatively diverse sample in terms of gender, race/ethnicity, grade in school, and EL status but didn't vary much in terms of self-reported HSGPA. Specifically:

- **Gender.** Female participants (twelve) were somewhat overrepresented relative to male participants (eight).
- **Race and ethnicity.** Most numerous were Asian participants not of Hispanic, Latino, or Spanish origin (nine), followed by Black or African American participants not of Hispanic, Latino, or Spanish origin (three) and White participants not of Hispanic, Latino, or Spanish origin (two). Together, these accounted for fourteen of twenty participants. Other races/ethnicities identified by participants (one each) were Asian/No response; Black or African American/ Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican; Native American or Alaska Native/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican; White/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican; White/Puerto Rican; and No response/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican. One racial category (Native Hawaiian or other Pacific Islander) wasn't represented, which constitutes a limitation on the study (see Section 5: Discussion).
- **Grade in school.** Students from both grade 11 (eleven) and grade 12 (nine) were represented.
- **Self-reported HSGPA.** Fourteen participants indicated an "A" HSGPA, four indicated a "B" HSGPA, and two didn't provide a response. This represents a limited range of high school achievement relative to the study's goal to be as representative as possible, within small sample size and self-selection limitations, of the EL subpopulation. While the sample is biased toward higher HSGPAs, this outcome should be considered in the context of grade inflation generally (e.g., Sanchez 2024), which suggests that we should expect to see fewer students overall with average-and-below HSGPAs.
- **EL status.** All participants reported being an English learner. Most participants (fourteen) reported often speaking English in their daily lives, while one participant reported sometimes doing so and five preferred not to respond. Participants indicated a range of languages known other than English. In terms of English language acquisition level, two participants reported level 1 ("I can understand familiar everyday expressions and very basic phrases in English"), one reported level 2 ("I can understand sentences and frequently used expressions in English"), seven reported level 3 ("I can understand the main points of clear texts on familiar subjects in English"), two reported level 4 ("I can understand the main ideas of complex texts in English"), two reported level 5 ("I can understand a wide range of demanding, longer texts in English"), one reported level 6 ("I can easily understand nearly any text in English"), and five opted not to respond.

Math

Following the same approach as for the Reading and Writing participant roster, table 3 displays the roster of Math participants.

Table 3. Math Participant Roster by Demographics and EL Status.

	Demographics						EL Status				
Part. ID	Gender	Race	Ethnicity	Home State	Grade in School	Self-Reported HSGPA	EL?	Freq. of English Communication	Language(s) Spoken at Home	Known Language(s) Besides English	English Language Acquisition Level
M8	Female	*NR*	Puerto Rican	TX	11	A+ (97–100)	Yes	Often	English and other	Spanish	6
M9	Female	White	Not of Hispanic, Latino, or Spanish origin	IL	12	A+ (97–100)	Yes	Often	English and other	Polish	4
M25	Female	Black or African American	Not of Hispanic, Latino, or Spanish origin	CA	11	B+ (87–89)	Yes	Often	English and other	Swahili	1
M29	Female	White	Not of Hispanic, Latino, or Spanish origin	FL	12	A+ (97–100)	Yes	Often	Another	Russian, Azerbaijani	5
M35	Female	Asian	Not of Hispanic, Latino, or Spanish origin	VA	11	A (93–96)	Yes	Often	English and other	Vietnamese	4
M36	Male	Asian	*NR*	TX	11	B+ (87–89)	Yes	Often	Another	Dari, Urdu	2
M41	Male	Asian	Not of Hispanic, Latino, or Spanish origin	MI	12	A+ (97–100)	Yes	Often	Another	Thai	3
M45	Female	*NR*	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	TX	11	A– (90–92)	Yes	Often	Another	Spanish	3
M48	Female	*NR*	Mexican	TX	12	A+ (97–100)	Yes	Often	Another	Spanish	2
M49	Female	Asian	Not of Hispanic, Latino, or Spanish origin	FL	11	A+ (97–100)	Yes	Often	English and other	Vietnamese	6
M50	Male	Asian	Not of Hispanic, Latino, or Spanish origin	PA	11	A (93–96)	Yes	*NR*	*NR*	Hindi	*NR*
M51	Male	White	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	NY	12	A (93–96)	Yes	Often	Another	Spanish	5
M52	Female	*NR*	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	12	A+ (97–100)	Yes	Often	Another	Spanish	5
M53	Male	Asian	Not of Hispanic, Latino, or Spanish origin	TX	11	A (93–96)	Yes	*NR*	*NR*	Hindi, French	*NR*
M54	Female	White	Mexican	TX	11	A+ (97–100)	Yes	Often	Another	Spanish	3
M55	Male	Black or African American	Puerto Rican	NY	12	A (93–96)	Yes	Rarely	Another	Spanish	4
M57	Female	*NR*	Mexican	MD	12	A+ (97–100)	Yes	Often	Another	Spanish	2
M58	Female	Asian	Not of Hispanic, Latino, or Spanish origin	FL	12	B+ (87–89)	Yes	Often	Another	Hindi	3

NR: No response

Definitions for Language(s) Spoken at Home ("In which language[s] do you typically speak at home?"):

English = Only in English

Another = Only in a language other than English

English and other = In English and one or more other languages

Definitions for English Language Acquisition Level ("Which of the following best describes your current level of English language acquisition?"):

1 = I can understand familiar everyday expressions and very basic phrases in English.

2 = I can understand sentences and frequently used expressions in English.

3 = I can understand the main points of clear texts on familiar subjects in English.

4 = I can understand the main ideas of complex texts in English.

5 = I can understand a wide range of demanding, longer texts in English.

6 = I can easily understand nearly any text in English.

Table 3 suggests that within the strictures of the small sample size (*n* = 18) and self-selection methodology used for this study, Math participants represented a somewhat diverse sample in terms of gender, race/ethnicity, grade in school, and EL status but didn't vary much in terms of self-reported HSGPA. Specifically:

- **Gender.** Female participants (twelve) were disproportionately represented relative to male participants (six).
- **Race and ethnicity.** Six participants identified as Asian not of Hispanic, Latino, or Spanish origin; two as White not of Hispanic, Latino, or Spanish origin; two as No response/Mexican; and two as No response/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican. Together, these made up twelve of the eighteen participants. Other races/ethnicities identified by participants (one each) were Asian/No response; Black or African American/Puerto Rican; Black or African American/Not of Hispanic, Latino, or Spanish origin; White/Mexican; White/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican; and No response/Puerto Rican. Two racial categories (Native Hawaiian or other Pacific Islander; Native American or Alaska Native) weren't represented, which constitutes a limitation on the study (see Section 5: Discussion).
- **Grade in school.** Students from grades 11 and 12 were equally represented (nine each).
- **Self-reported HSGPA.** Fifteen participants indicated an "A" HSGPA, and three indicated a "B" HSGPA. This represents a limited range of high school achievement relative to the study's goal to be as representative as possible, within small sample size and self-selection limitations, of the EL subpopulation. While the sample is biased toward higher HSGPAs, this outcome should be considered in the context of grade inflation generally (e.g., Sanchez 2024), which suggests that we should expect to see fewer students overall with average-and-below HSGPAs.
- **EL status.** All participants reported being an English learner. Most participants (fifteen) reported often speaking English in their daily lives, while one participant reported rarely doing so and two preferred not to respond. Participants indicated a range of languages known other than English. In terms of English language acquisition level, one participant reported level 1 ("I can understand familiar everyday expressions and very basic phrases in English"), three reported level 2 ("I can understand sentences and frequently used expressions in English"), four reported level 3 ("I can understand the main points of clear texts on familiar subjects in English"), three reported level 4 ("I can understand the main ideas of complex texts in English"), three reported level 5 ("I can understand a wide range of demanding, longer texts in English"), two reported level 6 ("I can easily understand nearly any text in English"), and two opted not to respond.

Coding and Analysis

CODING

The lead College Board researcher uploaded the interview transcripts generated by Vidlet into MAXQDA, a qualitative/mixed-methods research software package.

Reading and Writing and Math teams, using MAXQDA's cloud service, then coded each transcript against the previously defined required (Reading and Writing) / expected (Math) and optional behaviors associated with the question types' constructs. In cases in which transcripts were vague or ambiguous (e.g., the participant didn't verbalize the answer they selected or entered but had answered in Bluebook), the research team consulted the video recordings to confirm participant behaviors and answer choices.

Team members were also directed to code as "vignette candidates" any participant response that exhibited all required behaviors (Reading and Writing) / at least one expected behavior (Math) and that served to illustrate well-reasoned responses without significant errors, omissions, or uncorrected missteps. We elected to adopt a "case study" approach for the presentation of such vignettes in Section 4: Results, sharing transcript excerpts from a single participant in Reading and Writing and in Math and supplementing those excerpts with those from other participants when the case study participant failed to demonstrate adequate behaviors and/or failed to answer a given question correctly.

As a supplement to MAXQDA, the team concurrently recorded, in Microsoft Excel, whether each participant had answered each question correctly and exhibited each of the required/expected behaviors for the questions; these Excel spreadsheets served as the basis for calculating the statistics presented in Section 4: Results. The coding process resulted in approximately twenty-two hundred codes being assigned to thirty-eight participants' interactions with the thirty studied questions across Reading and Writing and Math.

ANALYSIS

The College Board researchers then analyzed the coded data to assess in various ways both participant and test question performance, as elicited from the think-aloud activity, as well as participant perceptions of their simulated test-taking experience, as elicited from postexperience interview questions.

1. **Participant performance** was analyzed in terms of the number and proportion of correctly answered questions for which participants demonstrated appropriate cognitive behaviors. Vignettes (transcript excerpts) from select participants were used to illustrate demonstrations of the cognitively complex thinking elicited by the test questions.
2. **Question performance** was analyzed in terms of the number and proportion of correctly answering participants who also demonstrated appropriate cognitive behaviors.
3. **Participant perceptions** of the think-aloud activity, in the form of responses to postexperience interview questions, were analyzed for both general themes and for any cases in which participants identified potential construct-irrelevant barriers to their success in the activity and to SAT Suite test taking more broadly.

Each of these approaches is discussed in turn below.

Participant Performance

Participant performance on each Reading and Writing or Math question was assigned a *performance level* (PL) from 1 to 5 based on two intersecting considerations: whether the participant answered the question correctly and whether appropriate behaviors were demonstrated.

Table 4 displays the definitions of the five performance levels in Reading and Writing and in Math.

Table 4. Participant Performance Level (PL) Definitions.

Performance Level	Definition	
	Reading and Writing	Math
1 (highest)	Answered correctly; demonstrated all required behaviors	Answered correctly; demonstrated at least one expected behavior
2	Answered correctly; demonstrated fewer than all required behaviors	*Not applicable; see below*
3	Answered correctly; demonstrated no other required behaviors	Answered correctly; demonstrated no expected behaviors
4	Answered incorrectly; demonstrated at least one other required behavior	Answered incorrectly; demonstrated at least one expected behavior
5 (lowest)	Answered incorrectly; demonstrated no other required behaviors	Answered incorrectly; demonstrated no expected behaviors

PL 2 was sometimes attainable in Reading and Writing and always unobtainable in Math given the previously discussed differences between required (Reading and Writing) and expected (Math) behaviors, as Math participants received a PL of 1 if they demonstrated at least one expected behavior. PL 2 was unobtainable in Reading and Writing when a given question type had only two required behaviors, one of which was (always) answering correctly.

In Section 4: Results, performance levels are displayed in figures, with each cell representing a participant-by-question interaction. PLs are indicated by number (1–5) and by supplementary color shading. Unobtainable PL 2s are indicated by a dash ("–").

Using these performance level findings, the research team calculated what this study refers to as the *participant differential*, or D_p, for each participant. Mathematically, D_p is represented by the following formulas:

$$\text{Reading and Writing: } D_p = \#AC - \#RB$$

$$\text{Math: } D_p = \#AC - \#EB$$

In these formulas, D_p is the participant differential, *#AC* is the total number of questions a given participant answered correctly, and *#RB* and *#EB* are, respectively, the number of correctly answered questions for which the participant also demonstrated all required behaviors (Reading and Writing) or at least one expected behavior (Math). D_p is always either zero or a positive integer except in the rare circumstance (not encountered in this particular study) in which a participant answered no questions correctly, in which case no "true" differential exists. In performance level terms, *#RB* and *#EB* represent PL 1.

Conceptually, D_p represents the "difference" between simply answering questions correctly and doing so while also exhibiting the cognitive behaviors intended by the test developers. D_p is thus a more appropriate and robust measure of participant performance than is the raw number of questions answered correctly because D_p, in essence, removes from consideration those questions that participants may have answered correctly by means other than those intended by the test makers (e.g., by random guessing or by finding a "shortcut" past the intended intellectual activity). Additionally, D_p considers only questions actually answered, meaning that unanswered questions have no meaningful effect.

Zero or low participant differentials are desirable, as ideally each participant answered questions correctly only by enacting the question types' constructs. Owing to the sometimes variable number of participants who answered each Reading and Writing or Math question, the threshold for a "good" differential was set at 70 percent or greater—meaning, for example, that if a participant answered all fifteen Reading and Writing or Math questions correctly, they would also have needed to have demonstrated all required behaviors on at least eleven of these questions (73 percent) to yield a "good" differential (in this example, 4 or lower). The "70 percent or greater" threshold is somewhat arbitrary, but it does represent a significant majority of correctly answered questions being responded to in ways that enact the question type–level constructs while at least partially accounting for the possibility that a given participant may well have understood how to "properly" answer a particular question but may simply have not verbalized one or more elements of doing so (essentially "underreporting" their skills and knowledge owing to the artificiality of the simulated testing experience and/or their lack of familiarity and comfort with thinking aloud).

To illustrate and concretize the cognitively complex thinking required to answer each of the studied test questions, the research team identified during coding cases in which participants exhibited exemplary (if not necessarily "perfect") reasoning in accordance with the question type's construct. These "vignettes" (transcript excerpts) are presented primarily in the form of a case study of a single participant as they answered each of the Reading and Writing or Math questions. For questions for which the case-study participant failed to demonstrate appropriate behavior(s), supplementary vignettes from other participants are provided.

Question Performance

The performance of the test questions themselves in the study was subjected to an analysis similar to that used for participant performance. To assess question performance, the research team calculated what this study refers to as the *question differential*, or D_q, which can be represented by the following formulas:

$$\text{Reading and Writing: } D_q = \#AC - \#RB$$

$$\text{Math: } D_q = \#AC - \#EB$$

In these formulas, D_q is the question differential, *#AC* is the total number of participants answering a given question correctly, and *#RB* and *#EB* are, respectively, the number of correctly answering participants who also demonstrated all required behaviors (Reading and Writing) or at least one expected behavior (Math). In performance level terms, *#RB* and *#EB* again represent PL 1.

Conceptually, D_q is closely analogous to D_p in that the former “discounts” from consideration instances in which participants correctly answered a given question without demonstrating appropriate cognitive behaviors. Zero to low differentials are again considered desirable, a result of no “true” differential could occur (but didn’t in this study) when no participant answered a given question correctly, and the same 70 percent-or-greater threshold for “good” differentials applies here, this time meaning that for each question, 70 percent or more of correctly answering participants also demonstrated all required behaviors/at least one expected behavior. Like D_p, D_q is concerned with the number of answered questions only, thus mitigating the effect of omitted responses.

Participant Perceptions

All participants were asked the following postexperience interview questions immediately after completing the think-aloud activity in Reading and Writing or Math:

1. Please tell me a bit about the experience you just had. What was it like to answer those questions?
2. How would you describe your general approach, in terms of strategies, for answering the questions?
3. Was there a particular type of question that you found especially easy to answer? If so, which one and why?
4. Was there a particular type of question that you found especially hard to answer? If so, which one and why?
5. Did you encounter anything in the questions that you had difficulty with given your comfort level with the English language? If so, what was it, and why was it difficult for you?
6. Is there anything about your test-taking experience today or about the test-taking strategies you used today that we haven’t talked about yet but that you’d like us to know?

Questions 1 and 6 were designed as open-ended prompts for participants to share anything on their minds about the think-aloud experience. Question 2 concerned general test-taking strategies used in the think-aloud activity. Questions 3 and especially 4 and 5 were more precisely targeted to elicit participant perceptions of potential construct-relevant and construct-irrelevant impediments to their successful performance in the activity.

Participants’ responses to these postexperience interview questions are summarized in Section 4: Results.

Section 4: Results

Reading and Writing

PARTICIPANT AND QUESTION PERFORMANCE

Participant and Question Performance Levels and Differentials

Figure 1 displays, as a single matrix, the Reading and Writing participant and question performance data derived from this study. The intended method of reading the figure is discussed immediately following.

Figure 1. Reading and Writing Participant and Question Performance Summary Matrix.

Part. ID	Question #															Performance by Level, by Participant						Participant Performance Summary		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**1**	**2**	**3**	**4**	**5**	***NR***	***#AC***	***#RB***	D_p
RW5	4	1	4	2	1	1	2	4	1	1	1	1	1	4	1	9	2	0	4	0	0	11	9	2 ✔
RW25	3	1	4	1	4	5	1	2	1	3	–	5	5	2	–	4	2	2	2	3	2	8	4	4 ✘
RW28	4	1	4	1	4	4	4	2	1	1	1	1	1	4	1	8	1	0	6	0	0	9	8	1 ✔
RW29	5	5	5	2	4	5	4	1	5	4	5	5	–	–	–	1	1	0	3	7	3	2	1	1 ✘
RW31	3	4	5	5	5	5	1	3	5	5	1	5	3	4	5	2	0	3	2	8	0	5	2	3 ✘
RW35	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	13	2	0	0	0	0	15	13	2 ✔
RW46	4	1	4	1	1	1	1	1	1	1	1	1	1	4	1	12	0	0	3	0	0	12	12	0 ✔
RW47	1	1	5	1	1	1	4	2	1	1	1	1	1	5	1	11	1	0	1	2	0	12	11	1 ✔
RW48	5	1	5	2	1	5	4	5	1	3	1	5	3	5	1	5	1	2	1	6	0	8	5	3 ✘
RW49	1	1	5	1	1	1	1	2	1	1	1	1	1	4	1	12	1	0	1	1	0	13	12	1 ✔
RW50	4	1	4	3	1	1	2	2	1	1	1	1	1	4	1	9	2	1	3	0	0	12	9	3 ✔
RW51	4	1	4	1	1	1	1	1	1	1	1	1	1	4	1	12	0	0	3	0	0	12	12	0 ✔
RW52	1	1	4	1	1	5	1	1	1	1	1	1	1	4	1	12	0	0	2	1	0	12	12	0 ✔
RW53	5	1	5	2	1	5	4	2	1	2	1	4	1	4	1	6	3	0	3	3	0	9	6	3 ✘
RW54	5	1	4	1	1	1	4	2	1	1	1	1	1	4	1	10	1	0	3	1	0	11	10	1 ✔
RW55	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1	14	0	0	1	0	0	14	14	0 ✔
RW57	4	1	1	1	1	1	2	2	4	1	1	1	1	1	1	11	2	0	2	0	0	13	11	2 ✔
RW59	1	1	4	1	4	4	1	1	1	1	1	1	1	4	1	11	0	0	4	0	0	11	11	0 ✔
RW60	5	1	5	1	1	5	2	2	1	2	1	1	1	4	1	8	3	0	1	3	0	11	8	3 ✔
RW61	5	1	5	1	1	5	1	1	1	1	1	5	1	–	–	9	0	0	0	4	2	9	9	0 ✔

Performance by Level, by Question

1	6	18	2	14	15	10	10	6	17	14	18	14	16	2	16
2	–	–	–	4	0	–	4	11	–	2	–	–	0	2	0
3	2	0	0	1	0	0	0	1	0	2	0	0	2	0	0
4	6	1	10	0	4	2	6	1	1	1	0	1	0	12	0
5	6	1	8	1	1	8	0	1	2	1	1	5	1	2	1
NR	0	0	0	0	0	0	0	0	0	0	1	0	1	2	3

Performance Level Legend

1 (highest): Answered correctly; exhibited all behaviors

2: Answered correctly; exhibited fewer than all other behaviors

3: Answered correctly; exhibited no other behaviors

4: Answered incorrectly; exhibited other behaviors

5 (lowest): Answered incorrectly; exhibited no other behaviors

Question Performance Summary

#AC	8	18	2	19	15	10	14	18	17	18	18	14	18	4	16
#RB	6	18	2	14	15	10	10	6	17	14	18	14	16	2	16
D_q	2 ✔	0 ✔	0 ✔	5 ✔	0 ✔	0 ✔	4 ✔	12 ✘	0 ✔	4 ✔	0 ✔	0 ✔	2 ✔	2 ✘	0 ✔

Summary Legend

#AC = # answered correctly

#RB = # answered correctly; demonstrated all other behaviors

D_p, D_q = Differentials (*#AC* – *#RB*); ✔ = criterion-passing differential (70%+), ✘ = criterion-failing differential (<70%)

In the top-left portion of the figure, participants are listed in the far-left column ("Part. ID") and questions in the topmost row ("Question #"). Each cell created by the intersection of a row and column represents the performance of a single participant on a given test question (i.e., a participant-by-question interaction). Five performance levels, numbered 1 through 5 (further identified with color shading) and defined in Section 3: Methodology, are used to indicate how a particular participant did on a particular question. A "1" in a cell represents the most successful outcome (a participant answering correctly and demonstrating all required behaviors), while a "5" represents the least successful outcome (a participant answering incorrectly and demonstrating no other required behaviors). A dash ("–") in one of these cells indicates that the participant didn't answer the question. (This could be because they ran out of time, attempted the question but didn't complete it, or, in rare cases, were misdirected from the question by the moderator.)

From this participant-by-question portion, the matrix can be read either horizontally for a summary of participant performance or vertically for a summary of question performance.

Participant Performance

The "Performance by Level, by Participant" sub-table (top center) shows the number of questions answered by each participant in terms of performance levels attained (including *NR* / no response). The "Participant Performance Summary" sub-table (top right) indicates the total number of questions each participant answered correctly (*#AC*), the number of questions each participant answered correctly while also demonstrating all required behaviors (*#RB*), and the participant differential (D_p), or the arithmetic difference between *#AC* and *#RB*. Cells in the "D_p" column include a symbol and are shaded to indicate whether a given participant differential met or exceeded (✔; blue) or fell below (✘; orange) the threshold for a "good" differential. Recall that Section 3: Methodology defines a good D_p as one indicating that at least 70 percent of a participant's correctly answered questions were responded to using all required behaviors, a statistic derived by dividing *#RB* by *#AC*.

Example: Participant Performance

Performance by Level, by Participant

1	2	3	4	5	*NR*
9	2	0	4	0	0

Participant Performance Summary

#AC	*#RB*	D_p
11	9	2 ✔

Participant RW5, per the top row in "Performance by Level, by Participant" sub-table, attained PL 1 (the most successful outcome) on nine questions, PL 2 on two questions, and PL 4 on four questions, and answered all questions (as indicated by the "0" in the "*NR* / no response" cell). Turning to the "Participant Performance Summary" sub-table, we find that RW5 answered a total of eleven questions correctly (*#AC*, calculated by adding together the number of question responses attaining PLs 1, 2, and 3) and answered nine of those questions correctly while also demonstrating all required behaviors (*#RB*, which is the same as the number in the "PL 1" cell in the "Performance by Level, by Participant" sub-table). This results in a participant differential (D_p) of 2, as 11 (*#AC*) minus 9 (*#RB*) equals 2. This D_p exceeds the threshold for a "good" differential, hence the checkmark and blue shading, as 9 (*#RB*) divided by 11 (*#AC*) equals .818, or 81.8 percent, which is above the 70 percent cutoff.

Question Performance

The "Performance by Level, by Question" sub-table (center left) shows for each test question the number of participants whose responses attained each of the five performance levels (plus *NR* / no response). A dash ("–") in cells for PL 2 indicates cases in which that level is unobtainable due to there being only two potential behaviors being evaluated for that question type. The "Question Performance Summary" sub-table (bottom left) indicates the total number of participants answering each question correctly (*#AC*), the number of correctly answering participants who also exhibited all required behaviors (*#RB*), and the question differential (D_q), or the arithmetic difference between *#AC* and *#RB*. Cells in the "D_q" row include a symbol and are shaded to indicate whether a given question differential met or exceeded (✔; blue) or fell below (✘; orange) the threshold for a "good" differential.

Findings

Participant Performance

As shown in the "Participant Performance Summary" sub-table of figure 1, fifteen of twenty participants (75 percent) met or exceeded the criterion for a good D_p, which provides evidence that these participants were able to adequately demonstrate cognitively complex thinking in line with the question types' constructs.

The performance of the remaining five participants failed to meet the criterion for a good D_p. For example, participant RW25 answered eight questions correctly and demonstrated all required behaviors for four of those questions, resulting in a D_p of 4, representing half (50 percent) of the correctly answered questions. While the performance of these participants fell below the criterion level, they were still able to demonstrate all required behaviors while answering correctly from 40 percent to 67 percent of the time, indicating that these participants were able to demonstrate cognitively complex thinking in line with the question types' constructs at least some of the time.

Question Performance

As shown in the "Question Performance Summary" sub-table of figure 1, thirteen of the fifteen studied Reading and Writing questions (87 percent) met or exceeded the criterion for a good D_q, which provides evidence that these questions are capable of eliciting cognitively complex thinking from EL students. The remaining two questions (questions 8 and 14) were still answered correctly by six and two participants, respectively, who also demonstrated all required behaviors (*#RB*; PL 1), suggesting that these questions, too, are capable of eliciting cognitively complex thinking from EL students, even if they didn't always during the study. Question 8 had a high differential of 12 but was still answered correctly by eighteen participants; by contrast, question 14 had a low differential of 2, but only four participants answered it correctly. As we return to in Section 5: Discussion, specific aspects of these two questions—the tendency of participants not to demonstrate a specific required behavior in question 8; the numerous sources of challenge in question 14—likely explain these criterion-failing D_qs.

PARTICIPANT PERFORMANCE VIGNETTES

Vignettes from participant performance on the examined Reading and Writing questions provide further evidence that EL participants were able to exhibit cognitively complex thinking in line with the questions types' constructs.

This section relies primarily on a case study approach, in which we follow a single participant, RW59, as he works through all fifteen Reading and Writing questions, succeeding on some and struggling with others. In the latter cases, RW59's vignettes are supplemented with those from participants who were more successful (i.e., attained PL 1). These supplements serve to show that even when the case study participant encountered difficulties with particular questions, other participants were able to answer correctly and demonstrate all required behaviors, suggesting that EL participants were able to demonstrate cognitively complex thinking in accordance with the question types' constructs at least some of the time.

Case Study: Participant RW59

Participants were considered good candidates for the case study approach when they met the following criteria:

- They placed their English language acquisition level in the moderate range (i.e., identified as level 3 or 4 on the six-point scale used in this study).
- They answered all fifteen Reading and Writing questions (e.g., didn't run out of time).
- They exhibited a good participant differential (D_p).

Example: Question Performance

14
4
1
0
1
0

19
14
5 ✔

Performance statistics for Reading and Writing question 4 are pictured above. The responses from fourteen participants attained PL 1 (the most successful outcome), those from four participants attained PL 2, that from one participant attained PL 3, and that from one participant attained PL 5. All participants answered the question, as indicated by the "0" in the "*NR* / no response" cell. Adding together the counts in PLs 1–3, we find that a total of nineteen participants answered the question correctly (*#AC*). Fourteen of these participants also demonstrated all required behaviors (*#RB*, which is the same as the number in the "PL 1" cell in the "Performance by Level, by Question" sub-table). Subtracting 14 (*#RB*) from 19 (*#AC*) yields a question differential (D_q) of 5. This D_q exceeds the threshold for a "good" differential, hence the checkmark and blue shading, as 14 (*#RB*) divided by 19 (*#AC*) equals approximately .737, or 73.7 percent, which is above the 70 percent cutoff.

Participant RW59, a male twelfth grader from New Jersey, met these conditions. He identified as Asian and not of Hispanic, Latino, or Spanish origin. He self-reported a high school GPA (HSGPA) of A+ (not uncommon among the sample), indicated that he often used English in his everyday life, typically spoke a language other than English at home, listed Mandarin as a known language besides English, and rated his English language acquisition level as 3 (“I can understand the main points of clear texts on familiar subjects in English”). RW59 answered eleven of the fifteen Reading and Writing questions correctly and demonstrated all required behaviors in all cases, resulting in a participant differential of 0 (100 percent), which exceeded the criterion for a good D_p.

Reading and Writing Question 1

Skill/Knowledge Testing Point	Words in Context
Performance Score Band	7
Stimulus Subject Area	Science
Stimulus Text Complexity	PSR (postsecondary readiness, grades 12–14)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that completes the passage with the most logical and precise word or phrase.
RW59 Performance Level	1

To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes, engineer Aroba Saleem and colleagues ________ the tendency of some metals’ internal magnetic fields to alter under stress: the team showed that such alterations can be measured from a distance and can reveal concentrations of stress in the pipes.

Which choice completes the text with the most logical and precise word or phrase?

A) hypothesized

B) discounted

C) redefined

D) exploited

Question 1, a hard (PSB 7) Words in Context question set in a highly challenging (PSR) science context, requires test takers to determine the word or phrase that completes the text (i.e., fills in the blank) in the most logical and precise way. The best answer (*key*) is choice D, as the engineers “exploited,” or made use of, “the tendency of some metals’ internal magnetic fields to alter under stress” to assess the “integrity of underground metal pipes” without digging up those pipes.

> [*Reads silently*] So it looks like [the engineers] are trying to assess something without, like, trying to damage it. And there are two engineers, multiple engineers, and they are trying to assess the tendency of the metals to alter under stress. And the word that, the answer choice—I

> think they're trying to ask me to fill in the blank with the most logical or precise word or phrases.
>
> [*Moderator asks participant to read aloud the question before going further; participant complies*]
>
> Answer choice A ["hypothesized"]. I think that one maybe can fit because I think this is typically used for the science field. But I would look at the rest to, just to make sure.
>
> And [choice] B ["discounted"]. It kind of means to, like, make it worse or, like, less quality. But I don't think that fits here; B is not correct.
>
> And [choice] C, "redefined." I think they tried to test but not to, like, improve the metal pipes. So C is not correct.
>
> And [choice] D, "exploited." I think maybe it's just as good. They are trying to see the tendency of some metals to alter under stress. So maybe D is better than A because A sounds like science, like research in the lab. So maybe it's not correct, but D seems to be, like, a better choice than [A].
>
> So my final answer choice is D.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW59 shows adequate passage comprehension (behavior 1) at a few points throughout his verbalization. For example, he begins his answering process by noting that "it looks like [the engineers] are trying to assess something without, like, trying to damage it." He provisionally rules in choice A, "hypothesized," on general "fit" grounds ("I think this is typically used for the science field") but later finds the best answer, choice D, to be preferable. He offers a definition that's vaguely resonant in tone but technically inaccurate when observing that "discounted," choice B, "kind of means to, like, make it worse or, like, less quality." He's less successful in providing a definition for choice C, "redefined," perhaps confusing it with the word "refined" ("I think they tried to test but not to, like, improve the metal pipes"), but still rules out this option. He doesn't offer up the meaning of choice D, "exploited," noting only that, comparatively, "hypothesized," choice A, sounds "like research in the lab" rather than a practical study. This line of reasoning isn't precise, but it's sufficient to allow him to correctly pick choice D as his answer (behavior 2).

Reading and Writing Question 2

Skill/Knowledge Testing Point	Text Structure and Purpose (Passage main purpose subtype)
Performance Score Band	3
Stimulus Subject Area	Literature
Stimulus Text Complexity	MID (middle school/junior high, grades 6–8)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that best states the main purpose of the passage.
RW59 Performance Level	1

Vignette and Transcription Notes

Broadly speaking, the vignettes throughout this report are verbatim representations of participants' verbalizations, and the College Board researchers verified the transcripts' accuracy and completeness against the sessions' video recordings. However, for the sake of readability, some minor alterations were made in the vignettes' presentation. Repetitions (e.g., "It's like, it's like") were cleaned up, "ums" and similar verbal hesitations were removed, and [bracketed] text was sometimes added by the researchers to clarify participants' points or to complete or correct quotations from test passages and answer choices.

The following text is adapted from Jean Webster's 1912 novel *Daddy-Long-Legs*. The narrator is a young college student writing letters detailing her weekly experiences.

[The college is] organizing the Freshman basket-ball team and there's just a chance that I shall make it. I'm little of course, but terribly quick and wiry and tough. While the others are hopping about in the air, I can dodge under their feet and grab the ball.

Which choice best states the main purpose of the text?

A) To compare basketball with other sports

B) To provide details of how to play basketball

C) To state how players will be chosen for the basketball team

D) To explain why the narrator thinks she might make the basketball team

Question 2, an easy (PSB 3) Text Structure and Purpose question set in a moderately challenging (MID) literature context, requires test takers to determine which answer choice best states the main purpose of the passage. The best answer is choice D, as the focus of the text is on the reasons the narrator thinks she'll make the basketball team: she's "terribly quick and wiry and tough" and, because of her small stature, can "dodge under [other players'] feet and grab the ball."

> I think there's a college. They're trying to start a freshman basketball team, and the author, she is trying to make it, but I don't think she's that confident about it. Or maybe she's trying to dodge under their feet and grab the ball. So I think she's trying to compare with others.
>
> So [choice] A I don't think is correct because I don't think she's trying to compare basketball to, like, other sports.
>
> And [choice] B, "provide details." I feel like the text is more about, like, herself, how she feels about it, but not like play, like, detail about basketball. So I don't think B is correct.
>
> And [choice] C, she's not really saying, like, what kind of the role that she'd be choosing for the basketball team. So I don't think C is correct.
>
> And [choice] D, "To explain why narrator thinks she might make [the basketball team]." I think this could be correct because she says here there's just a chance that she will make the team and, like, how she's, like, kind of better than others, like at the last sentence.
>
> So I think D is correct for this one.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW59 begins by demonstrating adequate passage comprehension (behavior 1) while summarizing the passage: "I think there's a college. They're trying to start a freshman basketball team, and the author, she is trying to make it, but I don't think she's that confident about it." He also shows passage understanding in the way he rules out the incorrect answers—choice A because "I don't think she's trying to compare basketball to, like, other sports," choice B because "the text is more about, like, herself, how she feels about [basketball]," and choice C because "she's not really saying, like, what kind of the role that she'd be choosing for the basketball team" (although RW59 mistakes choice C's reference to the method of selecting players for the team with the notion of establishing specific roles for players on the team). He further draws on passage understanding when evaluating choice D: ". . . she says here there's just a chance that she will make the team and, like, how she's, like, kind of better than others, like at the last sentence." Consistent with that reasoning, RW59 picks the best answer, choice D, as his response (behavior 2).

Reading and Writing Question 3

Skill/Knowledge Testing Point	Text Structure and Purpose (Part-whole relationships subtype)
Performance Score Band	7
Stimulus Subject Area	History/social studies
Stimulus Text Complexity	PSR (postsecondary readiness, grades 12–14)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that best describes the main function of the underlined portion in the passage as a whole.
RW59 Performance Level	4

More than 60% of journeys in Mexico City occur via public transit, but simply reproducing a feature of the city's transit system—e.g., its low fares—is unlikely to induce a significant increase in another city's transit ridership. As Erick Guerra et al. have shown, transportation mode choice in urban areas of Mexico is the product of a complex mix of factors, including population density, the spatial distribution of jobs, and demographic characteristics of individuals. System features do affect ridership, of course, but there is an irreducibly contextual dimension of transportation mode choice.

Which choice best describes the function of the underlined portion in the text as a whole?

A) It presents an objection to the argument of Guerra et al. about transportation mode choice in urban areas of Mexico.

B) It explains why it is challenging to influence transit ridership solely by altering characteristics of a transit system.

C) It illustrates the claim that a characteristic associated with high transit ridership in Mexico City is not associated with high transit ridership elsewhere.

D) It substantiates the assertion that population density, the spatial distribution of jobs, and demographic characteristics are important factors in transportation mode choice.

Question 3, a hard (PSB 7) Text Structure and Purpose question set in a highly challenging (PSR) history/social studies context, requires test takers to determine the main function of the underlined portion of the passage in terms of the passage as a whole. The best answer is choice B. The underlined portion—“there is an irreducibly contextual dimension of transportation mode choice”—restates the passage’s claim that “simply reproducing” an aspect of Mexico City’s transit system, such as its low fares, is “unlikely to induce a significant increase in another city’s transit ridership,” a claim supported in the passage by findings from Erick Guerra et al., who determined that “transportation mode choice in urban areas of Mexico is the product of a complex mix of factors.”

> I think they are trying to do a correlation. So, like, they are trying to—so one of the cities, they are trying to copy other cities’, like, a unique feature of their public transit system. They did some research on Mexico’s transportation method, and the last sentence, I think— [*Rereads question stem*] So I think that this uses the last sentence, so I have to read the whole paragraph again to see how it, like, affects the paragraph. And, OK, I see it’s a word, like a contextual dimension. So maybe it’s about how the writers feel after making a certain change in the transportation system.
>
> And choice A says, **“**It presents an objection to the argument [of Guerra et al.] about transportation mode choice [in urban areas of Mexico].” I don’t think it really makes, like, an objection, for choice A, since the argument is basically about how, like, the Mexicans’ transportation mode, like, this mix of different factors like this. It doesn’t make any objection.
>
> And [choice] B, “It explains why it is challenging to [influence transit ridership solely by altering characteristics of a transit system].” No, I don’t think there’s any challenges in here when it says it’s a contextual dimension. I don’t think so.
>
> [Choice] C, “It illustrates the claim [that a characteristic associated with high transit ridership in Mexico City is not associated with high transit ridership elsewhere].” I don’t think C is correct because— Wait, C could be correct because, like, the first sentence here, it says— [*Long pause*] No, C is not correct. It doesn’t say— Wait, it is “not associated.” Here, the first sentence, it says it is unlikely to induce this increase [in], you know, the city transit ridership.
>
> And [choice] D. Yeah, I don’t think it’s correct.
>
> So my final answer choice is C.
>
> *Participant RW59*

Participant RW59 answered the question incorrectly but did demonstrate a single required behavior, resulting in a PL of 4. RW59 exhibits adequate passage comprehension (behavior 1) when he at least partially summarizes the passage’s content: “So, like, they are trying to—so one of the cities, they are trying to copy other cities’, like, a unique feature of their public transportation system.” This summary doesn’t capture the full passage, but it shows a basic understanding of the content. RW59 correctly rules out choice A on the grounds that the underlined portion of the passage “doesn’t make any objection.” After this, though, RW59 rules out the best answer, choice B, because although he grasps that there’s a “contextual dimension” to transportation mode choice, he fails to recognize the “challenges” of influencing transit ridership discussed in the passage. He wavers on choice C, seemingly in part because he only belatedly attends to the use of “not” (as in “not associated”) in the option, but ultimately selects it because he apparently overinterprets the passage’s first sentence to mean that replicating

Mexico City's low public transit fares hasn't worked to increase ridership in other urban areas of Mexico.

Supplementary Vignette: Participant RW35

Participant RW35 answered question 3 correctly and also demonstrated both required behaviors, resulting in a PL of 1. RW35 was one of only two participants to have answered the question correctly, both of whom doing so while also demonstrating both required behaviors.

> OK, so, technically, the function of this sentence, it's more about— OK. So the sentence previously states some examples, and the first sentence states that, the fact that, like, in Mexico [City], most people, like, most journeys, occur via public transit. So, basically, this last sentence says that, like, that every situation is different, so there are a lot of factors to take into consideration. OK. So, going to the choices.
>
> [*Reads choice A*] I'm gonna go back.
>
> [*Reads choice B*] OK. I can, I need to, like, reread.
>
> [*Rereads part of passage*] OK. So the first and the last sentence have a connection. [*Returns to passage*] OK, so definitely [choice] A is not [correct] because it's not, like, an objection. All the sentences talk about the different contexts of people.
>
> [*Rereads choice B*] This might be.
>
> [*Reads choice C*] Nope, because, elsewhere, it's not, like, stated in any part of the text.
>
> [*Reads choice D*] They're important, "but there is an irreducibly contextual dimension of transportation mode choice."
>
> [*Rereads underlined portion*] So I think it's B. Wait, let me think again about D. [*Rereads choice D*] Maybe it's not D because it's, I've read some tips that say that, like, usually [incorrect] answers have recycled language. So I would not choose D because it says "transportation mode choice," and it's the same [language] as [in] the last sentence.
>
> So I would choose B.
>
> *Participant RW35*

RW35 demonstrates adequate passage comprehension (behavior 1) most clearly in her concise summation of the passage: "So, basically, this last sentence says that, like, that every situation is different, so there are a lot of factors to take into consideration." She correctly eliminates choice A because the underlined portion is "not, like, an objection" given that all the passage's sentences "talk about the different contexts of people." She further rules out choice C as unsupported by the passage. She wavers between choices B and D and ultimately blocks D on "testwise" grounds: "Maybe it's not D because it's, I've read some tips that say that, like, usually [incorrect] answers have recycled language." Her "tip" pays off in this case, as choice D is incorrect, primarily because the underlined portion doesn't "substantiate," or provide further support for, an assertion about the importance of three specific factors in transportation mode choice; rather, the underlined portion restates the centrality of context to a good understanding of such choice. Having gotten rid of choice D—by whatever means—RW35 selects the best answer, choice B, as her response (behavior 2).

Reading and Writing Question 4

Skill/Knowledge Testing Point	Command of Evidence: Quantitative (Table subtype)
Performance Score Band	4
Stimulus Subject Area	Science
Stimulus Text Complexity	SCO (upper secondary, grades 9–11)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Demonstrate an understanding of the table, including what the table as a whole as well as its various rows and columns represent. 3. Demonstrate an understanding of the relationship among the passage, the table, and the criterion set forth in the question's stem. 4. Select the answer choice that best meets the criterion set forth in the question's stem.
RW59 Performance Level	1

Partial List of Candidate Species for De-extinction

Common name	Scientific name	Became extinct
Huia	*Heteralocha acutirostris*	1907
Caribbean monk seal	*Monachus tropicalis*	1952
Passenger pigeon	*Ectopistes migratorius*	1914
Saber-toothed cat	*Smilodon*	11,000 years before present
Woolly mammoth	*Mammuthus primigenius*	6,400 years before present

The passage of time is among the many obstacles faced by scientists who are pursuing de-extinction efforts—that is, efforts to use breeding or a mixture of cloning and genetic engineering to bring back extinct species. Specifically, researchers are concerned that the longer a species has been extinct, the less likely it is that a suitable habitat still exists for that species. Among candidate species for de-extinction, this problem would be especially concerning for the ________

Which choice most effectively uses data from the table to complete the statement?

A) passenger pigeon (*Ectopistes migratorius*), which became extinct only a few years after the huia (*Heteralocha acutirostris*).

B) saber-toothed cat (*Smilodon*), which became extinct 11,000 years ago.

C) woolly mammoth (*Mammuthus primigenius*), which became extinct several thousand years before the saber-toothed cat (*Smilodon*).

D) Caribbean monk seal (*Monachus tropicalis*), which became extinct in 1952.

Question 4, a medium-difficulty (PSB 4) Command of Evidence: Quantitative question set in a challenging (SCO) science context, requires test takers to draw on both passage and table to complete the statement containing the blank with the most effective data from the table. The passage establishes that "the longer a species has been extinct, the less likely it is that a suitable habitat still exists for that species," thus making longer-extinct species progressively worse candidates for de-extinction efforts. Per the table, the saber-toothed cat (*Smilodon*) went extinct "11,000 years before present," making it the longest-extinct candidate in the table and making choice B the best answer.

> Answer choice A. [*Rereads answer choice*] So A is correct, but it might not fit the text.
>
> [Choice] B. It is also correct, according to the table, but it might not fit the text.
>
> [Choice] C. C is just not correct because one is about 6,400 years, and the other is 11,000 years. So it's not correct. The statement is just not correct.
>
> And [choice] D. So A, B, and D might be correct. [*Long pause*] So, and the text says the researchers "are concerned that the longer a species has been extinct, the less likely it is that a suitable habitat still exists for that species." So I think, so, yeah, I think maybe I can choose the one that has been gone for the most time. And so I think B is the best choice because that is the longest away from now.
>
> So my final answer choice is B.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW59 begins his verbalization by evaluating each of the answer choices against the data provided in the table (behavior 2). In the process, he concludes that choices A, B, and D convey accurate information from the table and therefore "might be correct" but that choice C "is just not correct" per the table, as that option reverses the order in which the woolly mammoth and saber-toothed cat went extinct. RW59 then returns to the passage to narrow down the remaining choices. He demonstrates both adequate passage comprehension (behavior 1) and clarity about the relationship among the passage, table, and question stem (behavior 3) by observing that "maybe I can choose the [species] that has been gone for the most time," since the passage notes that the longer a species has been gone, the less likely it is that there's still a suitable habitat for it. Realizing that choice B, the best answer, represents "the longest away from now," he selects that as his response (behavior 4).

Reading and Writing Question 5

Skill/Knowledge Testing Point	Command of Evidence: Textual
Performance Score Band	4
Stimulus Subject Area	Literature
Stimulus Text Complexity	SCO (upper secondary, grades 9–11)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Demonstrate an understanding of the relationship between the criterion set forth in the question's stem and the passage. 3. Select the answer choice that best meets the criterion set forth in the question's stem.
RW59 Performance Level	4

"The Yellow Wallpaper" is an 1892 short story by Charlotte Perkins Gilman. In the story, the narrator expresses mixed feelings about her surroundings: ________

Which quotation from "The Yellow Wallpaper" most effectively illustrates the claim?

A) "This wallpaper has a kind of sub-pattern in a different shade, a particularly irritating one, for you can only see it in certain lights, and not clearly then."

B) "By moonlight—the moon shines in all night when there is a moon—I wouldn't know it was the same paper."

C) "I'm really getting quite fond of the big room, all but that horrid [wall]paper."

D) "The color is repellant, almost revolting; a smouldering, unclean yellow, strangely faded by the slow-turning sunlight."

Question 5, a medium-difficulty (PSB 4) Command of Evidence: Textual question set in a challenging (SCO) literature context, requires test takers to determine which of the provided quotations from the short story "The Yellow Wallpaper" most clearly expresses the narrator's mixed feelings about her surroundings. The best answer is choice C, as it illustrates both the narrator's general appreciation for the room ("I'm really getting quite fond of the big room") and specific dislike of its "horrid" wallpaper.

> So I think I need, we need to read each answer choice and fill in the blank [with] which supports the claim, which is about mixed feelings about [her] surroundings.
>
> And [choice] A, "This wallpaper has a kind of sub-pattern in a different shade, a particularly irritating one, for you [can only see it in certain lights, and not clearly then]." So A might be because she's feeling different about the wallpaper, and she says she's kind of concerned about she can only see it in certain lights and not clearly then. So I think A could be one choice. I don't know about the rest.
>
> [Choice B,] "By moonlight—" No, I don't think B is the correct choice because it doesn't really say she has mixed feelings.

> And [choice] C. C, she's just mad about the wallpaper.
>
> And [choice] D, "The color—" D, I don't think is correct.
>
> I think maybe A [or] D is correct, but [*long pause*] maybe A is more correct because it says "you can only see it in certain lights" and just, like, more personal feelings involving the answer choice. I think A might be the more correct choice, this one.
>
> So my final answer is A.
>
> *Participant RW59*

Participant RW59 answered the question incorrectly but did demonstrate two required behaviors, resulting in a PL of 4. RW59 demonstrates adequate comprehension of the passage (behavior 1) and a clear understanding of the relationship between the passage and question stem (behavior 2) but nonetheless picks the wrong answer choice. He most clearly demonstrates passage (here, answer choice) understanding and a grasp of what the question is asking when he correctly rules out choice B on the grounds that "it doesn't really say she has mixed feelings." Although RW59 recognizes that the best answer choice must support the claim that the narrator has mixed feelings about her surroundings, he seems to misinterpret the thrust of both choice A, which is consistently negative in tone, and choice C, which does, in fact, express conflicting feelings, and incorrectly selects the former as his response.

Supplementary Vignette: Participant RW46

Participant RW46 answered question 5 correctly and also demonstrated all required behaviors, resulting in a PL of 1. RW46 was one of fifteen participants to have answered the question correctly, all of whom doing so while also demonstrating all required behaviors.

> We don't want to look at the claims. We wanna go straight to the text. [*Reads question*] OK. We're looking for something that has mixed feelings within the answer choices.
>
> Going to A. [*Reads choice A*] OK. This is something that's, there's only one emotion to this. It's irritation. So this one cannot be the answer 'cause it's not about the narrator having mixed feelings, 'cause obviously they're irritated by this, but they don't like it. It, so there has to be mixed feelings about this. So it cannot be A.
>
> [Choice] B. [*Reads choice B*] This one's a little confusing, but I'm gonna skip over it and go to C.
>
> [*Reads choice C*] OK. C, they're getting quite fond, and they're also being, getting horrid. Those are two contrasting feelings. They're a mix of feelings. So it could be C.
>
> [Choice] D. [*Reads choice D*] OK. So it's detailing a color, and it's strangely faded. But this is not what the narrator feels about this. So I'm going to eliminate D.
>
> I believe it's [choice] C because it has two feelings about the room and the wallpaper: it's quite fond, and it's also horrid. So my final answer is gonna be C.
>
> *Participant RW46*

RW46 demonstrates adequate passage—in this case, answer choice—comprehension (behavior 1) when working through the options, and he exhibits an understanding of the relationship between the passage and question stem (behavior 2) by observing that “we’re looking for something that has mixed feelings within the answer choices.” He correctly eliminates choice A because “there’s only one emotion to this.” He’s unable to effectively parse choice B, and he blocks choice D on the mistaken basis that D doesn’t express any of the narrator’s emotions (“this is not what the narrator feels about this”), whereas D, in fact, uses strongly negative emotive language to describe the wallpaper. In any event, RW46 correctly recognizes that the best answer, choice C, “has two feelings about the room and the wallpaper: it’s quite fond, and it’s also horrid” and selects this as his response (behavior 3).

Reading and Writing Question 6

Skill/Knowledge Testing Point	Transitions
Performance Score Band	5
Stimulus Subject Area	History/social studies
Stimulus Text Complexity	SCO (upper secondary, grades 9–11)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that completes the passage with the most logical transition.
RW59 Performance Level	4

> According to Duverger’s law, countries with single-ballot majoritarian elections for single-member districts tend to polarize into two-party systems, wherein dueling political parties consistently dominate the political system. ________ countries with proportional-representation electoral systems tend to support multi-partyism, under which power gets distributed among many political parties.
>
> Which choice completes the text with the most logical transition?
>
> A) Subsequently,
>
> B) Conversely,
>
> C) For instance,
>
> D) In other words,

Question 6, a medium-difficulty (PSB 5) Transitions question set in a challenging (SCO) history/social studies context, requires test takers to determine the most logical transition word or phrase to complete the sentence in the passage with the blank. The best answer is choice B, “conversely,” as the passage’s last sentence (the one containing the blank) contrasts proportional-representation electoral systems and multi-partyism with the single-ballot majoritarian elections for single-member districts and two-party systems mentioned in the passage’s first sentence.

> I think it's about a political party and the political system. And at the sentence after the blank, it seems to—the countries with proportional-representation electoral systems tend to support multi-partyism, where, like, their powers get distributed to multiple political parties. And the first sentence is also about two-party, but it's not, like, multiple-party. [*Long pause*]
>
> So for [choice] A, "subsequently." I think that A might be, like, something about, like, the second sentence is a result about the first sentence. [*Long pause*] I feel it's not, like, a result. It's different because the first one is just about two-party systems and the second one is multi-partyism. So I think A is not correct.
>
> And [choice] B, "conversely." I don't think this is correct. It's about something different. I think so.
>
> And [choice] C, "for instance," Like, to give us an example.
>
> [Choice] D, "in other words."
>
> So maybe C. I'm not too sure, but I think the final answer is C for this one.
>
> [*Moderator asks for clarification*]
>
> So I was stuck between C and D, but I feel D says "in other words," but it doesn't feel like it's "in other words" because it's talking about two different things. The first sentence talks about two-party, and the second one talks about multi-partyism, so maybe D is not correct, and I was left with answer choice C. So I think that one is the correct choice.
>
> *Participant RW59*

Participant RW59 answered the question incorrectly but did demonstrate a single required behavior, resulting in a PL of 4. RW59 demonstrates adequate passage comprehension (behavior 1) throughout his response, for example by noting that the passage is "about a political party and the political system" and defining "multi-partyism" as where "powers get distributed to multiple political parties." He also exhibits understanding of the passage's structure, observing that the two sentences each describe "something different." He correctly rules out choice A, "subsequently," because the second sentence is "not, like, a result" of the first sentence and ultimately choice D, "in other words," because the two sentences are "talking about two different things." After narrowing down to choices C and D, RW59 is "left with" C, which is incorrect. His failure here to select the best answer, choice B, seems likely to result from a lack of understanding of the word's definition, as "conversely" appropriately sets up the relationship between the two sentences, which RW59 otherwise correctly delineates.

Supplementary Vignette: Participant RW50

Participant RW50 answered question 6 correctly and demonstrated both required behaviors, resulting in a PL of 1. RW50 was one of ten participants to have answered the question correctly, all of whom doing so while also demonstrating both required behaviors.

> OK. So a transition word is needed. The first part [of the passage] talks about two-party systems, while the second part mentions multiparty systems, which are two different things.
>
> “For instance” [choice C] doesn’t work because that’s for an example. “Subsequently” [choice A] doesn’t fit either. “In other words” [choice D]. So it can’t be that because it’s different items.
>
> “Conversely” [choice B] seems to work because it contrasts two different ideas. I think the answer is choice B.
>
> *Participant RW50*

RW50 employs strong vocabulary and comprehension skills to decisively answer the question correctly. He demonstrates adequate passage comprehension (behavior 1) when asserting that “the first part [of the passage] talks about two-party systems, while the second part mentions multiparty systems, which are two different things.” This understanding sets him up well to find a contrastive transition word or phrase. He recognizes that “subsequently,” choice A, “doesn’t fit” the context, that “for instance,” choice C, “doesn’t work because that’s for an example,” and that “in other words,” choice D, is incorrect because the passage discusses two “different items,” meaning that the second sentence isn’t a restatement of the first. He then reasons that the answer should be “conversely,” choice B, “because it contrasts two different ideas” and then selects the best choice as his response (behavior 2).

Reading and Writing Question 7

Skill/Knowledge Testing Point	Rhetorical Synthesis
Performance Score Band	4
Stimulus Subject Area	Humanities
Stimulus Text Complexity	MID (middle school/junior high, grades 6–8)
Required Behaviors	1. Read and demonstrate comprehension of the student-produced notes. 2. Demonstrate an understanding of the relationship between the notes and the criterion set forth in the question’s stem. 3. Select the answer choice that best meets the criterion set forth in the question’s stem.
RW59 Performance Level	1

While researching a topic, a student has taken the following notes:

- In 1859, the novel *Adam Bede* was published in England.
- According to the novel's title page, the author's name was George Eliot.
- George Eliot was widely assumed to be a pseudonym.
- A pseudonym is a fake name used to conceal an author's identity.
- A woman named Mary Ann Evans later revealed herself as the novel's real author.

The student wants to identify the real author of *Adam Bede*. Which choice most effectively uses relevant information from the notes to accomplish this goal?

A) The real author of *Adam Bede* was Mary Ann Evans, who published the novel using the pseudonym George Eliot.

B) George Eliot, which *Adam Bede*'s title page indicated was the name of the novel's author, was widely assumed to be a pseudonym.

C) The title page of the novel *Adam Bede* indicated that the author's name was George Eliot.

D) A woman who had used a pseudonym to conceal her identity later revealed herself as the real author of *Adam Bede*.

Question 7, a medium-difficulty (PSB 4) Rhetorical Synthesis question set in a moderately challenging (MID) humanities context, requires test takers to select the answer choice that best uses relevant information from the student-produced "notes" (bulleted list of informational points, ostensibly gathered from research) to meet the question's criterion, which, in this case, is to identify the real author of *Adam Bede*. The best answer is choice A, as it clearly indicates that *Adam Bede*'s author was Mary Ann Evans, who used the pseudonym George Eliot when publishing.

> And from the notes, it says "In 1890, 1859, the novel [*Adam Bede*] was published in England." The student wants to identify the real author, and "the author's name was George Eliot." But it says it is "a fake name used to conceal an author's identity." The name is not correct. And it says "A woman named Mary Ann Evans later revealed herself as the novel's real author." So the student wants to identify the real author. OK.
>
> Answer choice A, "The real author of *Adam Bede* was Mary Ann Evans, who published [the novel using the pseudonym George Eliot]." A is correct, but I'll look at the rest.
>
> [Choice] B. B doesn't seem—the student wants to identify the real author. Even though it's correct, but it doesn't fit the question that much, right? So I don't think B is correct.
>
> [Choice] C. C is just an answer choice about, like, the author's name was George Eliot. So it's not like—just like B, it doesn't really fit the question.

> And [choice] D, “A woman who had [used a pseudonym to conceal her identity later revealed herself as the real author of *Adam Bede*].”
>
> So A and D kind of more fits the question, but A, like, just gives more details and the real name of the author. But D is also similar, but it doesn’t give that much details. “The student wants to identify the real author [of *Adam Bede*].” Yeah, the student wants to identify the real author. So I think more details will probably be better. So I think between A and D, I will choose A.
>
> So my final answer choice is A.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW59 begins by paraphrasing the information contained in the student-produced notes, thereby demonstrating adequate comprehension (behavior 1). He observes, for example, that “George Eliot” is “not correct,” or is a pseudonym for the real author. He also repeatedly invokes the intended relationship between the notes and the question stem (behavior 2), reminding himself that “the student wants to identify the real author” of *Adam Bede*. He’s quickly drawn to choice A, the best answer (“correct”) but considers the alternatives in turn. He figures out that all the answer choices correctly represent information from the notes but blocks choices B and C on the grounds that neither “fit[s] the question.” He ultimately prefers choice A to choice D because while he thinks that both “kind of more [fit] the question,” choice A “just gives more details and the real name of the author.” RW59 then selects the best answer, choice A, as his response (behavior 3).

Reading and Writing Question 8

Skill/Knowledge Testing Point	Rhetorical Synthesis
Performance Score Band	5
Stimulus Subject Area	Science
Stimulus Text Complexity	PSR (postsecondary readiness, grades 12–14)
Required Behaviors	1. Read and demonstrate comprehension of the student-produced notes. 2. Demonstrate an understanding of the relationship between the notes and the criterion set forth in the question’s stem. 3. Select the answer choice that best meets the criterion set forth in the question’s stem.
RW59 Performance Level	1

While researching a topic, a student has taken the following notes:

- Scientists have developed a “freeze-thaw” battery that can retain 92% of its charge after twelve weeks.
- The battery contains molten salt (a type of salt that liquifies when heated and solidifies at room temperature).
- When the salt is in a liquid state, energy flows through the battery.
- When the salt is in a solid state, energy stops flowing and is stored in the battery.
- The stored (frozen) energy can be used by reheating (thawing) the battery.

The student wants to specify how the salt enables energy storage. Which choice most effectively uses relevant information from the notes to accomplish this goal?

A) Scientists have developed a freeze-thaw battery that contains molten salt, which liquifies when heated and solidifies at room temperature.

B) The stored energy in a freeze-thaw battery, which contains molten salt, can be used by reheating the battery.

C) When the molten salt in a freeze-thaw battery solidifies at room temperature, energy stops flowing and can be stored in the battery.

D) Molten salt allows a freeze-thaw battery to retain 92% of its charge after twelve weeks.

Question 8, a medium-difficulty (PSB 5) Rhetorical Synthesis question set in a highly challenging (PSR) science context, requires test takers to, again, select the answer choice that best uses relevant information from the notes to accomplish the writer’s goal, which, in this case, is to specify how the salt in the freeze-thaw battery described in the notes enables energy storage. The best answer is choice C, as this option addresses how solidifying the battery’s molten salt, which occurs at room temperature, stops energy flow and thereby permits energy storage.

> So the student is researching a topic, and the student took the following notes, and “the student wants to specify how the salt enables energy storage.” “Which choice most effectively uses relevant information from the notes to accomplish this goal?” [*Inaudible*] OK. “Scientists have developed a [“freeze-thaw”] battery that can retain 92% of its charge after twelve weeks.” And the battery has a salt that liquefies when heated and solidifies at room temperature. “When the salt is in a liquid state, energy flows [through the battery].” And the fourth bullet point, it says “When the salt is in a solid state, energy stops flowing, and it [*sic*] is stored in the battery.” So the four bullet points fits the answer—I mean, fits the question the student has about energy storage. So now I’m going to look at the answer choices.

> So the answer choice A, “Scientists have developed a freeze-thaw battery that contains molten salt, which liquefies [when heated and solidifies at room temperature].” I don’t think A is correct; A just talks about how the salt works. So A is not correct.
>
> [Choice] B, “The stored energy in a freeze-thaw battery, which contains molten salt, can be used by reheating the battery.” Just talks about how the salt works, so it’s nothing about energy storage. So B is not correct.
>
> And [choice] C, “When the molten salt in a freeze-thaw battery solidifies at room temperature, energy stops flowing and can be stored in the battery.” So I think C is correct; it talks about energy being stored in the battery. And so the student also wants to know how the salt enables energy storage. I think C is correct.
>
> And [choice] D doesn’t feel correct. Just about how much percentage that got stored after twelve weeks.
>
> So I think my final answer choice is C.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW59 shows adequate notes comprehension (behavior 1) when he quotes and paraphrases the various bullet points. It’s not clear why he refers to “four bullet points” instead of five, but in any event he’s correct in that only the first four points relate directly to the criterion set forth in the question’s stem (“fits the question the student has about energy storage”; behavior 2). He correctly rules out choices A and B because they merely describe how the battery’s salt is affected by heating and cooling and choice D because it’s “just about how much percentage [of the charge] that got stored [i.e., retained] after twelve weeks.” He ends up selecting the best answer, choice C (behavior 3), because it “talks about energy being stored in the battery,” which aligns with the question stem’s criterion.

Reading and Writing Question 9

Skill/Knowledge Testing Point	Words in Context
Performance Score Band	4
Stimulus Subject Area	Science
Stimulus Text Complexity	PSR (postsecondary readiness, grades 12–14)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that completes the passage with the most logical and precise word or phrase.
RW59 Performance Level	1

According to a team of neuroeconomists from the University of Zurich, ease of decision making may be linked to communication between two brain regions, the prefrontal cortex and the parietal cortex. Individuals tend to be more decisive if the information flow between the regions is intensified, whereas they make choices more slowly when information flow is ________.

Which choice completes the text with the most logical and precise word or phrase?

A) reduced

B) evaluated

C) determined

D) acquired

Question 9, a medium-difficulty (PSB 4) Words in Context question set in a highly challenging (PSR) science context, requires test takers to select the most logical and precise word or phrase to fill in the blank in the passage. The best answer is choice A. "Reduced" most effectively completes the blank, as what's called for here is a word or phrase that logically concludes the passage's contrast between increased decisiveness when information flow between the prefrontal cortex and parietal cortex is intensified and decreased decisiveness when information flow between these two brain regions is lowered.

> So after reading, it's about neuroeconomics from the University of Zurich. They're trying to see how decision making is linked in to communication between two parts in the brain, to two cortexes in the brain. And it's kind of like a comparison, like in the second sentence: ". . . tend to be more decisive if the information flow between the regions is intensified." So I have to find the word that is, like, a contrast, like, the two parts in the second sentence.
>
> And answer choice A, "reduced." "Reduced" is correct because it is different from the "intensified." "Intensified," I think, means more, and "reduced" means less. So A is correct.
>
> [Choice] B, "evaluated." I don't think they are looking for something, so B is not correct.
>
> [Choice] C, "determined." I don't think it's related to "determined."
>
> [Choice] D, "acquired [*participant struggles to pronounce*]." And D I don't think is correct.
>
> So [choice] A is, I think, the best choice. So my final answer is A.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW59 shows adequate passage comprehension (behavior 1) through his paraphrase and structural analysis of the passage: "And it's kind of like a comparison, like in the second sentence. . . . So I have to find the word that is, like, a contrast, like, the two parts in the second sentence." He draws on both contextual and vocabulary knowledge to recognize that “intensified” is the word that needs to be contrasted with and finds that in choice A, “reduced”: “‘Intensified,’ I think, means more, and ‘reduced’ means less.” He’s less clear about why choices B, C, and D are incorrect, but the knowledge he gains from applying vocabulary and comprehension skills leads him to the best answer, choice A (behavior 2).

Reading and Writing Question 10

Skill/Knowledge Testing Point	Cross-Text Connections
Performance Score Band	4
Stimulus Subject Area	Humanities
Stimulus Text Complexity	SCO (upper postsecondary, grades 9–11)
Required Behaviors	1. Read and demonstrate comprehension of Text 1, including its point of view on the topic. 2. Read and demonstrate comprehension of Text 2, including its point of view on the topic. 3. Demonstrate an understanding of the fundamental relationship between the two passages in terms of topic, content, and/or point of view. 4. Select the answer choice that best meets the criterion set forth in the question’s stem.
RW59 Performance Level	1

Text 1
Graphic novels are increasingly popular in bookstores and libraries, but they shouldn’t be classified as literature. By definition, literature tells a story or conveys meaning through language only; graphic novels tell stories through illustrations and use language only sparingly, in captions and dialogue. Graphic novels are experienced as series of images and not as language, making them more similar to film than to literature.

Text 2
Graphic novels present their stories through both language and images. Without captions and dialogue, readers would be unable to understand what is depicted in the illustrations: the story results from the interaction of text and image. Moreover, Alison Bechdel’s *Fun Home* and many other graphic novels feature text that is as beautifully written as the prose found in many standard novels. Therefore, graphic novels qualify as literary texts.

Based on the texts, how would the author of Text 2 most likely respond to the overall argument presented in Text 1?

A) By asserting that language plays a more important role in graphic novels than the author of Text 1 recognizes

B) By acknowledging that the author of Text 1 has identified a flaw that is common to all graphic novels

C) By suggesting that the story lines of certain graphic novels are more difficult to understand than the author of Text 1 claims

D) By agreeing with the author of Text 1 that most graphic novels aren’t as well crafted as most literary works are

Question 10, a medium-difficulty (PSB 4) Cross-Text Connections question set in a challenging (SCO) humanities context, requires test takers to draw the most reasonable conclusion connecting the content of the two topically related passages presented. This involves comprehension of each passage separately as well as making the appropriate synthetic connection "bridging" the two passages. In this case, test takers are asked to determine how the author of Text 2 would most likely respond to the argument presented in Text 1. Text 1's premise is that graphic novels don't qualify as and therefore "shouldn't be classified" as literature because the words in a graphic novel are subordinate to the visuals in meaning making; Text 2, on the other hand, argues that "graphic novels qualify as literary texts" because the words are just as important as the visuals to comprehension and because the language used in some graphic novels is as beautiful as that in some traditional prose works. Given this, the author of Text 2 would most likely respond to the author of Text 1 as choice A, the best answer, does, by "asserting that language plays a more important role in graphic novels than the author of Text 1 recognizes." Note that, commensurate with its relative level of challenge, the question doesn't simply ask for a statement of each author's point of view but rather calls on test takers to focus on a specific part of the comparison the two authors implicitly draw between each other's views.

> Text 1, I think, means just that graphic novels are getting popular, but they shouldn't be known as literature because they are different and because they have pictures, images, and no languages. So they shouldn't be the same as literature, where it's all about language. And now for Text 2. [*long pause*] OK. Text 2 seem to agree that graphic novel is a literature because it says they present their stories through both language and images. And [Text 2] says, it gives an example of one of the, Alison [Bechdel's] work that graphic novels feature text that is as beautifully written as things from, in many standard novels. They also have text in them. So Text 1 is just, and Text 2 are just different arguments. So Text 2's author, they probably wouldn't agree, would disagree on what is said in Text 1.
>
> So the [choice] A here. [*Reads choice A*] I don't think A is correct because it doesn't really want to say [language is] more important. They just want to state that it is part of the graphic novels.
>
> [Choice] B. It's not just, like, a flaw in the graphic novels. B is also not correct.
>
> [Choice] C. [*Reads choice C*] C is not correct. It's not, like, difficult to understand [the story lines of certain graphic novels]. Text 2 just wants to say that they have both languages and images. So they're not, like, more difficult to understand.
>
> And [choice] D. [*Reads choice D*] D is also not correct because they don't really agree. So D is just completely wrong.
>
> C is also wrong because it's not more difficult to understand. B, there's no flaw. A—so maybe A can be correct because it doesn't really contradict what is said in Text 1 and Text 2. So maybe even though it doesn't really look like it's the best choice, but I think that's the only one that works.
>
> So my final answer is A.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW59 demonstrates adequate comprehension of Text 1 (behavior 1) when he notes that the text "means just that graphic novels are getting popular, but they shouldn't be known as literature because they are different and because they have pictures, images, and no languages," and he shows adequate comprehension of Text 2 (behavior 2) when he observes that "Text 2 seem to agree that graphic novel is a literature because it says they present their stories through both language and images." He captures the basic relationship between the two passages (behavior 3) when asserting that "Text 1 is just, and Text 2 are just different arguments" and that "Text 2's author, they probably wouldn't agree, would disagree on what is said in Text 1." He's initially dissatisfied with the best answer, choice A, on the grounds that Text 2 "doesn't really want to say [language is] more important"—presumably, more important than images—but rather "just want[s] to state that it is part of the graphic novels." In truth, choice A asserts that the author of Text 2 recognizes a larger role for language in graphic novels than does the author of Text 1—something RW59 doesn't seem ever to grasp. Nonetheless, he's able to use knowledge of both passages and the criterion in the question's stem to rule out the incorrect answer choices: choice B because there's no "flaw in the graphic novels" indicated in Text 2, choice C because Text 2 doesn't assert that the story lines of certain graphic novels are "more difficult to understand" than claimed by the author of Text 1, and choice D because the two authors "don't really agree." This leaves him with the best answer, choice A, which he selects (behavior 4) because he's eliminated the other options and because while it "doesn't really look like it's the best choice," it also "doesn't really contradict what is said in Text 1 and Text 2."

Reading and Writing Question 11

Skill/Knowledge Testing Point	Central Ideas and Details
Performance Score Band	3
Stimulus Subject Area	Literature
Stimulus Text Complexity	SCO (upper secondary, grades 9–11)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that best states the main idea of the passage or accurately states a detail from the passage.
RW59 Performance Level	1

The following text is adapted from Ann Petry's 1946 novel *The Street*. Lutie lives in an apartment in Harlem, New York.

The glow from the sunset was making the street radiant. The street is nice in this light, [Lutie] thought. It was swarming with children who were playing ball and darting back and forth across the sidewalk in complicated games of tag. Girls were skipping double dutch rope, going tirelessly through the exact center of a pair of ropes, jumping first on one foot and then the other.

©1946 by Ann Petry

Which choice best describes what is happening in the text?

A) Lutie is observing the appearance of the street at a particular time of day and the events occurring on it.

B) Lutie is annoyed by the noise of children playing games on her street.

C) Lutie is puzzled by the rules of certain children's games.

D) Lutie is spending time alone in her apartment because she doesn't want to interact with her neighbors.

Question 11, an easy (PSB 3) Central Ideas and Details question set in a challenging (SCO) literature context, requires test takers to generalize about the content presented in the passage. Choice A is the best answer. The background information presented in the question informs readers that Lutie, the passage's narrator, lives in a Harlem apartment. The passage itself suggests that Lutie is observing activities on the street from her apartment window at a particular time of day: the "sunset was making the street radiant," and the street was "swarming with children" playing various games, such as rope jumping.

> It looks like a really happy thing from the sunset and the children. They are playing around on the sidewalks and look, like, really happy, and Lutie, she says the street is nice in this light. OK. And the question asks, which choice matches what's happening? OK.
>
> And answer choice A. [*Reads choice A*] A could be right.
>
> [Choice] B. It's annoying. No, I don't think, she is not annoyed because it says the street is nice in this light. B is not correct.
>
> [Choice] C. She's "puzzled by the rules [of certain children's games]." I don't think she's puzzled by anything. She doesn't, she seems to be enjoying the game. It doesn't, like, really say in the text, so I don't think C is correct.
>
> [Choice] D. She is spending time in her home alone. No, she's not alone. I think she's probably outside.
>
> So I think the best choice—I said it was A, so my final answer choice is A.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW59 shows adequate passage comprehension (behavior 1) via his summary: “It looks like a really happy thing from the sunset and the children. They are playing around on the sidewalks and look, like, really happy, and Lutie, she says the street is nice in this light.” He’s able to rule out each incorrect answer choice directly: B is wrong because Lutie “is not annoyed,” C is wrong because “I don’t think she’s puzzled by anything,” and D is wrong because “she’s not alone” and is “probably outside” (though it’s possible she could be observing the street from her apartment window). Finally, he selects the best answer, choice A, as his response (behavior 2).

Reading and Writing Question 12

Skill/Knowledge Testing Point	Central Ideas and Details
Performance Score Band	6
Stimulus Subject Area	Humanities
Stimulus Text Complexity	PSR (postsecondary readiness, grades 12–14)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that best states the main idea of the passage or accurately states a detail from the passage.
RW59 Performance Level	1

Many literary theorists distinguish between *fabula*, a narrative’s content, and *syuzhet*, a narrative’s arrangement and presentation of events. In the film *The Godfather Part II*, the *fabula* is the story of the Corleone family, and the *syuzhet* is the presentation of the story as it alternates between two timelines in 1901 and 1958. But literary theorist Mikhail Bakhtin maintained that *fabula* and *syuzhet* are insufficient to completely describe a narrative—he held that systematic categorizations of artistic phenomena discount the subtle way in which meaning is created by interactions between the artist, the work, and the audience.

Which choice best states the main idea of the text?

A) Literary theorist Mikhail Bakhtin argued that there are important characteristics of narratives that are not fully encompassed by two concepts that other theorists have used to analyze narratives.

B) Literary theorist Mikhail Bakhtin claimed that meaning is not inherent in a narrative but is created when an audience encounters a narrative so that narratives are interpreted differently by different people.

C) The storytelling methods used in *The Godfather Part II* may seem unusually complicated, but they can be easily understood when two concepts from literary theory are utilized.

D) Narratives that are told out of chronological order are more difficult for audiences to understand than are narratives presented chronologically.

Question 12, a hard (PSB 6) Central Ideas and Details question set in a highly challenging (PSR) humanities context, requires test takers to determine the passage's main idea. The best answer is choice A, as the main focus of the question is Mikhail Bakhtin's view that *fabula* and *syuzhet* are "insufficient to completely describe a narrative." The passage defines the concepts of *fabula* and *syuzhet* and illustrates them with the example of *The Godfather Part II* but then questions these concepts' adequacy by citing Bakhtin's belief that "meaning [in art] is created by interactions between the artist, the work, and the audience."

> So there's two different things: narratives' content and narratives' arrangement and presentation of events. OK. So in the first film, it says it's just a story of a family. But the other movie is a "presentation of the story as it alternates between two timelines in 1901 and 1958." So the second one is more current. It shows presentation of events across a, across more than fifty years. And. OK. And it says [Bakhtin] believes the narrative's content are insufficient to completely describe the narrative.
>
> OK. So he feels that this kind of work with the presentation of events doesn't really show, like, the interactions between the artist. OK. So now we want to visit the answer choices.
>
> Choice A. "[Literary theorist Mikhail Bakhtin] argued that there are important characteristics [of narratives that are not fully encompassed by two concepts that other literary theorists have used to analyze narratives]." I don't think there's, like, any two concepts that other theorists have used to analyze narratives. So I don't think A is right.
>
> [Choice] B. [*Long pause*] B is not interpreted differently by different people because the text only talks about the person [Bakhtin] himself, how he feels [about] it. So it's not, like, different people.
>
> [Choice] C. *Godfather Part II*. It's not really complicated; that literally says so in the text. So C is not correct.
>
> [Choice] D. "Narratives that are told [out of chronological order are more difficult for audiences to understand than are narratives presented chronologically]." OK. So D just contradicts what is said.
>
> I don't know. Maybe A is correct because these things are, you know, insufficient to completely describe a narrative. So maybe A. Yeah, A could be good because there are these two concepts. So I think A is probably the best choice for this one.
>
> My answer choice is A.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW59's verbalization isn't without errors and missteps, but he ultimately selects the best answer. RW59 shows adequate—albeit imperfect—passage comprehension (behavior 1) when he describes *fabula* as "narratives' content" and *syuzhet* as "narratives' arrangement and presentation of events" and when he says, somewhat imprecisely, that "[Bakhtin] believes the narrative's content are insufficient to completely describe the narrative." He mistakenly seems to think that two films, rather than two dimensions of a single film, are being discussed in the passage, and he unreasonably rules out choice B on the grounds that “the text only talks about the person [Bakhtin] himself, how he feels [about] it,” whereas the passage actually contrasts Bakhtin’s views with those of other literary theorists. RW59 does, however, effectively eliminate choices C and D as incorrect because neither is supported by the passage. Having eliminated—on solid grounds or no—all the distractors, RW59 returns to the best answer, choice A (behavior 2), and decides it “could be good” because the passage does talk about two concepts that, in Bakhtin’s view, are inadequate tools with which to analyze narratives.

Reading and Writing Question 13

Skill/Knowledge Testing Point	Command of Evidence: Textual
Performance Score Band	4
Stimulus Subject Area	Science
Stimulus Text Complexity	SCO (upper secondary, grades 9–11)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Demonstrate an understanding of the relationship between the criterion set forth in the question’s stem and the passage. 3. Select the answer choice that best meets the criterion set forth in the question’s stem.
RW59 Performance Level	1

Fish whose DNA has been modified to include genetic material from other species are known as transgenic. Some transgenic fish have genes from jellyfish that result in fluorescence (that is, they glow in the dark). Although these fish were initially engineered for research purposes in the 1990s, they were sold as pets in the 2000s and can now be found in the wild in creeks in Brazil.

A student in a biology seminar who is writing a paper on these fish asserts that their escape from Brazilian fish farms into the wild may have significant negative long-term ecological effects. Which quotation from a researcher would best support the student’s assertion?

A) “In one site in the wild where transgenic fish were observed, females outnumbered males, while in another the numbers of females and males were equivalent.”

B) “Though some presence of transgenic fish in the wild has been recorded, there are insufficient studies of the impact of those fish on the ecosystems into which they are introduced.”

C) “The ecosystems into which transgenic fish are known to have been introduced may represent a subset of the ecosystems into which the fish have actually been introduced.”

D) “Through interbreeding, transgenic fish might introduce the trait of fluorescence into wild fish populations, making those populations more vulnerable to predators.”

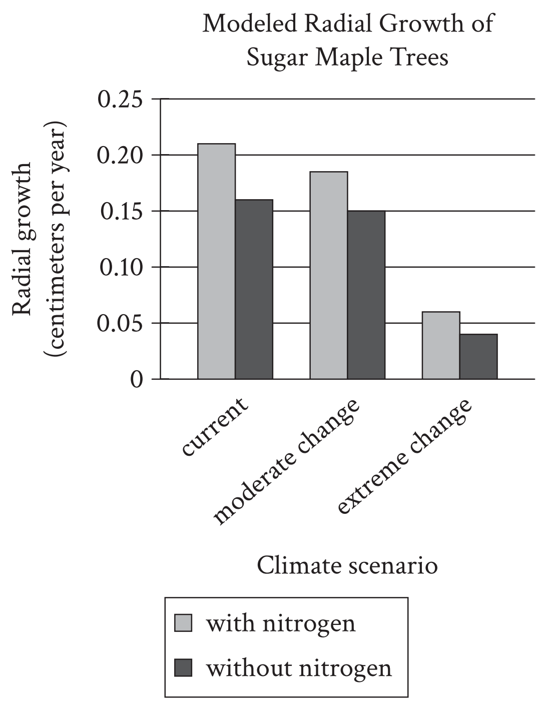
Question 13, a medium-difficulty (PSB 4) Command of Evidence: Textual question set in a challenging (SCO) science context, requires test takers to select the quotation from among the answer choices that best supports the student's claim that the escape from containment of transgenic fish "may have significant negative long-term ecological effects." The best answer is choice D. The passage defines the term *transgenic* as it relates to fish and brings up the example of fluorescent fish found in the wild in Brazilian creeks. Given this, choice D makes the most sense here, as it describes a tangible negative consequence of such fish escaping into the wild: By passing on their trait of fluorescence via breeding, these fish may make their populations more vulnerable to predators.

> So some fishes that have to have things that glow in the dark, and it seems pretty, it was used for engineering purposes, but they were sold as pets in the 2000s. And so they escaped from the farms to the wild and may have significantly negative long-term ecology effects. So.
>
> [Choice] A is not really correct because they just different stuff. It doesn't really show, like, a correlation of effects. [Choice] A is about females outnumbered males, but other numbers, they are the same, so it doesn't really show a pattern.
>
> And [choice] B. [*Reads choice B*] B might be correct, but it doesn't really support what the student is saying, that [the escape] might cause significant negative long-term ecology effects.
>
> And [choice] C. [*Reads choice C*] C might be correct because it does show side effects that we introduced may represent a subset of ecosystem[s].
>
> [Choice] D. [*Reads choice D*] I think D might be the best choice. Yes. [*Rereads choice D*] Yeah, I think D is probably the best choice because it does support that the fish, they have a new trait that makes them more visible in the dark, so if the wild fish, they get it, they might be more vulnerable to predators. I think my final answer choice is D.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW59 exhibits adequate passage comprehension (behavior 1) through his summary of the passage's content: "So some fishes that have to have things that glow in the dark, and it seems pretty, it was used for engineering purposes, but they were sold as pets in the 2000s. And so they escaped from the farms to the wild and may have significantly negative long-term ecology effects." He makes clear both here and throughout his verbalization that he understands that the goal is to best support the student's assertion that transgenic fish that escaped from Brazilian fish farms into the wild "may have significant negative long-term ecological effects" (behavior 2). He reasonably rules out choice A because the quotation "doesn't really show, like, a correlation of effects" and "doesn't really show a pattern." He decides that choice B may be true but that "it doesn't really support what the student is saying, that [the escape] might cause significant negative long-term ecology effects." He seems to have some trouble processing choice C, claiming that it "does show side effects that we introduced" when it merely contends that we don't know for certain how many ecosystems transgenic fish have been introduced into. Nonetheless, he selects the best answer, choice D (behavior 3), on the reasonable basis that it "does support that the fish, they have a new trait that makes them more visible in the dark, so if the wild fish, they get it, they might be more vulnerable to predators."

Reading and Writing Question 14

Skill/Knowledge Testing Point	Command of Evidence: Quantitative (Graph subtype)
Performance Score Band	7
Stimulus Subject Area	Science
Stimulus Text Complexity	PSR (postsecondary readiness, grades 12–14)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Demonstrate an understanding of the graph, including what the graph as a whole as well as its various components (e.g., bars) represent. 3. Demonstrate an understanding of the relationship among the passage, the graph, and the criterion set forth in the question's stem. 4. Select the answer choice that best meets the criterion set forth in the question's stem.
RW59 Performance Level	4



Inés Ibáñez and colleagues studied a forest site in which some sugar maple trees receive periodic fertilization with nitrogen to mimic the broader trend of increasing anthropogenic nitrogen deposition in soil. Ibáñez and colleagues modeled the radial growth of the trees with and without nitrogen fertilization under three different climate scenarios (the current climate, moderate change, and extreme change). Although they found that climate change would negatively affect growth, they concluded that anthropogenic nitrogen deposition could more than offset that effect provided that change is moderate rather than extreme.

Which choice best describes data from the graph that support Ibáñez and colleagues' conclusion?

A) Growth with nitrogen under the current climate exceeded growth with nitrogen under moderate change, but the latter exceeded growth without nitrogen under extreme change.

B) Growth without nitrogen under the current climate exceeded growth without nitrogen under moderate change, but the latter exceeded growth with nitrogen under extreme change.

C) Growth with nitrogen under moderate change exceeded growth without nitrogen under moderate change, but the latter exceeded growth without nitrogen under extreme change.

D) Growth with nitrogen under moderate change exceeded growth without nitrogen under the current climate, but the latter exceeded growth with nitrogen under extreme change.

Question 14, a hard (PSB 7) Command of Evidence: Quantitative question set in a highly challenging (PSR) science context, requires test takers to use data from the graph to best support the conclusion of Ibáñez and colleagues that "anthropogenic nitrogen deposition could more than offset" the negative impact of climate change "provided that change is moderate rather than extreme." Choice D is the best answer, as it accurately and appropriately compares growth without nitrogen (i.e., without "anthropogenic nitrogen deposition," or artificial fertilization) under the current climate to both growth with nitrogen under moderate climate change and growth with nitrogen under extreme climate change. These comparisons are relevant to supporting the researchers' claim because the researchers assert that using nitrogen fertilizer will "more than offset" the effects of moderate climate change but not those of extreme climate change. This claim is supported by data in the graph drawn from two comparisons: first, that growth without nitrogen under the current climate (dark gray bar above the heading "current") is exceeded by growth with nitrogen under moderate climate change (light gray bar above the heading "moderate change"), which indicates an offsetting of the effects of moderate climate change via the use of artificial fertilizer, and, second, that growth without nitrogen under the current climate exceeds growth with nitrogen under extreme climate change (light gray bar above the heading "extreme change"), which indicates that the effects of extreme climate change can't be offset by adding nitrogen.

> OK. So sugar maple trees receive periodic fertilization. Nitrogen, "to mimic the broader trend." [*Inaudible*] OK. All right. You have three different climate scenarios: the current one, the moderate one, and the extreme one.
>
> So it says [climate change] will negatively affect growth. [*Looks at graph*] Yes, it does. [*Long pause*] OK. That's good. OK. So it says "moderate" and "extreme." So I'm looking at the graph that is towards the right—not

> the current one, but the two block graphs on the right side here. And for answer choice A, "Growth with nitrogen under the current . . .," and so that's about current climate. So I think—
>
> So A, B, [or] D is just incorrect because it says "current climate" for the text here, they don't really talk about how currently the climate changes. So only C will be correct because that one doesn't talk about the current climate.
>
> So I think my final answer choice for this is C.
>
> *Participant RW59*

Participant RW59 answered the question incorrectly but did demonstrate two required behaviors, resulting in a PL of 4. RW59 shows an understanding that the passage deals with how climate change "will negatively affect growth" in sugar maple trees (behavior 1) and that the graph depicts "three different climate scenarios" and supports the passage's interpretation ("So it says [climate change] will negatively affect growth. Yes, it does"; behavior 2). However, he concludes that only choice C could be correct ("A, B, [or] D is just incorrect") because "that one doesn't talk about the current climate." He seems to believe that because the passage and graph address climate change, discussion of current conditions in the answer choices isn't germane. Accordingly, he selects choice C as his answer.

Supplementary Vignette: Participant RW57

Participant RW57 answered question 14 correctly and demonstrated all required behaviors, resulting in a PL of 1. RW57 was one of only two participants who answered the question correctly, both of whom doing so while also demonstrating all required behaviors.

> [*Reads choice A*] So this one, I don't think it's right because it's saying that moderate change has more offset and [deposition; *participant says "decomposition"*] when they have, like, the nitrogen. But this one is saying that without nitrogen is under extreme change has more. So I don't think it's right. It's comparing without and with. So it's, like, not the same category.
>
> And [choice] B. [*Reads choice B*] This one is also comparing without to with, so I'm not gonna choose this one.
>
> [Choice] C. [*Reads choice C*] So this one is also comparing with and without.
>
> [Choice] D. [*Reads choice D*] I think D is most correct because it compares to maple trees with the nitrogen [deposition]. And so we can see the difference that if there's more offset or less offset as the time passing. So I'm going to choose D.
>
> *Participant RW57*

As noted above, choice D is the best answer because it (and only it) includes the proper terms of comparison: growth without nitrogen under the current climate, growth with nitrogen under moderate climate change, and growth with nitrogen under extreme climate change. RW57 approaches this question chiefly conceptually rather than empirically and thus makes relatively light use of the specific data in the graph (behavior 2). She shows both adequate passage comprehension (behavior 1) and a clear understanding of the relationship between the passage, graph, and question stem (behavior 3) by the manner in which she reasons through each of the answer choices. The common theme in her approach is identifying the answer option with the correct terms of comparison. She rules out choices A, B, and C because each illogically compares "with nitrogen" and "without nitrogen" conditions, resulting in category errors. She ultimately identifies choice D, the best answer (behavior 4), as "most correct": ". . . it compares to maple trees with the nitrogen [deposition]. And so we can see the difference that if there's more offset or less offset as the time passing." In other words, choice D is the best answer because growth without nitrogen under the current climate is being compared to both growth with nitrogen under moderate climate change and growth with nitrogen under extreme climate change.

Reading and Writing Question 15

Skill/Knowledge Testing Point	Inferences
Performance Score Band	4
Stimulus Subject Area	History/social studies
Stimulus Text Complexity	MID (middle school/junior high, grades 6–8)
Required Behaviors	1. Read and demonstrate comprehension of the passage. 2. Select the answer choice that most logically completes the passage.
RW59 Performance Level	1

In dialects of English spoken in Scotland, the "r" sound is strongly emphasized when it appears at the end of syllables (as in "car") or before other consonant sounds (as in "bird"). English dialects of the Upland South, a region stretching from Oklahoma to western Virginia, place similar emphasis on "r" at the ends of syllables and before other consonant sounds. Historical records show that the Upland South was colonized largely by people whose ancestors came from Scotland. Thus, linguists have concluded that ________

Which choice most logically completes the text?

A) the English dialects spoken in the Upland South acquired their emphasis on the "r" sound from dialects spoken in Scotland.

B) emphasis on the "r" sound will eventually spread from English dialects spoken in the Upland South to dialects spoken elsewhere.

C) the English dialects spoken in Scotland were influenced by dialects spoken in the Upland South.

D) people from Scotland abandoned their emphasis on the "r" sound after relocating to the Upland South.

Question 15, a medium-difficulty (PSB 4) Inferences question set in a moderately challenging (MID) history/social studies context, requires test takers to complete the text (i.e., fill in the blank) with the most logical text-based inference. Choice A is the best answer. The passage establishes, first, that the "r" sound is sometimes strongly emphasized in English dialects spoken in Scotland; second, that English dialects in the Upland South of the United States carry the same emphasis; and, third, that the Upland South region was largely colonized by Scots. The most logical inference from this information is that the English dialects spoken in the Upland South gained their emphasis on the "r" sound from English dialects spoken in Scotland.

> OK. So it seems like the Upland South, they also have a similar accent of the English spoken in Scotland. "Have concluded that . . ." We have to find a conclusion for it, fill in the blank from the answer choices.
>
> So [choice A,] "The English dialects spoken [in the Upland South acquired their emphasis on the 'r' sound from dialects spoken in Scotland]." A is correct. "Spoken in the Upland South." And it's true that they probably got the sound from them because they were colonized largely by people who came from Scotland. So A could be correct.
>
> [Choice] B. [*Reads choice B*] B is just, like, a prediction. It's not a clear conclusion. B is probably not correct.
>
> [Choice] C. [*Reads choice C*] C is not "influences" because it's kind of like the, Scotland is spirited to Upland South rather than it was influenced by them. So C is not correct.
>
> And [choice] D is not correct. It didn't say they abandoned them.
>
> So the answer choice for this one is A. My final answer choice is A.
>
> *Participant RW59*

Participant RW59 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW59 demonstrates adequate passage comprehension (behavior 1) when he notes that "it seems like the Upland South, they also have a similar accent of the English spoken in Scotland" and that this region "probably got the sound from [the Scots] because they were colonized largely by people who came from Scotland." He reasonably eliminates choice B because "it's not a clear conclusion" supported by the passage, choice C because it reverses the direction of influence, and choice D because the passage "didn't say" that people in Scotland abandoned their emphasis on the "r" sound after relocating to the Upland South. Guided by this understanding, RW59 selects the best answer, choice A, as his response (behavior 2).

PARTICIPANT PERCEPTIONS

Following the think-aloud activity, Reading and Writing participants were asked a standardized set of six follow-up questions. An analysis of participants' responses to each of the questions follows.

General Impressions

1. Please tell me a bit about the experience you just had. What was it like to answer those questions?

In response to postexperience question 1, participants tended to describe the experience of the think-aloud activity as good, with a few characterizing it as difficult or challenging and a few others as “OK.” Several participants cited the novelty of thinking aloud; of those who made this observation, most found thinking aloud helpful, though at least one participant found it unhelpful. A few participants noted that, overall, the experience was similar to other SAT Suite testing they’ve done, while others noted that it differed markedly.

> I’d say it was really good. You know, I’m the type of person that likes to read to myself. So I’m a very quiet reader, you know, analyze the question in my head, and that’s how I usually answer questions. I did do some practice SAT questions in class and in summer school, so I’d say that the similarity is the same. And I also found that it was easier to really compare both of those. Overall, I’d say it was a really good experience. I think the SAT is going to become digital now, so, you know, having the experience of using the laptop and going on the app, going through each of the questions and answers, I think that would be really great in training for the actual test. *RW28*
>
> So, first of all, I liked how challenging it was. It was challenging because, for English second-language speakers or other third-language speakers, even if English is their second language, it’s really hard to read, even though they are really, really good. And they’ve been here for a long time. Like, if they read, they might stutter and have an accent, or they might not fully understand what they said because of their accent. So it was really challenging for me. *RW31*
>
> I think the experience was OK. This really makes me think—reading it out loud and actually going through each choice and trying to figure out why each answer choice is wrong. Because usually what happens on the test is because of time restrictions, I have to skim through, and if I find a choice that’s right, that’s usually what I go with without reading the rest and trying to find out why that choice is wrong. So I think this is a better way to get quicker and practice questions and just get faster at answering these questions. I quickly get through the test. *RW46*
>
> I kind of felt like, for most of them, I was kind of confused because for some of the answers, they were, like, kind of similar thing[s] for me. And I feel like, I felt overwhelmed with, like, some, like, the passages because, like, there was a lot of words, and I would also have to, like, reread to, like, figure out what type of answer would fit with, like, either the blank spaces or what they’re trying to say. *RW48*
>
> I think it was, honestly, a good experience for me too because I’ve never really thought out loud. I always think in my head. So I feel like this, at least this helped me realize that maybe I need to think about specific parts of the question or, like, highlight more or stuff like that rather than just skimming through everything in my head. So I do feel like it benefited me too, you know? *RW51*
>
> Well, of course, I liked it better than the actual test because I can express my thoughts while thinking. Because in the test, I’m just rushing through it. Like, sometimes I’m not really thinking because I know I have to finish fast. . . . It was nice expressing what I think while I answer a question. *RW52*

> So, for me, answering those questions was like any other test. Maybe there was, like, a question or two where I was thinking about it and just guessed. . . . But, other than that, it was easy. It wasn't bad. It was just a regular test for me. *RW53*

> It felt really different compared to when I actually took my SAT because [then] I didn't have the opportunity to read it out loud and then kind of just, like, let it stick to my, like, brain and understand the information completely. And by, I guess by reading it out loud, it made it easier to answer the questions. So that was one thing. But there are different words that are a little confusing for, you know, people like me, [an] English learner who doesn't understand. So it's, like, it confuses us a lot . . . *RW55*

> It was a bit harder to answer questions when I'm reading the question first 'cause I won't understand, really, the question if I'm reading it out loud. So I had to reread it in my brain, and so it was a bit harder to answer questions out loud. *RW60*

Strategies

2. How would you describe your general approach, in terms of strategies, for answering the questions?

Participants identified a range of general strategies in response to postexperience question 2. Most frequently mentioned were checking for "fit" between answer choices, question, and passage; rereading; using answer elimination; writing down one's own answer first before looking at the provided choices; reading the question before reading the passage (something participants were constrained from doing in the think-aloud activity); and circumventing the presented question order to answer the "shorter" Standard English Conventions questions first in order to save time for other, more comprehension-focused questions. Several participants mentioned skipping entirely or making minimal use of the student-produced notes included in Rhetorical Synthesis questions. While this approach works with the Rhetorical Synthesis questions included in the study [Reading and Writing questions 7 and 8], it's a risky strategy to employ more generally, as some such questions, due to the known prevalence of this "shortcut," include incorrect answer choices that misstate one or more aspects of the notes.

> My general approach, I feel like, is usually just trying to walk myself through every step, making sure I read the question properly and process what it's asking me. And then I, depending on what kind of question it is, I go to what I normally look at first—like, maybe a claim, or sometimes I just look at the [answer choices] immediately and see if it matches with the question. And then I just continue talking myself out until I reach my final answer. *RW5*

> I would see which answer made more sense based on what the passage was about, and I compared the questions with the answers. *RW25*

So, basically, my strategy was to look at the question, not look at the question, read through the passage, and then do process of elimination. That's how they taught it to me in class. So that's the same strategy I used when going through these questions. *RW28*

So my first step was to read. I'm not going to talk now about general people; I'm just going to say about myself. I used to be a person who just picked the answer because I was lazy. I would not read it and would just pick an answer sometimes. But, like, this year and last year, my sophomore and junior years, it was more challenging. I was like, OK, if I have to do this, first I read the context, then I read the question, and then I read the choices. But every day, I would just read the context of the reading again and again so I could understand it better, or I would come back to that question and answer it later on. It was hard to do critical thinking, and I'm not in that stage yet, I guess, because it was really hard. It was a lot to process in, like, two minutes. You can't process anything. So, yeah, that was hard. *RW31*

I have more challenges in the Math section, but in the Reading [and Writing section], I have got a lot of tips from the Princeton Review book. It wasn't that much of cost, so I could afford it. And I understood, like, the eight types of questions that you have that is like vocabulary, transitions, claims, purpose—all that. So basically, first, I would identify the question type and, afterwards, I would, like, choose how to approach it depending on the question type. Some questions require more highlighting than others. For example, the conclusion ones, I would usually highlight more, so I can, like, have the relationship between all the sentences. Other ones from, like, the vocabulary—and the vocabulary, the claims, and the purpose—I usually type down an answer before going to the answer choices. So that's actually pretty good. So I cannot get confused because sometimes the wording is pretty similar in the other choices. And that helps me a lot. Also, I would say that in the rhetorical ones—oh, the rhetoric ones. I would, I don't read the bullet points because I have read that it consumes you a lot of time so that if you read only the goal, you can check all the answers based on the goal. So I use that, and for the transition ones, I also highlight and write if the two sentences agree or disagree, and that helps me a lot to eliminate the answers. *RW35*

So I have different approaches for different questions. For the logically, like, logical words at the beginning of the module, like fill-in-the-blank [responses], I usually just don't look at the answer choices. So I recognize the question right away. It's asking me to find a word that completes the sentence. So I just go and try to figure, like, make my own word there. And then once I make my own word, I try to match it with the answer choices. And if I don't understand the words, which sometimes happens in the second module, I try to break each word down by using prefixes or suffixes in the answer choices. And then for logically completing the text, I look for certain things in the paragraph and just make a mental note of it. And then in the ending questions for Rhetorical Synthesis, I don't look at the reading. If it says this and this are similar, I just look for an answer choice that completes that. I find an answer choice that shows this-and-this are similar or this-and-this are different. *RW46*

Normally, I usually look at the question before I read the text. Then I read the text and keep the question in mind during my reading so I don't have to reread it. Then I pick the answer that looks the best. I look at the others to make sure that maybe there are two that look kind of similar, and I see which one looks the best. But sometimes it's pretty obvious because all the other answers are, like, not fitting the question except one of them. *RW47*

For the Reading [and Writing] section, I would go straight to number 15, question 15, and start with the grammar questions because those are the easier points to, like, collect because at the end of the day, you just have this really straight-out and flat-out very easy and limited amount of grammar rules you have to memorize to get those kind of questions right. So I just, like, get them out of the way quick and then collect all those points and then come back to the vocabulary questions. My, I think I have a decent vocabulary, but sometimes here and then I kind of find myself contemplating between two options or, like, three options on a bad day. So I would just, like, pick an answer choice and then go through it. I don't want to spend any more than forty seconds doing so. But for the reading [comprehension questions], I really want to get my logical reasoning act in place, so I would just focus on them more, like, emphasis. So I would save them for the last because, I mean, I could get them wrong, so why spend a lot of time when I can, like, go ahead and spend a lot of time on other questions that I know I can get right? *RW49*

For some questions, I used process of elimination to see which answers connected with the passage or didn't. For others, I looked for things not referenced in the passage and eliminated those. And some questions had extreme language, like "never" or "always," so I knew those were usually wrong answers. *RW50*

So I usually, if I was taking a full test, I would start with question 15, the Standard English Conventions [ones], because I could go through those faster and then have more time for the ones in the beginning. And the questions that ask me about the best word choice. I do feel like my strategy is not that great right now. But based on this [think-aloud activity], I had a, I think, I tried a new strategy today, actually, which is, like, plugging in the word to the sentence and reading how the meaning goes. So I feel like that strategy actually would probably help me more. So I think I'm gonna keep using that strategy. *RW51*

For the ones where you have to read the notes and then answer questions about the notes [Reading and Writing questions 7 and 8], I think it's better to read the question first and see if you can find the answer without the notes. If you can't, then go back to the notes—it saves time. For the main idea questions, I think it's good to look at the first or last sentence in the text—that's usually where the main idea is. Then you can paraphrase that in your head or make a note on the test and choose the answer that fits. *RW54*

> So I kind of—like, this [simulated] test, we have, like, unlimited time, I mean, like, more time than the real test. So on the real test, I would, like, try not to read all of them. I try to, like, skim through them, but this one, I had, like, more time, so I read all of the texts and answer choices, and I compare[d] each of them to make sure which one is correct. But on the real test, I don't think I have that much time for it. I probably just, when I find the correct choice, I'll just go to the next question instead of looking at all of [the options] again. So I think a strategy on the real test is, like, not to read this carefully or maybe just skim through the text and just skip through the answer choices. *RW59*

"Easy" Question Types

3. Was there a particular type of question that you found especially easy to answer? If so, which one and why?

Responses to postexperience question 3 tended to converge on certain question types or features of questions that participants associated with ease. The most frequently mentioned types or features were Rhetorical Synthesis questions (in part because of the aforementioned "shortcut" to answering them), questions in literature contexts, short questions, and questions in the fill-in-the-blank format.

> In my opinion, the Rhetorical Synthesis questions [Reading and Writing questions 7 and 8] are the easiest ones to answer because you don't really have to look at the reading notes. You just have to look for the keyword and try to find that keyword in the answer choices. So if it's asking if this research study is similar to this research study, you're just trying to find the answer in the text where they compare another and have, like, a similar objective within that answer, if that makes sense. *RW46*

> I think the ones where they ask, like, you know, the ones where they have the blank, like "the (blank) was upset" and you had to see which phrase would make the most sense. And the same with the wallpaper [Reading and Writing question 5] because it's kind of like you just have to look at what they're saying about the person and then choose the answer that matches why that person is acting that way. *RW47*

> The one that was really specific about the street [Reading and Writing question 11] was the easiest one for me. It was really clear about what the question was, so it was easy. *RW48*

> I feel like for me, one of the easiest ones is a transition word because I feel like it's easier to look at the first part, or the part before the blank, and then compare it in my head to the part after the blank, which helps me establish their relationship. And that's much easier to fill in the word in the middle. *RW51*

> I think usually the ones that I find easy are the bullet point questions [Reading and Writing questions 7 and 8]. Sometimes I don't even have to read all the bullet points to know which answer is correct. *RW52*

> I'm trying to think. Maybe the ones that were just fill-in-the-blank or something like that. OK. Those were easy. *RW53*

> I think the ones where they describe a piece of a book, like "The Yellow Wallpaper" [Reading and Writing question 5] and ask what's going to happen in the scene. They give you a description, and then you just match it to the right answer. Those are the easiest because it's already giving you a clue. *RW54*

> I think the one [that's] easy to answer is to those, like, the short claims, for example, from a book, and you need to find the example that is close to the claim. For those one[s], it's, like, you know what they want, and you know what is the claim already. You just need to find something to support it and find evidence, why it's supporting it . . . *RW57*

> I think for the word choice question, that is the word choice questions, I feel is kind of more straightforward without coming to, really, like, a long paragraph. It's just fill-in-the-blank. *RW59*

> I like the questions [that] are, like, a short story and [ask] "Which text do you think is from the short story?" or "What's the main idea of the passage?" or, like, especially like, "Which quotation is gonna be from a story?" It's very easy to answer because it just gives you a small piece, and then the answers are pretty obvious. *RW60*

"Hard" Question Types

4. Was there a particular type of question that you found especially hard to answer? If so, which one and why?

Broadly speaking, participants' responses to postexperience question 4 settled on a few question types and features that were associated with difficulty. Most frequently mentioned were questions with difficult vocabulary, questions with informational graphics, questions about main ideas and conclusions, and questions with longer and/or similarly worded answer choices. Reading and Writing question 14, which concerned nitrogen deposition in various climate change scenarios, ticked several of these boxes and was frequently called out as particularly challenging.

> I don't remember which one—like, the specific ones. I feel like the ones I usually struggle with the most are the ones with difficult vocabulary because I have to think twice as hard for what the definition is and what the purpose of the sentence is. *RW5*

> Some words were hard because I didn't know them since English isn't my first language. . . . I think the last question I answered, there were some words I didn't know. I just picked the one that seemed to fit better, but I think it might be wrong. *RW25*

> There were definitely some that were hard to answer, like that bar graph that we went through with the nitrogen and without the nitrogen [Reading and Writing question 14]. You know, the answer choices kind of tripped

me up because, you know, it's a long answer choice, and, you know, reading through it, I kind of got a bit confused, like, "Oh, yeah. Which one[s] should I compare?" You know, looking at the bar graph and then looking back at the text, it kind of made it a little hard to comprehend that. But, you know, after analyzing, I was able to choose the right answer. *RW28*

Usually, I have a lot of struggle with conclusion ones and main ideas. Well, for this one specifically, the main ideas wasn't that bad, but sometimes the main ideas and claims confuses me because sometimes the answer choices are pretty lengthy, and I can, like, sometimes get confused among the wording, and sometimes I usually, when I check my errors, I sometimes, like, fuse the answer choices. So it's a little tough. *RW35*

Something that's difficult to answer in the Bluebook are vocab and evidence-based questions where you're presented with a graph and you have to quickly analyze it and read the paragraph at the bottom. I find that difficult because of time constraints, so I usually save those for last and try to finish grammar or Rhetorical Synthesis first. *RW46*

We didn't do any of those types of questions here, but when I did the SAT test, the harder ones were like the Standard English–type questions because even in my other language, French, my first language, I don't know how to do that. Like, I don't know where the little points go and stuff like that. I just kind of guess. *RW47*

The one, I feel like the ones that are, like, the answers are, like, pretty similar with, like, the one or the two changes of the words [e.g., Reading and Writing question 14]. I found that kind of hard for me to, like, think about. *RW48*

I would say the questions that asked about [how to] most logically complete the text. These type of questions, really, the—on the ACT, there are a lot of advice saying don't go way beyond the passages. But these type of questions make you look out of the passages, which means, I mean, you're only restricted by your imaginations and what the choices are offering to you. So this can get a little bit confusing, and you can find yourself contemplating between one or two answer choices as they seem, as both of them are from the outside, as all of them are from the outside knowledge. So you might find yourself, like, second-guessing or even having to choose between three other choices that [are] completely similar. And the other type of questions that I find really, really hard are the quantitative evidence–type of questions, those graph questions, because they really take a lot of time. They need your logical reasoning; they need your utmost distinction. And using, when I use my strategy, I usually save them for last. No, not for one of the last question that I do, but I kind of skip around questions when I get to the second module. And so if I'm doing a reading question last, with the graph question last, [that] really has some sort of impact [on] the time crunch, especially because of how many, how much time that the type of question needs. So I find them really difficult to do under that time crunch. And the other type of questions are the main idea questions or the excerpts from literary works. I'm not a big fan of literary works, so I really have a hard time with those just because of how, like, abstruse the words, the wordings are, especially the poems, how abstruse the word that the literatures are, and those are not everyday English. So I find myself having trouble with this type of questions. *RW49*

Probably the dual-text questions. . . . Because it's more time consuming. There are two texts, and you have to compare both to figure out if it's one or the other, which takes more effort than a single passage. *RW50*

And, generally, I feel like the hardest ones for me are the ones, the scientific kind of questions, like the nitrogen question [Reading and Writing question 14]. It's just a little more difficult for me to understand. So whenever I'm going through those, I always mark, and I always skip it. Whenever I look at it, I just skip over it. I answer everything else and come back to that at the end so I have more time for it. But I feel like it's more difficult for me to understand. It's not the scientific terms that make it difficult to understand. But I feel like connecting their conclusion to the evidence that they gave to the answer choices is just a little complicated for me. *RW51*

Yeah, I felt like the one with the graph [Reading and Writing question 14] was a bit complex. It seemed like all the answers were similar, so that was confusing. The one about *The Godfather* [*Part II*] and the narrative [Reading and Writing question 12] was also a bit complex, but it was mostly the structure and the answers that made it challenging. It felt like there were two possible correct answers. *RW53*

Yeah, definitely the scientific chart questions. They're hard for me because there's a lot of information to process, like scientific terms, names of researchers that are hard to pronounce, and all of that fills up your head, so it's hard to focus. *RW54*

I kind of feel the—like the notes, like a student doing a research project and taking notes [Reading and Writing questions 7 and 8]. Those kinds of questions, like towards the end, are more difficult because sometimes you have to look carefully about, like, What is actually the student looking for? Like, the one we did earlier, like, sometimes that the answer choice, it might give you the correct choice, but it doesn't really fit what the student is looking for. So you look carefully for that type of question. *RW59*

Other ones with answer choices [that are] kind of similar. It's like you get to choose which ones are more correct than the other. Or sometimes when I didn't understand the passage, then that's when it became difficult for me to answer the question and answers. *RW61*

EL Status Impact

5. Did you encounter anything in the questions that you had difficulty with given your comfort level with the English language? If so, what was it, and why was it difficult for you?

When asked, via postexperience question 5, to identify factors involving their status as English learners that may have affected their think-aloud performance, participants frequently cited challenges with English vocabulary. Informational passages, passages with complex sentence structures, and longer passages were also cited by multiple participants. Roughly a quarter of participants indicated they encountered no particular challenges in answering the questions in the think-aloud activity that could be attributed to their status as English learners.

> Again, maybe it was just the vocabulary for the questions with more advanced vocabulary. I, and I just have to use context clues to figure [it] out. *RW5*
>
> I had no issues with that. I was able to comprehend everything. I was able to read through it, you know, [and] use the process of elimination and the strategies I learned in class. *RW28*
>
> Yeah, I think I have to talk about it generally for English speakers. I would talk about how vocabulary is really hard for us. Some of the vocabulary we would know would be like "we," "he," "she," "they," or maybe some advanced words we would know, but we can't understand all the vocabularies in the questions that you gave me. Like, it was really hard words. Second of all, . . . there are people in my school who take the ACT, but they are really new. I've been here for three years, so I can understand words better. But there are people who've been here for maybe four months, five months, or six months and totally don't know the words. So, vocabulary is really, really hard for me, generally, too. *RW31*
>
> I think that the very complex sentences are really tough because, like, I needed to reread them several times so I can really follow up the line and, like, yeah, find the relation between that and the other sentences. So maybe I think that sometimes also when you, like, repeat certain words constantly, like, in the same sentence, it's so confusing because, like, you, I lose track of it all. So that was pretty hard. *RW35*
>
> Since English is my third language, I have trouble with vocabulary. I have been learning prefixes in order to try to answer that question that pops up on the test. *RW46*
>
> Sometimes the grammar is a little bit hard to understand. And, like, when they use fancy words. Sometimes it's close to French because in French we have fancy words too, but sometimes I just don't know what they mean. *RW47*
>
> I would still bring the literary works because I love, like, I do love literature when it's in my mother tongue language. . . . But when it comes to English, I didn't have my time with reading literary works. My

> passion is for, in the medicine field, premed. And I really didn't enjoy reading [literature] in English because it's my third language. And I don't basically, I don't really have that much time to spend on reading the literature works for, by, in three types of languages. *RW49*

> Well, sometimes, when—well, it depends if I get a hard module—the fill-in-the-blank, like, the words: sometimes I don't know them. I don't know. I've never heard of that word, and I try to, well, look at the prefixes or the suffixes, but sometimes that doesn't even make sense. And sometimes, well, some words are similar to Spanish, but that doesn't, sometimes they don't make any sense either. So I just have to guess in those, in that scenario. *RW52*

> The vocabulary questions can be difficult if you're not comfortable with English. In this case, they weren't too bad, but in other practices I've taken, there have been questions where the vocabulary was really tough. It's hard to study for those because you can't learn every word in the English language. *RW54*

> Sometimes the harder part is, maybe it's the length of the passage, because in a long passage, like, when I first look at it, I'm, like, already tired to, like, actually look over the whole thing. So, like, my attention will go, like, everywhere. It's not like I'm not gonna really focus in reading on every single one. So I'm going to waste some time to read one more time and one more time again. And that's kind of distracting me from, like, reading [for understanding]. *RW57*

Final Comments

6. Is there anything about your test-taking experience today or about the test-taking strategies you used today that we haven't talked about yet but that you'd like us to know?

In response to postexperience question 6, most participants either didn't have additional final comments or reiterated themes raised via other interview questions. Among the unique comments were those that mentioned the importance of keeping a good pace and avoiding being distracted by extraneous material in test questions.

> It was a lot of experience reading a lot—like, reading it, reading it properly correctly, getting [an] understanding and then answering it, looking at the answers and the questions twenty-three times to get it right. *RW29*

> I usually also skip the hard questions. That's a very thing, a very useful thing I use. I have a personal order of difficulty. . . . And I use—no, I first answered the grammar questions, then the transition ones, rhetoric ones, and then I start again with the vocabulary and all the ones until about question 15 or 16, which again goes back to grammar. So I use that a lot. *RW35*

> I'd say the best strategy is to keep track of how long you're spending on each question. If it's more than a minute or so, make sure you're moving fast enough to finish on time. It's better to guess than to leave questions blank, since you don't lose points for guessing. . . . But I do struggle with that, because if I don't know the answer, I get stressed, and that affects how I do on the rest of the test. *RW54*
>
> I like to cross—I like to eliminate the answer options. That helped a lot. And then also highlighting context flows. Annotation is definitely, like, the major part of it. *RW55*
>
> I think the strategy is, like, I don't think it's, like, something new because it's always just read and analyze it and find the question out. I find the answer out. I think the only thing is that sometimes maybe just catch the key points—like, sometimes the name is not really important—and just catch what is the cause and what is the effect and why do you use that and catch some important verbs or important adjectives. *RW57*
>
> I think we have more time today. It's kind of the same approach [as usual]; I just had more time. So I was more careful today than when I take the real test. There's no, like, other different strategy. *RW59*

Math

PARTICIPANT AND QUESTION PERFORMANCE

Participant and Question Performance Levels and Differentials

Figure 2 displays, as a single matrix, the Math participant and question performance data derived from this study. An explanation of the intended method of reading the figure is provided in the corresponding subsection of the Reading and Writing results, above, although the following differences should be observed:

- For the Math domain, expected behaviors, rather than required behaviors, were defined to account for the fact that some Math questions are, by design, open to multiple, largely mutually exclusive solution paths.
- Because of the above difference, PL 2 was unobtainable by Math participants, as they were only expected to answer each question correctly and demonstrate at least one expected behavior. (For Reading and Writing, by contrast, PL 2 was attainable for questions with more than two required behaviors by participants who answered a given question correctly and demonstrated one or more additional required behaviors but not all such behaviors.)

Figure 2. Math Participant and Question Performance Summary Matrix.

Part. ID	Question #															Performance by Level, by Participant						Participant Performance Summary		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	5	*NR*	*#AC*	*#EB*	D_p
M8	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	14	–	0	1	0	0	14	14	0 ✔
M9	1	1	4	1	4	1	1	1	5	1	1	1	1	1	1	12	–	0	2	1	0	12	12	0 ✔
M25	1	1	3	3	5	4	1	1	4	1	1	1	4	1	5	8	–	2	3	2	0	10	8	2 ✔
M29	1	1	1	1	1	4	1	1	1	1	1	1	1	1	5	13	–	0	1	1	0	13	13	0 ✔
M35	1	1	1	5	4	5	1	1	1	1	1	1	1	1	1	12	–	0	1	2	0	12	12	0 ✔
M36	5	1	5	5	5	5	5	5	1	5	3	5	5	5	5	2	–	1	0	12	0	3	2	1 ✘
M41	1	1	1	1	1	1	1	1	1	1	1	1	1	4	1	14	–	0	1	0	0	14	14	0 ✔
M45	1	5	3	5	5	5	1	5	4	5	5	5	5	5	5	2	–	1	1	11	0	3	2	1 ✘
M48	1	1	5	1	4	1	1	1	1	1	1	1	1	1	1	13	–	0	1	1	0	13	13	0 ✔
M49	1	1	1	5	5	5	1	1	1	1	5	1	1	1	1	11	–	0	0	4	0	11	11	0 ✔
M50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15	–	0	0	0	0	15	15	0 ✔
M51	1	1	1	5	1	1	1	1	1	1	1	1	1	5	1	13	–	0	0	2	0	13	13	0 ✔
M52	1	1	1	1	4	5	1	1	1	1	1	1	1	1	1	13	–	0	1	1	0	13	13	0 ✔
M53	1	1	1	1	4	5	1	1	1	1	1	1	1	1	1	13	–	0	1	1	0	13	13	0 ✔
M54	1	5	5	5	4	5	1	1	1	1	4	4	5	5	5	5	–	0	3	7	0	5	5	0 ✔
M55	1	5	5	5	5	5	5	1	1	1	1	1	5	5	5	6	–	0	0	9	0	6	6	0 ✔
M57	1	1	1	5	4	1	1	1	1	1	1	5	1	1	1	12	–	0	1	2	0	12	12	0 ✔
M58	1	5	5	1	4	5	1	1	5	1	5	1	5	5	5	6	–	0	1	8	0	6	6	0 ✔

Performance by Level, by Question

1	17	14	10	9	4	7	16	16	14	16	13	14	12	11	11
2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
3	0	0	2	1	0	0	0	0	0	0	1	0	0	0	0
4	0	0	1	0	9	2	0	0	2	0	1	1	1	1	0
5	1	4	5	8	5	9	2	2	2	2	3	3	5	6	7
NR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Performance Legend

1 (highest): Answered correctly; exhibited 1+ expected behaviors
2: *Not applicable to Math*
3: Answered correctly; exhibited no expected behaviors
4: Answered incorrectly; exhibited 1+ expected behaviors
5 (lowest): Answered incorrectly; exhibited no expected behaviors

Question Performance Summary

#AC	17	14	12	10	4	7	16	16	14	16	14	14	12	11	11
#EB	17	14	10	9	4	7	16	16	14	16	13	14	12	11	11
D_q	0 ✔	0 ✔	2 ✔	1 ✔	0 ✔	0 ✔	0 ✔	0 ✔	0 ✔	0 ✔	1 ✔	0 ✔	0 ✔	0 ✔	0 ✔

Summary Legend

#AC = # answered correctly

#EB = # answered correctly; demonstrated 1+ expected behaviors

D_p, D_q = Differentials (*#AC* – *#EB*); ✔ = criterion-passing differential (70%+), ✘ = criterion-failing differential (<70%)

Findings

Participant Performance

As shown in the "Participant Performance Summary" sub-table of figure 2, sixteen of eighteen participants (89 percent) met or exceeded the criterion for a good participant differential (D_p), which provides evidence that these participants were able to adequately demonstrate cognitively complex thinking in line with the question types' constructs. Two participants had a criterion-failing differential; although only 1 in both cases, this differential failed the criterion owing to the fact that each participant answered only three questions correctly. These two participants were still able to demonstrate cognitively complex thinking on two-thirds of the (small number of) questions they did answer correctly, indicating that these participants were able to demonstrate cognitively complex thinking in line with the question types' constructs at least some of the time.

Question Performance

As shown in the "Question Performance Summary" sub-table of figure 2, all fifteen studied Math questions (100 percent) met or exceeded the criterion for a good question differential (D_q), which provides evidence that these questions are capable of eliciting cognitively complex thinking from English learners.

PARTICIPANT PERFORMANCE VIGNETTES

Case Study: Participant M9

Participant M9 was selected as the Math case study participant using the same criteria as outlined for the Reading and Writing case study. M9, a female twelfth grader from Illinois, identified as White and not of Hispanic, Latino, or Spanish origin. She self-reported a HSGPA of A+ (not uncommon among the sample), indicated that she often used English in her everyday life, typically spoke English and another language at home, listed Polish as a known language besides English, and rated her English language acquisition level as 4 ("I can understand the main ideas of complex texts in English"). M9 answered twelve of the fifteen Math questions correctly and demonstrated at least one expected behavior in all cases, resulting in a participant differential of 100 percent, which exceeded the threshold for a good D_p.

Math Question 1

Content Domain	Algebra
Skill/Knowledge Testing Point	Linear Inequalities: Identify
Performance Score Band	4
Stimulus Subject Area	Science
Question Format	MC
Expected Behaviors	1. Read and demonstrate comprehension of the context described. 2. Set up/identify a linear equation or inequality as described in the context.
M9 Performance Level	1

For a snowstorm in a certain town, the minimum rate of snowfall recorded was 0.6 inches per hour, and the maximum rate of snowfall recorded was 1.8 inches per hour. Which inequality is true for all values of *s*, where *s* represents a rate of snowfall, in inches per hour, recorded for this snowstorm?

A) $s \geq 2.4$

B) $s \geq 1.8$

C) $0 \leq s \leq 0.6$

D) $0.6 \leq s \leq 1.8$

Question 1, a medium-difficulty (PSB 4) multiple-choice Linear Inequalities: Identify question set in a science context, requires test takers to identify a linear inequality that represents the given context. The correct answer (*key*) is choice D. It's given that the minimum and maximum rates of snowfall recorded were 0.6 and 1.8 inches per hour, respectively. Therefore, the rate of snowfall, *s*, ranges from 0.6 to 1.8 inches per hour.

> So when I see a word problem like this, what I first do is just write down every number I see. So it's asking for *s* in inches per hour. The minimum rate of snowfall was 0.6 inches per hour, and the maximum was 1.8 inches per hour. Based on what it's telling me, *s* has to be in between 0.6 and 1.8 because the number has to be in between those two, since one is a minimum and one is a maximum. So my final answer choice is D.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. After reading and demonstrating comprehension of the context ("Based on what it's telling me, *s* [the rate of snowfall] has to be in between 0.6 and 1.8 [inches per hour]"; behavior 1), she's able to identify the correct answer (behavior 2) as choice D "because the number [*s*] has to be in between those two, since one is a minimum [rate] and one is a maximum [rate]." M9 fails to explicitly account for the fact that choice D includes both 0.6 and 1.8 themselves as possible values, and not just the numbers "in between," but this has no material impact on her question answering.

Math Question 2

Content Domain	Problem-Solving and Data Analysis
Skill/Knowledge Testing Point	Ratios
Performance Score Band	5
Stimulus Subject Area	Real-world topics
Question Format	MC
Expected Behaviors	1. Read and demonstrate comprehension of the context described. 2. Use the ratio and given information to set up and solve a proportion.
M9 Performance Level	1

At a particular track meet, the ratio of coaches to athletes is 1 to 26. If there are x coaches at the track meet, which of the following expressions represents the number of athletes at the track meet?

A) $\frac{x}{26}$

B) $26x$

C) $x+26$

D) $\frac{26}{x}$

Question 2, a medium-difficulty (PSB 5) multiple-choice Ratios question set in a real-world context, requires test takers to identify the expression that best represents the situation by either logically deducing this relationship from the context or through calculation by setting up a proportion. The correct answer is choice B. It's given that at a particular track meet, the ratio of coaches to athletes is 1 to 26. Logically, a test taker could determine from the context that the number of athletes at this track meet, given the provided ratio and x number of coaches, must be $26x$ (choice B), as the ratio indicates that there are twenty-six athletes for every coach. By calculation, a test taker could arrive at the same conclusion by setting up and solving the proportion $\frac{1\ \textit{coach}}{26\ \textit{athletes}} = \frac{x\ \textit{coaches}}{y\ \textit{athletes}}$, resulting in $y = 26x$, where y represents the number of athletes at the track meet.

> So the ratio is 1 to 26, which means that per 1 coach, there are 26 athletes. So there are x coaches at the track meet, which would mean that if there is 1 coach—I mean, there are 26 athletes, which would mean if there's 1 coach, it has to be $\frac{x}{26}$. Yeah, it has to be $\frac{x}{26}$ because of that. No, sorry, it would be $26x$ because based on the ratio of coaches, it would be 26 times the number of coaches there are. So my final answer choice is B.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated a single expected behavior, resulting in a PL of 1. She first reads and demonstrates comprehension of the context (behavior 1), observing that "the ratio is 1 to 26, which means that per 1 coach, there are 26 athletes." She then asserts, without much explanation, that the correct answer "has to be $\frac{x}{26}$," which is choice A. M9 then realizes her mistake and correctly concludes that "it would be $26x$" (choice B) "because based on the ratio of coaches, it would be 26 times the number of coaches there are." Given this reasoning, she selects the correct answer, choice B, as her response.

Math Question 3

Content Domain	Geometry and Trigonometry
Skill/Knowledge Testing Point	Circles
Performance Score Band	6
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Using the graph of a circle in the *xy*-plane, determine a possible *x*-value on the graph. 2. Identify the center of a circle in the *xy*-plane. 3. Identify the radius of a circle in the *xy*-plane. 4. Using the equation of a circle in the *xy*-plane, identify the domain of the circle.
M9 Performance Level	4

$$(x+4)^2+(y-19)^2=121$$

The graph of the given equation is a circle in the *xy*-plane. The point (a, b) lies on the circle. Which of the following is a possible value for *a*?

A) -16

B) -14

C) 11

D) 19

Question 3, a hard (PSB 6) multiple-choice Circles question outside of context, requires test takers to demonstrate an understanding of where the graph of a circle exists in the *xy*-plane by identifying a possible *x*-coordinate of a point that lies on that circle. The correct answer is choice B. The standard equation for a circle is $(x-h)^2+(y-k)^2=r^2$, where *h* and *k* represent, respectively, the *x*- and *y*-coordinates of the circle's center and where *r* represents the circle's radius. The equation given in the question is written in this standard form, meaning that the described circle's center is (−4, 19) and its radius (the square root of r^2) is 11. The domain of a circle, or set of all possible *x*-values within that circle's boundary, is represented by the inequality $h-r \leq x \leq h+r$, where *x* is the domain, *h* is the *x*-coordinate of the circle's center, and *r* is the circle's radius. For the given equation, the circle's domain is thus $-4-11 \leq x \leq -4+11$, or [−15, 7]. Choice B, −14, is the only offered value that lies within the domain bounded by −15 and 7 and thus the only possible value for *a* among the answer options. Alternatively, students could use a graphing calculator, such as the one built into Bluebook, to graph the equation of the circle, visually inspect where the circle exists in the *xy*-plane, and then identify the only possible value for *a* among the answer choices.

> So what I would do first is probably plug in the equation into my graphing calculator. So I'm typing it in exactly as it appears on the screen, and I'm just going to look for one of those points on the circle. I just have to zoom out—weird. OK. So it's not popping up, but I'll redo this. OK. I'm typing the equation into my graphing calculator again. OK.

> It looks like it does not work on my calculator, so I'll just do it by hand. Just by looking at this, I would probably try plugging in each number because a has to be an *x*-coordinate. Well, I know that the radius is 11 because $\sqrt{121}$ is 11. So, looking at that, answer choice C is eliminated; it can't be 11 because that can't be an *x*-value. I would also eliminate answer choice D because that would be the *y*-value, not the *x*-value. So, I would choose answer choice A because 4^2 is 16, and the *x*-value is the opposite sign. So my final answer choice is A.
>
> *Participant M9*

Participant M9 answered the question incorrectly but did demonstrate a single expected behavior, resulting in a PL of 4. M9 first tries to view the graph of the given equation in a graphing calculator but is unsuccessful with the device. From there, M9 attempts to "do it by hand" by "plugging in each number because a has to be an *x*-coordinate." It doesn't seem that she fully commits to this strategy, but she does recognize that "a has to be an *x*-coordinate." She properly identifies the radius of the circle as 11 ("$\sqrt{121}$ is 11"; behavior 3) and dismisses the corresponding answer option, choice C, "because that can't be an *x*-value." While her rationale isn't precisely why choice C is incorrect, this move allows her to investigate the other choices. M9 next eliminates choice D "because that would be the *y*-value, not the *x*-value" and commits to choice A as her response "because 4^2 is 16, and the *x*-value is the opposite sign." Presumably, she alludes here to the circle's center, $(-4, 19)$, which has an *x*-coordinate of -4 and a *y*-coordinate of 19. M9 doesn't give any consideration to choice B, the correct answer here, in part due to the attractiveness of the distractors, which include –16 as an option.

Supplementary Vignette: Participant M41

Participant M41 answered question 3 correctly and demonstrated three expected behaviors, resulting in a PL of 1. M41 was one of twelve participants who answered the question correctly and one of ten participants doing so while also demonstrating at least one expected behavior.

> For this one, I know that a is the *x*-coordinate of the point that lies on the circle. I'm gonna graph this out by putting in the equation we're given. From the graph, you can see that the center of the circle has an *x*-coordinate of -4, and the radius is 11. So the possible *x*-values of the point [(*a*, *b*)] that lie[s] on this circle would be from -15 to 7. So a has to be between -15 and 7, and that would be [choice] B, -14.
>
> *Participant M41*

Participant M41 recognizes that "*a* is the *x*-coordinate of the point that lies on the circle" and then graphs the circle using Bluebook's built-in graphing calculator. He then determines that "the center of the circle has an *x*-coordinate of -4, and the radius is 11" (behavior 3). He doesn't mention the *y*-coordinate of the circle's center, presumably due to his focus on a representing an *x*-coordinate. M41 concludes that "the possible *x*-values of the point [(*a*, *b*)] that lie[s] on this circle would be from -15 $[-4-11]$ to 7 $[-4+11]$" (behavior 4). From the graph, along with a fundamental understanding of circles, M41 is able to determine that –14 is a possible value of *a* (behavior 1) and selects the correct answer, choice B, as his response.

Math Question 4

Content Domain	Advanced Math
Skill/Knowledge Testing Point	Nonlinear Functions: Rewrite
Performance Score Band	7
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Use the graph of an exponential function to determine a minimum value. 2. Demonstrate an understanding of key features of the graph of an exponential function. 3. Demonstrate an understanding that exponential functions don't have relative extrema.
M9 Performance Level	1

Which of the following functions has(have) a minimum value at -3?

I. $f(x) = -6(3)^x - 3$

II. $g(x) = -3(6)^x$

A) I only

B) II only

C) I and II

D) Neither I nor II

Question 4, a hard (PSB 7) multiple-choice Nonlinear Functions: Rewrite question outside of context, requires test takers to demonstrate an understanding of minimum value in relation to exponential functions. The correct answer is choice D. Exponential functions continuously increase or decrease and therefore don't have a minimum (or maximum) value. Test takers may simply recall and apply this characteristic, or they could graph both functions to visually make this observation.

> So a minimum value at -3. I would probably go back to graphing it. I'll try using this one again. I'm just going to graph both of them and see if they have a minimum value. OK. Just by graphing the first function [function *f*], it does not have a minimum at -3; it goes down to negative infinity. So I would automatically cross out [choices] A and C. Now I'll graph the second one [function *g*], and it also goes down to negative infinity. So, seeing that both of them go down to negative infinity, they don't have a minimum value at -3. My final answer choice is D.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated a single expected behavior, resulting in a PL of 1. M9 graphs function *f* using her own handheld graphing calculator and notes that the function "goes down to negative infinity," allowing her to eliminate choices A and C, which include function *f*, since the function "does not have a minimum at -3." Again by graphing, M9 is able to determine that function *g* "also goes down to negative infinity." She concludes that as "both of [the functions] go down to negative infinity, they don't have a minimum value at -3" (behavior 1) and selects the correct answer, choice D, as her response.

Math Question 5

Content Domain	Problem-Solving and Data Analysis
Skill/Knowledge Testing Point	Percentages
Performance Score Band	7
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Convert percentages greater than 100 to decimals. 2. Write an equation to compute an increase to a quantity by a percentage greater than 100. 3. Solve a linear equation. 4. Logically eliminate multiple-choice distractors (incorrect answers) by size of numbers relative to given information and the question asked.
M9 Performance Level	4

The result of increasing the quantity x by 400% is 60. What is the value of x?

A) 12

B) 15

C) 240

D) 340

Question 5, a hard (PSB 7) multiple-choice Percentages question outside of context, requires test takers to demonstrate an understanding of a percentage increase greater than 100. The correct answer is choice A. Four hundred percent is equivalent to $\frac{400}{100}$, or 4. Therefore, increasing quantity x by 400% can be represented by the expression $x + 4x$, or $5x$. It's given that the result of increasing a certain quantity, x, by 400% is 60. Therefore, $5x = 60$, which when solved yields $x = 12$.

> So whenever I see a problem with percentages, I always just check with my answer choices. It's saying increasing x by 400% is 60. First, I'll just plug in: 12×4, which is, like, 400%, is 48, so it's not [choice] A. 15×4 is 60, so I'll check the other ones, but just by that, it's probably [choice] B. I'm just taking the numbers [in the answer choices] and multiplying them by 4 because it's asking for an increase of 400%. So I ended up with 15; 400% of 15 is 60. So my final answer choice is B.
>
> *Participant M9*

Participant M9 answered the question incorrectly but did demonstrate a single expected behavior, resulting in a PL of 4. M9's attempt at "just taking the numbers [in the answer choices] and multiplying them by 4" on the grounds that the question is "asking for an increase of 400%" demonstrates a lack of clear understanding of how to determine a percentage increase. Proceeding through the answer options and multiplying each by 4, "which is, like, 400%" (behavior 1), leads M9 to the incorrect answer of choice B. Her decision is prefaced by her assertion that "400% of 15 is 60," which isn't the question being asked.

Supplementary Vignette: Participant M50

Participant M50 answered question 5 correctly and demonstrated three expected behaviors, resulting in a PL of 1. M50 was one of four participants who answered the question correctly, all of whom doing so while also demonstrating at least one expected behavior.

> So the answer to increasing quantity *x* by 400% to get 60 would be 12. Because when you start off with 12, increasing it by 400% is essentially multiplying it by 5 instead of 4. When something is increased by 100%, it doubles the value. So, for 100%, it would be 24; for 200%, it would be 36; for 300%, it would be 48; and for 400%, it would be 60. So the value of *x* has to be 12, choice A.
>
> *Participant M50*

Participant M50 quickly determines the correct answer, choice A, and in the process concisely explains how he reached this conclusion: "Because when you start off with 12, increasing it by 400% is essentially multiplying it by 5 instead of 4" (behaviors 1 and 2). Next, using 12 (choice A) to demonstrate, he progresses through what it would mean for that quantity to be increased by 100% ("it doubles the value"; 24), 200% (36), 300% (48), and finally 400% (60). From this line of reasoning, he logically eliminates the distractors (behavior 4), determines that "the value of *x* has to be 12," and selects the correct answer, choice A.

Math Question 6

Content Domain	Advanced Math
Skill/Knowledge Testing Point	Nonlinear Functions: Make Connections
Performance Score Band	7
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Make connections between the equation of a quadratic function and its *x*-intercepts. 2. Rewrite a quadratic equation in a form that facilitates identifying unknown values. 3. Given certain pieces of information, recognize characteristics of the unknown values of a quadratic function.
M9 Performance Level	1

The function f is defined by $f(x) = ax^2 + bx + c$, where a, b, and c are constants. The graph of $y = f(x)$ in the xy-plane passes through the points $(7, 0)$ and $(-3, 0)$. If a is an integer greater than 1, which of the following could be the value of $a + b$?

A) -6

B) -3

C) 4

D) 5

Question 6, a hard (PSB 7) multiple-choice Nonlinear Functions: Make Connections question outside of context, requires test takers to draw connections between a quadratic function with unknown constants and its two given *x*-intercepts. Test takers must also be capable of handling a fair amount of algebraic computation as well as understand the significance of an unknown constant being called out as a specific type of number. The correct answer is choice A. It's given that function *f* passes through the points (7, 0) and (−3, 0). Substituting 7 for *x* and 0 for $f(x)$ and also −3 for *x* and 0 for $f(x)$ in the function $f(x)=ax^2+bx+c$ yields the equations $49a+7b+c=0$ and $9a-3b+c=0$. It follows that $49a+7b=9a-3b$. Combining like terms in this equation gives $40a=-10b$, or $-4a=b$. To find $a+b$, substituting $-4a$ for b gives $a-4a$, or $-3a$. So $a+b$ is equivalent to $-3a$, which is a multiple of −3. Since it's given that a is an integer greater than 1, when a is 2, then $a+b=-3a=-3(2)=-6$.

> So, just by looking at this problem, the function it's giving me is a quadratic function, so I would write that down first. They gave me two points, $(7, 0)$ and $(-3, 0)$, so I also just write those down. a is bigger than 1, and it's asking for $a+b$. Just by that information, I would first find the slope of the two points it gave me. So that would be $\frac{0-0}{-3-7}$, which is $\frac{0}{-10}$, which is 0, meaning the slope is 0. That doesn't give me much information. I'll try to guess and check by plugging in the answer choices [in]to the actual function and seeing what makes the most sense. But I'm kind of lost. OK. I'm gonna try to plug in *x* [in]to the actual function. I'll plug in 7 into the function, which gives me $y=49a+7b+c=0$. Then I'll plug -3 into the function, which gives me $9a-3b+c=0$. Now that I have two equations, I'll try elimination. I'll multiply the bottom equation by -1 to eliminate *c* and solve for *b* and a. So I just eliminated *c*, which gives me $40a+10b=0$. I can take out a 10, which is $10(4a+b)=0$. Solving for *b*, I can divide by 10, which makes $4a+b=0$. Then $-4a$ goes to the other side, so $b=-4a$. Now I have *b*; I just need to find a. To find a, I'll do the same process: $4a=-b$, divided by 4, so $a=-\frac{b}{4}$. Then, plugging in $a=-\frac{b}{4}$ back into my other equation [$b=-4a$], $b=-4\left(-\frac{b}{4}\right)$, which is $b=\frac{4b}{4}$, which gives me $b=b$, which does not make sense to me. I think I went wrong somewhere. I'm gonna go back to the original equations and try to eliminate a different variable instead of c. OK. Going back, I'll try to get rid of a different variable. I have $49a+7b+c=0$ and $9a-3b+c=0$. So I'm going to multiply the top equation [$49a+7b+c=0$] by 3 to try to get the *b*'s to be the same, and then multiply the bottom [equation; $9a-3b+c=0$] by 7. Multiplying [the top equation] by 3 gives $147a+21b+3c=0$ and [multiplying the bottom equation by 7 gives] $63a-21b+7c=0$, which eliminates my *b*'s. Then $147+63$ is $210a+10c=0$. I'll solve for a: $210a=-10c$, and 10 divided by 210 is $\frac{1}{21}$, so $a=-\frac{1}{21}c$. So it looks like a is probably a negative number. I'm leaning towards answer choices A and B, but even if I plug a back into *b*, I still have two variables in one equation, so I can't fully solve for a or *b*. However, they said a is greater than 1, which actually means *b* has to be a negative number because

> $b=-4a$. So b has to be some bigger negative number, which would mean the answer could be [choices] A or B. If I just plug in a guess of 2 for a—sometimes I can just get lucky with numbers. If $a=2$ and I plug that back in here $\left[a=-\frac{b}{4}\right]$, $2=-\frac{b}{4}$, then $b=-8$ and $a=2$. That gives me $a+b=-6$. So my best guess is answer choice A.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. She begins her solution path by first investigating the slope of the graph of $y=f(x)$ using the two given points: "So that would be $\frac{0-0}{-3-7}$, which is $\frac{0}{-10}$, which is 0, meaning the slope is 0." She observes that "that doesn't give me much information"; the calculation was performed correctly but doesn't contribute to solving the problem. Next, M9 attempts to substitute the given points, (7, 0) and (−3, 0), for x and y into the given equation, which results in $49a+7b+c=0$ and $9a-3b+c=0$ (behavior 2). She then proposes elimination: "I'll multiply the bottom equation by −1 to eliminate c and solve for b and a." It's not clear whether her decision to eliminate c arises out of convenience (given its coefficient of 1) or strategy (the question asks for the value of $a+b$), but, regardless, this allows M9 to correctly conclude that $b=-4a$. However, her efforts to "find a" result in the dead end of "$b=b$," which she notes "does not make sense to me." Believing she "went wrong somewhere," she "go[es] back to the original equations" to "try to eliminate a different variable instead of c." In doing so, she correctly concludes that $a=-\frac{1}{21}c$ but quickly determines that she still has "two variables in one equation," which means she "can't fully solve for a or b." Continuing to persevere, M9 realizes that the question stipulates that "a is greater than 1"—more precisely, that a is an integer greater than 1—which "actually means b has to be a negative number because $b=-4a$" (behavior 3). This realization allows her to place the final piece of the puzzle: "If $a=2$ and I plug that back in here $\left[a=-\frac{b}{4}\right]$, $2=-\frac{b}{4}$, then $b=-8$ and $a=2$. That gives me $a+b=-6$." She then selects choice A, the correct answer, as her response.

Math Question 7

Content Domain	Algebra
Skill/Knowledge Testing Point	Linear Functions: Identify
Performance Score Band	2
Stimulus Subject Area	Science
Question Format	MC
Expected Behaviors	1. Read and demonstrate comprehension of the context described. 2. Set up/identify a linear equation or inequality as described in the context.
M9 Performance Level	1

A veterinarian recommends that each day a certain rabbit should eat 25 calories per pound of the rabbit's weight, plus an additional 11 calories. Which equation represents this situation, where *c* is the total number of calories the veterinarian recommends the rabbit should eat each day if the rabbit's weight is *x* pounds?

A) $c = 25x$

B) $c = 36x$

C) $c = 11x + 25$

D) $c = 25x + 11$

Question 7, an easy (PSB 2) multiple-choice Linear Functions: Identify question set in a science context, requires test takers to identify a linear equation in two variables that represents the given context. The correct answer is choice D. It's given that a veterinarian recommends that each day a certain rabbit eat 25 calories per pound of the rabbit's weight, plus an additional 11 calories. If the rabbit's weight is *x* pounds, then the total number of calories, *c*, can be written as $c = 25x + 11$.

> So it told me that *c* is the total number of calories the veterinarian recommends the rabbit should eat each day if the rabbit is a certain amount of pounds. A veterinarian recommends that usually a certain rabbit eats 25 calories times the rabbit's weight, which would give me $25x$, plus an additional 11 calories on top of that. That completely eliminates [choices] A and B because they both don't include anything about 11, which leaves [choices] C and D. Since it's saying 25 [calories] per pound, that has to mean 25 times *x* plus an additional 11. So my final answer is D.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. M9 demonstrates comprehension of the context (behavior 1), noting, for example, that "*c* is the total number of calories the veterinarian recommends the rabbit should eat each day if the rabbit is a certain amount of pounds." She determines that "25*x*," with *x* representing the rabbit's weight, in pounds, "plus an additional 11 calories on top of that" accurately expresses the amount of calories to be consumed daily. From this, M9 eliminates choices A and B because neither includes the additional 11 calories per day recommended by the veterinarian. Ruling out choice C based on the contextual understanding that 25 represents the number of calories recommended per pound of the rabbit's weight and not a fixed amount of additional calories, she concludes that $c = 25x + 11$ is the proper equation here (behavior 2) and selects the correct answer, choice D, as her response.

Math Question 8

Content Domain	Geometry and Trigonometry
Skill/Knowledge Testing Point	Measure of Angles in a Triangle
Performance Score Band	3
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Demonstrate an understanding of the triangle sum theorem. 2. Use logic to determine the maximum value of an angle in a triangle given the measure of one of the other angles.
M9 Performance Level	1

In $\triangle RST$, the measure of $\angle R$ is 63°. Which of the following could be the measure, in degrees, of $\angle S$?

A) 116

B) 118

C) 126

D) 180

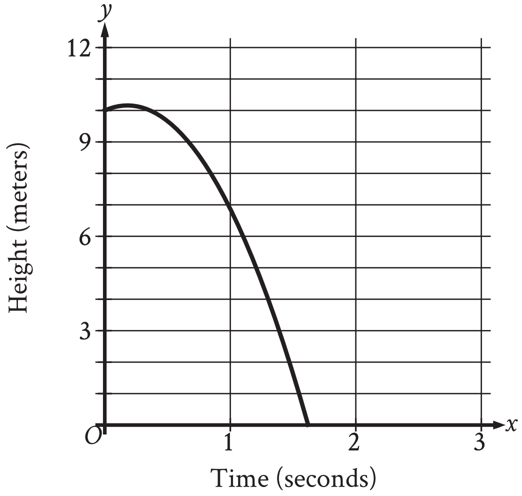
Question 8, an easy (PSB 3) multiple-choice Measure of Angles in a Triangle question outside of context, requires test takers to demonstrate an understanding of the triangle sum theorem, the concept that the sum of all interior angles of a triangle is 180°. The correct answer is choice A. For $\triangle RST$, it's given that the measure of $\angle R$ is 63°. Therefore, by the triangle sum theorem, the sum of the measures of $\angle S$ and $\angle T$ is $(180 - 63)^\circ$, or 117°. This means that the measure of $\angle S$ must be less than 117°. Of the given answer options, only choice A, 116, is less than 117 and therefore could be the measure, in degrees, of $\angle S$.

> So, just by looking at this, automatically [choices] C and D have to be gone because a triangle is 180°, meaning an angle can't be 180°, and $180 - 126 = 54^\circ$, which doesn't make sense. A triangle can't have angles that add up to more than 180[°], so it can't be [choice] C either. So I'm between [choices] A and B. In the triangle, $\angle R$ is 63[°], meaning $180 - 63 = 117$[°]. This means it can't be 118° [choice B] because there has to be space for two angles. So, by elimination, my answer choice is A.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. M9 begins her successful approach to this question by demonstrating command of the triangle sum theorem (behavior 1): "[The sum of the measures of the interior angles of] a triangle is 180°." This understanding leads her to quickly eliminate choices C and D. She rules out the former because subtracting 126° (choice C's proposed measure for $\angle S$) from 180° (the sum of the measures of the triangle's interior angles) yields 54° as the sum of the measures of $\angle R$ and $\angle T$, an impossibility considering that the given measure of $\angle R$ itself is 63°; she rules out the latter because 180° (choice D) represents the sum of the measures of the triangle's three interior angles, not a possible measure of a single interior angle of the triangle. Recognizing that ∠*R*'s measure is given as 63°, M9 determines that the combined measure of ∠*S* and ∠*T* is therefore 117° and that logically only 116°, choice A, could be a possible measure of ∠*S* itself among the provided answer options because "there has to be space for two angles" and 118°, choice B, is greater than the sum of the measures of ∠*S* and ∠*T* (behavior 2). She thus selects the correct answer as her response.

Math Question 9

Content Domain	Advanced Math
Skill/Knowledge Testing Point	Nonlinear Functions: Interpret
Performance Score Band	4
Stimulus Subject Area	Science
Question Format	MC
Expected Behaviors	1. Read and demonstrate comprehension of the context described. 2. Identify the *x*-intercept of a graph of a quadratic function. 3. Interpret the context of an *x*-intercept of the graph of a quadratic function.
M9 Performance Level	5



A competitive diver dives from a platform into the water. The graph shown gives the height above the water *y*, in meters, of the diver *x* seconds after diving from the platform. What is the best interpretation of the *x*-intercept of the graph?

A) The diver reaches a maximum height above the water at 1.6 seconds.

B) The diver hits the water at 1.6 seconds.

C) The diver reaches a maximum height above the water at 0.2 seconds.

D) The diver hits the water at 0.2 seconds.

Question 9, a medium-difficulty (PSB 4) multiple-choice Nonlinear Functions: Interpret question set in a science context, requires test takers to interpret a key feature of the graph of a quadratic function in terms of the context. The correct answer is choice B. The *x*-intercept of a graph is the point at which a graph intersects the *x*-axis, which, in the given graph, represents time, in seconds. The given graph intersects the *x*-axis between $x = 1$ and $x = 2$. In context, this means that the diver hits the water (reaches 0 on the *y*-axis, which represents height, in meters, above the water) between 1 and 2 seconds after diving from the platform, making choice B the best interpretation of the graph's *x*-intercept.

> So it's asking for the best interpretation of the *x*-intercept based on time. Looking at the graph, the height in meters is above the water, so it looks like this person is jumping off a diving board at 10 meters [and] into the water. Just looking at this, answer choice D is gone because it doesn't make sense. The diver doesn't hit the water at 0.2 seconds. Answer [choice] A is also eliminated because the diver reaches a maximum height before 1.6 seconds, so it can't be A. So I'm between [choices] B and C. The diver's maximum height above the water is 0.2 [choice C]; that could be an answer. But B also seems right to me: The diver hits the water at 1.6 seconds. Since I don't know for sure when the diver hits the water because it could be after 1.6 seconds, and the maximum on the graph is at around 0.2 seconds, my final answer is [choice] C.
>
> *Participant M9*

Participant M9 answered the question incorrectly and didn't demonstrate any expected behaviors, resulting in a PL of 5. After trying to visualize the situation ("it looks like this person is jumping off a diving board at 10 meters [and] into the water"), M9 dismisses choice D because "the diver doesn't hit the water at 0.2 seconds" and choice A because "the diver reaches a maximum height before 1.6 seconds." Next, she debates between choices B and C, noting that both seem correct and without making reference to the graph's *x*-intercept, the proper interpretation of which is the focus of the question. Her reasoning for ultimately picking choice C, an incorrect option, over choice B, the correct option, seems to be that somehow the moment at which the diver reaches maximum height above the water can be determined from the graph ("the maximum on the graph is at around 0.2 seconds") but that the moment at which the diver hits the water can't be ("I don't know for sure when the diver hits the water because it could be after 1.6 seconds"). In any case, choice C, though an accurate assertion per the graph, isn't a reasonable interpretation of the graph's *x*-intercept, which represents the moment in time, in seconds, at which the diver hits the water.

Supplementary Vignette: Participant M48

Participant M48 answered question 9 correctly and demonstrated three expected behaviors, resulting in a PL of 1. M48 was one of fourteen participants who answered the question correctly, all of whom doing so while also demonstrating at least one expected behavior.

> So if *y* is how many meters above the water it is, then when $y = 0$, it is 0 meters above the water, meaning that it is touching the water. So, since the *x*-axis is where $y = 0$, that means that the *x*-intercept of the

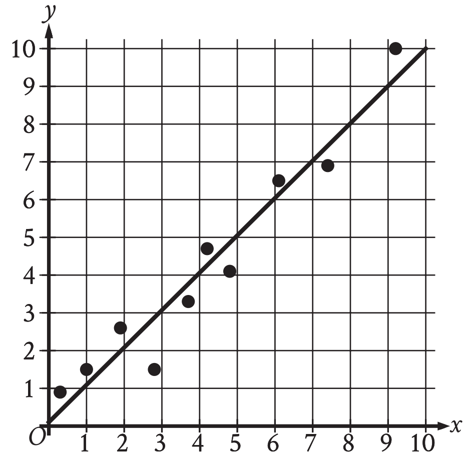
> graph is when the diver actually touches the water. So "The diver reaches a maximum height above the water at 1.6 seconds" [choice A]. That is not true because the maximum height is when the *y*-value is greatest, which is a little bit more than 10 [meters]. So "The diver hits the water at 1.6 seconds" [choice B]. That is what I believe to be correct, which is B. Answer [choice] C, "The diver reaches a maximum height above the water at 0.2 seconds." That is possibly true, but that is not the *x*-intercept, so that is not what we need. So C would be incorrect. So "The diver hits the water at 0.2 seconds" [choice D]. That is incorrect because we know that since the water is the *x*-axis, it hits [the water] at 1.6 seconds. So we know the answer is [choice] B.
>
> *Participant M48*

Participant M48 begins by demonstrating an understanding of the overall context (behavior 1) as well as a conceptual and context-based understanding of what the graph's *x*-intercept represents (behaviors 2 and 3): "So if *y* is how many meters above the water it is, then when $y = 0$, it is 0 meters above the water, meaning that it is touching the water. So, since the *x*-axis is where $y = 0$, that means that the *x*-intercept of the graph is when the diver actually touches the water." She rules out choices A and D as factually incorrect per the graph and choice C as accurate but irrelevant ("that is not the *x*-intercept"). Determining, per the graph, that "the water is the *x*-axis" and that the diver "hits [the water] at 1.6 seconds," M48 selects choice B, the correct answer, as her response.

Math Question 10

Content Domain	Problem-Solving and Data Analysis
Skill/Knowledge Testing Point	Scatterplot
Performance Score Band	4
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Understand that the data points in a scatterplot represent actual values and that the line of best fit represents predicted values. 2. Understand that for the actual *y*-values in a scatterplot to be greater than the predicted *y*-values, the data points will have to be above the line of best fit.
M9 Performance Level	1

The scatterplot shows the relationship between two variables, *x* and *y*. A line of best fit for the data is also shown.



For how many of the 10 data points is the actual *y*-value greater than the *y*-value predicted by the line of best fit?

A) 3

B) 4

C) 6

D) 7

Question 10, a medium-difficulty (PSB 4) multiple-choice Scatterplot question outside of context, requires test takers to understand what a line of best fit represents in a scatterplot. The correct answer is choice C. In conceptual terms, any data point located above a scatterplot's line of best fit has a *y*-value greater than that predicted by the line. For the given scatterplot, six of the data points are positioned above the line of best fit.

> The line of best fit is basically just the average of the *y*-values. So I would start by just counting the dots above the line of best fit. There are 6 dots above the line of best fit, so my final answer choice is C.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. M9 begins by asserting that "the line of best fit is basically just the average of the *y*-values" (behavior 1). This is an oversimplification, as a scatterplot's line of best fit is designed to minimize the distance between data points and the line, thus providing predicted values, but she's nevertheless able to successfully determine the correct answer, choice C, "by just counting the dots above the line of best fit" (behavior 2). Even though M9 doesn't precisely articulate how the line of best fit is derived, she still displays clear command of the skill being assessed.

Math Question 11

Content Domain	Problem-Solving and Data Analysis
Skill/Knowledge Testing Point	Probability
Performance Score Band	4
Stimulus Subject Area	Real-world topics
Question Format	MC
Expected Behaviors	1. Calculate, express, or interpret the probability of an event. 2. Apply the understanding that the sum of probabilities of all possible outcomes of an event is 1. 3. Determine an unknown number using probability and the context described.
M9 Performance Level	1

At a movie theater, there are a total of 350 customers. Each customer is located in either theater A, theater B, or theater C. If one of these customers is selected at random, the probability of selecting a customer who is located in theater A is 0.48, and the probability of selecting a customer who is located in theater B is 0.24. How many customers are located in theater C?

A) 28

B) 40

C) 84

D) 98

Question 11, a medium-difficulty (PSB 4) multiple-choice Probability question set in a real-world context, requires test takers to determine an unknown quantity using probability and given information. The correct answer is choice D. Per the context, each of 350 customers is located in one of three theaters, A, B, or C. It's further given that the probability of randomly selecting a customer located in theater A is 0.48 and that the probability of randomly selecting a customer located in theater B is 0.24. Therefore, the probability of randomly selecting a customer located in either theater A or theater B is $0.48 + 0.24$, or 0.72. As the sum of probabilities of all possible outcomes of an event is 1, it follows that the probability of randomly selecting a customer located in theater C is $1 - 0.72$, or 0.28. This means there are $(0.28)(350)$, or 98, customers located in theater C.

> It gave me two probabilities. So, first, I'll calculate the customers in theater A by multiplying 350 by 0.48, which gives me 168 customers in theater A. Then, for theater B, I'll multiply 350 by 0.24, which is 84 customers. To find how many are in theater C, I'd subtract, $350 - 168 - 84$, which gives me 98 people in theater C. So my final answer is D.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. She begins her approach by observing that the probability of randomly selecting a customer who is located in one of the three theaters is equivalent to the proportion of customers in that theater (behavior 1): "So, first, I'll calculate the customers in theater A by multiplying 350"—the total number of customers across all three theaters—"by 0.48"—the given probability of randomly selecting a customer who is located in theater A—"which gives me 168 customers in theater A. Then, for theater B, I'll multiply 350 by 0.24"—the given probability of randomly selecting a customer who is located in theater B—"which is 84 customers." At this point, simply subtracting these results from the total of 350 customers allows M9 to determine the unknown number of customers located in theater C (behavior 3): "To find how many are in theater C, I'd subtract, $350-168-84$, which gives me 98 people in theater C." Having made this determination, M9 selects the correct answer, choice D, as her response.

Math Question 12

Content Domain	Advanced Math
Skill/Knowledge Testing Point	Nonlinear Equations: Solve
Performance Score Band	5
Stimulus Subject Area	None
Question Format	SPR
Expected Behaviors	1. Set a quadratic equation equal to zero. 2. Apply an understanding of the zero-product property. 3. Solve a quadratic equation algebraically. 4. Solve a quadratic equation graphically.
M9 Performance Level	1

$$(d-30)(d+30)-7=-7$$

What is a solution to the given equation?

Question 12, a medium-difficulty (PSB 5) student-produced response Nonlinear Equations: Solve question outside of context, requires test takers to solve a quadratic equation, which in this case yields two distinct solutions. Correct answers are −30 and 30, though (as indicated by "a solution" as well as the overall test section directions) test takers are expected (and allowed) only to supply one such correct answer. To solve this equation algebraically, students could add 7 to both sides of the given equation. This gives $(d-30)(d+30)=0$. The zero-product property states that a product of two factors is equal to 0 if and only if at least one of the factors is 0. Therefore, $d-30=0$ or $d+30=0$. It follows that $d=30$ or $d=-30$. Another reasonable algebraic approach would be to multiply the binomials and combine like terms, resulting in the equation $d^2=900$. Applying the square root property, which states that if $x^2=c$, then $x=\pm\sqrt{c}$, to this equation gives $d=30$ or $d=-30$. This quadratic equation could also be solved graphically by entering the given equation into a graphing calculator (using *x* instead of *d*) and applying the understanding that the two vertical lines produced represent the distinct solutions to the equation.

> First, I'd add 7 to both sides [of the equation] to separate the variable, d, so $-7+7=0$. I have $(d-30)(d+30)=0$. This gives me two answers, $d=30$ and $d=-30$. Just to be sure, I'm factoring out again: $d^2+30d+900=0$. Now I'll try to find what factors into 30 from 900. After some calculations, I can't narrow it down further. I'll just go with my initial answer, $d=30$.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated three expected behaviors, resulting in a PL of 1. She begins her solution path by adding 7 to both sides of the equation so that it's set equal to zero (behavior 1), which results in $(d-30)(d+30)=0$. M9 doesn't mention the zero-product property by name but applies an understanding of it (behavior 2) by concluding that solving for d (behavior 3) "gives me two answers, $d=30$ and $d=-30$." Seemingly in an attempt to gain further confidence in her answers, she investigates an alternate solution path ("I'm factoring out again"). She makes a mistake when multiplying the binomials $(d-30)(d+30)$, resulting in the equation $d^2+30d+900=0$. When she's unable to factor the left side of this equation ("after some calculations, I can't narrow it down further"), she resorts to one of the answers she obtained from her initial solution path: "I'll just go with my initial answer, $d=30$."

Math Question 13

Content Domain	Algebra
Skill/Knowledge Testing Point	Linear Equations in Two Variables: Make Connections
Performance Score Band	5
Stimulus Subject Area	None
Question Format	SPR
Expected Behaviors	1. Rewrite a linear equation into an appropriate form to identify the slope of a graph. 2. Perform numerical calculations involving fractions and/or decimals. 3. Calculate the slope of a graph from two points on the graph.
M9 Performance Level	1

What is the slope of the graph of $y=\frac{1}{3}(29x+10)+5x$ in the xy-plane?

Question 13, a medium-difficulty (PSB 5) student-produced response Linear Equations in Two Variables: Make Connections question outside of context, requires test takers to determine the slope of the graph of a line given the equation for that line. The correct answer is $\frac{44}{3}$. A linear equation can be written in the form $y=mx+b$, where m is the slope of the graph of the line. To rewrite the given equation in this form, students could distribute the $\frac{1}{3}$ to the grouped binomial, which gives $y=\frac{29}{3}x+\frac{10}{3}+5x$. Combining like terms gives $y=\frac{44}{3}x+\frac{10}{3}$. Therefore, the slope is $\frac{44}{3}$. In Bluebook, students can validly enter this answer fractionally as

44/3 or as the decimals 14.66 or 14.67. (Either of these decimal answers would be acceptable, as the instructions provided for SPR questions state “If your answer is a **decimal** that doesn’t fit in the provided space, enter it by truncating or rounding at the fourth digit.”)

> So it’s asking for the slope. So what I would do first is try to condense this equation. So I would, I’m gonna factor out [distribute] $\frac{1}{3}$ into the parenthesis $29x+10$. So $\left(\frac{1}{3}\right)(29)$ gives me $\frac{29}{3}x+\frac{10}{3}+5x$. And to get rid of the 3, I’m going to multiply the whole equation by 3. So it’s $3y=29x+10+15x$, which gives me $3y=44x+10$. Now divide by 3 to single out the *y*, which is $y=\frac{44}{3}x+\frac{10}{3}$, which gives me $\frac{44}{3}$ [as the slope], which doesn’t simplify. So the slope of the, this equation is $\frac{44}{3}$.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. After correctly distributing $\frac{1}{3}$ to $29x$ and 10, M9 decides to “multiply the whole equation $\left[y=\frac{29}{3}x+\frac{10}{3}+5x\right]$ by 3” to “get rid of the 3 [in the fraction denominators]” (behavior 2). She combines like terms on the right side of the equation and then “divide[s] by 3 to single out the *y*” (behavior 2). With the equation now in the slope-intercept form $y=mx+b$, where m is the slope of the graph of the line, she concludes that “the slope of the, this equation is $\frac{44}{3}$” (behavior 1), which she enters as 44/3.

Math Question 14

Content Domain	Geometry and Trigonometry
Skill/Knowledge Testing Point	Scale Factor and Area
Performance Score Band	6
Stimulus Subject Area	None
Question Format	MC
Expected Behaviors	1. Apply an understanding of how applying scale factor to side lengths affects the areas of similar rectangles. 2. Calculate the area of similar rectangles using two possible side lengths. 3. Logically eliminate multiple-choice distractors (incorrect answers) by size of numbers relative to given information and the question asked.
M9 Performance Level	1

Rectangles *ABCD* and *EFGH* are similar. The length of each side of *EFGH* is 6 times the length of the corresponding side of *ABCD*. The area of *ABCD* is 54 square units. What is the area, in square units, of *EFGH*?

A) 9

B) 36

C) 324

D) 1,944

Question 14, a hard (PSB 6) multiple-choice Scale Factor and Area question outside of context, requires test takers to understand how a scale factor applied to one rectangle affects the area of a similar rectangle. The correct answer is choice D. If x represents the length, in units, of the base of rectangle *ABCD* and y represents its height, in units, then the area of rectangle *ABCD* is xy square units. It's given that each side of similar rectangle *EFGH* is 6 times the length of the corresponding side of rectangle *ABCD*. Therefore, $6x$ represents the length, in units, of the base of rectangle *EFGH*, $6y$ represents its height, in units, and $(6x)(6y)$, or $36xy$, square units represents its area. It's also given that the area of rectangle *ABCD* is 54 square units; therefore, $xy = 54$. Substituting 54 for xy in the expression $36xy$ yields $(36)(54)$, or 1,944, square units as the area of rectangle *EFGH*.

> So, first, I draw the rectangles. Because, since they're similar, that makes my life a lot easier by being able to just correspond each letter to each other. So it gave me that each, the length of each side on $EFGH$ is 6 times the length of [the corresponding side of] $ABCD$. So that means each side on $ABCD$ is x and for [each corresponding side of] $EFGH$ is $6x$, and the area of $ABCD$ is 54 [square units]. So since they are rectangles, it means two of the sides are the same, and then the other two sides are the same. So my options for $ABCD$, my options are 9 × 6. One could be 9, one could be 6. Another option is—I'm just trying to figure out factors of 54—18 and 3 and 27 and 2—or what the, like, size can be for $ABCD$. So looking at this, I don't think it's 27 and 2. So I'm gonna just put in 6 and 9. So that means $(6)(6) = 36$ and then $(9)(6) = 54$ are the side lengths of rectangle $EFGH$, which means $(36)(54) = 1{,}944$, which is one of my answer choices. So I'm just gonna leave that there for now, and I'm gonna also do 18 and 3 to see what I get there. Since $(6)(3) = 18$ and $(6)(18) = 108$, and then $(108)(18) = 1{,}944$. So because of that, my answer choice is D.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated a single expected behavior, resulting in a PL of 1. In an attempt to visualize the geometry question being asked, M9 begins by drawing the rectangles and then proceeds to restate the given information in her own words: "So that means each side on *ABCD* is x and for [each corresponding side of] *EFGH* is $6x$, and the area of *ABCD* is 54 [square units]. So since they are rectangles, it means two of the sides are the same, and then the other two sides are the same." Since the area of rectangle *ABCD* is given as 54 square units, M9 is able to assign possible values for the lengths of the sides of this rectangle: "One could be 9, one could be 6. Another option is—I'm just trying to figure out factors of 54—18 and 3 and 27 and 2." After deciding to use 6 and 9 as possible values for the lengths of the sides of rectangle *ABCD*, she multiplies each of these values by 6 given that we're told the length of each side of *EFGH* is 6 times the length of the corresponding side of *ABCD*: "So that means $(6)(6) = 36$ and then $(9)(6) = 54$ are the side lengths of rectangle *EFGH*." Next, M9 calculates the resulting area, in square units, of *EFGH* (behavior 2): ". . . which means $(36)(54) = 1{,}944$, which is one of my answer choices." In an apparent attempt to gain confidence in her answer, she performs the same calculations with two different factors of 54 and achieves the same result: ". . . I'm gonna also do 18 and 3 to see what I get there. Since $(6)(3) = 18$ and $(6)(18) = 108$, and then $(108)(18) = 1{,}944$." She then selects the correct answer, choice D, as her response.

Math Question 15

Content Domain	Algebra
Skill/Knowledge Testing Point	Systems of Two Linear Equations in Two Variables: Solve
Performance Score Band	6
Stimulus Subject Area	None
Question Format	SPR
Expected Behaviors	1. Fluently eliminate a variable in a system of two linear equations. 2. Identify the solution to a linear system from its graph. 3. Solve for a multiple of the value of *x*.
M9 Performance Level	1

$$5y = 10x + 11$$
$$-5y = 5x - 21$$

The solution to the given system of equations is (x, y). What is the value of $30x$?

Question 15, a hard (PSB 6) student-produced response Systems of Two Linear Equations in Two Variables: Solve question outside of context, requires test takers to work with a system of two linear equations in determining a multiple of the value of *x*. The correct answer is 20. Adding the two equations in the system gives $0 = 15x - 10$. Adding 10 to both sides of this equation yields $15x = 10$. The value of $30x$ can be found by multiplying both sides of this equation by 2. Therefore, $30x = 20$.

> So I can just solve this by either substitution or elimination. And I'm gonna go with elimination because it's easier since my *y*'s already have a 5 in front of them. So I'm gonna swap the *y*'s and the constant number. So I'm gonna get $-11 = -5y + 10x$ and then $21 = 5y + 5x$. This allows me to get rid of *y*. So I have $21 - 11 = 10$. So $10 = 15x$. Now I'll divide by 15 so I can find *x*. So $x = \frac{10}{15}$, which is $\frac{2}{3}$, and to get the actual answer, you have to multiply 30 times my *x*. So $(30)\left(\frac{2}{3}\right)$ gives me 20. So my final answer is 20.
>
> *Participant M9*

Participant M9 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. After deciding to use the elimination method to solve this problem "because it's easier since my *y*'s already have a 5 in front of them," M9 elects to "swap the *y*'s and the constant number," resulting in the equations $-11 = -5y + 10x$ and $21 = 5y + 5x$. She next adds the two equations, eliminating the terms that include y (behavior 1), which gives $10 = 15x$. She then proceeds to solve this equation for x (behavior 3): "So $x = \frac{10}{15}$, which is $\frac{2}{3}$." She concludes that "to get the actual answer, you have to multiply 30 times my x," which results in the correct answer, 20.

PARTICIPANT PERCEPTIONS

Following the think-aloud activity, Math participants, like their Reading and Writing counterparts, were asked a standardized set of six follow-up questions. An analysis of participants' responses to each of the questions follows.

General Impressions

1. Please tell me a bit about the experience you just had. What was it like to answer those questions?

Postexperience question 1 responses revealed diverse reactions to the think-aloud activity, with many participants mentioning the novelty of being asked to verbalize their mathematical thought processes. Several participants found this verbalization beneficial, noting that it helped them identify mistakes, while at least one other found expressing complex mental processes verbally distracting. Some participants mentioned already being familiar with the mathematical content and presentation style, making reference to previous standardized test experiences. A few participants identified challenges with question comprehension, with one student specifically noting "some words that were not very understandable" and another characterizing certain problems as "trick questions meant to confuse students." Emotive responses spanned from negative ("I was very nervous") to positive ("It was fun"), with many comparing the experience to actual testing conditions, noting differences in approach and pressure between the two settings.

> I like the experience. I like, I find it easier getting to my answer talking through my, like, thoughts, but, I mean, since you can't do that on the actual test, the actual test is a lot harder, but since I was able to, like, talk through my thinking, it made me realize, like, what mistakes I was making. So I like the experience. *M9*

> Some of [the questions] were OK, but some of the questions were very confusing because they asked you a question, and [then] there were some words that were not very understandable. Some of them didn't really make sense to me. *M45*

> The questions themselves, I feel like, are all things that I've seen before, so they weren't particularly hard. Sometimes they did trip me up because I didn't pay close enough attention. And so I wanted to not get the question over with, but I wanted to explain every step of the question. And so that sometimes maybe tripped me up or made me think that I was doing one thing when I was doing another thing. And so explaining it definitely made me need to slow it down a little. And so I wouldn't say it was hard, but it definitely took a little bit more brainpower than I was used to. *M48*

> Well, some of the stuff I learned, but I haven't reviewed it yet. So I don't remember some of the questions. I might not be able to remember how to do them, but I usually do Desmos and then look at the answer choices or think in a logical way. I think they're OK. They're not too bad. *M49*

> It was fun. I don't get to talk too much about what I think. So it was fun, you know, doing something different, approaching the questions in a different way. So a lot of people can understand what I am trying to do. But it is also limiting because the simple fact of me saying things trying to, you know, express myself and how can I approach certain questions, and it is limited to me because I do a lot of process in my head, and getting stuck on one simple process just to talk about it and explain it to other people gets me stuck. *M51*

> I was very nervous because sometimes, like, I don't know the topic, so I'm just guessing, and it's a little bit embarrassing. But so it was good. It was OK. I felt comfortable. *M55*

> Well, I was hoping that I was gonna get the English and reading section [when participating in the study] because I hate math so much. But I thought it was fun. It was definitely different to answer it by thinking out loud, but I thought that it was more helpful when I actually vocalized what I was thinking because it made me process my answers. *M57*

Strategies

2. How would you describe your general approach, in terms of strategies, for answering the questions?

In response to postexperience question 2, participants described various approaches to math problem-solving. Many emphasized careful question analysis and thorough reading, while technology tools, particularly the built-in Desmos graphing calculator, emerged as prominent resources. Participants mentioned relying on prior mathematical knowledge and adapting their approach based on question type. Visual techniques were important to certain participants, who preferred writing down their work. When facing challenging problems, participants described specific tactics such as skipping difficult questions to revisit later or employing process-of-elimination strategies. Pacing strategies varied, with some participants prioritizing efficiency while others advocated for a more methodical approach to ensure accuracy.

> Yeah, so my approach is to do the question as efficiently as possible, whether that includes using the answer choices as kind of a guide on how to navigate the question or to use the—like, basically, if I don't know how to do the question, I can, of course, reference the answer choices, or I can also think, like, What other options are there? I've also kind of learned to ask myself, like, Is there a trick in this question? And also to read really carefully. *M8*

> If I, before I, like, go into a test, I tell myself that, like, no matter the outcome, everything will turn out fine. So going into the test, I go into

> it, like, stress free, and I just do what I can. And if I get really stuck on a question, I skip it, and then I always come back. And sometimes I end up actually getting it because you have to, I have to, like, warm up my brain first. So usually when I come back, then I can figure out my answer. *M9*

> Well, I really, I think I looked at, I read the question, and then I looked at the answer choices, and then I solved, and then I just went one by one to see which [choice] was the correct one. *M45*

> I like to write everything down, and even if I don't need to do it on my calculator, I like to have it just to reassure myself that it's the correct answer whatever I'm doing. So I would say it's definitely very visual because I like to see what I'm doing on my paper. *M48*

> Anything that I could use the Desmos [graphing calculator] for, and you get some stuff that way. Yeah. So things I can't remember. So what is the biggest thing? Writing down information. I like visualizing things. *M52*

> So definitely, I think, you know, using Desmos as heavily as possible just because it saves so much time that I've been using in school so much for math sections. Especially I just try to go to that, but if not, then I do have a piece of paper with me as well where I'm writing it down. Just like, you know, for that triangle question [Math question 8]. It wasn't really practical for me to just go to Desmos all the way. Sometimes visualizing it first is much better as well. *M53*

"Easy" Question Types

3. Was there a particular type of question that you found especially easy to answer? If so, which one and why?

When asked, via postexperience question 3, about the types of think-aloud activity questions they found easy to answer, participants most frequently identified algebra questions as particularly accessible, especially equation-solving tasks and systems of equations. Some participants appreciated questions with minimal text, which they felt allowed them to focus directly on the mathematical task. Graph interpretation questions were highlighted as straightforward, particularly when variables were clearly defined. Some participants reported finding specific topic areas easier, including probability, linear equations, and geometry problems involving triangles.

> I think these types of questions are quite straightforward when they don't have a lot of words associated with them. You can get right to the point and answer the question. When there's a lot of words, it takes more time, and you have to read the question, which makes it more difficult for me to stay focused with long paragraphs. *M8*

> Probably the basic algebra questions, like the last algebra question, though, like the system of equations [Math question 15]. Probably because I'm just strong in that area. I've always just really liked algebra. *M9*

> The question with probability—the one about customers and theater rooms A, B, and C [Math question 11]. That was direct for me because with probabilities, the answer is just 1. You can just make an estimate, so it was a bit easier. *M25*

> When they asked [us] to read the graph, it's pretty straightforward because, like, the variable given to us, like x and y, is, like, self-explanatory. So when you, like, when they asked [us] to explain what does—like, one of the questions was, like, the [one about the] diver [Math question 9]. And then they give us, like, access, like, the time, and then y is, like, the height? So when the height reaches zero, it's technically like you hit the water. So that's, like, really obvious. *M35*

> The linear equations are pretty easy for me because there are only two components: the slope and the y-intercept. *M41*

> I feel like I like algebra questions. Like, what I just did. Maybe, I don't know, I feel, like, the triangle question about finding the degree [Math question 8]. The content was easier. *M57*

"Hard" Question Types

4. Was there a particular type of question that you found especially hard to answer? If so, which one and why?

When discussing the most challenging questions in response to postexperience question 4, participants consistently identified problems containing extensive text as difficult to navigate, with some noting that text-heavy problems were overwhelming even when the underlying mathematics wasn't complex. Questions involving multiple variables and/or unknown constants, especially in quadratic expressions, were frequently cited as problematic, with many participants noting struggles to determine relationships between these variables and constants. Function-related problems emerged as another significant area of reported difficulty, with participants expressing discomfort with functional notation and application. Some participants found geometry problems with a limited amount of given information challenging, particularly when these questions required visualization or multistep reasoning. Time constraints appeared to be a factor in perceived difficulty, with participants indicating that questions requiring deeper thinking become especially challenging under testing conditions.

> I believe there were some, like, even the diving one [Math question 9], which had a long paragraph, and then there was the theater one with the probabilities [Math question 11]. I think it's easy to get lost with such a long problem, even though mathematically it's not very complicated. *M8*

> The hardest one was for sure the quadratic formula one [Math question 6]. That one was really complicated because instead of two variables, which is what I'm usually working with, it was three variables that kind of, like, [I] couldn't figure out what, like, how to solve for all three variables in the time, like, I would be usually given. *M9*

> The one about triangle RST [Math question 8]. There wasn't much information. They just gave me one angle and asked me for the other angle, and it was kind of hard. *M25*

> The question where—I think I still remember the question—it was, like, *x*, wait, $ax^2 + bx + c$, where a, b, and c are constants, and then, What is $a + b$ [Math question 6]? *M35*

> The ones that are too wordy, especially when I am reading out loud—the ones that are too wordy and have a lot of, you know, a lot of, well, they're not complicated words, but they try to elude you, leading to fancy terms, that their fancy terms that make you think that it is actually one thing, but it is instead another, and it gets you confused about all that, makes you waste more time that way. *M51*

> The circle of radius 11 [Math question 3], because, one, the wording and, yeah, I was just, like, the content was hard for me. But maybe some of them were just lengthy, and there was a lot of space to mess up. *M57*

EL Status Impact

5. Did you encounter anything in the questions that you had difficulty with given your comfort level with the English language? If so, what was it, and why was it difficult for you?

When discussing, in response to postexperience question 5, how their comfort level with English affected their test-taking experience, most participants indicated that language didn't present significant barriers to their mathematical problem-solving. Some participants stated they had no difficulties with the English components of the questions; however, certain participants acknowledged challenges with text-heavy problems that required additional processing time or mental translation. A few mentioned occasional vocabulary difficulties or confusion with complex linguistic constructions. While most expressed confidence in their English comprehension, those who did encounter language barriers noted that these challenges added complexity to their problem-solving processes.

> I mean, sometimes the word problems can confuse me when I'm reading through them sometimes. Like, I just read it too fast. I have an issue with this—like, reading and talking too fast, and I can't comprehend it. So I have to reread it a bunch of times. So, I mean, I had an issue with that, but other than that, it was fun. *M9*

> Not really. I can say all the questions were clear. *M41*

> I think definitely the solutions, or when it asks for solutions and things like that, just because I'm not that used to vocab. You know, I think if you can tell, I kind of have an accent but not an accent at the same time, but also, like, being a non-native English speaker, there are some gaps. *M53*

> Well, it's, like, I'm not a, not like a, an English native speaker. Like, so when I read it, I have to, like, translate it in my mind. So it takes me time,

> and sometimes, well, like, you have to, like, be competitive in this test to, like, maximize all your time. So it was difficult for me to, like, translate in my mind and then going to solve the problem. *M55*
>
> Not really. I think they were all pretty, like, they're common terms, so not anything that, like, someone who doesn't know English is not gonna know. *M58*

Final Comments

6. Is there anything about your test-taking experience today or about the test-taking strategies you used today that we haven't talked about yet but that you'd like us to know?

When asked, via postexperience question 6, whether there was anything additional about their test-taking experience or strategies they wanted to share, one participant emphasized the importance of having confidence regardless of skill level, while others highlighted how using the built-in Desmos graphing calculator could significantly reduce the time spent on questions.

> I don't know, I feel like a good strategy, like I said before, it's just going into, like, questions being confident, no matter your skill level—just, you just got to be confident in yourself because confidence is kind of what drives you to success. *M9*
>
> So, as I said before, sometimes there are just obvious things, like formulas in there, that we already know that allow us to find a solution quicker instead of sitting and writing everything down, and especially with the Desmos [graphing calculator], it can be a great help in reducing the amount of time spent on the question. *M29*
>
> I just wanna say that, like, the SAT seems easier than [in] the previous year or like decades because we have Desmos now. So nothing is solved by hand. Like, way easier. I mean, easier means that I have a higher grade, like, I can get a higher score, so I'm happy, but it's kind of unfair for, like, you know, people who have already taken it. *M35*
>
> Yeah, but there are definitely some high-level questions in there that require a tremendous amount of critical thinking. But people coming out of Algebra II should be OK, you know, because the whole thing is that the SAT is, like, junior-year Algebra II and geometry. *M52*

Section 5: Discussion

Reading and Writing

PARTICIPANT PERFORMANCE

Participant performance levels on individual Reading and Writing test questions used in this study were determined by College Board subject matter experts, who compared transcripts of student verbalizations of their thinking aloud during their question answering to lists of required cognitive behaviors associated with a given question's type (e.g., Central Ideas and Details). Participants who both answered particular questions correctly and demonstrated all required behaviors were assigned the highest performance level (1), while participants who answered incorrectly, failed to demonstrate other required behaviors, or both were assigned lower performance levels. A participant differential (D_p) was then calculated for each participant. This differential was determined by subtracting from the total number of correctly answered questions the number of questions for which all required behaviors were demonstrated. This differential was considered "good" if it represented at least 70 percent of correctly answered questions being so answered while the participant demonstrated all required behaviors associated with the question's type.

Fifteen of twenty Reading and Writing participants (75 percent) met or exceeded the threshold for a good D_p, providing evidence that EL students are able to demonstrate cognitively complex thinking in line with the question types' constructs. The remaining participants had differentials ranging from 1 to 4. (For example, participant RW29 had a D_p of 1, but because he answered only two questions correctly and only one of those while demonstrating all required behaviors, his performance didn't meet the 70 percent threshold.) Even participants with a criterion-failing D_p, though, were still able to demonstrate cognitively complex thinking by demonstrating all required behaviors on 40 to 67 percent of the questions they answered correctly. In general, these results offer evidence that EL students are able to exhibit cognitively complex thinking in line with the question types' expectations.

QUESTION PERFORMANCE

A question differential (D_q) was similarly calculated for each of the fifteen Reading and Writing questions used in this study. This differential represents the arithmetic difference between the number of participants who answered a given question correctly and the number who also demonstrated all required behaviors associated with the question's type (i.e., attained PL 1). A "good" D_q for a particular question was set at 70 percent or more of all correctly answering participants also demonstrating all required behaviors via their verbalizations.

Thirteen of the study's fifteen Reading and Writing questions (87 percent) met or exceeded the threshold for a good D_q. One of the remaining two questions had a very high differential of 12, while the other had a low differential of 2. The former question—Reading and Writing question 8—was answered correctly by eighteen of twenty participants, while the latter question—Reading and Writing question 14—was answered correctly by only four participants, and six and two participants, respectively, were able to attain PL 1 on those questions.

Although eighteen of twenty participants (90 percent) answered Reading and Writing question 8, a medium-difficulty (PSB 5) Rhetorical Synthesis question set in a highly challenging (PSR) science context, correctly, only six of eighteen (33 percent) did so while demonstrating all required behaviors. The chief impediment was a failure to demonstrate behavior 1, which required showing adequate comprehension of the bulleted-list notes that test takers are supposed to synthesize relevant information from into a single statement that satisfies the question stem's criterion: Only the six participants who attained PL 1 (i.e., answered correctly and demonstrated all required behaviors) exhibited behavior 1, while the remaining twelve participants who answered the question correctly (thereby attaining PL 2 or 3) failed to exhibit this behavior. This suggests that many participants essentially "skipped" the notes and worked the question solely from the answer choices, a supposition supported by a few participants noting in their postexperience interviews that they avoided or made only limited use of the notes when answering. While that strategy may have worked with this particular Rhetorical Synthesis question, it should be noted that, largely in response to College Board having observed this behavior over time, some questions of this type include answer choices that misstate information in the bulleted list of notes, meaning that ignoring the notes comes at some peril to test takers. Interestingly, Reading and Writing question 7, also of the Rhetorical Synthesis type, had a criterion-passing D_q, with ten of fourteen correctly answering participants also demonstrating all required behaviors, including behavior 1.

Reading and Writing question 14, a hard (PSB 7) Command of Evidence: Quantitative question set in a highly challenging (PSR) science context, was simply hard for the participants in the study, with only four participants answering it correctly and only two of those also demonstrating all required behaviors. In their postexperience interview comments, participants both frequently cited as sources of challenge factors that question 14 has, including difficult vocabulary, a fairly complex informational graphic, and long, similarly worded answer choices, and mentioned question 14 specifically as a question they struggled with.

Nonetheless, six and two participants attained PL 1 on questions 8 and 14, respectively, suggesting that these questions, too, are capable of eliciting cognitively complex thinking from EL students at least some of the time. The overall findings support the claim that the presented Reading and Writing questions are capable of eliciting cognitively complex thinking in line with the question types' constructs from EL students.

PARTICIPANT PERFORMANCE VIGNETTES

Participant performance vignettes (transcript excerpts) exhibiting highly successful (PL 1) outcomes in line with question types' constructs were obtained for all fifteen Reading and Writing questions, providing further evidence that the questions are capable of eliciting cognitively complex thinking from EL students.

PARTICIPANT PERCEPTIONS

Reading and Writing participants gave generally positive assessments of their think-aloud experience (postexperience question 1). As test-taking strategies (postexperience question 2), they called out checking for "fit" between answer choices, question, and passage; rereading; using answer elimination; writing down their own answer first before looking at the provided answer choices; reading the question before reading the passage (something they were kept from doing by the think-aloud protocol); and skipping around to answer "shorter" questions before approaching longer questions that they believed took more time and effort.

In terms of question types or features of questions they associated with ease (postexperience question 3), participants mentioned finding Rhetorical Synthesis questions, questions in literature contexts, short questions, and questions in the fill-in-the-blank format least challenging. Rhetorical Synthesis questions were cited in part because of some participants' belief—only partially correct, as noted above—that it's unnecessary to read the bulleted-list notes to answer questions of this type correctly. Participants tended to consider hard (postexperience question 4) those questions that included what they felt was difficult vocabulary, that incorporated informational graphics, that focused on main ideas and conclusions, and that had longer and/or similarly worded answer choices. Reading and Writing question 14, which dealt with sugar maple growth under various climate change scenarios, ticked all those boxes and was often cited as particularly challenging.

When asked about how their status as English learners affected their performance in the test-taking activity (postexperience question 5), participants often mentioned challenges with English vocabulary as well as difficulties with questions whose passages are set in informational contexts, passages with complex sentence structures, and longer passages. Around a quarter of participants reported finding no particular challenges attributable to their EL status arising from the test questions. Most participants merely reiterated prior observations when asked for final thoughts (postexperience question 6).

The think-aloud methodology used in this study also shows some signs of reactivity. Recall that, per the discussion in Section 2: Literature Review, *reactivity* is the concept that study conditions themselves change the behavior they're meant to observe. We find some indications of this when participants, in their postexperience interviews, mention benefitting from (or being distracted by) reading aloud and call attention to the fact that they had to read the passage before the question. This level of reactivity is probably inevitable given the inherent artificiality of the think-aloud method, and, while worthy of note, a necessary tradeoff for the insights gained into otherwise fugitive cognitive processes.

Math

PARTICIPANT PERFORMANCE

Sixteen of eighteen Math participants (89 percent) met or exceeded the threshold for a good participant differential (D_p), thereby providing evidence that EL students are capable of demonstrating cognitively complex thinking in line with the question types' constructs. The two other participants each had a low differential of 1 but still failed to meet the criterion owing to the fact that they each answered only three questions correctly and only two of those while also demonstrating at least one expected behavior; they were nonetheless able to demonstrate cognitively complex thinking on two-thirds of the (small number of) questions they did answer correctly. In general, these results offer evidence that EL students are able to exhibit cognitively complex thinking in line with the question types' expectations.

QUESTION PERFORMANCE

All fifteen of the Math questions used in this study (100 percent) met or exceeded the criterion for a good question differential (D_q). This finding supports the claim that the presented Math questions are capable of eliciting cognitively complex thinking in line with the question types' constructs from EL students.

PARTICIPANT PERFORMANCE VIGNETTES

Participant performance vignettes (transcript excerpts) exhibiting highly successful (PL 1) outcomes in line with question types' constructs were obtained for all fifteen Math questions, providing further evidence that the questions are capable of eliciting cognitively complex thinking from EL students.

PARTICIPANT PERCEPTIONS

Math participants expressed a wide range of sentiments, from positive to negative, about the think-aloud activity (postexperience question 1), with some finding talking through their thought processes novel, interesting, or helpful and others finding the experience distracting or unnerving. They identified a range of strategies (postexperience question 2) that included reading carefully and breaking down questions, making (sometimes heavy) use of a graphing calculator, drawing on prior mathematical knowledge, visualizing problems and writing down important information, varying answering approaches based on question type, skipping difficult questions to return to later, and eliminating incorrect answer choices.

Participants tended to describe as easy (postexperience question 3) those questions that involved algebra, questions with minimal text, and questions concerning graph interpretation, and they tended to describe as hard (postexperience question 4) those questions that incorporated a great deal of text, questions involving multiple variables and/or unknown constants, function-related questions, and geometry questions in which a small amount of information is provided and for which visualization and/or multistep reasoning are required.

Most participants in the Math segment of the study indicated that their EL status (postexperience question 5) had no significant effect on their question-answering ability, but some participants did note extra challenges posed by questions with extensive text, challenging vocabulary, and/or complicated syntax. Final comments (postexperience question 6) tended to reiterate previously made points.

As in the Reading and Writing segment, some signs of reactivity to the think-aloud method are discernible from Math participants' comments about perceiving either benefit or detriment from being asked to solve problems while verbalizing their thought processes.

General Discussion

Results from this cognitive lab study involving EL students can be summarized and evaluated quantitatively and qualitatively.

QUANTITATIVE RESULTS SUMMARY: PARTICIPANT AND QUESTION DIFFERENTIALS

Table 5 summarizes the quantitative analyses performed as part of this study in terms of participant (D_p) and question differentials (D_q).

Table 5. Participant and Question Differentials, by Test Section.

	Differential Type	
Test Section	Participant (D_p)	Question (D_q)
Reading and Writing	15 of 20 (75%)	13 of 15 (87%)
Math	16 of 18 (89%)	15 of 15 (100%)

In terms of D_p, 75 percent of Reading and Writing participants (*n* = 20) and 89 percent of Math participants (*n* = 18) met or exceeded the threshold for "good" differentials, which were set at the level of participants demonstrating all required behaviors (Reading and Writing) or at least one expected behavior (Math) for at least 70 percent of the questions they answered correctly. In terms of D_q, 87 percent of the Reading and Writing (*n* = 15) and 100 percent of the Math questions (*n* = 15) met or exceeded the threshold for "good" differentials, which were set at the level of at least 70 percent of correctly answering participants also demonstrating all required behaviors (Reading and Writing) or at least one expected behavior (Math). We judge these to be good results, particularly in Math.

QUALITATIVE RESULTS SUMMARY: PARTICIPANT PERFORMANCE VIGNETTES AND PARTICIPANT PERCEPTIONS

From analysis of individual participant transcripts, we were able to obtain vignettes exhibiting PL 1—the study's highest—from all fifteen Reading and Writing and all fifteen Math questions. The fact that all Reading and Writing and Math questions analyzed for this study were able to elicit both correct answers and appropriate behaviors from EL students—and under the artificial condition of a think-aloud procedure—we regard as additional evidence that these questions are performing as intended in eliciting cognitively complex thinking, including from EL students.

Participants' perceptions of the study test questions and the think-aloud activity more broadly, as elicited by a standardized set of six postexperience interview questions, coalesced into a few themes that typically applied to both test sections, except as noted:

- The act of thinking aloud while answering test questions had an impact on participants, although opinions were divided regarding whether verbalizing helped them by sharpening their focus and concentration on the task at hand or distracted them from that task.
- Commonly identified test-taking strategies included close reading, answer elimination, and working first on questions they perceived as easier and/or less time consuming and saving harder and/or more time-intensive questions for later in the sequence.
- Participants found shorter questions and those with less text easier to work with; conversely, they found longer questions and those with more text, challenging vocabulary, and complex sentence structure harder to answer.
- Reading and Writing participants were more likely than their Math counterparts to identify challenges to testing rooted in their English language acquisition level, and fewer Reading and Writing than Math participants asserted that their EL status had no discernible impact on their question-answering performance.

STUDY LIMITATIONS

Several study limitations should be kept in mind when evaluating the results heretofore presented.

The first and most important is small sample size. While typical for cognitive lab/think-aloud studies such as this, small sample sizes (*n* = 20 for Reading and Writing; *n* = 18 for Math) limit the generalizability of findings and increase the risk that idiosyncratic variables impact results. We've attempted to ameliorate such concerns by including diverse (and well-documented) samples within the constraints of the study design, but this study shouldn't be taken as a definitive analysis of the performance of and challenges faced by EL students in large-scale assessment but rather as one set of data and conclusions complementing the work of many other researchers. As a corollary to the above, this study does include shortcomings with respect to full representation of the EL student population. Notably, members of some racial/ethnic groups are absent altogether, and higher-achieving students, as indicated by self-reported high school GPA (HSGPA), are overrepresented in the samples, but the latter may reflect both grade inflation (Sanchez 2024) and self-selection bias, as we'd expect relatively few academically low-achieving students to volunteer to participate in a study of their test-taking performance.

Second, as was discussed extensively throughout this report, the think-aloud methodology itself, though frequently employed for studies of cognition and generally well regarded, entails a (greater or lesser) degree of artificiality. That U.S. secondary students aren't routinely asked to think aloud to a stranger while they attempt to answer sometimes very challenging questions almost goes without saying. Moreover, while we sought to make the question-answering experience as authentic as possible (e.g., using actual practice test questions, minimizing probes and prompts), it was, fundamentally, an artificial experience under observation. As is intuitively obvious and as responses to the postexperience interview questions make clear, participants to greater or lesser extents altered their typical test-taking approach to accommodate the study format. Notably, the methodology compelled them to begin each question by reading the stimulus, whereas some, in a more naturalistic setting, may have preferred to begin by reading any multiple-choice options first, say, or by examining an included informational graphic. Ultimately, we deem this degree of artificiality as a necessary, inevitable compromise, an exchange of some degree of verisimilitude for the yielded insights into cognitive processes that would otherwise remain hidden. As we detailed in Section 2: Literature Review, the think-aloud methodology, within well-understood constraints and with appropriate safeguards, remains one of the best and only ways in which to peer into otherwise occluded cognitive processes in essentially real time and with minimal retrospective or inferential biases. At the same time, methodological concerns regarding veridicality, reactivity, and demand-induced bias (Kirk and Ashcraft 2001) can’t and shouldn’t be dismissed.

Finally, as we noted in Section 3: Methodology, technical constraints required that we use a preexisting SAT practice test form as the source for the questions we asked participants to respond to during the think-aloud activity. To minimize the risk that participants would have previously engaged with these questions in their own test preparation, we selected a practice test that was relatively new, in the linear format (whereas students are encouraged to practice in-platform with a digital adaptive practice test, the SAT Suite’s standard format, unless they expect to test on paper for accommodations or other reasons), and in the middle of the sequence of practice forms (based on the assumption that the typical student would start their preparation with either the lowest-numbered [oldest] or highest-numbered [newest] practice tests). This concern about prior exposure to the questions on the part of participants seems to have been theoretical rather than actual: no participant in either Reading and Writing or Math gave verbal evidence of having previous experience with any of the questions, and their performance profiles aren’t suggestive of such experience either.

Section 6: Conclusion

This report details the results of a verbal protocol study conducted by College Board, with support from vendor Vidlet, Inc., involving samples of high school juniors and seniors who are English learners (ELs) thinking aloud as they worked through sets of SAT Suite Reading and Writing and Math questions. The research goals of the study were, first, to ascertain, via qualitative and quantitative means, whether these EL students were able to demonstrate cognitively complex thinking in line with the question types' constructs and college and career readiness requirements and, second, to explore whether participants' performance on the questions or their postexperience reflections on the think-aloud activity would uncover any construct-irrelevant barriers to their success on such questions, and in particular barriers not already addressed by the provision of testing supports.

With regard to the first goal, the study's findings support the conclusion that EL students are capable of demonstrating cognitively complex thinking via their responses to SAT Suite Reading and Writing and Math test questions. With regard to the second goal, no clear indications of construct-irrelevant barriers residing in the test sections' designs or delivery method were identified.

References

Al-Maani, Alaa, Bara'ah AlAbabneh, Bassil Mashaqba, and Anas Huneety. 2024. "Investigating Second Language Learning Strategies Using Think Aloud Protocols: Evidence from Jordanian EFL Learners." *Eurasian Journal of Applied Linguistics* 10 (2): 12–22. **https://ejal.info/article-view/?id=724**.

Atman, Cynthia J., and Jennifer Turns. 2001. "Studying Engineering Design Learning: Four Verbal Protocol Studies." In *Design Knowing and Learning: Cognition in Design Education*, edited by Charles M. Eastman, W. Michael McCracken, and Wendy C. Newstetter. Elsevier.

Bainbridge, Lisanne, and Penelope Sanderson. 1995. "Verbal Protocol Analysis." In *Evaluation of Human Work: A Practical Ergonomics Methodology*, 2nd ed., edited by John R. Wilson and E. Nigel Corlett, 169–201. Taylor and Francis.

Bettman, James R., and C. Whan Park. 1980. "Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis." *Journal of Consumer Research* 7 (3): 234–48. **https://www.jstor.org/stable/2489009**.

Biggs, Stanley F., and Theodore J. Mock. 1983. "An Investigation of Auditor Decision Processes in the Evaluation of Internal Controls and Audit Scope Decisions." *Journal of Accounting Research* 21 (1): 234–55. **https://doi.org/10.2307/2490945**.

Bolton, Ruth N. 1993. "Pretesting Questionnaires: Content Analyses of Respondents' Concurrent Verbal Protocols." *Marketing Science* 12 (3): 280–303. **https://www.jstor.org/stable/184025**.

Botsas, George. 2017. "Differences in Strategy Use in the Reading Comprehension of Narrative and Science Texts Among Students with and Without Learning Disabilities." *Learning Disabilities: A Contemporary Journal* 15 (1): 139–62. **https://files.eric.ed.gov/fulltext/EJ1141985.pdf**.

Bowles, Melissa A., and Kacie Gastañaga. 2022. "Heritage, Second, and Third Language Learner Processing of Written Corrective Feedback: Evidence from Think-Alouds." *Studies in Second Language Learning and Teaching* 12 (4): 675–96. **https://doi.org/10.14746/ssllt.2022.12.4.7**.

Branch, Jennifer L. 2001. "Junior High Students and Think Alouds: Generating Information-Seeking Process Data Using Concurrent Verbal Protocols." *Library and Information Science Research* 23 (2): 107–22. **https://doi.org/10.1016/S0740-8188(01)00065-2**.

Branch, Jennifer L. 2013. "The Trouble with Think Alouds: Generating Data Using Concurrent Verbal Protocols." In *Proceedings of the Annual Conference of CAIS / Actes du Congrès Annuel de l'ACSI*. University of Alberta Library. **https://doi.org/10.29173/cais8**.

Cho, Byeong-Young, Lindsay Woodward, and Dan Li. 2018. "Epistemic Processing When Adolescents Read Online: A Verbal Protocol Analysis of More and Less Successful Online Readers." *Reading Research Quarterly* 53 (2): 197–221. **https://www.jstor.org/stable/26622508**.

College Board and HumRRO. 2020. *The Complex Thinking Required by Select SAT Items: Evidence from Student Cognitive Interviews.* College Board. **https://satsuite.collegeboard.org/media/pdf/sat-cognitive-lab-report.pdf**.

College Board. 2024a. *The Cognitively Complex Thinking Required by Select Digital SAT Suite Questions.* College Board. **https://satsuite.collegeboard.org/media/pdf/digital-sat-cognitive-lab-report.pdf**.

College Board. 2024b. *Assessment Framework for the Digital SAT Suite*, version 3.01 (August 2024). College Board. **https://satsuite.collegeboard.org/media/pdf/assessment-framework-for-digital-sat-suite.pdf**.

College Board. 2025a. *The Cognitively Complex Thinking Required by Select SAT Suite Questions: Evidence from Students with Specific Learning Disorders Affecting Reading (Dyslexia)*. College Board. **https://satsuite.collegeboard.org/media/pdf/digital-sat-cognitive-lab-report-sldr.pdf**.

College Board. 2025b. *The Cognitively Complex Thinking Required by Select SAT Suite Questions: Evidence from Students with Attention Deficit Hyperactivity Disorder (ADHD).* College Board. **https://satsuite.collegeboard.org/media/pdf/digital-sat-cognitive-lab-report-adhd.pdf**.

Council of Europe. n.d. "Global Scale—Table 1 (CEFR 3.3): Common Reference Levels." Common European Framework of Reference for Languages (CEFR). **https://www.coe.int/en/web/common-European-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale**.

Deshpande, Divya S., Paul J. Riccomini, Elizabeth M. Hughes, and Tracy J. Raulston. 2021. "Problem Solving with the Pythagorean Theorem: A Think Aloud Analysis of Secondary Students with Learning Disabilities." *Learning Disabilities: A Contemporary Journal* 19 (1): 23–47. **https://files.eric.ed.gov/fulltext/EJ1295343.pdf**.

Ericsson, K. Anders, and Herbert A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data*, rev. ed. MIT Press.

Goos, Merrilyn, and Peter Galbraith. 1996. "Do It This Way! Metacognitive Strategies in Collaborative Mathematical Problem Solving." *Educational Studies in Mathematics* 30 (3): 229–60. **https://www.jstor.org/stable/3482842**.

Haffer, Ann G. 1990. “Beginning Nurses’ Diagnostic Reasoning Behaviors Derived from Observation and Verbal Protocol Analysis.” EdD diss., University of San Francisco. ProQuest 9117892.

Isenberg, Daniel J. 1986. “Thinking and Managing: A Verbal Protocol Analysis of Managerial Problem Solving.” *Academy of Management Journal* 29 (4): 775–88. **https://www.jstor.org/stable/255944**.

Johnstone, Christopher J., Nicole A. Bottsford-Miller, and Sandra J. Thompson. 2006. *Using the Think Aloud Method (Cognitive Labs) to Evaluate Test Design for Students with Disabilities and English Language Learners.* Technical Report 44. University of Minnesota, National Center on Educational Outcomes. **https://files.eric.ed.gov/fulltext/ED495909.pdf**.

Johnstone, Christopher, Kristi Liu, Jason Altman, and Martha Thurlow. 2007. *Student Think Aloud Reflections on Comprehensible and Readable Assessment Items: Perspectives on What Does and Does Not Make an Item Readable.* Technical Report 48. University of Minnesota, National Center on Educational Outcomes. **https://files.eric.ed.gov/fulltext/ED499410.pdf**.

Kirk, Elizabeth P., and Mark H. Ashcraft. 2001. “Telling Stories: The Perils and Promise of Using Verbal Reports to Study Math Strategies.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27 (1): 157–75. **https://doi.org/10.1037/0278-7393.27.1.157**.

Kletzien, Sharon Benge. 1991. “Strategy Use by Good and Poor Comprehenders Reading Expository Text of Differing Levels.” *Reading Research Quarterly* 26 (1): 67–86. **http://www.jstor.com/stable/747732**.

Leow, Ronald P., and Kara Morgan-Short. 2004. “To Think Aloud or Not to Think Aloud: The Issue of Reactivity in SLA Research Methodology.” *Studies in Second Language Acquisition* 26 (1): 35–57. **https://psycnet.apa.org/record/2004-11297-002**.

Lundberg, Gustav. 1984. “Protocol Analysis and Spatial Behavior.” *Geografiska Annaler, Series B, Human Geography* 66 (2): 91–97. **https://doi.org/10.2307/490719**.

Magliano, Joseph P., and Keith K. Millis. 2003. “Assessing Reading Skill with a Think-Aloud Procedure and Latent Semantic Analysis.” *Cognition and Instruction* 21 (3): 251–83. **https://www.jstor.org/stable/3233811**.

Montague, Marjorie, and Brooks Applegate. 1993. “Middle School Students’ Mathematical Problem Solving: An Analysis of Think-Aloud Protocols.” *Learning Disability Quarterly* 16 (1): 19–32. **https://doi.org/10.2307/1511157**.

Nguyen, Lemai, and Graeme Shanks. 2007. “Using Protocol Analysis to Explore the Creative Requirements Engineering Process.” In *Information Systems Foundations: Theory, Representation, and Reality*, edited by Dennis N. Hart and Shirley D. Gregor. Australian National University Press.

Nisbett, Richard E., and Timothy DeCamp Wilson. 1977. “Telling More Than We Can Know: Verbal Reports on Mental Processes.” *Psychological Review* 84 (3): 231–59. **https://doi.org/10.1037/0033-295X.84.3.231**.

Özcan, Zeynep Çiğdem, Yeşim Imamoğlu, and Vildan Katmer Bayraklı. 2017. "Analysis of Sixth Grade Students' Think-Aloud Processes While Solving a Non-Routine Mathematical Problem." *Kuram Ve Uygulamada Eğitim Bilimleri [Journal of Educational Sciences: Theory and Practice]* 17 (1): 129–44. **https://jestp.com/menuscript/index.php/estp/article/view/492/444**.

Özkubat, Ufuk, and Emine Rüya Özmen. 2021. "Investigation of Effects of Cognitive Strategies and Metacognitive Functions on Mathematical Problem-Solving Performance of Students with or Without Learning Disabilities." *International Electronic Journal of Elementary Education* 13 (4): 443–56. **http://dx.doi.org/10.26822/iejee.2021.203**.

Pressley, Michael, and Peter Afflerbach. 1995. *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading.* Erlbaum.

Russo, J. Edward, Eric J. Johnson, and Debra L. Stephens. 1989. "The Validity of Verbal Protocols." *Memory and Cognition* 17 (6): 759–69. **https://doi.org/10.3758/BF03202637**.

Sanchez, Edgar I. 2024. *Changes in Predictive Validity of High School Grade Point Average and ACT Composite Score After the COVID-19 Pandemic.* ACT, Inc. **https://www.act.org/content/dam/act/secured/documents/R2328-Changes-in-Predictive-Validity-of-HSGPA-and-ACT-Composite-Score-After-COVID-19-2024-09.pdf**.

Sanchez, Edgar, and Richard Buddin. 2016. *How Accurate Are Self-Reported High School Courses, Course Grades, and Grade Point Average?* ACT, Inc. **https://www.act.org/content/dam/act/unsecured/documents/5269-research-report-how-accurate-are-self-reported-hs-courses.pdf**.

Stratman, James F., and Liz Hamp-Lyons. 1994. "Reactivity in Concurrent Think-Aloud Protocols: Issues for Research." In *Speaking About Writing: Reflections on Research Methodology*, edited by Peter Smagorinsky. Sage.

Suto, W. M. Irenka, and Jackie Greatorex. 2008. "What Goes Through an Examiner's Mind? Using Verbal Protocols to Gain Insights into the GCSE Marking Process." *British Educational Research Journal* 34 (2): 213–33. **https://www.jstor.org/stable/30032828**.

Taylor, K. Lynn, and Jean-Paul Dionne. 2000. "Accessing Problem-Solving Strategy Knowledge: The Complementary Use of Concurrent Verbal Protocols and Retrospective Debriefing." *Journal of Educational Psychology* 92 (3): 413–25. **https://doi.org/10.1037/0022-0663.92.3.413**.

Vessey, Iris. 1986. "Expertise in Debugging Computer Programs: An Analysis of the Content of Verbal Protocols." *IEEE Transactions on Systems, Man, and Cybernetics* 16 (5): 621–37. **https://doi.org/10.1109/TSMC.1986.289308**.

Yayli, Demet. 2010. "A Think-Aloud Study: Cognitive and Metacognitive Reading Strategies of ELT Department Students." *Eurasian Journal of Educational Research* 38 (Winter 2010): 234–51. **https://www.researchgate.net/publication/286547114_A_Think-Aloud_Study_Cognitive_and_Metacognitive_Reading_Strategies_of_ELT_Department_Students**

Appendix

Exhibit 1: Recruitment Solicitation

College Board is seeking a number of high school juniors and seniors to participate in an upcoming research study. Participants will meet one-on-one virtually (via Zoom) with a moderator, who will walk them through an activity and ask follow-up questions. The activity involves reading, thinking aloud through, and answering a series of digital SAT questions in either Reading and Writing or Math and answering some follow-up interview questions. Our goal is to better understand how students interact with our test questions. This activity will take approximately 90 minutes for each student to complete; on successful completion, participants will receive a $150 gift card.

To be eligible to participate, students must

- be either high school juniors or seniors;
- have previously taken the SAT, PSAT/NMSQT, or PSAT 10 tests from College Board in either paper and pencil or digital format;
- have uninterrupted access to an appropriate digital device (desktop computer, laptop computer, tablet; *not* a phone) with a camera; a private space in which to participate virtually in the activity; and an uninterrupted internet connection robust enough for stable videoconferencing;
- commit to spending approximately 90 minutes in working through test questions and answering follow-up interview questions from the moderator to the best of their ability, on a day and at a time mutually agreeable to the moderator and participant; and
- be willing and able to share as much of their thought processes as possible with the moderator while answering test and interview questions.

Participants from all school achievement levels are encouraged to apply. Participants will **not** be evaluated on whether they answer the study's test questions correctly, and participation in this activity will **not** generate a test score, nor will it affect any prior SAT, PSAT/NMSQT, and/or PSAT 10 scores participants may have.

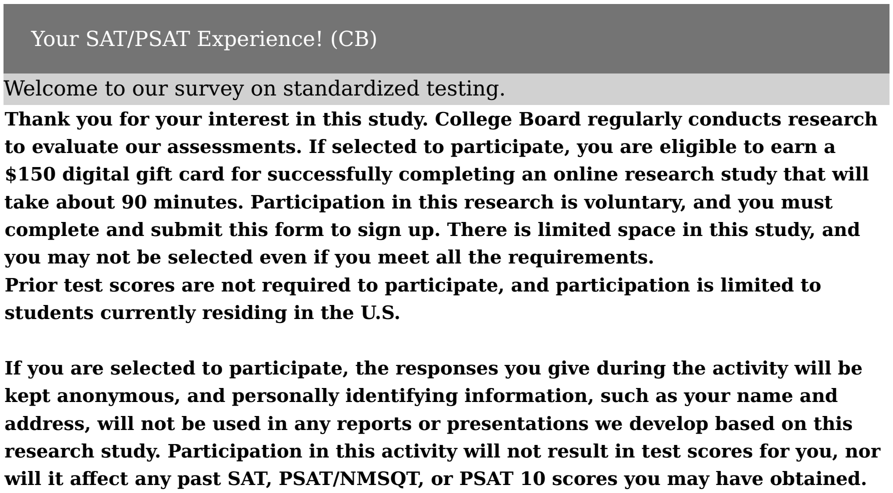
College Board will assign participants to either a reading and writing or a math activity. Participants selected for the math activity should also have access to scratch paper and pencils/pens for use in answering test questions; in addition, they should either be comfortable with the Desmos graphing calculator, which is available as part of the activity, or have their own approved calculator available. For information on acceptable handheld calculators, please visit **https://satsuite.collegeboard.org/sat/what-to-bring-do/calculator-policy**.

This study is for research purposes. Participants' names and other personally identifying information will **not** be used in reports and presentations College Board produces. Sessions will be recorded.

Students (or a parent/guardian, if the student is under 18 years of age) must complete a consent form to participate. This consent form describes the study and its purposes as well as how participants' data will be collected, used, and kept anonymous.

On successful completion of the activity, each participant will receive a $150 gift card, which can be deposited in a bank, deposited into PayPal, or redeemed at one of numerous businesses selected by the participant from a list provided by College Board. Participants may opt out of answering any question or participating in the activity at any time, but successful completion is required to receive the gift card.

Exhibit 2: Recruitment Screener (Survey)

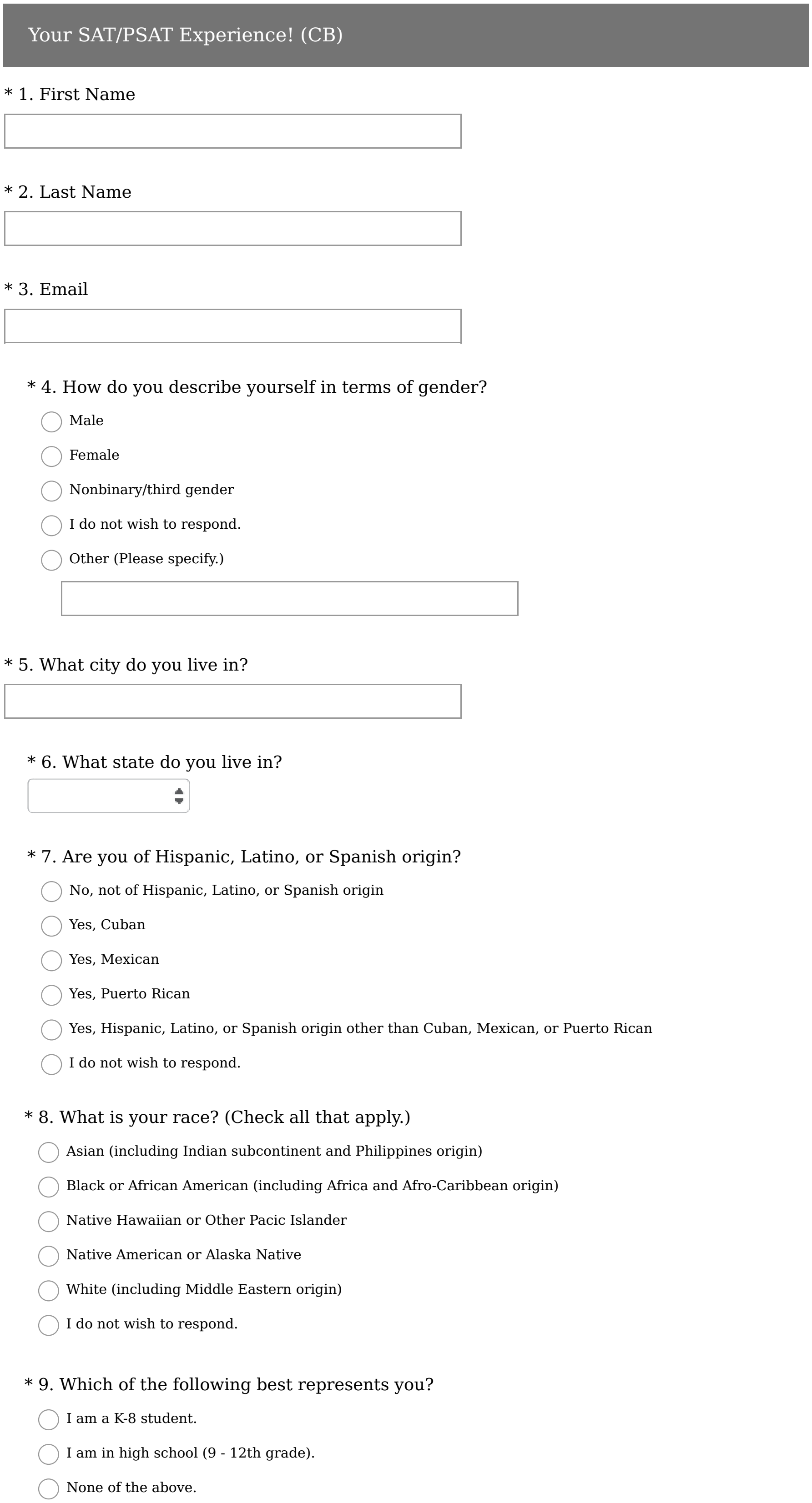


Your SAT/PSAT Experience! (CB)

Welcome to our survey on standardized testing.

Thank you for your interest in this study. College Board regularly conducts research to evaluate our assessments. If selected to participate, you are eligible to earn a $150 digital gift card for successfully completing an online research study that will take about 90 minutes. Participation in this research is voluntary, and you must complete and submit this form to sign up. There is limited space in this study, and you may not be selected even if you meet all the requirements.
Prior test scores are not required to participate, and participation is limited to students currently residing in the U.S.

If you are selected to participate, the responses you give during the activity will be kept anonymous, and personally identifying information, such as your name and address, will not be used in any reports or presentations we develop based on this research study. Participation in this activity will not result in test scores for you, nor will it affect any past SAT, PSAT/NMSQT, or PSAT 10 scores you may have obtained.



Your SAT/PSAT Experience! (CB)

* 1. First Name

* 2. Last Name

* 3. Email

* 4. How do you describe yourself in terms of gender?

○ Male

○ Female

○ Nonbinary/third gender

○ I do not wish to respond.

○ Other (Please specify.)

* 5. What city do you live in?

* 6. What state do you live in?

* 7. Are you of Hispanic, Latino, or Spanish origin?

○ No, not of Hispanic, Latino, or Spanish origin

○ Yes, Cuban

○ Yes, Mexican

○ Yes, Puerto Rican

○ Yes, Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican

○ I do not wish to respond.

* 8. What is your race? (Check all that apply.)

○ Asian (including Indian subcontinent and Philippines origin)

○ Black or African American (including Africa and Afro-Caribbean origin)

○ Native Hawaiian or Other Pacific Islander

○ Native American or Alaska Native

○ White (including Middle Eastern origin)

○ I do not wish to respond.

* 9. Which of the following best represents you?

○ I am a K-8 student.

○ I am in high school (9 - 12th grade).

○ None of the above.

Education

* 10. What grade are you in? **Please select the grade level you will be in for the upcoming 2024/2025 school year.**

○ 9

○ 10

○ 11

○ 12

* 11. What is the name of your current school?

* 12. Select your high school grade point average (HGPA).

○ A+ (97–100)

○ A (93–96)

○ A- (90–92)

○ B+ (87–89)

○ B (83–86)

○ B- (80–82)

○ C+ (77–79)

○ C (73–76)

○ C- (70–72)

○ D+ (67–69)

○ D (65–66)

○ E/F (Below 65)

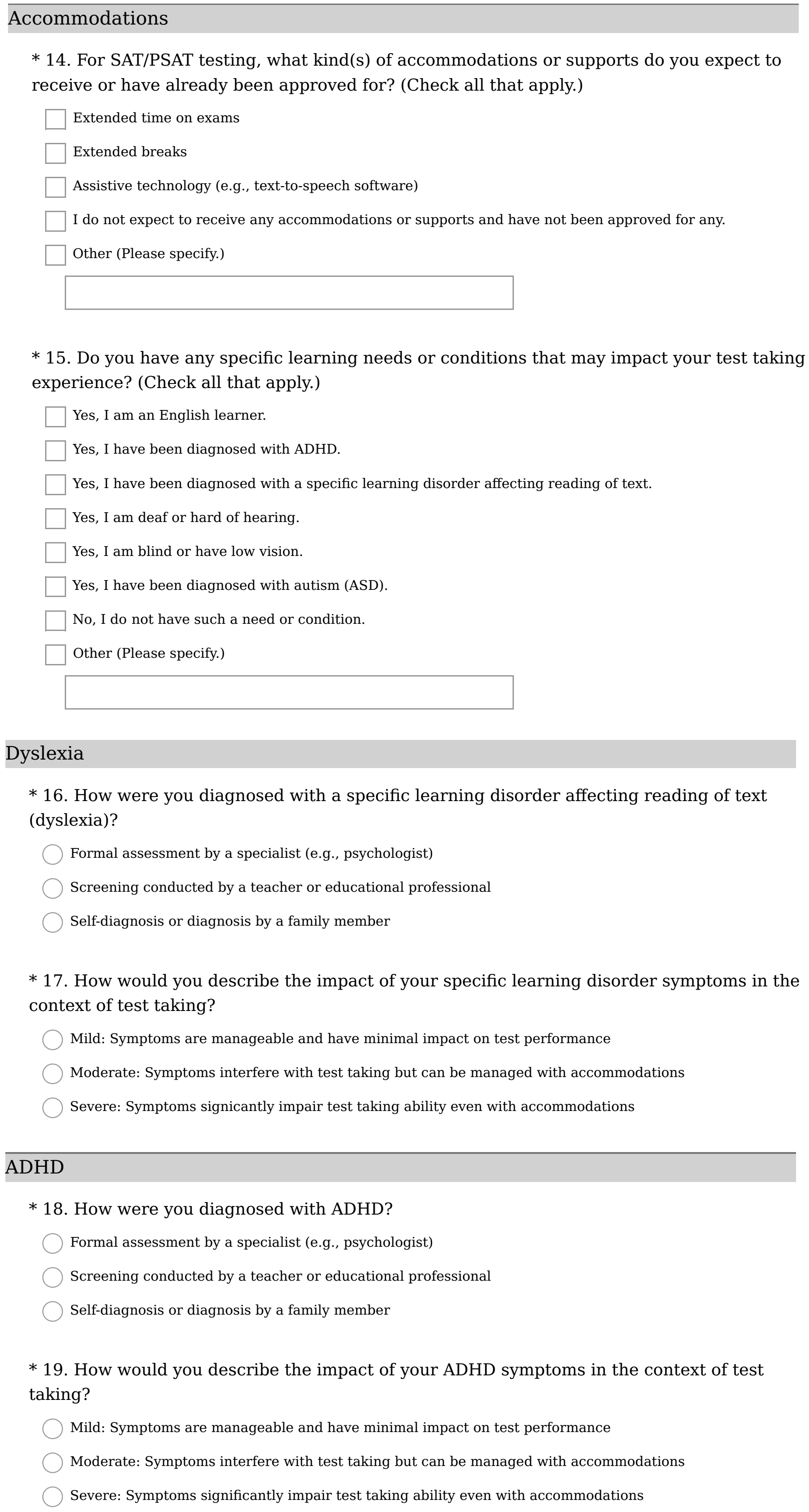
○ I do not wish to respond.

* 13. Do you expect to receive or have you previously been approved for accommodations or supports for SAT/PSAT testing?

Examples of accommodations or supports can include =
- Extended time
- Extended breaks
- Assistive technology

○ Yes

○ No



Accommodations

* 14. For SAT/PSAT testing, what kind(s) of accommodations or supports do you expect to receive or have already been approved for? (Check all that apply.)

□ Extended time on exams

□ Extended breaks

□ Assistive technology (e.g., text-to-speech software)

□ I do not expect to receive any accommodations or supports and have not been approved for any.

□ Other (Please specify.)

* 15. Do you have any specific learning needs or conditions that may impact your test taking experience? (Check all that apply.)

□ Yes, I am an English learner.

□ Yes, I have been diagnosed with ADHD.

□ Yes, I have been diagnosed with a specific learning disorder affecting reading of text.

□ Yes, I am deaf or hard of hearing.

□ Yes, I am blind or have low vision.

□ Yes, I have been diagnosed with autism (ASD).

□ No, I do not have such a need or condition.

□ Other (Please specify.)

Dyslexia

* 16. How were you diagnosed with a specific learning disorder affecting reading of text (dyslexia)?

○ Formal assessment by a specialist (e.g., psychologist)

○ Screening conducted by a teacher or educational professional

○ Self-diagnosis or diagnosis by a family member

* 17. How would you describe the impact of your specific learning disorder symptoms in the context of test taking?

○ Mild: Symptoms are manageable and have minimal impact on test performance

○ Moderate: Symptoms interfere with test taking but can be managed with accommodations

○ Severe: Symptoms signicantly impair test taking ability even with accommodations

ADHD

* 18. How were you diagnosed with ADHD?

○ Formal assessment by a specialist (e.g., psychologist)

○ Screening conducted by a teacher or educational professional

○ Self-diagnosis or diagnosis by a family member

* 19. How would you describe the impact of your ADHD symptoms in the context of test taking?

○ Mild: Symptoms are manageable and have minimal impact on test performance

○ Moderate: Symptoms interfere with test taking but can be managed with accommodations

○ Severe: Symptoms significantly impair test taking ability even with accommodations

Language

* 20. How often do you communicate in English in your daily life?

○ Often

○ Sometimes

○ Rarely

* 21. In which language(s) do you typically speak at home?

○ Only in English

○ Only in a language other than English

○ In English and one or more other languages

* 22. Which language(s) other than English do you know well? (Check all that apply.)

□ Arabic

□ Mandarin/Cantonese

□ Spanish

□ Vietnamese

□ None

□ Other (Please specify.)

* 23. Which of the following best describes your current level of English language acquisition?

○ I can understand familiar everyday expressions and very basic phrases in English.

○ I can understand sentences and frequently used expressions in English.

○ I can understand the main points of clear texts on familiar subjects in English.

○ I can understand the main ideas of complex texts in English.

○ I can understand a wide range of demanding, longer texts in English.

○ I can easily understand nearly any text in English.

* 24. Participants who are English learners may ask a family member or friend to act as a translator for all or part of the activity. Arranging for such a translator is optional and solely the responsibility of the participant.

Would you plan to use a translator during the interview session?

○ Yes, I would plan to use a translator.

○ No, I would not plan to use a translator.

Your Standardized Testing Experience

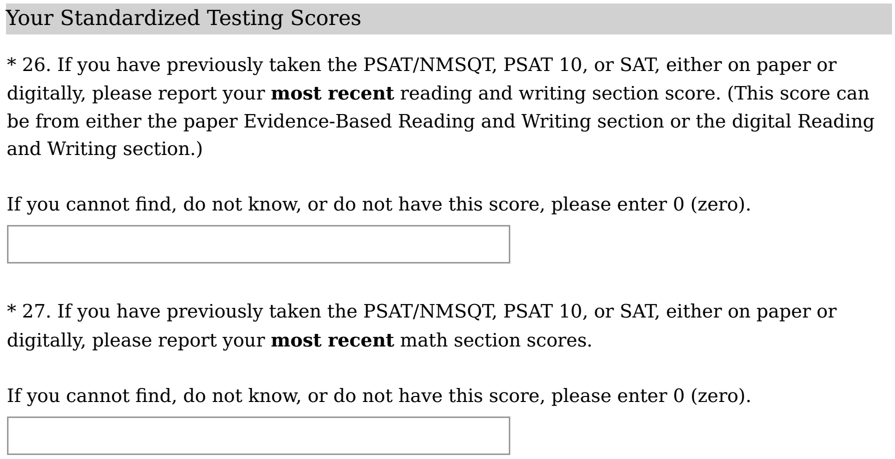
* 25. Which of the following College Board tests, if any, have you taken most recently?

Prior PSAT/NMSQT, PSAT 10, or SAT scores are **NOT** required for eligibility to participate in this study.

○ SAT

○ PSAT/NMSQT or PSAT 10

○ I have not taken any of these tests.



Your Standardized Testing Scores

* 26. If you have previously taken the PSAT/NMSQT, PSAT 10, or SAT, either on paper or digitally, please report your **most recent** reading and writing section score. (This score can be from either the paper Evidence-Based Reading and Writing section or the digital Reading and Writing section.)

If you cannot find, do not know, or do not have this score, please enter 0 (zero).

* 27. If you have previously taken the PSAT/NMSQT, PSAT 10, or SAT, either on paper or digitally, please report your **most recent** math section scores.

If you cannot find, do not know, or do not have this score, please enter 0 (zero).

Exhibit 3: Consent Form

CollegeBoard

Student Research Group Agreement

By signing this agreement, the student identified below ("**Student**"), with consent of their parent/guardian ("**Parent/Guardian**") if the student is under eighteen years of age, agrees to Student's participation in SAT Question Interviews, a research study for College Board ("**Study**"). The Study involves the Student providing feedback to College Board on SAT questions, including but not limited to, providing feedback via a screen-sharing session with a College Board researcher where students may be asked questions or provide feedback about how they answer SAT questions. The study will be conducted entirely online. The activity will take no more than an hour and a half, and on successful completion of the activity, payment will be made via digital payment platform, Tremendous. Student will receive a link from Tremendous to the email address provided which can be used to redeem payment in the form of a bank transfer, PayPal deposit, or a gift card of choice—Tremendous has over 300 gift card options.

Student and Parent/Guardian hereby give their full and complete permission to College Board and its agents to photograph, record (audio and video) Student's participation ("**Images**"). Student and Parent/Guardian grant College Board and its designees, affiliates, agents, subcontractors, and licensees (collectively, "**College Board**") the right to use, transcribe, edit, reproduce, broadcast, publish, exhibit, publicize, and otherwise distribute, without compensation to Student and Parent/Guardian, any Images, along with Student responses, statements and comments Student makes during or in connection with the Study (together with the Images, "**Information**"). The rights hereby granted to College Board are perpetual and worldwide.

Any Images will be stored securely consistent with College Board policies and only College Board personnel involved in the Study and related research and product development will access the recordings. Images will be kept for one year and then securely destroyed. Transcriptions will be kept for two years and then securely destroyed.

Student and Parent/Guardian acknowledge that College Board will rely on this permission and that College Board, in its sole discretion, may decide whether or not to use the Information. Student and Parent/Guardian will not assert a claim that the use of the Information is a violation of Student rights. Student and Parent/Guardian further understand and agree that they hereby waive all rights and claims to ownership of the College Board materials in which the Information may appear.

As the session will include use of live video during the screen-sharing session, please be mindful of your background including, for example, avoid having other individuals in the room, secure any personal items and information from view of the camera and other similar safeguards the Student and Parent/Guardian may wish to consider in their discretion, understanding and acknowledging that the researcher will be able to view the Student's background through the Student's camera.

In addition, Student and Parent/Guardian acknowledge that any information and materials that is disclosed or otherwise made available to Student and Parent/Guardian in connection with the Study ("**Confidential Information**") is highly confidential and proprietary to College Board and agree (i) to keep it strictly confidential, (ii) not to disclose to or discuss with any third party, and (iii) not to use for any purpose other than to participate in the Study.

Student and Parent/Guardian understand that College Board is offering to pay Student based on the research activity a US $150 gift card, provided that such payment is permissible under applicable laws and regulations, and the policies and regulations of my employer, if any. Student and Parent/Guardian acknowledge and agree that College Board is not, and that Student and Parent/Guardian is responsible for determining whether Student and/or Parent/Guardian institution's policies and regulations or applicable laws and regulations preclude the Student from participating in the Study or receiving such payment. Student and Parent/Guardian will not consider this agreement an offer to provide this payment if Student and/or Parent/Guardian is prohibited from accepting such payment.

This Student Research Group Agreement is the full and complete understanding between College Board, Student, and Parent/Guardian. Student and Parent/Guardian each represent they have had adequate time to read this document carefully and to ask any questions that they may have.

Please Print:

Name of Participant Signature Date

Name of Parent/Guardian Signature Date

Student Street Address, City, State

Student Email address

Exhibit 4: Interview Session Training Questions

Note: The following questions were used for participant training purposes prior to the formal start of the think-aloud activity. Session moderators demonstrated thinking aloud for one question using the script included below, after which they gave participants one or (at the moderators' discretion) two questions on which to practice thinking aloud. The training portion of sessions was neither recorded nor analyzed.

READING AND WRITING

Moderator Demonstration Question and Script

> The Younger Dryas was a period of extreme cooling from 11,700 to 12,900 years ago in the Northern Hemisphere. Some scientists argue that a comet fragment hitting Earth brought about the cooling. Others disagree, partly because there is no known crater from such an impact that dates to the beginning of the period. In 2015, a team led by Kurt Kjær detected a 19-mile-wide crater beneath a glacier in Greenland. The scientists who believe an impact caused the Younger Dryas claim that this discovery supports their view. However, Kjær's team hasn't yet been able to determine the age of the crater. Therefore, the team suggests that ________
>
> Which choice most logically completes the text?
>
> A) it can't be concluded that the impact that made the crater was connected to the beginning of the Younger Dryas.
>
> B) it can't be determined whether a comet fragment could make a crater as large as 19 miles wide.
>
> C) scientists have ignored the possibility that something other than a comet fragment could have made the crater.
>
> D) the scientists who believe an impact caused the Younger Dryas have made incorrect assumptions about when the period began.

Reading this passage and question, it looks like I'm being asked to figure out how best to fill in the blank with something that makes the most sense in context.

I'm now looking at the answer choices and trying to figure out which is the best answer here. I'm looking for something that logically completes the text.

Choice A says, "It can't be concluded that the impact that made the crater was connected to the beginning of the Younger Dryas." That makes sense because the passage says that the team "hasn't yet been able to determine the age of the crater," so there's still some doubt about whether this crater is even what the team suspects it is. The word "however" also makes me think that Kjær is trying to keep other scientists from jumping to conclusions.

So I like choice A, but I want to look at the other choices before making my decision.

Choice B, “It can’t be determined whether a comet fragment could make a crater as large as 19 miles wide.” This doesn’t make as much sense to me, because the passage doesn’t say anything that would suggest there’s any doubt about whether the crater was made by a comet fragment, only about how old the crater is.

Choice C, “Scientists have ignored the possibility that something other than a comet fragment could have made the crater.” This one seems wrong for the same basic reason choice B was: the passage doesn’t suggest that there’s real doubt about whether the crater was made by a comet fragment.

And choice D, “The scientists who believe an impact caused the Younger Dryas have made incorrect assumptions about when the period began.” No, it’s not this either. The passage doesn’t tell us there’s any real debate about when the Younger Dryas began. There’s a date range, but it’s just presented as a fact. And the passage doesn’t suggest that scientists have made mistakes about dating the period itself. Kjær just seems to want other scientists not to assume that the crater they found is old enough to support some scientists’ hypothesis about how the Younger Dryas started.

So I’ll select answer choice A.

Notice how when I was thinking aloud, I didn’t try to simply summarize what I did after I was done answering. Instead, as I approached this question, I told you exactly what I was thinking as I thought it. I first read the passage and the question aloud and then explained what I thought the question was asking, how I went about answering the question, and why I came up with the answer that I did. I want you to do the same sort of thing when you read and answer test questions today.

Any questions or concerns?

Participant Practice Questions

“The Bet” is an 1889 short story by Anton Chekhov. In the story, a banker is described as being very upset about something: ________

Which quotation from “The Bet” most effectively illustrates the claim?

A) “Then the banker cautiously broke the seals off the door and put the key in the keyhole.”

B) “It struck three o’clock, the banker listened; everyone was asleep in the house and nothing could be heard outside but the rustling of the chilled trees.”

C) “The banker, spoilt and frivolous, with millions beyond his reckoning, was delighted at the bet.”

D) “When [the banker] got home he lay on his bed, but his tears and emotion kept him for hours from sleeping.”

Celebrated Tewa potter Maria Martinez (1887–1980) made her signature all-black ceramic vessels using a heating technique called reduction firing. This technique involves smothering the flame surrounding the clay vessel. ________ the vessel takes on a shiny, black hue.

Which choice completes the text with the most logical transition?

A) On the contrary,

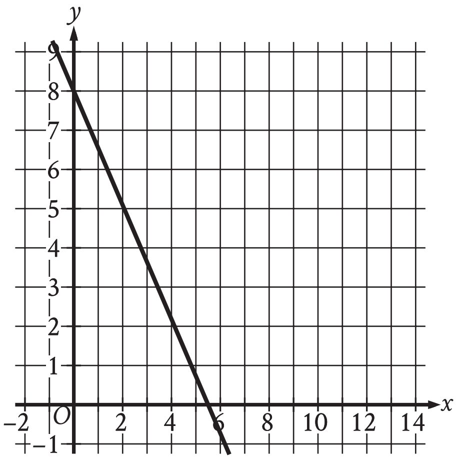
B) For example,

C) Previously,

D) As a result,

MATH

Moderator Demonstration Question and Script



The graph of the linear function f is shown, where $y = f(x)$. What is the y-intercept of the graph of f?

A) (0, 0)

B) $\left(0, -\frac{16}{11}\right)$

C) $(0, -8)$

D) (0, 8)

This is a question where I need to understand what a y-intercept of a graph is. A y-intercept of a graph is a point where the graph crosses the y-axis. I'm told this is a linear function, so I know there is only one y-intercept. From the graph, it appears the line crosses the y-axis at the point (0, 8). Since this is a multiple-choice question, choice D is probably my answer.

Let me check the other choices, though. Choice A, (0, 0), isn't right. (0, 0) is the point where the x-axis intercepts the y-axis. I'm not sure where choices B or C even come from, as (0, negative 16 over 11) and (0, negative 8) don't make any sense here, given the graph we're presented with. So I'm going with my first answer, choice D.

Notice how when I was thinking aloud, I didn't try to simply summarize what I did after I was done answering. Instead, as I approached this question, I told you exactly what I was thinking as I thought it. I first read the passage and the question aloud and then explained what I thought the question was asking, how I went about answering the question, and why I came up with the answer that I did. I want you to do the same sort of thing when you read and answer test questions today.

Any questions or concerns?

Participant Practice Questions

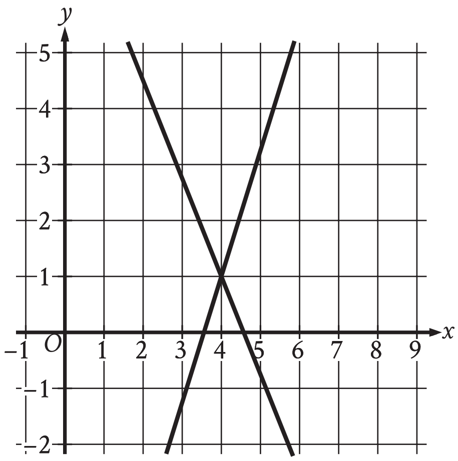
If $4x - 28 = -24$, what is the value of $x - 7$?

A) −24

B) −22

C) −6

D) −1



The graph of a system of linear equations is shown. The solution to the system is (x, y). What is the value of x ?