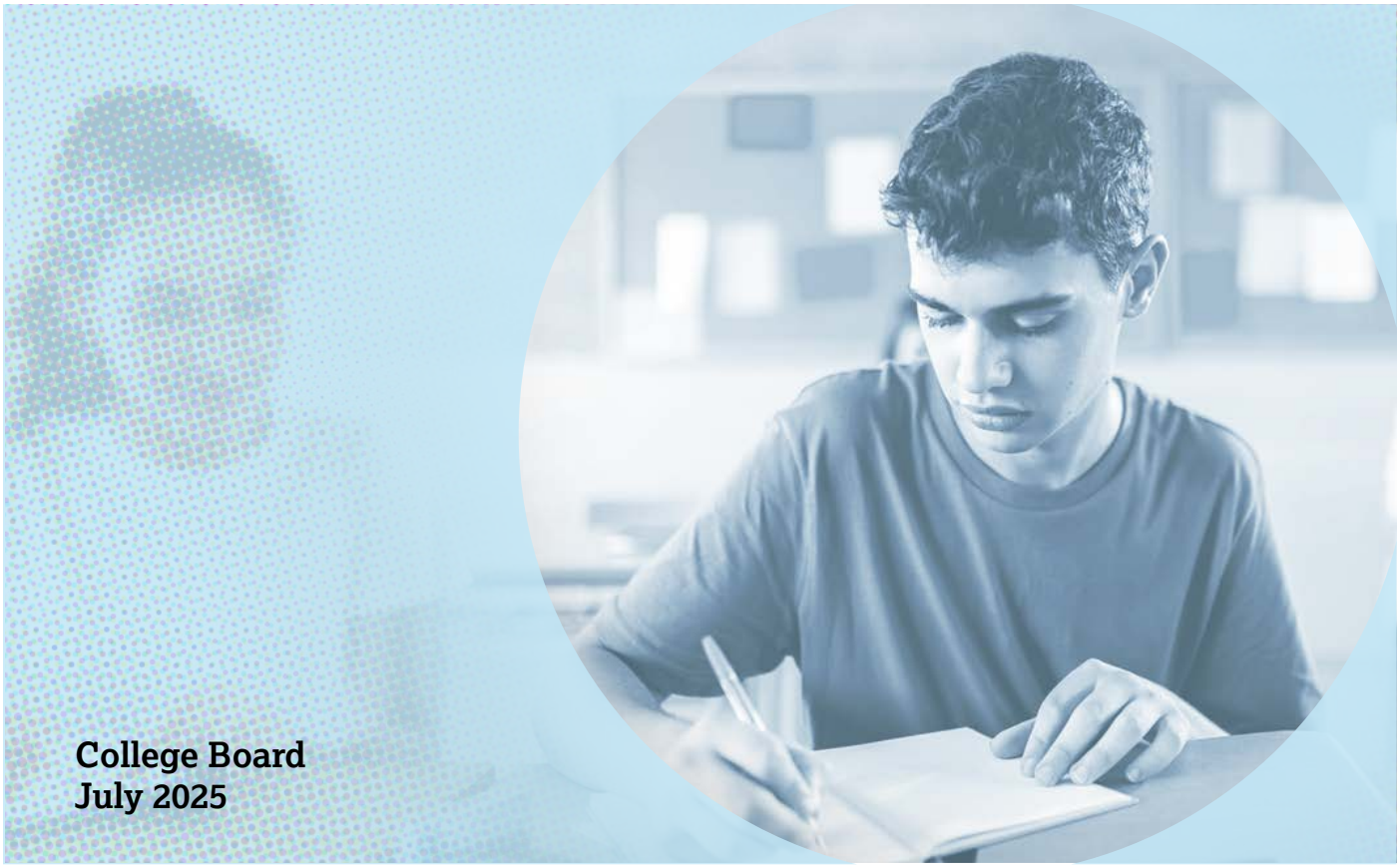
CollegeBoard

The Cognitively Complex Thinking Required by Select SAT® Suite Questions

Evidence from Students with Attention Deficit Hyperactivity Disorder (ADHD)



College Board
July 2025

The Cognitively Complex Thinking Required by Select SAT Suite Questions:

Evidence from Students with Attention Deficit Hyperactivity Disorder (ADHD)

July 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 Students with Attention Deficit Hyperactivity Disorder (ADHD).* 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**.

Contents

Tables

Figures

Executive Summary

This report documents the findings of a think-aloud (cognitive lab) study conducted with students with attention deficit hyperactivity disorder (ADHD) 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 students with ADHD 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-four high school juniors and seniors who indicated having ADHD and met other criteria were selected to participate in the Reading and Writing segment of the study, while an additional twenty-one 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, self-reported high school GPA (HSGPA), and self-reported ADHD impact on their test-taking ability.

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/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 when available 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 when available 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.** Nineteen of twenty-four Reading and Writing participants (79 percent) and thirteen of twenty-one Math participants (62 percent) met or exceeded the threshold for a good D_p, providing evidence that students with ADHD 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 questions and for fourteen of fifteen Math questions, providing additional support for the claim that students with ADHD can demonstrate cognitively complex thinking via SAT Suite test questions.
- **Question performance.** Twelve of fifteen Reading and Writing questions (80 percent) and twelve of fifteen Math questions (80 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 students with ADHD.
- **Participant perceptions.** No clear evidence of construct-irrelevant barriers not already addressed by the provision of testing accommodations 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.

The study's positive outcomes respecting students with ADHD must be contextualized with the understanding that the results assume these students have access to appropriate accommodations during testing, including extra time/extra breaks and possibly assistive technology, such as text-to-speech.

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 having attention deficit hyperactivity disorder (ADHD) 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 with ADHD support the conclusion that select SAT Suite test questions are capable of eliciting cognitively complex thinking from students with ADHD 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 the ability of students with ADHD 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 accommodations, 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 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 whenever possible 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 the small number of 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. In addition, participants in either the Reading and Writing or Math segment could use a third-party screen reader. (At the time the cognitive interviews were conducted [2024], Bluebook didn't have a native text-to-speech option as an alternative to the use of screen readers; this feature was added in 2025.) Participants in the Math segment also had access to a set of common formulas and could make use of scratch paper.

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), and students who were English learners (College Board 2025b). The present study reports the results of the study involving students who have ADHD.

The *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, text revision (DSM-5-TR) (American Psychiatric Association 2022, 69) observes that people diagnosed with ADHD have issues with "inattention and/or hyperactivity/impulsivity" that can manifest in varying ways. Inattention can be exhibited via lack of careful attention to details and the frequent making of seemingly careless mistakes; struggles with remaining attentive, maintaining focus, staying organized, or keeping track of materials during work or leisure activities; distractibility; avoidance of sustained mental work; and forgetfulness. Hyperactivity/impulsivity can be exhibited via fidgeting, inappropriate movement (e.g., during school), an inability to work or play quietly, excessive talk, difficulties with turn taking, and verbal and physical interruptions of others. While these symptoms don't directly prevent students from demonstrating on the SAT Suite the literacy and numeracy skills and knowledge they've acquired, as some conditions may, they can, without appropriate accommodations, depress performance. Students prone to being inattentive to details, for example, may miss crucial qualifiers in test questions; easily distracted students may not give their full attention and effort to each question; impulsive students may struggle with testing for a sustained period under timed conditions; and so on.

As part of the sample selection screener (see appendix), prospective participants were asked to indicate whether they had ADHD (and/or had a specific learning disorder affecting reading or were an English learner). If the answer was "yes," they were further asked to indicate how they were diagnosed with ADHD (formal assessment by a specialist, screening conducted by a teacher or educational

professional, or self-diagnosis or diagnosis by a family member) and to describe the impact of their ADHD symptoms in the context of test taking (mild, moderate, or severe, with provided operational definitions discussed subsequently in this report). Students who answered "yes" to the ADHD question and met other selection criteria (see next subsection) were considered eligible for this study, and no further documentation or other evidence of their condition was requested or collected. This approach minimized medical privacy concerns but did raise the possibility that one or more participants would self-identify as having ADHD when they didn't merely to participate in the study and receive its incentive. To militate against this possibility, the screener didn't specify that students with ADHD were being sought, and the initial query about having ADHD 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.

Prospective participants were also asked whether they had received or expected to receive accommodations as part of SAT Suite testing and, if so, to identify them (extended time on exams, extended breaks, assistive technology [e.g., text-to-speech software], or other). This is important because the provision of appropriate testing accommodations for universally designed exams is a key means by which fairness on the SAT Suite is ensured for students with disabilities (College Board 2024b). Any response to these two questions, including ones indicating that students hadn't received or didn't expect to receive accommodations, was considered acceptable for sample selection.

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 had ADHD.
- 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

[1] One Reading and Writing participant (RW24) failed to provide HSGPA and was inadvertently included in the study.

so was not a requirement, and as all students (with one exception 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.

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 ADHD-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 who were English learners (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 ADHD condition, including whether they have ADHD (always "yes"), which SAT Suite testing accommodations they have already received or expected to receive, and a rating of the impact of their ADHD symptoms on test taking (mild, moderate, severe; definitions discussed below).

Table 2. Reading and Writing Participant Roster by Demographics, SAT Suite Accommodations Status, and Self-Reported ADHD Impact on Test-Taking Ability.

	Demographics						SAT Suite Accommodations		
Part. ID	Gender	Race	Ethnicity	Home State	Grade in School	Self-Reported HSGPA	Received/ Expected?	Type(s)	Self-Reported ADHD Impact
RW1	Female	Black or African American	Not of Hispanic, Latino, or Spanish origin	MO	12	B+ (87–89)	Yes	ET, EB	Mild
RW2	Female	White	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	11	A+ (97–100)	Yes	ET, EB	Moderate
RW3	Female	Black or African American	Not of Hispanic, Latino, or Spanish origin	GA	11	B (83–86)	Yes	ET	Mild
RW6	Male	White	Not of Hispanic, Latino, or Spanish origin	MI	11	A– (90–92)	Yes	ET, EB	Moderate
RW7	Male	*NR*	Not of Hispanic, Latino, or Spanish origin	WA	12	B+ (87–89)	Yes	ET	Mild
RW9	Male	Asian	Not of Hispanic, Latino, or Spanish origin	ID	11	B+ (87–89)	Yes	EB	Mild
RW10	Male	White	Not of Hispanic, Latino, or Spanish origin	MN	11	B (83–86)	Yes	ET, EB	Moderate
RW12	Male	Black or African American	Not of Hispanic, Latino, or Spanish origin	AL	12	B+ (87–89)	Yes	ET	Moderate
RW15	Female	*NR*	Puerto Rican	MN	11	A+ (97–100)	Yes	ET	Moderate
RW16	Male	White	Not of Hispanic, Latino, or Spanish origin	NJ	11	B+ (87–89)	Yes	ET	Moderate
RW17	Female	White	Not of Hispanic, Latino, or Spanish origin	PA	11	B– (80–82)	Yes	ET	Moderate
RW18	Female	White	Not of Hispanic, Latino, or Spanish origin	FL	12	A– (90–92)	No	–	Mild
RW19	Female	White	Not of Hispanic, Latino, or Spanish origin	AR	12	B (83–86)	Yes	ET	Severe
RW20	Female	*NR*	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	11	A (93–96)	Yes	ET, EB	Moderate
RW21	Female	White	Not of Hispanic, Latino, or Spanish origin	WA	12	A (93–96)	Yes	ET, EB	Moderate
RW24	Female	White	Puerto Rican	FL	11	*NR*	Yes	ET, EB	Moderate
RW26	Male	White	Not of Hispanic, Latino, or Spanish origin	GA	12	B+ (87–89)	Yes	ET	Moderate
RW27	Female	Asian	Not of Hispanic, Latino, or Spanish origin	TX	11	A (93–96)	No	–	Moderate
RW33	Male	Black or African American	Not of Hispanic, Latino, or Spanish origin	MD	12	D+ (67–69)	Yes	ET	Moderate
RW37	Female	White	Not of Hispanic, Latino, or Spanish origin	TX	11	B (83–86)	Yes	ET, AT	Moderate
RW38	Male	White	Not of Hispanic, Latino, or Spanish origin	TX	11	A– (90–92)	Yes	ET	Mild
RW42	Male	White	Not of Hispanic, Latino, or Spanish origin	FL	12	A (93–96)	Yes	ET	Moderate
RW44	Female	Asian	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	WA	12	A– (90–92)	Yes	ET, EB	Severe
RW45	Male	White	Mexican	TX	12	A (93–96)	Yes	ET	Severe

NR: No response

SAT Suite Accommodations Types:

ET = Extended time

EB = Extended breaks

AT = Assistive technology (e.g., text-to-speech)

Definitions for Self-Reported ADHD Impact on Test-Taking Ability:

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

Table 2 suggests that within the strictures of the small sample size (*n* = 24) 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, self-reported HSGPA, and ADHD symptom impact on test-taking ability. Specifically:

- **Gender.** An approximately equal proportion of female (thirteen) and male students (eleven) participated.
- **Race and ethnicity.** Most numerous were White participants not of Hispanic, Latino, or Spanish origin (eleven), followed by Black or African American participants not of Hispanic, Latino, or Spanish origin (four), and Asian participants not of Hispanic, Latino, or Spanish origin (two). Together, these accounted for seventeen of twenty-four participants. Other races/ethnicities identified by participants (one each) were Asian/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; White/Mexican; No response/Puerto Rican; No response/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican; and No response/Not of Hispanic, Latino, or Spanish origin. 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 both grade 11 (thirteen) and grade 12 (eleven) were represented.
- **Self-reported HSGPA.** Eleven participants indicated an "A" HSGPA, eleven indicated a "B" HSGPA, one indicated a "D" HSGPA, and one 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 ADHD 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.
- **ADHD symptom impact.** Most participants (fifteen) indicated that their ADHD symptoms had a moderate impact on their test-taking ability; six participants indicated a mild impact, and three participants indicated a severe impact.

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, SAT Suite Accommodations Status, and Self-Reported ADHD Impact on Test-Taking Ability.

	Demographics						SAT Suite Accommodations		
Part. ID	Gender	Race	Ethnicity	Home State	Grade in School	Self-Reported HSGPA	Received/ Expected?	Type(s)	Self-Reported ADHD Impact
M1	Male	*NR*	Not of Hispanic, Latino, or Spanish origin	GA	12	B– (80–82)	Yes	ET, EB	Moderate
M2	Male	Asian	Not of Hispanic, Latino, or Spanish origin	VA	11	A– (90–92)	Yes	ET	Moderate
M3	Male	*NR*	Puerto Rican	NY	11	B (83–86)	Yes	ET, EB	Moderate
M5	Female	Black or African American	Not of Hispanic, Latino, or Spanish origin	LA	11	B+ (87–89)	No	–	Mild
M6	Male	*NR*	Mexican	NC	12	C+ (77–79)	Yes	EB	Moderate
M7	Female	White	Not of Hispanic, Latino, or Spanish origin	VA	11	A+ (97–100)	Yes	ET, EB, AT	Moderate
M8	Female	White	Not of Hispanic, Latino, or Spanish origin	PA	12	B– (80–82)	Yes*	ET	Moderate
M12	Female	White	Not of Hispanic, Latino, or Spanish origin	KY	11	A– (90–92)	Yes	ET	Moderate
M14	Female	Asian	Not of Hispanic, Latino, or Spanish origin	TX	12	A (93–96)	Yes	ET	Mild
M15	Male	*NR*	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	12	A+ (97–100)	Yes	ET	Mild
M17	Female	White	Not of Hispanic, Latino, or Spanish origin	MI	12	C (73–76)	No	–	Moderate
M18	Female	White	Not of Hispanic, Latino, or Spanish origin	WA	12	A+ (97–100)	Yes	ET	Moderate
M19	Male	White	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	NJ	11	C– (70–72)	Yes	ET, EB	Moderate
M23	Male	White	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	FL	12	B+ (87–89)	Yes	ET, EB	Severe
M30	Female	White	Not of Hispanic, Latino, or Spanish origin	NJ	12	B+ (87–89)	Yes	ET	Moderate
M31	Female	*NR*	Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican	CA	12	B– (80–82)	No	–	Moderate
M37	Male	White	Not of Hispanic, Latino, or Spanish origin	OR	12	C+ (77–79)	No	–	Mild
M42	Female	Asian	Not of Hispanic, Latino, or Spanish origin	NC	12	B (83–86)	Yes	ET	Severe
M43	Female	White	Not of Hispanic, Latino, or Spanish origin	GA	12	A (93–96)	Yes	ET, EB	Moderate
M59	Female	Asian	*NR*	TX	12	A+ (97–100)	Yes	ET, EB	Severe
M61	Female	White	Not of Hispanic, Latino, or Spanish origin	NC	12	A (93–96)	Yes	EB	Moderate

* Participant M8 replied "no" to the "Received/expected SAT Suite accommodations?" question but listed an expected/received accommodation.

NR: No response

SAT Suite Accommodations Types:

- ET = Extended time
- EB = Extended breaks
- AT = Assistive technology (e.g., text-to-speech)

Definitions for Self-Reported ADHD Impact on Test-Taking Ability:

- 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

Table 3 suggests that within the strictures of the small sample size (*n* = 21) and self-selection methodology used for this study, Math participants represented a somewhat diverse sample in terms of gender, race/ethnicity, grade in school, self-reported HSGPA, and ADHD symptom impact on test-taking ability. Specifically:

- **Gender.** Female participants (thirteen) were somewhat disproportionately represented relative to male participants (eight).
- **Race and ethnicity.** Nine participants identified as White and not of Hispanic, Latino, or Spanish origin; three as Asian and not of Hispanic, Latino, or Spanish origin; two as White and of Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican; and two as No response/Hispanic, Latino, or Spanish origin other than Cuban, Mexican, or Puerto Rican. Together, these made up sixteen of the twenty-one participants. Other races/ethnicities identified by participants (one each) were Asian/No response; Black or African American/Not of Hispanic, Latino, or Spanish origin; No response/Puerto Rican; No response/Mexican; and No response/Not of Hispanic, Latino, or Spanish origin. 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 both grade 11 (six) and grade 12 (fifteen) were represented.
- **Self-reported HSGPA.** Nine participants indicated an "A" HSGPA, eight indicated a "B" HSGPA, and four indicated a "C" HSGPA. This represents a somewhat 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 ADHD subpopulation. While the sample is somewhat 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.
- **ADHD symptom impact.** Most participants (fourteen) indicated that their ADHD symptoms had a moderate impact on their test-taking ability; four participants indicated a mild impact, and three participants indicated a severe impact.

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. In the single case (in Math) in which no participant answered a given question correctly and exhibited appropriate behaviors, no supplementary vignette was incorporated.

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-four hundred codes being assigned to forty-five 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 when available 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 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 when available.

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 applied 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 that you have ADHD? 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 #														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RW1	5	1	4	1	1	5	1	2	1	4	1	4	4	5	1
RW2	4	1	1	1	1	1	2	2	1	1	1	1	1	4	1
RW3	5	1	5	2	5	5	4	4	1	1	1	4	5	3	4
RW6	4	1	1	1	1	1	4	1	1	1	1	4	1	4	1
RW7	3	1	1	1	1	5	1	1	1	1	1	4	1	4	1
RW9	4	1	5	1	1	1	2	2	1	1	1	1	1	3	1
RW10	1	5	5	–	1	5	4	2	1	1	1	1	4	2	4
RW12	5	1	5	2	1	5	4	5	5	3	3	1	1	2	1
RW15	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1
RW16	5	1	5	1	1	5	2	1	1	1	1	5	4	5	1
RW17	5	1	5	2	4	5	5	4	5	5	1	1	3	5	4
RW18	4	1	4	1	1	1	2	1	1	1	1	1	1	1	1
RW19	3	1	5	4	1	5	4	2	1	1	1	1	1	4	1
RW20	5	1	3	2	1	3	5	5	1	1	1	3	1	4	4
RW21	5	1	3	1	1	5	5	5	3	2	1	3	4	3	1
RW24	1	1	4	1	1	1	1	1	1	1	1	1	1	4	3
RW26	4	1	4	4	1	5	4	1	1	1	–	5	4	4	1
RW27	4	1	1	1	1	1	1	2	1	4	1	1	1	2	1
RW33	1	1	4	4	1	4	4	2	1	4	1	4	–	–	–
RW37	5	1	4	4	1	5	2	4	5	5	1	5	1	5	1
RW38	4	1	4	1	1	1	1	1	1	1	1	4	1	–	–
RW42	4	1	4	1	1	1	1	1	1	1	1	1	1	1	1
RW44	1	1	1	–	1	1	1	1	1	2	1	1	1	2	3
RW45	4	1	1	1	1	1	1	2	1	1	1	4	1	1	1

Performance by Level, by Participant						Participant Performance Summary		
1	2	3	4	5	*NR*	*#AC*	*#RB*	D_p
7	1	0	4	3	0	8	7	1 ✔
11	2	0	2	0	0	13	11	2 ✔
4	1	1	4	5	0	6	4	2 ✘
11	0	0	4	0	0	11	11	0 ✔
11	0	1	2	1	0	12	11	1 ✔
10	2	1	1	1	0	13	10	3 ✔
6	2	0	3	3	1	8	6	2 ✔
5	2	2	1	5	0	9	5	4 ✘
13	2	0	0	0	0	15	13	2 ✔
8	1	0	1	5	0	9	8	1 ✔
3	1	1	3	7	0	5	3	2 ✘
12	1	0	2	0	0	13	12	1 ✔
8	1	1	3	2	0	10	8	2 ✔
6	1	3	2	3	0	10	6	4 ✘
5	1	4	1	4	0	10	5	5 ✘
12	0	1	2	0	0	13	12	1 ✔
6	0	0	6	2	1	6	6	0 ✔
11	2	0	2	0	0	13	11	2 ✔
5	1	0	6	0	3	6	5	1 ✔
5	1	0	3	6	0	6	5	1 ✔
10	0	0	3	0	2	10	10	0 ✔
13	0	0	2	0	0	13	13	0 ✔
11	2	1	0	0	1	14	11	3 ✔
12	1	0	2	0	0	13	12	1 ✔

Performance by Level, by Question

1	5	23	7	14	22	11	8	9	20	16	22	12	16	4	16
2	–	–	–	4	0	–	6	9	–	2	–	–	0	4	0
3	2	0	2	0	0	1	0	0	1	1	1	2	1	3	2
4	9	0	8	4	1	1	7	3	0	3	0	7	5	7	4
5	8	1	7	0	1	11	3	3	3	2	0	3	1	4	0
NR	0	0	0	2	0	0	0	0	0	0	1	0	1	2	2

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	7	23	9	18	22	12	14	18	21	18	23	14	17	11	18
#RB	5	23	7	14	22	11	8	9	20	16	22	12	16	4	16
D_q	2 ✔	0 ✔	2 ✔	4 ✔	0 ✔	1 ✔	6 ✘	9 ✘	1 ✔	2 ✔	1 ✔	2 ✔	1 ✔	7 ✘	2 ✔

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*
7	1	0	4	3	0

Participant Performance Summary

#AC	*#RB*	D_p
8	7	1 ✔

Participant RW1, per the top row in "Performance by Level, by Participant" sub-table, attained PL 1 (the most successful outcome) on seven questions, PL 2 on one question, PL 4 on four questions, and PL 5 (the least successful outcome) on three 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 RW1 answered a total of eight questions correctly (*#AC*, calculated by adding together the number of question responses attaining PLs 1, 2, and 3) and answered seven 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 1, as 8 (*#AC*) minus 7 (*#RB*) equals 1. This D_p exceeds the threshold for a "good" differential, hence the checkmark and blue shading, as 7 (*#RB*) divided by 8 (*#AC*) equals .875, or 87.5 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, nineteen of twenty-four participants (79 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 RW3 answered six questions correctly and demonstrated all required behaviors for four of those questions, resulting in a D_p of 2, representing two-thirds (67 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 half to two-thirds 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, twelve of the fifteen studied Reading and Writing questions (80 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 students with ADHD. The remaining three questions (questions 7, 8, and 14) were still answered correctly by a minimum of four participants who also demonstrated all required behaviors (*#RB*; PL 1), suggesting that these questions, too, are capable of eliciting cognitively complex thinking from students with ADHD, even if they didn't always during the study. The question differentials for the three criterion-failing questions were high, at 6 (question 7), 9 (question 8), and 7 (question 14), but these questions were also answered correctly by a moderate to high number of participants (fourteen, eighteen, and eleven, respectively. In terms of top-level characteristics, the three questions yielding higher-than-desirable D_qs have no immediately obvious commonalities besides being Reading and Writing questions. While question 14 was hard (PSB 7), questions 7 and 8 were of medium difficulty (PSBs of 4 and 5, respectively). While questions 8 and 14 had passages that were in the highest text complexity band (PSR), question 7 had a passage in the lowest (MID). While the passages for questions 8 and 14 were set in science contexts, the passage for question 7 was set in a humanities context. While questions 7 and 8 were both of the Rhetorical Synthesis type, question 14 was a Command of Evidence: Quantitative question. While question 14 included an informational graphic, questions 7 and 8 didn't. Given the moderate to high rates at which these questions were answered correctly (reflected by *#AC*) and the generally high rates at which participants demonstrated strong content mastery (reflected by *#AC* across questions) and ability to verbalize cognitively complex thinking (reflected by *#RB* and D_q across questions), it seems most likely that there's something about the three criterion-failing questions that raised particular issues for participants, a topic we return to in Section 5: Discussion.

PARTICIPANT PERFORMANCE VIGNETTES

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

Example: Question Performance

14
4
0
4
0
2

18
14
4 ✔

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, and those from four participants attained PL 4, while two participants didn't answer the question, as indicated by the "2" in the "*NR* / no response" cell. Adding together the counts in PLs 1–3, we find that a total of eighteen 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 18 (*#AC*) yields a question differential (D_q) of 4. This D_q exceeds the threshold for a "good" differential, hence the checkmark and blue shading, as 14 (*#RB*) divided by 18 (*#AC*) equals approximately .778, or 77.8 percent, which is above the 70 percent cutoff.

This section relies primarily on a case study approach, in which we follow a single participant, RW19, as she works through all fifteen Reading and Writing questions, succeeding on some and struggling with others. In the latter cases, RW19'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 participants with ADHD were able to demonstrate cognitively complex thinking in accordance with the question types' constructs at least some of the time.

Case Study: Participant RW19

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

- They indicated their ADHD condition had at least a moderate impact on their test-taking ability, indicating that they've observed at least some negative impact from their condition on their ability to take tests.
- They'd received or they expected to receive accommodations as part of SAT Suite testing.
- They answered all fifteen Reading and Writing questions (e.g., didn't run out of time).
- They exhibited a good participant differential (D_p).

Participant RW19, a female twelfth grader from Arkansas, met these conditions. She identified as White and not of Hispanic, Latino, or Spanish origin. She self-reported a high school GPA (HSGPA) of B, indicated that she'd received or she expected to receive an extra time accommodation in SAT Suite testing, and characterized her ADHD symptoms as severe, meaning (per the definition provided in the screener) that her ADHD condition had significant impact on her test-taking ability even with appropriate testing accommodations. RW19 answered ten of the fifteen Reading and Writing questions correctly and demonstrated all required behaviors for eight of those ten, resulting in a participant differential of 2 (80 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.
RW19 Performance Level	3

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.

> OK. I'm gonna read this again. "To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes, engineer Aroba Saleem and colleagues [*reads in answer choice A*] hypothesized the tendencies of [some] metal[s'] . . ." Yeah, I don't even know what that word means, honestly. So I'm not sure if that would be it though.
>
> And then "To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes, engineer Aroba Saleem and colleagues [*reads in answer choice B*] "discounted the tendency . . ." I don't think that's it either.
>
> "To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes, engineer Aroba Saleem and colleagues [*reads in answer choice C*] redefined the tendency of some metals' internal magnetic fields to alter under stress." OK. Not sure if it's that one either.
>
> "To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes, engineer Aroba Saleem and colleagues [*reads in answer choice D*] exploited the tendency . . ." OK. I think I'm gonna go with choice D, "exploited."
>
> *Participant RW19*

Participant RW19 answered the question correctly but didn't demonstrate the other required behavior, resulting in a PL of 3. RW19 uses the common test-taking strategy of "plugging" the various answer options into the blank in the passage's sentence. She suggests that she doesn't know what "hypothesized" (choice A) means, which makes her uncertain whether it's

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 best answer ("so I'm not sure if that would be it though"). Other than ruling out choices B and C and ruling in choice D, the key, on grounds seemingly of general "fit," RW19 doesn't provide a description of her thinking, thus failing to demonstrate adequate passage comprehension.

Supplementary Vignette: Participant RW33

Participant RW33 answered question 1 correctly and demonstrated both required behaviors, resulting in a PL of 1. RW33 was one of seven participants to have correctly answered the question and one of five to have done so while also demonstrating both required behaviors.

> "To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes, engineer Aroba Saleem and colleagues [*reads in choice A*] "[hypothesized; *participant says "hypnotized"*] the tendencies [*sic*] of some metals—"
>
> [*reads in choice B*] He "discounted the tendencies of some metals—"
>
> [*reads in choice C*] He "redefined the tendencies of [some] metals—"
>
> [*reads in choice D*] He "exploited the tendencies of some metals—"
>
> "Engineer Aroba Saleem and colleagues exploited the tendencies of some metals' internal magnetic fields—" He, colleagues, "Engineer Aroba Saleem and colleagues redefined the tendencies of some metals' internal magnetic fields—"
>
> I don't think it's A because—well? "Aroba Saleem and colleagues hypothesized the tendencies—" "To demonstrate that the integrity of underground metal pipes can be assessed without unearthing the pipes," he "exploited—"
>
> It couldn't be "hypothesized" because they're not, they're taking action—'cause they're demonstrating. It's not a question. "Discounted" doesn't sound like it should be there at all. "Engineer Aroba Saleem and colleagues discounted—"
>
> I think "redefined" or "exploited."
>
> ". . . engineer Aroba Saleem and colleagues redefined the tendencies—"
>
> I think it's "exploited" because "redefined" would mean, like, they're making something new, and all they're trying to do is to demonstrate that [the pipes] can be [assessed; *participant says "assembled"*] without unearthing the pipes. I think it's D.
>
> *Participant RW33*

Participant RW33 also uses the common "plug-in" strategy to "read in" the various answer choices. Although he initially mispronounces choice A, "hypothesized," as "hypnotized," he later self-corrects. RW33 demonstrates clear passage comprehension (behavior 1) when he rules out "hypothesized" because "they're taking action—'cause they're demonstrating. It's not a question" and eliminates "redefined" (choice C) because that word "would mean, like, they're making something new, and all they're trying to do is to demonstrate that [the pipes] can be [assessed] without unearthing the pipes." Having already discounted "discounted" (it "doesn't sound like it should be there at all"), RW33 settles on the best answer, choice D (behavior 2). Although RW33 doesn't provide a working definition for "exploited," he nonetheless provides passage-based reasons for ruling out choices A and C.

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.
RW19 Performance Level	1

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."

> Choice A, "To compare basketball with other sports." That is definitely not it because it does not compare [to] any other sport, and it does not mention any other sport besides basketball. So I would definitely cross choice A off.
>
> Choice B, "To provide details of how to play basketball." Yeah, it's definitely not that one either because it does not have any tips or anything of how you play.
>
> And then choice C, "To state how players will be chosen for the basketball team." Again, it does not say anything about, like, how people are chosen for the team, just about her, like, saying why she might get in it.
>
> And then choice D, "To explain why the narrator thinks she might make the basketball team." And, yeah, I just said that, so I am going to go with choice D, "To explain why the narrator thinks she might make the basketball team."
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. She demonstrates adequate passage comprehension (behavior 1) in several places, noting, for example, that the passage "does not compare [to] any other sport, and it does not mention any other sport besides basketball" and "does not have any tips or anything of how you play" but is "just about her, like, saying why she might get" on the team. Finding choice D, the best answer, a reasonable match to her own conclusion about the contents of the passage, RW19 selects that as her 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.
RW19 Performance Level	5

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."

> Choice A, "It presents an objection to the argument of Guerra et al. about transportation mode choice in urban areas of Mexico." Might have to read this whole thing [passage] again because— [*Rereads passage and question stem*]
>
> Choice A, "It presents an objection to the argument of Guerra et al. about transportation choice in urban areas of Mexico." That one may be it, but I'm going to go ahead and read the other ones to see.
>
> Choice B, "It explains why it is challenging to influence transit ridership solely by altering characteristics of a transit system." Not sure about that one either. But let's go to choice C.
>
> [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." Not sure about that. But let me read the last one.
>
> Choice D, "It substantiates the assertion that population density, the spatial distribution of jobs, and demographic characteristics are important factors in transportation mode choice."
>
> [*Rereads the question stem*] The underlined portion—I'll just read this whole sentence [of the passage]: "System features do affect ridership, of course, but there is an irreducibly contextual dimension of transportation mode choice." Then— [*Long pause*] I'm gonna go with choice D, "It substantiates the assertion that population density, the spatial distribution of jobs, and demographic characteristics are important factors in transportation mode choice."
>
> *Participant RW19*

Participant RW19 answered the question incorrectly and didn't demonstrate either required behavior, resulting in a PL of 5. Her verbalization doesn't provide any clear indication of why she chose D or ruled out the other answer options.

Supplementary Vignette: Participant RW7

Participant RW7 answered question 3 correctly and also demonstrated both required behaviors, resulting in a PL of 1. RW7 was one of nine participants to have answered the question correctly and one of seven to have done so while also demonstrating both required behaviors.

> [*Reads passage*] ". . . As Erick Guerra et al.—" I'm confused about what "et al." means. I'm assuming that it's a name, but I think there might be some confusion on "Erick Guerra et al." I don't know what that is. [*Continues reading passage*]
>
> I'm just gonna reread the underlined thing because I feel like that's what the question is gonna be about. ". . . there is an irreducibly contextual

dimension of transportation mode choice." "Irreducibly" means, like, you cannot reduce it. You cannot make it less. "Contextual dimension of transportation mode choice." So, this is—I feel like this is a lot of big words that could be turned into little words and be understood just fine. The question said, the writer of the question sounds like they're trying to sound smart, and it makes it really hard to understand what they're saying. But I'm guessing that this is just to show if we can understand hard words.

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

Option A, "It presents an objection to the argument of Guerra et al.—" So, now I know that his name is "et al." "—about transportation mode choice in urban areas of Mexico." I'm a little confused by this question. So ". . . but simply reproducing a feature . . ." What does it say? Et al. says, "As Erick Guerra et al. have shown, transportation mode choice in urban areas of Mexico is the product of a complex mix of factors." So, he's saying it's a complex mix of factors. Option A is saying "objection to the argument," but this really doesn't object to what he's saying. So, I'd throw out A.

Option B, "It explains why it is challenging to influence transit ridership solely by altering characteristics of a transit system." I gotta read B again because it's very wordy. It's very verbose. "It explains why it is challenging to influence transit ridership solely by altering characteristics of a transit system." This, I'm, this sounds like an—now that I read it again and I understand it—this sounds like an answer, like a good answer, because [the passage] says "there is an irreducibly contextual dimension of transportation mode choice." Basically, I think what that means is that there are much more factors than just making the fares cost less. So I think it might be B.

Option C, "It [illustrates the claim] that a characteristic associated with high transit ridership in Mexico [City] is not associated with high [transit] ridership elsewhere." It's a little vague. I wouldn't use C.

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

So I am torn between B and D. I'm just going to reread the underlined [portion]. ". . . there is an irreducibly contextual dimension of transportation mode choice." It is substantiating—on rereading [choice] D—it is substantiating the assertion of population density. These are important factors in transportation mode choice. Are important factors. I would say D is saying too much for that underlined sentence. I would say, I would go with B, "It explains why it is challenging to influence transit ridership solely by altering characteristics of a transit system."

Participant RW7

While reading the passage, RW7 exhibits "some confusion" over the meaning of "et al.," which he ultimately concludes is a name. While imprecise, his working definition is sufficient (and the issue trivial enough in context) to allow him to

continue successfully. RW7 demonstrates adequate passage comprehension (behavior 1) at two main points. First, drawing on prior knowledge, he defines "irreducibly" as "you cannot make it less." Second, he reasonably paraphrases the underlined portion ("there is an irreducibly contextual dimension of transportation mode choice") and determines its implication in context, observing that "there are much more factors than just making the fares cost less." RW7 demonstrates strong comprehension and problem-solving skills throughout his approach to the question. For example, he correctly rules out choice A by noting that the underlined portion "really doesn't object to what [Guerra et al. is] saying." He also blocks choice D on the grounds that it's "saying too much for the underlined sentence," which, in context, likely means that "substantiates" is more than the underlined portion accomplishes. After having been "torn between [choices] B and D," he finally decides on choice B, the best answer (behavior 2), drawing on his prior restatement of the underlined portion.

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.
RW19 Performance Level	4

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.

> [*Rereads passage*] OK. So I think that we have to basically find out which [species] has been extinct the longest, because in the passage, it's saying, like, time is an obstacle for the scientists because if [the species] are gone too long, then the harder it is, like, to have a habitat for them, 'cause you can't just, like, re-create nature. So it'd probably be choice C, "The woolly mammoth [*Mammuthus primigenius*], which became extinct several thousand years before the saber-toothed cat [*Smilodon*]," because it seems to be the one that has been gone the most, I believe. Yeah.
>
> *Participant RW19*

Participant RW19 answered the question incorrectly but did demonstrate a single required behavior, resulting in a PL of 4. RW19 demonstrates adequate comprehension of the passage (behavior 1) by observing that "we have to basically find out which [species] has been extinct the longest." Her response, choice C, is superficially consistent with that reasoning, as she asserts that, of the four species presented in the table and answer choices, the woolly mammoth "seems to be the one that has been gone the most, I believe." Choice C does imply that the woolly mammoth became extinct before the saber-toothed cat, but this interpretation is incorrect per the table. According to the table, the species (of the five presented) that has been extinct the longest is, in fact, the saber-toothed cat, which became extinct 11,000 years before the present, as opposed to the mammoth, which became extinct only 6,400 years before the present. RW19's response suggests that she didn't attend closely to the table and instead relied mainly on the information provided (as it turns out, partially incorrectly) by the answer choices. Nevertheless, RW19 shows partial enactment of the question type's construct.

Supplementary Vignette: Participant RW1

Participant RW1 answered question 4 correctly and demonstrated all required behaviors, resulting in a PL of 1. RW1 was one of eighteen participants to have answered the question correctly and one of fourteen participants to have done so while also demonstrating all required behaviors.

> [*Long pause after reading answer choices; moderator prompts her to think aloud*] I was just looking at the, I'm looking at the chart right now, the list of— I don't think it's [choice A] passenger pigeon because it says it became extinct in 1914, and that wasn't that long ago. So there will be, they will still have a suitable habitat that still exists for the species 'cause it wasn't that long ago [that it went extinct]. And [choice] B [saber-toothed cat], I think it's B because it says it became extinct 11,000 years before present, and [the passage] said "the longer a species has been extinct, the less likely it is that a suitable habitat still exists for that species." And that would be, it's been a long time. So, basically. I don't think it's [choice] D [Caribbean monk seal]. I don't think it's [choice] C [woolly mammoth] because, yeah, I don't think it's C 'cause the saber-toothed cat has been, became extinct more longer than the woolly mammoth. So I don't think it's C. And [choice D], I don't think it's Caribbean monk seal because it wasn't that long ago from the, yeah.
>
> *Participant RW1*

Participant RW1 demonstrates adequate passage comprehension (behavior 1) when she uses the fact that the text says "the longer a species has been extinct, the less likely it is that a suitable habitat still exists for that species" to pick choice B, the best answer (behavior 4) ("it's been a long time" since the saber-toothed cat went extinct). On similar grounds, she correctly rules out choices A ("it says [the passenger pigeon] became extinct in 1914, and that wasn't that long ago") and D ("it wasn't that long ago"). She most clearly exhibits adequate table comprehension (behavior 2) when ruling out choice C, the woolly mammoth: Answer choice C wrongly asserts that the mammoth went extinct before the saber-toothed cat, but RW1 correctly observes that, per the table, the saber-toothed cat is the longest extinct of the five species listed. It seems likely that RW1 didn't notice the factual error in choice C but instead relied on the table to obtain accurate data. Through her rationale for picking the best answer and ruling out the distractors, RW1 also clearly shows an understanding of the relationship between the passage, table, and question stem (behavior 3).

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.
RW19 Performance Level	1

"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.

> [*Rereads passage*] OK. So in the passage [*actually the question stem*], it says she has mixed feelings.
>
> So 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." This one, it's not all that bad but not all that good. So I kind of like this one, but I'm gonna read these ones again.
>
> Choice B, "By moonlight—the moon shines in all night when there is a moon—I wouldn't know it was the same paper." Again, it's not, like, dissing it that much, but it's also not, like, happy about it, I'd assume.
>
> Choice C, "I'm really getting quite fond of the big room, all but that horrid [wall]paper." This one talks about everything, actually. Like, I just realized [the question stem] says "her surroundings." So this one actually might be it.
>
> Choice D, "The color is repellant, almost revolting; a smoldering, unclean yellow, strangely faded by the slow-turning sunlight." That one is definitely not it because that one is full-on, like, "no" about her surroundings.
>
> So I would choose choice C, "I'm really getting quite fond of the big room, all but that horrid [wall]paper."
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW19 demonstrates adequate comprehension of the passage (in this case, the answer choices; behavior 1) by noting that choice C, the best answer, "talks about everything," meaning, in context, that that option captures effectively the notion of "mixed feelings," the tie that binds the question's stem and keyed response together. She states earlier that the question stem specifies that the narrator "has mixed feelings," indicating that she has an understanding of the question-passage relationship (behavior 2). She hesitates a bit on choices A and B, perceiving some degree of mixed assessment in them (e.g., "Again, it's [choice B is] not, like, dissing it that much, but it's also not, like, happy about it") but determines that choice C is a better fit with "mixed feelings" and that choice D "is definitely not it because that one is full-on, like, 'no' about her surroundings." RW19 ultimately correctly selects choice C as her answer (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.
RW19 Performance Level	5

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.

> OK. "According to Duverger's law, countries with single-ballot majoritarian—" [*Struggles to pronounce*] I don't know. "—elections for single-member districts tend to polarize into two-party systems, wherein dueling political parties consistently dominate the political system."
>
> OK, Choice A. [*Reads in choice A*] "Subsequently, countries with proportional-representation electoral systems tend to support multi-partyism, under which power gets distributed among many political parties." That one is OK. Not sure, but I'll read the other ones with that.
>
> [*Rereads first sentence of passage*] [*Reads in choice B*] "Conversely, countries with proportional-representation electoral systems tend to support multi-partyism, under which power gets distributed among many political parties." OK.
>
> And then. [*Rereads first sentence of passage*] [*Reads in choice C*] "For instance—" Yeah, I don't even have to read the rest of that to know that that is not it. And I would say the same with choice D ["in other words"] because they are both meaning almost the same.
>
> So I would go with choice A, "subsequently."
>
> *Participant RW19*

Participant RW19 answered the question incorrectly and didn't demonstrate either required behavior, resulting in a PL of 5. At no time in her response does RW19 demonstrate adequate comprehension of the passage. She notes that choices A and B are "OK" but provides no details as to why she feels this way. She correctly rules out choice C ("yeah, I don't even have to read the rest of that to know that that is not it") on unspecified grounds. She then wrongly asserts that choice D is incorrect because choice C is and because "they are both meaning almost the same"; in fact, "for instance" signals an example is forthcoming, while "in other words" signals a simpler paraphrase of previously presented information or ideas. She then incorrectly selects choice A as her answer.

Supplementary Vignette: Participant RW15

Participant RW15 answered question 6 correctly and demonstrated both required behaviors, resulting in a PL of 1. RW15 was one of twelve participants to have answered the question correctly and one of eleven participants to have done so while also demonstrating both required behaviors.

> I don't—OK, I'll just cross them off, I guess. So, like, [choice] D, it's not really "in other words" because the second statement isn't, like, reiterating the first statement. [Choice A,] "Subsequently," this one doesn't really work because it's not like the second one is the result, the second statement isn't, like, the result of the first statement. [Choice C,] "For instance," the second thing isn't, like, an example of the first one.
>
> So I guess it would be [choice B,] "conversely," because it's showing, like—it's, like, comparing the two and showing, like, the differences between them.
>
> *Participant RW15*

RW15 demonstrates both strong passage comprehension (behavior 1) and vocabulary skills by the manner in which she eliminates the incorrect answer

options. She correctly rules out choice A, "subsequently," on the basis that "the second statement isn't, like, the result of the first statement"; choice C, "for instance," because "the second [statement] isn't, like, an example of the first one"; and choice D, "in other words," on the grounds that "the second statement isn't, like, reiterating the first statement." RW15 also expresses a general sense of the meaning of choice B, "conversely," noting that it's needed in the blank to signal "the differences" between countries with single-ballot majoritarian elections for single-member districts and countries with proportional-representation elections, which tend to resolve into two-party systems and multi-party systems, respectively. With this clarity, she selects the best answer to the question, choice B (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.
RW19 Performance Level	4

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.

> [*Reads choice A without comment*]
>
> Choice B, "George Eliot, which *Adam Bede*'s title page indicated was the name of the novel's author, was widely [assumed; *participant says "presumed"*] to be a pseudonym." Hmmm. "The student wants to identify the real author [of *Adam Bede*]." Yeah, choice B is not it because it's not, like, he's not going to find out the author's real name just by assuming that they used a fake name in the first place, 'cause he already knew that. But, yeah.
>
> Choice C, "The title page of the novel *Adam Bede* indicated that the author's name was George Eliot." Again, that is not it because that's not, like, it's not using his information to actually find the real author.
>
> Choice D, "A woman who had used a pseudonym to conceal her identity later received herself as the real author of *Adam Bede*."
>
> The question again: "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?" [*Rereads bullet-list notes*]
>
> I would have to go with choice D, "A woman who had used a pseudonym to conceal her identity later revealed herself as the real author of *Adam Bede*."
>
> *Participant RW19*

Participant RW19 answered the question incorrectly but did demonstrate two required behaviors, resulting in a PL of 4. RW19 demonstrates adequate comprehension of the notes (behavior 1) in two main ways. First, she rules out choice B on the grounds that it doesn't add anything to uncovering the real author of *Adam Bede*, as the (hypothetical) student "already knew" that George Eliot was a pseudonym. Second, she rules out choice C because "it's not actually using [the student's] information to actually find the real author," an indication that she realizes the notes identify *Adam Bede*'s real author. RW19 also shows understanding of the notes and the criterion set forth in the question's stem (behavior 2), as she makes clear that the best answer to the question should identify the "real author." Despite all this, RW19 ends up selecting choice D as her answer, even though that option, like the ruled-out choices B and C, doesn't name Mary Ann Evans as the true author of *Adam Bede*, as choice A, the best answer, does. Nevertheless, RW19 shows partial enactment of the question type's construct.

Supplementary Vignette: Participant RW44

Participant RW44 answered question 7 correctly and demonstrated all required behaviors, resulting in a PL of 1. RW44 was one of fourteen participants who answered the question correctly and one of eight participants who did so while also demonstrating all required behaviors.

> [*Starts reading bullet-list notes*] Now I'm going to read the question—or what the question is asking—instead of finishing the notes, and then I'll go back. "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?" [*Rereads question stem*; *returns to notes*]
>
> So it looks like Mary Ann [Evans] is the author's, the real author of *Adam Bede*. And then it's asking which [answer choice] says this, so I'm going to go through them.
>
> [Choice] A, "The real author of *Adam Bede* was Mary Ann Evans, who published the novel using the pseudonym George Eliot." I think it's probably that one, but I'm going to go through the other ones, just in case.
>
> [Choice 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." I don't think it's this one because they don't say "Mary Ann Evans."
>
> [Choice C,] "The title page of [the] novel *Adam Bede* indicated that the author's name was George Eliot." I don't think it's that one because it doesn't say "Mary Ann."
>
> [Choice D,] "A woman who had [used; *participant says "to use"*] a pseudonym to conceal her identity later revealed herself as the real author of *Adam Bede*." Well, I don't think it's that one either because it asks to identify who it was—so, like, say the name—and none of them says her name except for [choice] A.
>
> So I think it's [choice] A, "The real author of *Adam Bede* was Mary Ann Evans, who published the novel using the pseudonym George Eliot."
>
> *Participant RW44*

RW44 demonstrates adequate comprehension of the notes (behavior 1) by observing that "it looks like Mary Ann [Evans] is the author's, the real author of *Adam Bede*." She quickly inclines to choice A, the best answer ("I think it's probably that one") but systematically eliminates the other answer choices to confirm. RW44 correctly surmises that the question asks test takers "to identify who [the real author of *Adam Bede*] was—so, like, say the name" (behavior 2) and reasonably blocks choices B, C, and D because "none of them says her name except for [choice] A." On this basis, she selects choice A, the best answer, as her 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.
RW19 Performance Level	2

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.

> Choice A, "Scientists have developed a freeze-thaw battery that contains molten salt, which liquefies when heated and solidifies at room temperature." That does not have anything to do with the question of how, like, the student wants to specify how the salt enables the energy storage. So that's just no.
>
> Choice B, "The stored energy in a freeze-thaw battery, which contains molten salt, can be used by reheating the battery." That does not say how the salt, like, does it. So no to that one too.
>
> 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." Hmmm, maybe.
>
> Choice D, "Molten salt allows a freeze-thaw battery to retain 92% of its charge after twelve weeks." No.
>
> OK. So I'm gonna go with 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."
>
> *Participant RW19*

Participant RW19 answered the question correctly but demonstrated only two required behaviors, resulting in a PL of 2. RW19 shows a clear understanding of the intended relationship between the notes and the question's stem (behavior 2) in the process of correctly ruling out choice A on the grounds that it "does not have anything to do with the question of how, like, the student wants to specify how the salt enables the energy storage" and choice B on the grounds that it "does not say how the salt, like, does it." However, although she does properly select choice C as the best answer (behavior 3), she doesn't offer a clear indication of having used comprehension of the notes to make her choice. Given that RW19 picked the best answer, we could assume that she understood and used the notes, as was apparent for question 8, but it's possible that she merely determined that choice C best aligned with the question stem's criterion. In any case, RW19 shows partial enactment of the question type's construct.

Supplementary Vignette: Participant RW42

Participant RW42 answered question 8 correctly and demonstrated all required behaviors, resulting in a PL of 1. RW42 was one of eighteen participants to have correctly answered the question and one of nine participants to have done so while also demonstrating all required behaviors.

> All right. Another [question], I guess, of this format. I'm gonna read the notes first.
>
> - "Scientists have developed a 'freeze-thaw' battery that can retain 92% of its charge after twelve weeks." So I guess a powerful battery.
> - "The battery contains molten salt (a type of salt that liquefies when heated and solidifies at room temperature)." So this is more of a, technical details—how, what the battery is made [of].
> - "When the salt is in a liquid state, energy flows through the battery."

- "When the salt is in a solid state, the [*sic*] energy stops flowing and is stored in the battery." So this identifies the mechanism of the battery. Then—
- "The stored (frozen) energy can be used by reheating (thawing) the battery." This would describe the use of the battery.

[*Reads question stem*]

[Choice A,] "Scientists have developed a freeze-thaw battery that contains molten salt, which liquefies when heated and solidifies at room temperature." "The student wants to specify how the salt enables energy storage." This doesn't—I mean, it states, I guess, a property of the battery but doesn't necessarily, I guess, describes [*sic*] how it enables energy storage, so I don't think it's A.

[Choice B,] "The stored energy in a freeze-thaw battery, which contains molten salt, can be used by reheating the battery." This does, I guess, describe the mechanism behind it. Yeah, it describes the mechanism behind it. I mean, it fits the question, but, I mean, there could potentially be a better answer. So I'm going to move on. I feel like the vibes are off with these. So I'm gonna move on, see if there's anything else.

[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." This directly answers the question, [and] especially does it better than B, in my opinion.

And then [choice D,] "Molten salt allows a freeze-thaw battery to retain 92% of its charge after twelve weeks." This just states a property; it can't answer the question.

My final answer choice is C.

Participant RW42

RW42 shows strong notes comprehension (behavior 1) by paraphrasing the various bullet points as he moves through them ("so I guess a powerful battery"; "so this is more of a, technical details—how, what the battery is made [of]"; "so this identifies the mechanism of the battery"; "this would describe the use of the battery"). He articulates the intended relationship between the notes and the question stem (behavior 2) by indicating that the student wants the information used from the notes to specify "how [the battery] enables energy storage." This understanding allows him to rule out choices A and D on the grounds that each merely describes a "property" of the battery, not the mechanism by which the battery stores energy. He ultimately decides that choice C, the best option, "directly answers the question, [and] especially does it better than B, in my opinion" and picks C as his response (behavior 3).

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.
RW19 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.

Choice A, "reduced." "Individuals tend to be more decisive if the information flow between the regions is intensified, whereas they make, they may—whereas they make choices more slowly when the [*sic*] information flow is [*reads in choice A*] reduced." Maybe.

Choice B. "Individuals tend to be more decisive if the information flow between the regions is intensified, whereas they make choices more slowly when the information flow is [*reads in choice B*] evaluated." Eh, no.

Choice C, "determined." "Individuals tend to be more decisive if the information flow between the regions is intensified, whereas they make choices more slowly when the information flow is [*reads in choice C*] determined." Again, no, "determined" is definitely not the word for that.

> And then choice D, "acquired." "Individuals tend to be more decisive if the information flow between the regions is intensified, whereas they make choices more slowly when the information is [*reads in choice D*] acquired."
>
> For this one, I'm definitely gonna go with choice A because it, choice A, is "reduced." "Individuals tend to be more decisive if the information flow between the regions is intensified, whereas they make choices more slowly when the information flow is reduced." And that one makes the most sense.
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW19 again implements a "plug in" strategy, reading in each of the answer choices where the blank appears in the passage. She rules out choice B ("eh, no") seemingly on general "fit" grounds but rules out choice C ("'determined' is definitely not the word for that") more clearly on contextual grounds. In selecting choice A, the best answer (behavior 2), RW19 indicates a sufficient sense of passage comprehension (behavior 1) by noting that "reduced" "makes the most sense" in this sharply contrastive context.

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.
RW19 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.

Choice A, "By asserting that language plays a more important role in graphic novels than the author of Text 1 recognizes." That could be true, but let's go to other choices.

Choice B, "By acknowledging that the author of Text 1 has identified a flaw that is common to all graphic novels." Not sure about that one.

Choice C, "By suggesting that the story lines of certain graphic novels are more difficult to understand than the author of Text 1 claims." Maybe.

And choice D, "By agreeing with the author of Text 1 that most graphic novels aren't as well crafted as most literary works are." Not sure about that one.

OK. [*Rereads Text 1*] For me, that one, they're trying to, like, say [graphic novels] are definitely more related to film and not literature, as the person in Text 2 is saying they are literature. OK.

[*Rereads question stem*] Yeah, their overall argument in Text 1 is that graphic novels have more images and not text or language, so they're not, like, literature books, I guess. Not sure if I'm explaining that right.

[*Rereads choice A*] I do agree that it would probably be this one: choice A, "By asserting that language plays a more important role in graphic novels than the author of Text 1 recognizes."

Participant RW19

Participant RW19 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. RW19 exhibits adequate comprehension of both Text 1 (behavior 1) and Text 2 (behavior 2) and the basic relationship between the two (behavior 3) when summarizing the gist of each ("For me, that one [Text 1], they're trying to, like, say [graphic novels] are definitely more related to film and not literature, as the person in Text 2 is saying they are literature"). Based on this understanding, she chooses choice A, the best answer (behavior 4), as her response ("I do agree that it would probably be this one").

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.
RW19 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.

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.

> Choice A, "Lutie is observing the appearance of the street at a particular time of day and the events occurring on it." That could be it.
>
> So choice B, "Lutie is annoyed by the noise of children playing games on her street." She did not seem annoyed at all, so I'm going to cross that one off.
>
> Choice C, "Lutie is puzzled by [the] rules of certain children's games." That is also not it 'cause nowhere in this passage does she seem confused or puzzled.
>
> Choice D, "Lutie is spending time alone in her apartment because she doesn't want to interact with her neighbors." Again, that's definitely not it because she's outside, I'm sure.
>
> My choice is choice A, "Lutie is observing the appearance of the street at a particular time of day and the events occurring on it."
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. She demonstrates adequate passage comprehension (behavior 1) on at least two occasions: first, when ruling out

choice B on the grounds that Lutie "did not seem annoyed at all" and, second, when ruling out choice C because "nowhere in this passage does [Lutie] seem confused or puzzled." In blocking choice D because Lutie is "outside, I'm sure," RW19 makes a reasonable but not conclusively textually supported inference, as Lutie could instead have been observing the children from her apartment window. Having eliminated the other answer choices, she selects choice A, the best answer (behavior 2), as her response.

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.
RW19 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."

> Choice 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." Maybe.
>
> Choice 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." Not sure about that one either.
>
> Choice 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." Not sure either. I might have to read that again probably.
>
> Choice D, "Narratives that are told out of chronological order are more difficult for audiences to understand than are narratives presented chronologically."
>
> [*Rereads passage and question stem*]
>
> [*Rereads choice A*] Maybe.
>
> [*Rereads choice B*; *rereads first sentence of passage*] OK. Yeah, this question has got me a little stuck. Maybe, it's just a lot of words. [*Rereads question stem*; *rereads passage and question stem*] [*Long pause*]
>
> OK, so I think I get it a little bit more.
>
> [*Rereads choice A*] [*Briefly interrupted*] [*Rereads choice B*] I don't think it's that one because he said that both of these, like, narrative things don't have enough—like, they aren't sufficient to completely describe a narrative. Wait. "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." OK. [*Long pause*] OK. I'm kind of guessing on this one because I'm really stuck between choice A [*rereads*] and then choice B [*rereads*]. I'm going to go with choice 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."
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. Though laborious and involving much rereading (likely not aided by a brief interruption), RW19's approach to this question is successful. She demonstrates adequate passage comprehension (behavior 1) when she initially eliminates choice B ("both of these, like, narrative things [*fabula* and *syuzhet*] don't have enough—like, they aren't sufficient to

completely describe a narrative"). She admits to being somewhat unsure ("I'm kind of guessing on this one") but ultimately selects choice A, the best answer, as her response (behavior 2).

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.
RW19 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

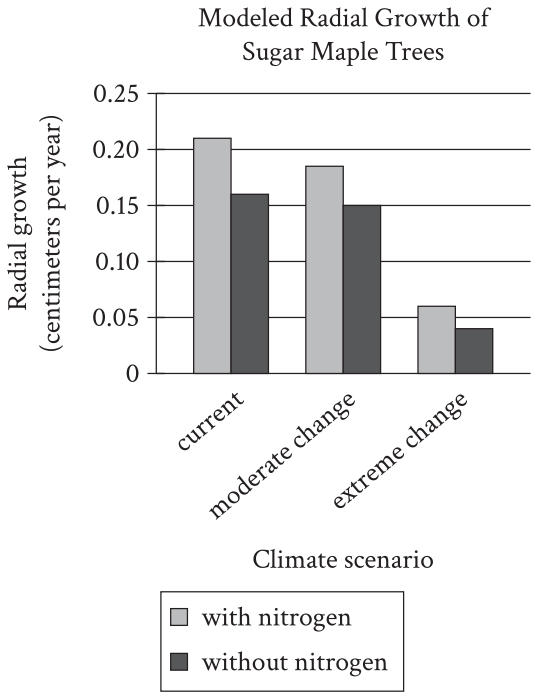
fluorescent fish found in the wild in Brazilian creeks. Given this, choice D makes the most sense here, as it describes a tangible negative consequence of such fish escaping into the wild: By passing on their trait of fluorescence via breeding, these fish may make their populations more vulnerable to predators.

> Choice 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." Not really something negative.
>
> So choice 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." That could possibly be it.
>
> Choice 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." OK.
>
> Choice D, "Through interbreeding, transgenic fish might introduce the trait of fluorescence into wild fish populations, making those populations more vulnerable to predators."
>
> I do believe it's choice D because here [in the question stem] it says "A student in a biology seminar who is writing a paper on these fish asserts that their escape from the Brazilian fish farms into the wild may have significant negative long-term ecological effects," and choice D [says], "Through interbreeding, transgenic fish might introduce the trait of fluorescence into wild fish populations, making those populations more vulnerable to predators." And that is definitely a negative long-term ecological effect.
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated all required behaviors, resulting in a PL of 1. Early on in her verbalization, RW19 demonstrates an understanding of both the passage (behavior 1) and the criterion linking the passage and question stem (behavior 2), noting that choice A is incorrect because its information is "not really something negative," whereas the question stem asks for evidence that transgenic fish's escape from Brazilian fish farms may have had major negative long-term ecological effects. She determines that choice D "is definitely a negative long-term ecological effect" and properly selects that option as her response (behavior 3).

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.
RW19 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.

> [*Rereads passage and question stem*] I need to read that [part of the passage] again. "Although they found that climate change would negatively affect growth, they concluded that anthropogenic nitrogen deposition could more than offset the effect provided that change is moderate rather than extreme." OK.
>
> So choice 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." No, I don't think that one's it.
>
> Choice B, "Growth without nitrogen under the current climate exceeded growth without nitrogen under moderate change—" OK. [*Consults figure*] It definitely surpassed that.

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

Choice D, "Growth with nitrogen under moderate change exceeded growth without nitrogen under the current climate—" Wait. OK. What's nitrogen under moderate? [*Consults figure*] "—exceeded growth without nitrogen [under the current climate]—" [*Consults figure*] It did. "—but the latter exceeded growth with nitrogen under extreme change." True. OK. I think I get what it is asking me now, actually.

So choice C, "Growth with nitrogen under moderate change—" [*Consults figure*] With nitrogen, moderate, with, without—wait. OK. "Growth with nitrogen under moderate change exceeded growth without nitrogen under moderate [change]—" [*Consults figure*] Yeah, it did exceed it. "—but the latter exceeded growth without nitrogen under extreme change." [*Consults figure*] True.

Choice B, "Growth without nitrogen under the current climate—" [*Consults figure*] Growth without, current. "—exceeded growth without nitrogen under moderate change, but the latter exceeded growth with nitrogen under extreme change." [*Consults figure*] True.

Choice A, "Growth with nitrogen under the current—" [*Consults figure*] With, OK. "—current climate exceed growth with nitrogen under moderate change—" [*Consults figure*] Yes, that is true. "—but the latter exceeded growth without nitrogen under extreme change." [*Consults figure*] No[?]. OK.

Choice B, by the way: "Growth without nitrogen under the current climate exceeded growth without nitrogen—" [*Consults figure*] Wait, current climate, without. Yeah. "—but the latter exceeded growth with nitrogen under extreme change." [*Consults figure*] Yes. OK.

"Although they found that climate change would negatively affect growth, they concluded that anthropogenic nitrogen deposition can more than offset that effect provided that change is moderate rather than extreme." OK.

Choice C, [*consults figure*] "Growth with the [*sic*] nitrogen under moderate change exceeded growth without nitrogen under moderate change, but the latter exceeded growth—" OK.

Yeah, I'm gonna go with choice C, "Growth with nitrogen under moderate change exceeded growth without nitrogen under moderate change, but the latter exceeded growth without nitrogen under extreme change."

Participant RW19

Participant RW19 answered the question incorrectly but did demonstrate one required behavior, resulting in a PL of 4. RW19 clearly uses data in the graph (behavior 2) in an attempt to vet the accuracy of the comparisons set up in the individual answer choices, but otherwise she doesn't exhibit a clear grasp of the passage's content or the relationship between passage, graph, and question stem. Each of the pairs of comparisons in the individual answer choices is accurate per the graph, but only choice D, the best answer, is responsive to the question's

criterion, which is to provide evidence from the graph for the passage's conclusion that "although [the researchers] 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." She ultimately incorrectly selects choice C as her response. Nonetheless, RW19 shows partial enactment of the question type's construct.

Supplementary Vignette: Participant RW18

Participant RW18 answered question 14 correctly and demonstrated all required behaviors, resulting in a PL of 1. RW18 was one of eleven participants who answered the question correctly and one of only four participants who did so while also demonstrating all required behaviors.

> [*Refers to graph*] "With nitrogen," "without nitrogen," "current with nitrogen"—it only goes down, like, this much, so I'll probably use that as my data, the difference between the "current with nitrogen" and the "moderate change with nitrogen."
>
> [Choice 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." No, that doesn't make sense. [*Refers to graph*] That doesn't make sense because the, like, latter exceeded growth without the nitrogen. Like, it just doesn't— ". . . the latter exceeded growth without nitrogen under extreme change." Yeah, but that doesn't really matter if that's not showing the fact that there's any, like, change, any offset provided with the nitrogen. It's just saying that the more extreme it gets, the worse it gets, which actually does the opposite 'cause it's showing that [*refers to graph*] if moderate change is better than extreme change, even with nitrogen, it wouldn't support the conclusion.
>
> [Choice 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." No, [*refers to graph*] 'cause they're comparing growth without nitrogen to growth with nitrogen in different subsets. It wouldn't make sense to compare one with nitrogen and current and then without nitrogen and extreme because that just—or vice versa—because it wouldn't properly identify the effects of the nitrogen. So that can't be the answer.
>
> [Choice] C, "Growth with nitrogen under moderate change exceeded growth without nitrogen under moderate change, but the latter exceeded growth without nitrogen under extreme change." [*Refers to graph*] "Growth with nitrogen under moderate change"— right here—"exceeded growth without nitrogen under moderate change—" [*Refers to graph*] That's correct. "—but the latter exceeded growth without nitrogen under extreme change." "The latter exceeded growth without nitrogen—" What is "the latter"? [*Refers to graph*] Oh, "moderate change" and "extreme change." I'm kind of confused in the wording of this answer, so I'm gonna skip it for now and see if [choice] D makes sense.
>
> [Choice D,] "Growth with nitrogen under moderate change exceeded growth without nitrogen under the current climate—" [*Refers to graph*]

> "Under moderate change." That could be a viable suggestion. ". . . but the latter exceeded growth with nitrogen under extreme change." [*Refers to graph*] "The latter exceeded growth with nitrogen under extreme change." OK, so I'm a bit confused with all these wordings of "the latter." I'm trying to figure out what "the latter" necessarily refers to.
>
> [*Rereads portion of passage*] ". . . nitrogen deposition could be [*sic*] more than offset [that effect] provided the change is moderate rather than extreme." So suggesting good about moderate and bad about extreme. "With nitrogen." I think [choice] D could be the answer 'cause it talks about, [the passage] says "more than offset" and [choice D] says "growth with nitrogen under moderate change"—[*refers to graph*] which is right here—"exceed[ed] growth without nitrogen under the current climate change [*sic*]," which shows that with added nitrogen in a moderate change, it more than offsets the effects because the growth is actually exceeding. ". . . but the latter exceeded growth with nitrogen under extreme change." [*Refers to graph*] OK, OK. I think it's saying that this is, like, the moderate change without nitrogen exceeded the extreme change with nitrogen, which, if that is the case, it would support that if it's extreme, it doesn't offset. So I would go with answer D.
>
> *Participant RW18*

RW18 exhibits both perseverance and strong problem-solving skills as she works through the question. She demonstrates adequate passage comprehension (behavior 1) and a clear understanding of the passage-graph-stem relationship (behavior 3) by noting that the best answer must show "the fact that there's any, like, change, any offset provided with the nitrogen." By this, she refers to the central claim of the passage: "Anthropogenic nitrogen deposition could more than offset [the effect of climate change] provided that change is moderate rather than extreme." Early on in her response, RW18 recognizes that the best answer must make the following comparisons in order to support the passage's claim: first, it must show that growth with nitrogen under moderate change exceeded growth without nitrogen under the current climate (thereby showing that nitrogen deposition could "more than offset" the effect of moderate climate change), and, second, it must show that growth without nitrogen under the current climate exceeded growth with nitrogen under extreme change (thereby showing that nitrogen deposition couldn't "more than offset" the effect of extreme climate change). Pulling on this understanding and making frequent use of data in the graph to draw comparisons between and among conditions (behavior 2), RW18 selects choice D, the best answer, as her response (behavior 4). She seems to have gotten a bit turned around while working through the second half of choice D, referring to growth without nitrogen under moderate change exceeding growth with nitrogen under extreme change (whereas the proper term of comparison for growth with nitrogen under extreme change is growth without nitrogen under the current climate), but this seems likely to have been a verbal slip, suggested by her repeated mentions of frustration with the answer choices' use of "the latter," and doesn't affect her best-answer selection.

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.
RW19 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.

> Choice A, "the English dialects spoken in the Upland South acquired their emphasis on the 'r' sound from dialects spoken in Scotland." Maybe, but let's read the others.
>
> Choice B, "emphasis on the 'r' sound will eventually spread from English dialects spoken in the Upland South to dialects spoken elsewhere." I do not think that's it because it's not really talking about the future or anything.

> Choice C, "the English dialects spoken in Scotland were influenced by dialects spoken in the Upland South." Maybe.
>
> And then choice D, "people from Scotland abandoned their emphasis on the 'r' sound after relocating to the Upland South." No.
>
> So, OK. [*Rereads passage*] ". . . Thus, linguists have concluded that—" [Choice] A. [*Rereads A*] [*Rereads passage*] ". . . Thus, linguists have concluded that—" C, "The English dialects spoken in Scotland were influenced by dialects spoken—" Wait, that's, like, completely opposite of what A said. I just noticed that.
>
> So I'm gonna actually go with choice A, "the English dialects spoken in the Upland South acquired their emphasis on the 'r' sound from dialects spoken in Scotland."
>
> *Participant RW19*

Participant RW19 answered the question correctly and demonstrated both required behaviors, resulting in a PL of 1. RW19 demonstrates adequate passage comprehension (behavior 1) most clearly in ruling out choice B ("I do not think that's it because [the passage is] not really talking about the future or anything"). She also implicitly exhibits passage comprehension when ruling out choice C apparently on the (correct) basis that the option reverses the sequence of events described in the passage, a fact she misses on an earlier reading ("I just noticed that"). Recognizing that choice A, the best answer, logically completes the conclusion asserted by the linguists referred to in the passage, RW19 selects that as her 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?

Based on their responses to postexperience question 1, participants had varied but typically neutral or mixed reactions to the think-aloud activity. Some participants did report a generally good, satisfying experience, while one characterized the experience as "frustrating." Interestingly, several participants noted that the act of reading passages and questions aloud enhanced their concentration and focus over reading silently, although one participant observed finding it "distracting." One participant compared the digital-adaptive version of the SAT favorably to the now-retired paper version, saying that the shorter passages of the former cut down on distractions. Modally, participants were neutral in sentiment toward the activity and tended to mention that some questions were easy for them while others were more difficult.

But, in overall, on my experience, I think that I've found a lot of ease reading through. Even the charts were easy to navigate. . . . But I think in all, everything, I think that it was pretty easy for me. I think this is as easy as the SAT can be. *RW2*

It was kind of different because I've never really answered questions out loud like that before. It was pretty nice because it helped me think more. *RW3*

Like, a lot of them, it was kind of mixed. Like, I think some of them I could get down pretty soon, like pretty early on, and then just, OK, my answer that I originally thought it would be. And then some, especially the *Godfather* one [Reading and Writing question 12], those took a while. I think just—especially it's using a lot of—it's a little bit using more complex language that could kind of trip you up a lot. *RW6*

It wasn't anything super crazy. It was just, like, a standardized test. In terms of difficulty, I don't think they were too hard. Towards the end, though, there were a lot of different types of questions, like ones about transition words or the meaning of a phrase. Those took me a while, especially because the way the answers were presented was confusing. They were all very similar in terms of wording, so it kept getting mixed up in my head. *RW15*

It was, let me see, frustrating. *RW17*

It was actually not that hard. I was a little bit stuck on some of them because I didn't understand the meaning of some words or sometimes couldn't pronounce some of the names they used in those. But other than that, I think I felt pretty good about answering them. *RW19*

It was a bit—for me, it was a bit more distracting to answer them out loud because I'm mostly thinking in my head. So, like, I'll speak to myself, but knowing I'm talking to someone else on the other side kind of distracted me a bit. I knew what to say in my mind, but I couldn't pronounce it correctly, stuff like that. So it was a bit more—like, the questions were pretty easy, but answering and talking in person to someone else is a bit harder. *RW20*

I thought that—I don't know. Usually, for me, the SAT has just never been, like, my strongest point. So I was struggling a little bit, but I thought they weren't too bad just because, like, the texts are a lot shorter than the paper version. So that helped me a little bit with the distractions. But, yeah, other than that, it was fine. *RW27*

I thought most of them were easy to follow. Some of them had a whole bunch of information and, like, words that I don't know. And I feel like with the amount of words that I didn't know on there, I would be wasting a lot of time in a dictionary. But most of the time, I thought they were easy to follow; some of them were just really wordy, and some answer choices were very similar to others. *RW37*

> I would say most of them weren't too hard. Like, it seemed like they were really easier—easier to moderate questions. Some of them were time-consuming, maybe, especially since I was reading them out loud, but not necessarily difficult. *RW45*

Strategies

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

In response to postexperience question 2, participants offered a range of strategies they used to answer the study's questions and to take tests in general. Among the most commonly mentioned strategies were reading the question before reading the passage (which participants were constrained from doing per the think-aloud protocol), rereading, highlighting (using Bluebook's annotation tool), coming up with one's own answer before looking at answer choices, drawing on content knowledge (noted in reference to vocabulary and transitions questions), eliminating incorrect answer choices (including using Bluebook's answer elimination tool), mentally "plugging in" answer options in blank-completion questions, and relying on a sense of "fit" between one's answer choice and the stimulus material.

> Well, I go by adjectives. So if it tells me how something, someone is very upset, I'll look for the most negative-analysis answer on that. Or if, you know, they're looking for the biggest sum, I'll look for the biggest amount of something if it gave me a chart. But that's, like, generally my thing. Like, if they give me an adjective, I'll kind of go into there and make my assumption and then, like, I'll put, and then there was another one. I think in general, though, it just takes me like a little bit longer. Like, I just have to, like, read through word for word. Like, I have to read, like, twenty-three times to try and get it through and, like, actually answer the question. I also look for, like, what could be, like, the best answer because, you know, it's always, like, the best response. There could be good responses there, but I want to look for the ones, like, closest to the thing. So I try to look for the ones that closely match up with the question, closely match up to the text, and yeah. *RW2*

> So I guess what I typically do is I'd read through it. But I think that I like reading the question before [the passage]; sometimes [that helps] a lot so I could pick up more clues in the text, especially if it's a longer one. You can, you don't necessarily retain all the information, but that—well, if it's longer, but if you have—if you're looking for something specifically from the question, as you're reading it, you can kind of pick up on it a little bit better. . . . And then I'd also, I would just read the graphs after. I mean, reading the graphs after the text is more—I'd say it's helpful because if you're reading the graph before, it can maybe confuse you a little bit. . . . I guess just making also just whatever kind of sounds right. Not necessarily, like, what rolls off the tongue better but just saying, like, with those blank questions, just reading it all together, you can kind of see what matches a little bit better. . . . Saying it out loud definitely helps. *RW6*

Well, for the ones where you just have to fill in the blank for one word, it's just know the definition of that word and know how that word is connecting the sentence before that word and the sentence after that word, and kind of just see which word helps it connect to it. The other one, the other strategy I have is summarizing for the answers—for the questions with long answers—is basically summarizing that first paragraph and then answering the second—answering the second paragraph. And it was right—summarizing the first paragraph and then summarizing the answers and seeing which ones are more synonymous. The most synonymous answer is the one that I click on. *RW7*

. . . [T]he way I usually answer is, I'll read one [answer choice], and I'll see if it contradicts; if it doesn't, it doesn't. If I see that's the one for sure, I'll just, I'll skip through the other ones just to make sure. *RW9*

I think I would just first think of what would make sense in my head beforehand, kind of like with the question of the saber-toothed tiger [Reading and Writing question 4]. That way, if I see it, I know it's there. And then if something doesn't make sense, just go back and read until it starts to make sense. Like, the one with the climate change [Reading and Writing question 14]. I was very confused, but then when I went back and read, I noticed that it also talked about extreme [climate change] not being worth it. Then that makes sense. I think with lots of them too, it's just which answer seems the strongest or which makes the most sense. For example, with some of the word replacements, I know if it says "for instance," there's an example, which means that if there's not an example [in the passage], then something like that can't be, like, make sense. So just kind of thinking logically in that way. *RW18*

When it came to fill-in-the-blanks, I would fill it in and then read the entire thing. If it didn't make any sense, then I would just be like, "No, that's not the right one." I would look at the text once, like, read the passage, read the questions, then read the passage again because then I'll see the question in the passage and be like, "Oh, that one makes more sense, that one." And then afterwards, after I finish answering the questions, I'll read the passage again just in case. *RW20*

Well, I, at first I kind of see if I even understand what it's talking about, try to see maybe if I have any insight about it that I already know. But if they give me, like, certain things and they tell me certain things—like, key clues that says in some of the answers—I could put them together and kind of give me, like, a context clue of this would make the most sense. And this is what was actually said in the text. So just kind of just, like, reading it and make sure I'm understanding what I'm reading because if I don't understand what I'm reading, it's very hard for me to answer. I'll just have to guess. *RW26*

I like to highlight and annotate. It's a bit hard to use the text typing annotation in Bluebook. I also like crossing out answers, as it helps me visually narrow them down. I generally take a lot of the information, figure out which answers seem right or wrong, and make an informed

guess, narrowing it down to at least two options. I was able to narrow it down to two or one choice in most cases today. *RW42*

So I would say, if it's asking about the content—like, the main idea, for example, or, like, "How would the author of Text 2 respond to the claim in Text 1?"—for those I would actually read the whole thing because I need to know what they're saying. But, for the notes ones [Reading and Writing questions 7 and 8], where they're talking about notes, I feel like usually the question is asking for a certain way to oppose the given information, or describe it. And so, usually, the information in it is true and the answers are true. So I know I just need to find which one is doing what the question is asking. So I usually don't have to look at the notes. . . . And then, like I said, for grammar or transition words, I usually don't read the whole thing unless it's short. If I read the question [and] I know I only need to fill in the blank or something like that, I only read part of it. *RW45*

"Easy" Question Types

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

In response to postexperience question 3, participants consistently mentioned that they found questions with short passages and questions in the blank-completion format to be (at least relatively) easy for them. These factors seemed to carry across skill/knowledge testing points such that it didn't so much matter whether the question was about vocabulary, transitions, or Standard English conventions, only that the passages and questions were brief and had a blank to fill in. After those, the most common response was that no questions or question types were particularly easy, whether because participants found them all equally easy, all equally difficult, or all of average difficulty. The Rhetorical Synthesis question type [Reading and Writing questions 7 and 8] were cited by two participants, in part because they believed it was unnecessary to read the bulleted-list notes given their belief that the questions could be answered solely from the question stem and answer choices. It's important to note here that while questions 7 and 8 are susceptible to this approach, other Rhetorical Synthesis questions in the SAT Suite pool include one or more answer choices that misstate the information in the notes, making those participants' approach riskier in practice.

I think word replacements because especially if you just read it, if it sounds like it makes sense, it probably makes sense. If it sounds like it doesn't make sense, well, then that's probably not the right answer. So I'd say those are probably the easiest. *RW1*

The grammar-based ones, the ones that will tell you, like, "What is the perfect word to put in the sentence or fill in the blank?" Those are the ones that I find the most easy, and that's because they're just super short in length and they're just really, like, easy to see, like, visually. *RW2*

> The easy ones are always the ones that are—it's, like, the one with the student notes [Reading and Writing questions 7 and 8] and you just have to kind of, like, summarize that. Those are pretty easy to me just because I think, like, the answer choice is pretty much in front of you, and they ask for a very specific thing most of the time. *RW27*

> The questions where there were student notes and you had to synthesize a claim were easier for me. You don't even have to look at the text [notes] to get the answer right, though I did to double-check. In a real SAT situation, I might not even look at the student notes and just choose the answer that fits the question. *RW42*

"Hard" Question Types

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

Participants' responses to postexperience question 4 were similarly consistent, as the most common factors participants associated with question difficulty were longer passages and the inclusion of tables and graphs, both of which add to task complexity and information load. Not surprisingly, Reading and Writing question 14, which focuses on the radial growth of sugar maple tress under various climate change scenarios and with and without the introduction of nitrogen fertilizer, was cited by numerous participants as the hardest in the activity, as it included a fairly complex graph, a relatively long passage, and answer choices that were phrased similarly to one another. Several participants, by contrast, couldn't come up with a "hard" question type.

> I think graph questions are the hardest to answer just because there's so much information that it's hard to know where you're supposed to look. And then oftentimes, like, it's very science heavy, and there's a lot of complicated terms and conclusions. I'd say those are probably the hardest, especially coming to the information in the graphs, because it all looks so similar but it's all just as important. I feel like it's easy to get lost in a graph or the paragraph under, or easy to forget what happened in the paragraph under, while you're reading the graph or vice versa. *RW1*

> I think there, you know, the *Godfather* one [Reading and Writing question 12] and then also the one in Mexico—the Mexico City transit one [Reading and Writing question 3]. I think when it's just talking about a lot, it's harder to kind of really zone in on what's the important parts of the text, especially when it's using complex language. *RW6*

> Yeah, the nitrogen one [Reading and Writing question 14]. I don't really know much about, like, nitrogen and stuff, and it's, like, boring. And I really didn't know what was going on. Plus, it was already, like—we were already, like—how long? And it was like, what's the duration of this [session]? It was already some time. So I was kind of already kind of—it wasn't—it was, like, "the wrong thing at the wrong time" kind of thing.

It was too far in [the question sequence]. I feel like they should do the harder questions first, when your mind is fresh, instead of, like, kind of wearing out your mind with a bunch of easy questions. They just do—should do, like, the few hard questions first. *RW7* (Note that, unlike Math questions, Reading and Writing questions aren't ordered from easiest to hardest but rather follow a fixed order in which questions of the same types are presented together.)

Definitely the ones with a lot of words. Sometimes when I'm reading, my eyes or whatever will get mixed up in all the words, and I'll have to go back to the start and try to read it again. Also, when speaking out loud, some of the names and words were hard to pronounce, but that's just it. *RW19*

I think transition questions are always hard for me, just because sometimes I don't even know what the transition word means, which I should study or I guess that's typical. The ones that give you tables are probably harder for me. I feel like all of the answer choices are really similar. They just kind of rearrange the wording a little bit to make it more confusing. I think just being able to narrow down the answer is harder just because the questions—the answer choices are similar. *RW24*

When they had especially, like, I would say the graph plus the long paragraph [Reading and Writing question 14], that does not help—like, a paragraph long. OK, fine. But then the graph too, then you kind of have to, like, look inside of the paragraph to understand what the graph is talking about. It's kind of confusing as well that, so I would say that's probably the hardest one. *RW26*

Yeah, like the ones with charts, because they're always, like—it's because sometimes I feel like they're asking one thing, and then they give you a chart that shows a different thing, and I get kind of distracted because I'm like, "Which one is it?" And then sometimes the ones with, like, the—what is it, like, author two, where they have, like, two passages, and then they ask you, like, "And what would the other one say to this author?" [Reading and Writing question 10]. And those were kind of confusing as well. *RW27*

I think the one that I spent a really long time on with the graph [Reading and Writing question 14]. I think that one was pretty hard just because the wording and the answer choices was like, it is really, like, kind of repetitive and also the way that they said—in the answer choices, it would give you, like, it would say, like, "the moderate change was better than the, like, extreme change." And then it would say, like, "but the latter was blank, like, better than this." I thought, like, "the latter" part was confusing just because, like, I couldn't, like, process that without, like, trying to, like, put it together. And I had to, like, reread it a whole ton of times, and it was like all the answer choices were, like, kind of repetitive. So I had to, like, read them so many times to understand, like, what they were all really saying. Yeah, I just think that one was kind of confusing overall. And then also the one about the *Godfather* [Reading and Writing

> question 12], where it was talking about the literary elements, just because I think that the first two [answer choices] were, or the—I think I was between A and B, and I think they were, they both, like, fit pretty well. *RW44*

ADHD Symptom Impact

> 5. Did you encounter anything in the questions that you had difficulty with given that you have ADHD? If so, what was it, and why was it difficult for you?

In response to postexperience question 5, participants identified a range of ADHD-related impacts on their test-taking ability. Most frequently mentioned were difficulties with longer texts; longer and similarly worded answer choices; challenging, unfamiliar vocabulary, including scientific names for species and other terms in languages other than English; graphs; passage pairs; and general speed and accuracy of text processing. The common thread tying most of these comments together was the struggle to sustain focus and persist through tasks, an unsurprising outcome given that all participants had ADHD. Small font size was mentioned by two participants, a reflection of the fact that not everyone is aware that the ability to zoom in is a universal tool in Bluebook.

> I'm not entirely sure because it's hard to know what's just hard and what's hard specifically for me because of that. I think it's also different too because reading out loud does help keep the line of thought, and you have to focus. So I feel like it's hard to tell if there were specific things that gave me trouble because of that. I mean, maybe the whole graph thing being too much information is connected to that and maybe not. I'm not entirely sure because, again, I don't know what's simply difficult and what's difficult because of that. *RW1*

> I think sometimes when I read, I'll just kind of—I know you could probably pick up on a little bit when I was reading out loud. I'd kind of fill in words and say them differently. . . . I think for the most part, I can still—like, when I'm reading them in my head, I get through all the, I get all the words down and everything, but I think sometimes just, especially not thinking I need to be more inclusive into what I'm thinking about . . . *RW6*

> So the very verbose ones—ones with a lot of words. My brain saw a lot of words, and it was, like, I want to think about something else. Those—like, when my speech was a little slurred, you could see that's when my ADHD was kicking in. *RW7*

> I think that one of them, that same one, the graph one [Reading and Writing question 14]. I just wanted to stop. It was getting in my mind—like, trying to get up and take a break. But, like, it was just annoying. It was just annoying. Like, when I don't get something, it's just like, "Oh, I have to read it again," like the point of ADHD—like, you want to go, like, get it done. So, like, you want to just get up, like, do it, like, you don't

want to wait. So it was annoying, like, reading it all. Like, if you keep reading it and reading it. Also, all the reading was just a lot. *RW9*

Mainly the ones that had different languages, like different types of names [e.g., Reading and Writing question 4]. And then the questions that had really long answers were just kind of confusing to me, especially the one that was about the climate [Reading and Writing question 14]. That was the one that I had the most problem with. *RW12*

In general, I have trouble holding on to what I just read. I often have to go back and reread because I read the words but don't always process what they're saying. For questions about which word fits best, it's not too hard because I can just go back and check. But for questions about the main point of a passage, I have to reread a few times to fully understand it. *RW15*

Kind of like having those two texts back-to-back [Reading and Writing question 10]. Like, again, I'm reading it and then getting the texts mixed up. *RW20*

Usually, it's just like them giving me similar words that kind of throw me off in the answer choices. So, like, I don't know. They would put, like, "conversely" and "in contrast" next to each other. Like, I think I remember seeing that in one of my practice tests at one point. And then usually, like, anything with a lot of text is just kind of difficult for me because I don't know, like, where to look, and then I kind of get lost in the passage. It's a little distracting. *RW27*

Given my ADHD, well, one of the things is that I do like to break the questions down. And like I said, if I'm in a time crunch, I might not always have the opportunity to do that. So, definitely, especially, like, when taking the actual SAT, time-and-a-half or more accommodations are especially helpful. *RW42*

I think, like, my main issue is that, like, a lot of times, I will just read the text and, like, not really, like, fully read it—like, I'll just be, like, looking at the words, like, and I'll be, like, going through it. Like, I'm actually like, remember—like, I could read the whole thing, but then if someone, like, asked me what it was about, I could not remember a single detail. So I think, like, actively reading is, like, my main issue. Then besides that, I think that's, like, mainly it, and then I, like, I kind of take a really long time on every question because of that. *RW44*

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?

Most participants didn't substantively respond to postexperience question 6, which solicited final, general responses to the think-aloud activity. Those who did answer tended to bring up themes discussed previously, including the need to find ways to maintain focus and persist during testing. One participant cited the benefits of reading aloud while another indicated that the protocol's requirement for reading the passage first was not something they'd do going forward, suggesting some methodological reactivity.

> Well, like I said, I do think that there should be an option for you to space out the text—or there's this font that's actually been added, I think back in 2022 or maybe more recent, that it capitalizes each word before—like, each word, like, the letter of the first letter of each word, and then it kind of shows you more clearly what things mean. So I'm—I don't know. That would help [as] kind of like an accommodation. But, overall, I think that all the tools are already there, like, really. *RW2*

> I feel like the experience was definitely different in terms of me thinking out loud. I guess I was able to, like, hear myself and how I was determining to get to the right answer. I feel like towards the end, I was talking so much, I was, like, OK, I'll read this giant paragraph one more time. But I think it's a little helpful to read out loud just because in a real test, you can't do that because it'll distract other people, of course. *RW24*

> I don't know. Maybe I'm gonna try reading the questions first next time because I did find that much easier to be able to, you know, identify, like, what I need to pick out in the text. *RW38*

> I think, I mean, I guess, like, I said, with those student notes questions [Reading and Writing questions 7 and 8], I generally don't even read the notes. *RW42*

> Like, I think I tend to, on the one with climate change [Reading and Writing question 14]. Those, I usually don't struggle with the English questions too much. Or, like, I know in the old SAT when it was reading and English, where one was passages and then there was grammar and vocab. I found in the English reading part, I'd get closer to running out of time, even with my extended time. I think the first time I took it, I either didn't get to the last few, or it was super close—I don't remember. And I know I struggled with the same thing on the ACT. I've taken it a few times, and I know I run over on the reading, whereas in the English section, I'm fine. But on those longer ones that are really difficult to understand, even though I think I can figure it out because I understand it, I end up getting stuck on them and don't realize how much time has gone by. And then, after all that time, I might get the answer—like, most of the time I do, given my scores and everything—but it either leaves

> me so close to running out of time at the end or it affects my ability to handle the last few questions, even though they might not be as hard. If I had skipped it, I probably would have done way better. . . . I think part of it is ADHD, because in other areas of my life, I get time blind. I'm, like, "Whoa, all that time went by," because I get so focused on that one question, trying to figure it out. I struggle with that, especially in math, where I'm close to solving the question, but it's so lengthy or has so many steps. I'll end up trying to finish it instead of skipping and coming back. That's what happened with the climate change question—it was difficult for me. . . . Especially if they're more in the middle of your exam rather than toward the end—it can prevent you . . . I would rather have all those lengthy questions at the end so that I can get through the less time-consuming ones first and then tackle the harder ones at the end. *RW45*

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 # 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M1	5	5	5	5	5	3	5	5	1	1	–	–	–	–	–
M2	1	5	5	3	4	5	1	4	1	5	5	5	1	5	–
M3	1	1	5	1	4	5	1	4	1	1	1	5	5	5	5
M5	1	1	3	3	1	3	1	1	1	1	1	3	5	–	–
M6	3	5	5	5	5	5	1	5	5	3	5	5	5	5	–
M7	5	5	3	5	5	5	5	5	1	3	3	1	5	5	5
M8	5	5	5	5	5	5	5	5	4	5	5	5	–	–	–
M12	1	1	5	5	4	5	1	1	1	1	1	1	4	5	1
M14	1	1	5	5	4	5	1	4	1	1	1	1	1	5	1
M15	1	1	1	1	4	5	1	1	1	1	1	1	1	1	1
M17	5	1	5	1	1	5	1	3	4	5	1	5	5	5	5
M18	1	1	3	3	3	5	1	1	1	1	1	1	1	1	1
M19	5	3	3	1	5	3	1	1	1	5	5	5	5	5	5
M23	1	1	1	5	5	5	1	1	1	1	1	1	5	4	4
M30	1	1	1	5	4	4	1	1	1	1	1	1	1	1	1
M31	1	3	5	5	5	5	5	5	1	5	4	5	5	5	5
M37	1	1	5	1	1	5	1	1	1	1	1	1	1	1	5
M42	3	3	3	5	3	5	1	5	1	5	3	5	5	5	4
M43	1	1	4	5	1	5	1	1	1	1	1	1	1	5	1
M59	1	5	1	1	4	4	1	1	1	1	1	1	4	5	4
M61	1	1	1	1	4	5	1	1	1	1	1	1	1	5	1

Performance by Level, by Participant

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

Participant Performance Summary

#AC	*#EB*	D_p
3	2	1 ✘
5	4	1 ✔
7	7	0 ✔
12	8	4 ✘
3	1	2 ✘
5	2	3 ✘
0	0	–
9	9	0 ✔
9	9	0 ✔
13	13	0 ✔
6	5	1 ✔
14	11	3 ✔
7	4	3 ✘
9	9	0 ✔
12	12	0 ✔
3	2	1 ✘
12	12	0 ✔
7	2	5 ✘
11	11	0 ✔
9	9	0 ✔
12	12	0 ✔

Performance by Level, by Question

1	14	12	5	7	4	0	17	11	18	13	13	11	8	4	7
2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
3	2	3	5	3	2	3	0	1	0	2	2	1	0	0	0
4	0	0	1	0	8	2	0	3	2	0	1	0	2	1	3
5	5	6	10	11	7	16	4	6	1	6	4	8	9	13	6
NR	0	0	0	0	0	0	0	0	0	0	1	1	2	3	5

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	16	15	10	10	6	3	17	12	18	15	15	12	8	4	7
#EB	14	12	5	7	4	0	17	11	18	13	13	11	8	4	7
D_q	2 ✔	3 ✔	5 ✘	3 ✔	2 ✘	3 ✘	0 ✔	1 ✔	0 ✔	2 ✔	2 ✔	1 ✔	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, thirteen of twenty-one participants (62 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. Seven participants had a criterion-failing differential, while an eighth had no "true" differential because she answered no questions correctly. With the exception of participant M5, who answered twelve questions correctly, the participants with criterion-failing differentials (or no differential) were among the lowest-performing students on this activity, as judged by raw question-answering success, as they answered zero (one participant), three (three participants), five (one participant), or seven questions correctly (two participants); by contrast, participants with criterion-passing differentials answered as few as five to seven questions correctly in only three cases. The seven participants who failed the criterion (excluding the participant with no true differential) had differentials ranging from 1 to 5 and were able to demonstrate cognitively complex thinking on roughly a quarter to two-thirds of the (typically 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, twelve of the fifteen studied Math questions (80 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 students with ADHD. The remaining questions had differentials of 2, 3, and 5, and all were answered correctly by a minimum of three participants (and as many as ten) who also demonstrated at least one expected behavior, suggesting that these questions, too, were capable of eliciting cognitively complex thinking and that these higher-than-desirable differentials may be attributable in part to some participants' relative lack of think-aloud verbalization skill or experience. While the three criterion-failing questions assessed differing testing points, they were all hard (PSB 6 or 7) multiple-choice questions outside of context.

PARTICIPANT PERFORMANCE VIGNETTES

Case Study: Participant M43

Participant M43 was selected as the Math case study participant using the same criteria as outlined for the Reading and Writing case study. M43, a female twelfth grader from Georgia, identified as White and not of Hispanic, Latino, or Spanish origin. She self-reported a HSGPA of A, indicated that she'd received or she expected to receive extra time and extra breaks accommodations as part of SAT Suite testing, and described her ADHD symptoms as moderate. M43 answered eleven of the fifteen Math questions correctly and demonstrated at least one expected behavior in every case, resulting in a participant differential of 0 (100 percent), which exceeded the criterion 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.
M43 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.

> First off, I know that [choices] A and B are incorrect because we need to take into account both of the numbers [0.6 and 1.8], and A is just adding them, and B is only counting one of them. So A and B are out for me. [Choice] C is not in the correct range, which leads me to [choice] D, which states that the snowfall [rate] is greater than or equal to 0.6 inches [per hour], which is the minimum [rate], and less than or equal to 1.8 inches [per hour], which is the maximum [rate].
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. She quickly dismisses choices A and B due to their lack of including all necessary values needed to represent the situation. She also dismisses choice C by confirming it "is not in the correct range [of values]" to represent the situation. By demonstrating comprehension of the context (behavior 1), M43 is able to correctly identify the linear inequality described in the context (behavior 2) as choice D, "which states that the snowfall [rate] is greater than or equal to 0.6 inches [per hour], which is the minimum [rate], and less than or equal to 1.8 inches [per hour], which is the maximum [rate]."

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.
M43 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 $\frac{1\ coach}{26\ athletes} = \frac{x\ coaches}{y\ athletes}$, resulting in $y = 26x$, where y represents the number of athletes at the track meet.

> So, the first thing that comes to mind is this is a ratio of 1 to 26, so it's going to be x to $26x$.
>
> The number of athletes would be $26x$.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated a single expected behavior, resulting in a PL of 1. After reading and comprehending the context (behavior 1), she immediately concludes that the correct answer would be $26x$ (choice B): Per the context, the ratio of coaches to athletes is "1 to 26": given x coaches, the ratio is "going to be x to $26x$," making the number of athletes $26x$.

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.
M43 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.

> Obviously, this is a circle that we're taking into account. I believe we need to go ahead and find the radius. 121 is the radius squared, I believe, so the radius would be 11. I think it's possible. The radius is 11. Circles are not nice. I believe what the 4 and 19 are representing is the height and width, or this is the center. I got to 11 is the radius. I don't think it's [choice] A; I don't really know where A is coming from. I don't think it's

> [choice] B; I don't think it should be negative. So I'm kind of thinking between 11 [choice C] and 19 [choice D]. I'm going to go with 19 because 11 is the radius, so I don't believe that'd be part of the center, which I think is what it's looking for.
>
> *Participant M43*

Participant M43 answered the question incorrectly but demonstrated one expected behavior, resulting in a PL of 4. Knowing that the graph of this equation is a circle, M43 focuses on determining its radius and center. After correctly identifying the radius of this circle ("121 is the radius squared, I believe, so the radius would be 11"; behavior 3), she then turns her attention to identifying the center of the circle: "I believe what the 4 and 19 are representing is the height and width, or this is the center." In addition to the fact that her use of the terms "height" and "width" suggests an incomplete understanding of the situation, she also neglects to consider the signs in the equation, which would alter the placement of the coordinates of the center of the resultant circle. Regardless, M43 only seems to use the center's *y*-coordinate (19) and the circle's radius (11) to assess the given answer choices. After dismissing choices A and B because she doesn't "think it should be negative," M43 concludes that choice D, 19, is correct "because 11 is the radius, so I don't believe that'd be part of the center, which I think is what it's looking for." In fact, the question is seeking a possible *x*-coordinate of a point, (*a*, *b*), that lies on the circle given the circle's equation.

Supplementary Vignette: Participant M30

Participant M30 answered question 3 correctly and demonstrated a single expected behavior, resulting in a PL of 1. M30 was one of ten participants who answered the question correctly and one of five participants who did so while also demonstrating at least one expected behavior.

> So, for me at least, I would just input the answers into the question and see what could work. Or, in this case, I could also use Desmos to help me out here. So I would put in the formula, and as you can see, I'm using the Desmos calculator, and I'm inputting the formula into Desmos to help me visualize a little bit better. So there we go. We have our circle, and now we have our values, which we can basically input if a and b are x and y. For example, -16 [choice A] does not work because x does not go past -15. So -16 doesn't work. And, for example, 11 [choice C] and 19 [choice D] do not work because the x-values on the coordinate plane are not in the circle, which means that the only value that would work here would be -14 [choice B].
>
> *Participant M30*

Participant M30 initially suggests and then quickly dismisses the idea of "input[ing] the answers into the question and see[ing] what could work." She instead elects to use Bluebook's built-in graphing calculator to visually inspect where the graph of the circle exists in the *xy*-plane. Before reviewing this graph, she acknowledges that "*a* and *b* are *x* and *y*," offering further clarity into what she should be searching for in the graph. She talks through each of the answer choices provided and concludes in turn that −16 (choice A) "does not work because *x* does not go past −15" and that 11 (choice C) and 19 (choice D) "do not work because the

x-values on the coordinate plane are not in the circle" (behavior 1). Of the options provided, only -14 could be a value of a, so M30 correctly selects choice B as her 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.
M43 Performance Level	5

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 I guess one way we could do this is by plugging in -3. 3^{-3} would be -27—or, no, -3. That would make it a fraction. It would be $-6(\frac{1}{27})$. Then $\frac{-6}{27} - 3$. That's not really coming out to be a clean number. 6^{-3} is 36. [*Long pause*] I'm gonna start off—I'm actually gonna go back and set the first one to 0. Yeah, so I have, after setting it to 0, I've started simplifying it a bit, or moving it around. I have $3 = -6(3)^x$. I'm going to divide both sides by -6, so $-\frac{1}{2} = 3^x$. Let me go to the g function next. Set that equal to 0. Divide by -3, and I—after setting both equal to 0, I believe g is the one that has a minimum value of -3. No, no, no, no, no. Yes, yes. g has a minimum value of -3 because if $x = 0$, then that would make the multiplication side 1, and then that would be -3, which should be as small as you could get it.
>
> *Participant M43*

Participant M43 answered the question incorrectly and didn't demonstrate any expected behaviors, resulting in a PL of 5. M43's initial strategy is to substitute -3 for x in the given functions. While doing so for function f, she states that "3^{-3} would be -27" but quickly catches her mistake: ". . . or, no, -3. That would make it a fraction. It would be $-6\left(\frac{1}{27}\right)$." Unable to draw a conclusion from this step, she moves to function g. After again substituting -3 for x, she makes a similar mistake in stating that "6^{-3} is 36." But instead of using or correcting this information, M43 pivots to an alternate approach of setting function f equal to 0 and manipulating the equation, resulting in $-\frac{1}{2}=3^x$. Then she moves to function g, sets it equal to 0, and divides both sides of the equation by -3. It's unclear what prompts the response "No, no, no, no, no. Yes, yes." Seemingly, she has pivoted her thought process again and focused on what happens when $x=0$. By saying "that would make the multiplication side 1," M43 is presumably referring to $g(x)=-3(6)^0=-3(1)=-3$. At this point, she incorrectly concludes "then that would be -3, which should be as small as you could get it."

Supplementary Vignette: Participant M61

Participant M61 answered question 4 correctly and demonstrated two expected behaviors, resulting in a PL of 1. M61 was one of ten participants who answered the question correctly and one of seven participants who did so while also demonstrating at least one expected behavior.

> So I am going to use Desmos again because that's just the quickest way to do it. Type these both in real quick. I'm just gonna see what those look like—the blue and green. So I'm just looking at [$x=$] -3 to see where that lies on the graph. Because none of these are parabolas and it just is continuously going down. Let's see. Do any of them lie on –3? Because they just keep continuously going down since it's exponential. I would say that neither of them have a minimum value at -3 just because it keeps decreasing, and it's not a parabola. So D is my final answer.
>
> *Participant M61*

Participant M61 first graphs both of the given functions in Bluebook's built-in graphing calculator and then investigates the graphs, specifically at $x=-3$ (behavior 1). There didn't appear to be anything of interest to her at this location on the graphs, but she then acknowledges that "none of these are parabolas," although that doesn't definitively rule out that neither has a minimum value at -3. Later, M61 connects that the functions "just keep continuously going down since it's exponential," helping solidify her conclusion that neither of the given functions has a minimum value at -3 (behavior 3).

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.
M43 Performance Level	1

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 $x = 12$.

> I'm gonna start by trying out some of the numbers. I won't go through it all, but off the bat, I can tell that it's not [choices] C or D because we are decreasing the value to get back to the quantity of x since we originally increased. 240 and 340 are increasing the number since the number we're trying to decrease from is 60. So we're left with 12 [choice A] and 15 [choice B]. For 12, I would do $12 + 4(12)$, which is 48, and that does give 60. And $15 + 4(15)$ is 75. Right. So that's not correct. That leaves me with [choice] A.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. She begins by taking a close look at the options and how they conceptually align with the question. In doing so, M43 is able to logically rule out choices C and D: "... because we are decreasing the value to get back to the quantity of x since we originally increased. 240 and 340 are increasing the number since the number we're trying to decrease from is 60" (behavior 4). In

other words, she's presumably suggesting that the correct answer must be less than 60. To determine whether the correct answer is 12 or 15, the other remaining choices, M43 sets up and evaluates the expressions $12+4(12)$ and $15+4(15)$, where 4 represents the percentage increase of 400% (behavior 1). She doesn't specifically write an equation to solve for the quantity of x, but she's nevertheless able to conclude that the correct answer is 12 by simplifying the constructed expressions.

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.
M43 Performance Level	5

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 $a+b=-3a=-3(2)=-6$.

> The first thing I'm going to do is think of a way to rewrite this by pulling out the important things. I think it's important to note that we're working with the quadratic formula. It's important to pull out the graph function and the points being used in the function. So what, we can go ahead and set $y = ax^2 + bx + c$. We know where $y = 0$; we have x-intercepts at 7 and -3, which leads me to believe this is a sort of parabolic shape since it's passing through the x-axis twice. Let's see. If we are taking x, rise over run. If we, a is an integer greater than 1. We know it's not -6 [choice A]. If a begins as an integer greater than 1, there's no number here that would get it lower than that. I think we have a difference of 10 going through there. The center of the parabola would be at 2. So the y-intercept is 2, to keep that in mind. Maybe I think I must start off by looking through my answer choices a little more, seeing where they're coming from. [Choice] C is the first point added to the second point. [Choice] D, I'm not sure where it comes from. [Choice] B is just taking one of the x-values. I'm going to go with C here, I think.
>
> *Participant M43*

Participant M43 answered the question incorrectly and didn't demonstrate any expected behaviors, resulting in a PL of 5. M43 wasn't able to successfully establish a proper entry point for the question. Although some of her observations are correct—for example, "We know where $y = 0$; we have x-intercepts at 7 and -3, which leads me to believe this is a sort of parabolic shape since it's passing through the x-axis twice"—they don't lead to anything substantive. Other comments indicate false or unsupported assumptions. At one point, for example, she observes that "the center of the parabola would be at 2" and "so the y-intercept is 2," but the given information doesn't directly allow for such specific conclusions to be made about key features of a possible graph of f. In any event, M43 persists and decides to investigate the given answer choices: "[Choice] C is the first point added to the second point. [Choice] D, I'm not sure where it comes from. [Choice] B is just taking one of the x-values." At this point, without providing any further justification, she incorrectly selects choice C, 4.

In this study, question 6 performed as the most difficult Math question, which is not only reflected by its PSB of 7 (the scale's highest) but also by the fact that no participant answered it correctly while exhibiting any expected behaviors and only three answered it correctly at all.

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.
M43 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 I'm going to keep that in mind: $25w$ is the amount of calories, with w being the weight in pounds, and then $25w + 11$ because we need to add on those extra 11 calories. [*Rereads question stem*] My equation is $25w + 11$. I'm just going to change w to x to get $c = 25x + 11$ so that it meets the question's format.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. After demonstrating comprehension of the context (behavior 1), M43 constructs an expression to represent the situation: "$25w$ is the amount of calories, with w being the weight in pounds, and then $25w + 11$ because we need to add on those extra 11 calories" (behavior 2). After restating the question stem, she replaces w with x, sets the equation equal to c "so that it meets the question's format," and selects D as her answer choice.

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.
M43 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

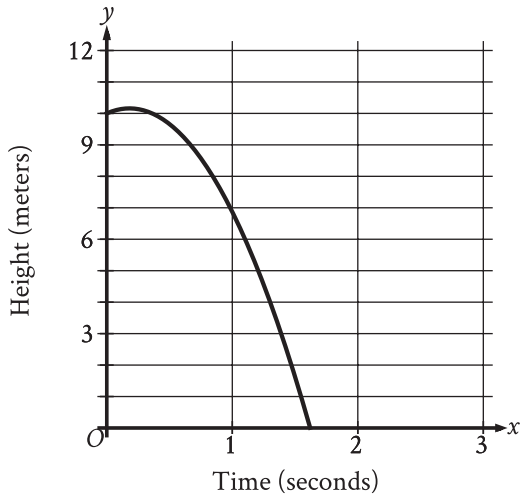
Question 8, an easy (PSB 3) multiple-choice Measure of Angles in a Triangle question outside of context, requires test takers to demonstrate an understanding of the triangle sum theorem, the concept that the sum of all interior angles of a triangle is 180°. The correct answer is choice A. For $\triangle RST$, 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 $(180 - 63)°$, 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 $\angle S$.

> We'll need to keep in mind that it's gonna have to be less than 180° when it's added together since that's the total in a triangle. $63 + 116$ is 179. $118 + 63$ would be over the number we want, which is 181. So I automatically know that [choice] B is out. [Choice] C is also out, and [choice] D is out because they're all too high. The only one that would get us to a number less than 180 is 116.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. She begins her successful approach to this question by demonstrating command of the triangle sum theorem (behavior 1): ". . . it's gonna have to be less than 180° when it's added together since that's the total in a triangle." Next, M43 logically uses the given angle measure, 63°, and adds to it the values of each of the answer choices, assessing whether the resultant sums would be less than 180° (behavior 2). She determines that choice A, 116°, results in a sum less than 180° ("$63 + 116$ is 179"), while choices B, C, and D are "over the number we want" and "all too high." She ultimately correctly selects choice A as her answer.

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.
M43 Performance Level	1



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.

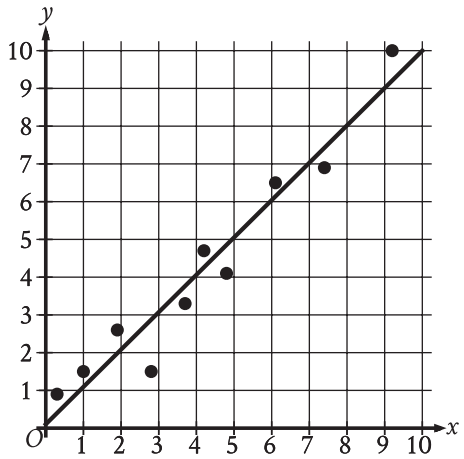
> The x-axis is the time when the height reaches 0. So the x-intercept would be where the diver has finished diving—when they've landed in the water. It looks like that is somewhere between 1 and 2 seconds. It would have reached its maximum at the peak of this exponential. It's not 0.2 seconds because the diver needs to be at the water. The diver hits the water at 1.6 seconds. That's what makes the most sense to me. The other two don't make any sense; I don't know where "0.2 seconds" [choices C and D] is coming from.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated all three expected behaviors, resulting in a PL of 1. After summarizing the context and how it relates to the graph ("So the *x*-intercept would be where the diver has finished diving—when they've landed in the water"; behaviors 1 and 3), she explains that the *x*-intercept "is somewhere between 1 and 2 seconds" (behavior 2). At one point, M43 incorrectly (and inconsequentially) refers to "this exponential," but she quickly moves on to properly dismiss choices with "0.2" (C and D): "It's not 0.2 seconds because the diver needs to be at the water." Given the graph and with the knowledge that the diver hits the water at some point between 1 and 2 seconds, she ultimately selects the correct answer, choice B.

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.
M43 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.

> We're looking to see which points are above the y, where they land above it. So that would be [*gestures*] here, here, here, here, here, and here. So that's six points above the line. I'm getting 6.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated both expected behaviors, resulting in a PL of 1. She makes exceptionally quick work of this question, demonstrating a clear understanding that a scatterplot's line of best fit represents predicted values and not the values of actual data points (behavior 1). In doing so, M43 counts the six points "above" the line of best fit—that is, the number of instances in which the actual *y*-values would be greater than those predicted by the line of best fit (behavior 2)—and selects the correct answer, choice C.

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.
M43 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.

> From knowing 0.48, we have 48% of the customers in theater A and 24% of the customers in theater B. [0.]48 + [0.]24 gives us [0.]72, which leaves 28% for theater C. Let me go ahead and plug this into my calculator: $(350)(0.28) = 98$. So there will be 98 customers located in theater C.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated all three expected behaviors, resulting in a PL of 1. While she uses the terms "probability" and "percentage" interchangeably, it's clear she understands in this context that the probability of a customer in a theater being selected at random is the same as the percentage of customers in that theater (behavior 1). After M43 adds the two given probabilities, 0.48 and 0.24, resulting in a sum of 0.72, she concludes by the concept of complementary events that this "leaves 28% for theater C" (behavior 2). In her final calculation, she correctly finds 28% of the total number of customers, 350, to be 98 (behavior 3) and selects choice D.

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.
M43 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.

> We're using variable d here. It looks like we are subtracting 7 from one side, so we go ahead and add 7 from the left side to the right side. We'll actually get 0 over there. So, it's $(d-30)(d+30)=0$. It will be whatever makes each individual answer 0. In this case, it could be both a negative or a positive answer. [The SPR directions say] that if there's more than one correct [answer to a question in the SPR format], enter only one. So I'm just going to put "30" in there. [The question stem] also says "a solution," so I guess that kind of tells you. But that will cause this one to be 0, and then it will come out as 0.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated three expected behaviors, resulting in a PL of 1. She efficiently and effectively navigates an algebraic solution path for this quadratic equation (behavior 3) by adding 7 to both sides, setting it equal to 0 (behavior 1). In doing so, and given the factored form of the right-hand side of the equation, M43 takes advantage of the zero-product property: "It will be whatever makes each individual answer 0" (behavior 2). Noting that "in this case, it could be both a negative or a positive answer" and recognizing that she can and only needs to enter one solution, she elects to answer "30" since "that will cause this one to be 0, and then it will come out as 0."

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.
M43 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.")

> The first thing I want to do is simplify this a bit. $\frac{1}{3}$ will go into—we don't really need to worry about the 10 as much, but I do want to distribute the $\frac{1}{3}$ to 29. So $\frac{1}{3}(29x)+5x$ is equal to $\frac{29}{3}x+\frac{15}{3}x$. So $\frac{29}{3}+\frac{15}{3}$ will give us $\frac{44}{3}x$. So $\frac{44}{3}$ is equal to m, which is the slope.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. While she doesn't fully communicate her plan, it's clear that M43 is intentional with her actions in simplifying the right-hand side of the linear equation so that the coefficient of x will then represent the slope of the graph (when in $y=mx+b$ form) (behavior 1). This is made evident when she states that "we don't really need to worry about the 10 as much," since it wouldn't be included in the desired value. She carefully navigates the fractions (behavior 2), combines like terms, and concludes that "$\frac{44}{3}$ is equal to m, which is the slope." She enters "44/3" as her answer in Bluebook.

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.
M43 Performance Level	5

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*.

> Since each side is 6 times bigger, the whole thing would be 36 times bigger. So (36)(54) would be 1,944, It's feeling a little big. Let me see. 6 times the length of the corresponding side. I think it's actually just gonna be 6 times bigger because it's already squared in the other one. So it'll be 324.
>
> *Participant M43*

Participant M43 answered the question incorrectly and didn't demonstrate any expected behaviors, resulting in a PL of 5. M43 starts her solution path flawlessly: "Since each side is 6 times bigger, the whole thing would be 36 times bigger." She grows doubtful about this outcome, however, stating that the product of 1,944 is "feeling a little big." It's not clear why she believes this to be the case, but it's a strong enough feeling (especially given that "1,944" is an available answer choice) that she reconsiders her interpretation of the effects of a scale factor on area. M43 ultimately convinces herself that "it's actually just gonna be 6 times bigger because it's already squared in the other one" and commits to an incorrect answer of choice C, 324.

Supplementary Vignette: Participant M37

Participant M37 answered question 14 correctly and demonstrated a single expected behavior, resulting in a PL of 1. M37 was one of four participants who answered the question correctly, all of whom doing so while also demonstrating at least one expected behavior.

> So two rectangles. The length of each side of *EFGH* is 6 times the length of the corresponding side of *ABCD*. So *E* is 6 times *A*, *F* is 6 times *B*, et

> cetera. The area of *ABCD* is 54 square units. What is the area in square units of *EFGH*? I think it would just be 6 times 54 because if you multiply them out, no matter what, that is fixed, which you should end up with the final product anyways. So 54 times 6. Yeah. But, actually, there's—since it's a square, and each side is 6 times more, there's 2 times 2 sides that are 6 times more. So we actually need to do it times 6 again, which leaves me the answer 1,944, which excludes [choices] A, B, and C and leaves me [choice] D.
>
> *Participant M37*

Participant M37 first rereads the question, clarifying the relationship between rectangles *EFGH* and *ABCD*: "So *E* is 6 times *A*, *F* is 6 times *B*, et cetera." He initially believes that the correct answer would simply be (6)(56) (which is 324, the area of rectangle *ABCD* times the given scale factor) "because if you multiply them out, no matter what, [the scale factor] is fixed, which you should end up with the final product anyways." After a momentary hesitation, M37 correctly comes to a realization: "since it's a square, and each side is 6 times more, there's 2 times 2 sides that are 6 times more. So we actually need to do it times 6 again" (behavior 1). On multiplying his previous answer, 324, by 6, he identifies the correct answer—choice D, 1,944—and dismisses the others.

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*.
M43 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 what I'll start off doing is by getting one side by itself. The thing I'm seeing immediately is just, take the first equation and divide both sides by 5 to get $y = 10x + \frac{11}{5}$, which would give us y—or we don't. Yeah, we simplified a bit. $y = 2x + \frac{11}{5}$. And then I can plug it into the next equation, say $-5\left(2x + \frac{11}{5}\right) = 5x - 21$. So now we only have one variable. That will simplify to $-10x - 11 = 5x - 21$. I'm going to add 11 to both sides to get $-10x = 5x - 10$. I am going to subtract $5x$ from the, both sides now to get $-15x = -10$. And then I will divide both sides by -15, and that will give me an answer of $x = \frac{10}{15}$. And then I will multiply that by 30, which will give me $\frac{300}{15}$. And I don't believe I have to simplify that. [*Attempts to enter into Bluebook but is unable*] So I guess I do. 300 divided by 15 equals 20.
>
> *Participant M43*

Participant M43 answered the question correctly and demonstrated two expected behaviors, resulting in a PL of 1. While her substitution approach may not be the most efficient, it proves effective in solving for x. After solving for y in the first equation and substituting $2x + \frac{11}{5}$ for y in the second equation, M43 fluently solves this equation for x (behavior 1). Observing that $x = \frac{10}{15}$, she "multipl[ies] that by 30, which will give me $\frac{300}{15}$" (behavior 3). She realizes this isn't an enterable value—it has too many characters—and continues to simplify, getting the correct answer of 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?

Responses to postexperience question 1 elicited mainly neutral or mixed reactions to the think-aloud activity, with smaller proportions describing the experience as generally good and easy or generally hard. The Math questions themselves and the Bluebook app drew some praise, but one participant called out what she felt was unduly tricky wording in the questions. Some participants noted that they either didn't like math, hadn't taken math for a while, and/or had forgotten previously learned math concepts or how to apply them. Many participants mentioned that the think-aloud experience deviated to at least some extent from how they'd normally go about testing, and of those who expressed an opinion about the impact of reading and thinking aloud, participants were evenly split between those who felt it helped them focus, answer correctly, and spot mistakes and those who found it distracting or intimidating.

Some of them were a little easy to understand, and some of them were a little hard to understand. *M1*

I think those questions were pretty difficult for me, maybe because I haven't attended school in two months. This might serve as inspiration or advice for me to start studying earlier before the school year starts. *M2*

I'm used to having long questions because before I moved to New York, I was in [location], and we had, like, ninety-question assignments. So I'm used to doing, like, very long-form test[s]. So it's not, like, the most insane thing, but it's also, like, I hate math, so I kinda—even doing, like, more than ten is annoying. *M3*

It was challenging. I didn't understand most of the questions. The format's OK; the format's not bad. So that is my experience. *M7*

Reading is not my best. What is that called? Like, I'm not that good at reading. So when I read stuff it, like, gets, like, mixed up in my brain a little. So it's a little confusing for me when I'm reading it because then it's just like, "What did I just read?" It's like, I don't take in information, so I have to keep rereading it and going over it, but it wasn't that bad. Just it, it just, my reading just got me confused. I, like, confused myself with some of the questions. *M8*

I think some of the questions were very easy, but the wording—like, you could tell that they are purposely making it more complicated for you to understand. The answer is in the question, but it's almost like they want the students to master how to understand the question rather than having the knowledge for it, I guess. But the tools were very helpful. The graphing calculator was very helpful, and the reference sheet, even though I didn't get to use it as much, I think it provides all the good information as well. The format is pretty clear. *M14*

So it was kind of different than when I would actually be taking the test because I had to say everything out loud. So my thoughts were kind of in a different order, and it was weird for me because I was trying to analyze them as I was reading the question. So it's just a different way of answering them because I feel like I would just go straight into analyzing it if I was actually taking the test. So it adds an extra factor when you're kind of reading it out loud. But in explaining it because that kind of takes kind of your mind to kind of explain what you're doing. So it's hard to really go into it. But it was pretty different to, like, actually say it out loud. But I feel like the experience was pretty good—like, I got used to it pretty fast, and it became easier to solve these questions. It was good. *M15*

So, first of all, the platform itself, it's so concise; it's so easy to navigate through answering the questions. I think they targeted each point, each concept of algebra, geometry, all those sorts of things. All the concepts, I think they did; this test did a really good job at that. . . . I think I love how it wasn't insanely hard. I think it's very, it measures a person's level of math very well. It was— With some questions, obviously, I forgot some of the concepts from Algebra 1, let's say, for some of the questions; that's

totally on me. But the questions were great. They were easy to answer, and I think it was very, you know, easy and enjoyable, better than the written ones, the written SAT. *M17*

I feel for a lot of it, it's just that I have to take up a lot of my older knowledge. And so a lot, specifically with geometry, it's just more of, a lot of it is stuff that I did when I was younger. And so now that I'm in, like, calculus and stuff, it becomes a little bit harder to try and remember that type of math. And so even though the questions, they seem easy, sometimes it's just, like, hard trying to remember the names and everything of it. So it's, like, conceptually, I find it more or less kind of easy, but remembering the specific steps is hard. So yeah, it was kind of, it was a little bit tricky, but in trying to explain my steps—because I know at least how to do it—but explaining it is like, "Uh . . ." *M30*

It was all right. I'm very, like—so this whole thing is for, like, [College Board] getting to see how other people test and stuff. I think I was very aware that I knew that people were, like, actually, like, watching what I was doing. So I don't know. I think that was a little bit overwhelming even though it wasn't graded for. . . . I also haven't done math in a very long time. So I forgot almost everything, and then I zoned out a lot reading the questions. That's why I kept rewrote, like, kept re-say[ing] them over and over again, because I forgot. I don't know why I forgot. It's always like that. And then some of the questions just came off as very hard to read, but it wasn't bad. I just, I was, I don't, I'm not good at math. So there's that. *M42*

Strategies

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

In response to postexperience question 2, participants cited various general approaches to working the Math questions presented to them during the think-aloud activity. The most commonly mentioned strategies included drawing on prior (and sometimes old) content knowledge, breaking down questions into more manageable parts, relying heavily or even to exclusion on a calculator, and using some degree of answer elimination. Multiple participants mentioned using scratch paper (available both during the activity and operational testing) as an aid for summarizing or visualizing questions. A small number of participants admitted frankly to lacking appropriate strategies, guessing in at least some cases, or using trial and error.

I think my strategies were lacking, so it definitely encourages me to develop more effective strategies to solve these questions. *M2*

I kind of had to go based off memory of the equations because a lot of them, I don't fully remember. Like, I'm not great with most linear equations, so I kind of had to go based off the context of the situation. So, like, if it was a system of equations, I had to remember, "OK. Is it followed in this order of, do I get rid of the *y* first, or do I have to separate

the whole number or the *x*-value?" So it's kind of just, like, process of elimination in my own head. And that's, like, the times I'll be silent. I would have had to, like, think about "OK. Is it this, this, and this?" So it was, like, too much to think about at one time to kind of put into words. *M3*

I say I go by reading it first, [and] then, whatever number it has, I add it up together—like, the first number or the second number. And then once I add it to a number, I get an answer, [and] then I go by that. *M5*

My general approach was going to the calculator. I didn't even try to write them down because I did not understand, but, yeah, it was just the calculator. *M7*

So, first, I looked at the question and then at all the answer choices. I tried to see if they absolutely worked. I would try and eliminate the ones that didn't fit, leaving me with one or two answers. Then, if I didn't know, I'd pick the one I thought was the best. If I knew, I'd solve the problem. *M12*

Basically, summarizing or, for each question, I feel like I need to read the question and then write my own version of what the question means to myself, in my own language, just to make it easier to start thinking about how to solve the problem. It's almost like a puzzle in the question itself, trying to understand what they are asking specifically. Then, if I'm making an educated guess, I also do a process of elimination to see what answer makes sense, what doesn't, and why. *M14*

If it's an earlier question, I try to answer them as fast as possible using the answer choices because they usually, those are usually the simpler questions. And if it's really hard, I just like trying random things, really, and try to get towards that solution because at one point it'll just click—like, "Oh, that's how I do it." But I kind of have to, like, experiment; trial and error is a big part of, like, my process. *M15*

So, OK. So the first thing that I do, I read the question just so I know what concept I'm tackling, just so that I can get a feel of the question and what it's asking of me. And then I go reread the sentence, the word equation, to dismantle it. So, the first time, I approach it like this: The first time, I'm just trying to get a grasp of it, see what it's asking of me. And then I go over it once or twice if I need to, just to really dismantle it and extract all the details that I need. And then I'm gonna write down on a piece of paper what I'm visualizing in my mind. *M17*

It's mostly just going, I'm just going off of what I remember. And if I don't remember it, the Desmos calculator does help a lot in terms of graphing and being able to help me visualize. Like, if it's, like, a graphing formula, then I can type it in, and it helps me visualize and, like, helps me remember how to do it rather than having to remember specific formulas that I can't remember. *M30*

Kind of, I start by reading the question and then kind of taking, like, looking for what numbers they are and kind of scanning for keywords,

because my brain kind of takes a long time to process things. So, yeah, I'm kind of just looking for keywords to try to understand what's happening in the problem. *M37*

I feel like, honestly, I didn't really know what I was, what I even thought I was gonna be doing when I came into today. And then when we were doing, like, the little practice review and stuff, I try to, I guess, understand—like, I don't know, use a tactic or something like it myself to use one. But then as soon as we, like, started the test and stuff, it kind of just all went out of it, and I just started talking and then, like, my mind when I'm doing a test, I don't really do tactics and do, like, process of elimination. I just start to panic and think, like, "This all looks weird," and that's how it was for me. Like, everything just looked scary, and I was just giving up already. So there's that. I think in the middle of the test—like, more towards the end probably on some of them—I did use process [of] elimination, but those are also the ones that I think I understood more. But if I didn't understand it, I didn't care. *M42*

I tried to relate it to something that I've seen before or automatically pick out what the question is asking about and not get distracted by the details. I tried to pick out all the information that I thought I was going to need and then sort of logic my way through it after I knew that much. *M43*

My strategy for answering the question is that—like, for me, I will lose my, like, I can't read something long without reading all of it. Like, if I read something, like, in English, if it's a big passage, I'm only reading, like, one sentence per paragraph, and that's what I work with. *M59*

"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 identified a wide range of elements that participants associated with question ease. Among the factors mentioned multiple times were questions with graphs, questions set in context, and questions involving slope determination—although the former two were raised by other participants as sources of difficulty (see postexperience question 4, below). A few participants noted heavily reliance on Bluebook's built-in graphing calculator (or similar handheld tool) and a preference for questions they perceived as easily answerable via such technology.

I feel that the graph questions were easier for me because they provided a more visual way of solving the equations. *M2*

Most likely, the word problems and finding, like, *x*-values for those. It's super simple because you are given so much more information than, like, a multiple-choice question where they just give you the equation and you have to figure it out yourself. Like, having to, when it['s] with a word problem, you get to assemble the equation by the info that you're given

and then solve it from there and try to find a specific goal, instead of "Oh, just find the *x*-value for this." *M3*

The type of the questions that were easiest to me were those asking for the slope, because all I had to do was enter the equation that was already there and go on to determine the slope. *M6*

Probably the ones with the graphs. . . . Well, I could just look at the graph and not have to solve anything. *M12*

So I kind of find system of equation problems really easy to solve for me because I know the multiple strategies, the system of equations. And I don't think word forms are as easy for me, but I think system [of] equations are really easy for me. *M15*

Questions that had graphs and visuals I liked because it felt like I could look at it. Questions that were shorter and, like, weren't word problems I liked more because it felt like there weren't all these variables to distract me with; it was just numbers. It wasn't like "movie theaters" [Math question 11], and then I was thinking about movies, and then I was thinking about *Despicable Me 4*, which is the last movie that my family went to go see. I wasn't there. I didn't see it. Well, I'm getting distracted even now, but, like, so they're a lot easier for me. Something else. Oh, yeah, and answers that I can easily, like, check—like "What is the solution to the quadratic equation?" Those ones I can check. *M18*

I would say all of them were very easy. I just, I know, like, it's with HC [honors classes?], or I guess, I mean, I don't know, I'm assuming, I guess we'll just say kids who have a, who take advanced classes overcomplicate things a lot because I'm so used to having to do multiple steps to find something for this. And then with this, you find that, and I, it's become, like, such a—I don't know—like, such a habit, and it's a bad one because for a lot of these questions, they're very simple, and I'm just overcomplicating it to such an extent and wasting a lot of time. *M23*

The algebra ones were particularly easy just because I have a lot of experience with algebra, kind of. So they were, they felt a lot easier. *M37*

My favorite questions are, like, the ones where you can just, like, put them in Desmos and then, like, get a visual of it to see, like, what you're even working with, like the circle [Math question 3]. Like, if I do a circle and I gave it the coordinates, like, it wouldn't make any sense to me. But since there's a graph, I can, like, easily see like what I'm answering, I guess. *M59*

I think, usually, like, the find-the-solution questions are pretty easy just because I usually just plug it into Desmos. And I think for any of the questions that, like, I know the skills on how to plug it into Desmos, those are pretty easy—like, the diameter of a circle or, like, you know, questions like those. *M61*

"Hard" Question Types

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

Responses to postexperience question 4 consistently pointed to several features of questions that at least some participants associated with difficulty: in-context problems, longer questions, the presence of graphs, the inclusion of many variables and constants, and problems involving geometry and systems of equations.

> Most likely the graphic questions, because those are usually the ones where you had to find *y* and *x*, so it's a lot more difficult to kind of figure out "OK, How do I separate *y* from this, from the main equation?"—or not "separate," but how to get rid of the number that's next to *y* while also keeping the rest of the equation. Especially the ones with parentheses, with, like, "OK, it's, like, 5 plus *x* to the power of 2 plus 7 plus *x* to the power of 2." Yeah, I hate those questions. I hate them. *M3*

> It was the one with the equation of the circle, like $ax + bx + c$ [Math question 3]. I think I might have learned that, but I might have forgotten how to solve it. *M12*

> I think some of the word problems get harder for me. Like, the ones where it's, like, asking which equation is right, this or that, that or this; or if there's too many variables involved, it just becomes a bit confusing. *M15*

> I would say these ones [questions in the student-produced response format] because I feel like multiple-choice helps me think better, in my opinion. *M19*

> Geometry. Not all of geometry—like, stuff with angles I'm still very familiar with because I did take a trigonometry class more recently. But more or less stuff with geometry, I'm still, like, I don't remember all too well since I did take it a while back. And yeah, that's the thing that I do struggle with because I took it in eighth grade, and I am going into twelfth grade now. *M30*

> Yes. Some of them, it was like I knew exactly what I could be doing but I just couldn't remember certain things to give me that answer, so I would just give up and guess. And then [for] some of them, I did kind of know, and then I was hoping I would have gotten it right. Like, I hoped I used the right tactic. And then some of them, I just had no idea what they were talking about or what it was asking me to do, so I just guessed. Because if I don't know it now, I don't know how I would remember in the next minute or so. *M42*

> The circle question threw me for a loop for a second, which is terrible. I'm in calculus; I shouldn't be getting upset over circles. But I'm just not much of a geometry person, I guess. So that might just be me personally; that's not something I really retain, I guess. *M43*

> The really, like, wordy ones where—on the real SAT, like, I took the August one in the Math module two, [and,] like, the last three, four questions, they were so wordy and since it's—like, I get extra time, right? So I'm taking, like, a 3.5-hour, four-hour test, and at the very, like, last five minutes, it's, like, four paragraphs. It's like, of course I'm gonna get them all wrong now. So it's like a little bit with, like, the wordiness of the Math questions. It's, like, I can't, it takes too much, like, stamina, I don't know. I can't do it. *M59*

ADHD Symptom Impact

5. Did you encounter anything in the questions that you had difficulty with given that you have ADHD? If so, what was it, and why was it difficult for you?

Participants responding to postexperience question 5 cited various aspects of Math questions that played into their ADHD symptoms, including questions of longer length and higher information density and questions set in context. The difficulty of dealing with distractions was an oft-repeated theme, and individual participants reported challenges such as persevering through monotony, maintaining appropriate pace, recognizing key details, and blocking out extraneous information.

> I think what made it difficult for me was the monotony of it. It was very boring and continuous, which was challenging for me to stay focused. *M2*

> Yeah, probably the same type of questions [over and over] because it, like, jumbles my brain a little, and it's hard to kind of keep focus and remember specific things without having to constantly write it down. *M3*

> The questions that gave me a hard time were those asking me to find the area of a rectangle or any shape. I felt it should have been easy, but sometimes it takes me a bit longer. I feel like I need to answer quickly to avoid the perception that I'm taking too long. *M6*

> I think it was difficult for me to not rush it—like, not fly through the questions. There wasn't a specific one that stood out; it was just all confusing. *M7*

> I feel like it's more of my reading, and it confuses me and that I just, like, get off track. *M8*

> I feel like the more wordy, lengthy questions were a lot harder for me, even though the easier questions were about transforming the words into an equation. It was easy mentally, but it was hard for me to really dissect the question and understand what they're asking, rather than just having an equation question with just the numbers itself and asking me. That was almost easier because it was straightforward instead of adding an extra step to understand the question itself. *M14*

> So I think some of the word problems that give a lot of information, I just kind of, I think maybe it makes me, like, kind of get distracted as

I'm reading this problem with so many multiple variables. So I have to write it down for myself to kind of, like, take in all of it at once. But when there's a lot of things involved in a problem, and, like, it kind of confuses me because it just makes it a bit more complicated, and I have to think. I don't know; I just get distracted as I'm trying to piece every part of this solution together or every part of this question together to find that solution. *M15*

So when a word [problem] has a plethora of words and when it's hard to dismantle, I get a mental block. When I'm overwhelmed with words or when I'm trying to sort things out in my mind, I just get a mental block, and my mind just stops working. And then I have to try really hard to get my mind running. So when there's just longer—not even longer word [problems]—it's just the super-packed, complicated ones that just give me a mental block. *M17*

. . . [A] lot of times, I will get distracted about what the actual question is asking. I think maybe, like, the reason I get stuck with systems of equations is because I take it and I write it on a piece of paper, and that's where I do all the math, and I don't look at the screen. And I usually will find myself, like, trying to solve for something different than what the question is asking me to solve for, just because I know I can, and that's really easy for me to get distracted with. I think it was easier before when [the SAT] was, like, not digital. I've never taken a nondigital SAT—I've only taken nondigital PSATs—but it was still easier because the question is, like, right in front of you, and if you need any scratch paper, you can, like, write on the booklet, and, like, it's not, like, a separate thing from the question. *M18*

It's mostly phrasing. The tests often, like, it just—the phrasing of questions, they'll add in a lot of, like, unnecessary word choices that confuses me. Like, there was that one question where it was like saying, oh yeah, like "$y = f(x)$" [Math question 6]. Like, it was, like, adding this additional information that I already knew. But the way it [was] phrased, it made me, like, kind of confused because it will say, like, even though it was just saying "$y = f(x)$," the way it was phrased, it was just like it was, as if it was saying, "Well, this additional formula also equals this formula" to have to do the extra math for it. *M30*

. . . [L]ike, whenever it comes to a test, it just, I always zone out [on] every single question. Like, whenever I don't understand a question, I just zone out and start reading it over and over again. I could probably read it twenty times, and I still wouldn't understand what it's saying or what it's asking me to do. And also, like, repeating a lot; I could tell I was doing that. That was, like, just an—I don't know. That was me going blank because I didn't even know what I was thinking about at that point. *M42*

It's difficult to say what is caused by [ADHD]. Believe it or not, it's always been like this. So it's kind of hard to tell what's caused by it and what isn't. I would say I have a tendency to jump around. For some of the more wordy questions, I would read them with the wrong inflection or

> something like I would be jumping around. Before reading it out loud, I really kind of had to look at a couple of them and make sure I knew what I was doing. That's why I have to pull the information out, because otherwise I'll skip over really important things. *M43*

> I think the question about the ring [probably Math question 3, about a circle] was very hard. . . . Yeah, it was hard for me because in geometry questions like that, I think it's a lot easier when, like, I see the triangle or I see the rectangle—like, the shape's given to me. Like, it's hard for me to, like, see it [otherwise]. That's why I usually try and draw it out on my paper. But it's just, like, it's hard for me to imagine, like, those sort of, like, shapes and what the side length is like. *M61*

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?

Most participants didn't have additional, final comments about the think-aloud activity in response to postexperience question 6. Among those who did, several commented on the Bluebook testing platform, variously noting that it was easy to use, cut down on distractions, and was helpful in providing access to the Desmos graphing calculator. One participant was distressed by how hard the second module in the Math section is relative to the first, and another participant observed that the think-aloud activity was obviously different from a standard testing experience.

> I relied heavily on Desmos. Like, about, honestly, I think I would have passed on, like, three, like ten different questions if I did not have Desmos. *M3*

> I use, like, a calculator, and sometimes I use my hands for the count stuff while I think in my head by adding stuff. *M5*

> I think you guys know that, like, speaking, saying everything out loud, kind of changes the way you take the test, and that's, like, a big variable in it. It's just, like, a different way of—you know, I feel like you're thinking differently if you're just saying it all out loud. *M15*

> I really, really like this platform. It's amazing. It's very concise. There's minimal distraction. It's very minimalist. They give you the question to you right over—I mean, they give you the question straight up. It's very concise. It's very short. It's very straightforward. There's no flip-flopping thing. No, it's very—I love the platform. It's minimally distracting. It very, it helps a lot with my ADHD, yes. *M17*

> . . . I would say the one thing I've always had a hard time understanding in the last other SAT [I took] before is having to type out the equation. I wish you could just copy and paste it because I'm thinking, like, what would be, what, why you wouldn't be allowed to do that. Because I just, I know there's, like, especially when you get into the harder section of

Math. Like, Math [module] one is usually, it's fairly easy and simple. Maybe, like, maybe one or two questions or, yeah. But when you get into that harder Math section, there's, there'll be a lot of functions, and you have to type in certain stuff. I just wish you could copy and paste it into Desmos because typing just wastes time. And I mean, if the test is testing for standardized knowledge, it doesn't need to—like, I guess there's no knowledge in typing. You know what I mean? Like, I just wish you could copy and paste it because it just saves a lot of time. *M23*

I feel like it's overall, I feel like, with the digital SAT, it's definitely an improvement in terms of from the paper SAT because I did take the paper PSAT, and I do like the addition of the Desmos calculator and just in general being able to utilize a calculator for the entire section. However, what I do dislike especially is the fact that of how much harder it gets. Like, if you do well in the first [module of the Math section], how much harder it gets for the second [module]. So I remember when I was taking the SAT earlier, the first [module] was fairly easy for Math. It was pretty easy. But then the second [module], it increased in difficulty. And even if it was, like, the questions themselves weren't, like, super difficult, it was the phrasing of the questions that made it difficult. Like, that's, like, specifically with phrasing, because at least for me, I can understand the question fairly easily if it's just saying "OK, you do this, you do that." But when they add, like, certain word choices, it makes it, like, harder for me to process the question, and it starts confusing me a lot more. *M30*

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.

Nineteen of twenty-four Reading and Writing participants (79 percent) met or exceeded the threshold for a good D_p, providing evidence that students with ADHD are able to demonstrate cognitively complex thinking in line with the question types' constructs. The remaining participants had differentials ranging from 2 to 5. (For example, participant RW12 had a D_p of 4, and because he answered only nine questions correctly and only five 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 half to two-thirds of the questions they answered correctly. In general, these results offer evidence that students with ADHD 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.

Twelve of the study's fifteen Reading and Writing questions (80 percent) met or exceeded the threshold for a good D_q. The remaining three questions exhibited high differentials of 6, 7, and 9. These questions were answered correctly at reasonable rates—by fourteen, eleven, and eighteen participants, respectively, suggesting that the issue isn't strictly or solely one of appropriate content knowledge. Moreover, as noted in Section 4: Results, these criterion-failing questions don't seem to have much in common in terms of passage text complexity, passage subject area, question difficulty, or presence/absence of informational graphics.

However, data from participant performance and perceptions give us some strong clues as to why many participants struggled to verbalize cognitively complex thinking while answering these questions. First, in terms of the two criterion-failing Rhetorical Synthesis questions (7 and 8), participants often failed to demonstrate adequate notes comprehension (behavior 1; eleven times across twenty-four participants for question 7, thirteen times for question 8). This suggests that many participants essentially "skipped" the notes and worked the questions solely from the answer choices, a supposition supported by a few participants noting in their postexperience interviews that they typically don't read the notes. While this strategy may have worked with these two Rhetorical Synthesis questions, 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. Second, question 14, the Command of Evidence: Quantitative question about the impact of various climate change scenarios on sugar maple growth, had many of the factors participants associated in their postexperience interviews with question challenge, including a longer passage (and a high-complexity one), the presence of a somewhat complex informational graphic, and longer answer choices that have only minor (but critical) wording differences.

Nonetheless, a minimum of four participants attained PL 1 on these questions, suggesting that these questions, too, are capable of eliciting cognitively complex thinking from students with ADHD 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 students with ADHD.

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 students with ADHD.

PARTICIPANT PERCEPTIONS

Reading and Writing participants gave generally neutral or mixed assessments of their think-aloud experience (postexperience question 1). They called out a few test-taking strategies (postexperience question 2); these most prominently included reading questions before reading passages (which participants were constrained from doing by the think-aloud protocol), rereading, highlighting, coming up with one's own answer before looking at answer choices, drawing on content knowledge, eliminating incorrect answer choices, mentally "plugging in" answer options in blank-completion questions, and relying on a sense of "fit" between one's answer choice and the stimulus material. In terms of question types they found particularly easy to answer (postexperience question 3), participants generally identified two factors contributing to ease: short passages and the blank-completion question format. Conversely, participants tended to consider hard (postexperience question 4) those questions that had longer stimulus passages, that involved informational graphics (tables and graphs), and that had answer choices whose wording had minor (but still important) variations. 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 to identify ADHD-related issues that impacted their performance on the think-aloud activity (postexperience question 5), participants frequently cited issues rooted in their condition, including difficulties with longer texts; longer and similarly worded answer choices; challenging, unfamiliar vocabulary; graphs; passage pairs; and text processing speed and accuracy, with all of these tying back to challenges with maintaining focus and persisting through tasks. When offered—which was relatively seldom—final comments (postexperience question 6) tended to emphasize previously raised matters.

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 benefiting from (or being distracted or intimidated 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

Thirteen of twenty-one Math participants (62 percent) met or exceeded the threshold for a good participant differential (D_p), thereby providing evidence that students with ADHD are capable of demonstrating cognitively complex thinking in line with the question types' constructs. Seven of the remaining participants had differentials ranging from 1 to 5, while the eighth participant lacked a true differential because she didn't answer any questions correctly. The seven participants with criterion-failing differentials were able to demonstrate

cognitively complex thinking on roughly a quarter to two-thirds of the questions they did answer correctly, suggesting that relative lack of experience or comfort with thinking aloud may have contributed to their failure to meet or exceed the threshold. In general, these results offer evidence that students with ADHD are able to exhibit cognitively complex thinking in line with the question types' expectations.

QUESTION PERFORMANCE

Twelve of the fifteen Math questions used in this study (80 percent) met or exceeded the criterion for a good question differential (D_q). The remaining questions had differentials of 2, 3, and 5, and from three to ten participants answered each of these questions correctly while also demonstrating at least one expected behavior (i.e., attained PL 1). All three criterion-failing questions were hard (PSB 6 or 7) multiple-choice questions outside of context. Three other hard (PSB 6 or 7) Math questions had criterion-passing differentials, indicating that students with ADHD can demonstrate cognitively complex thinking on questions from even this topmost difficulty band and lending credence to the possibility that the higher-than-desirable differentials are attributable, at least in part, to some students' ability to verbalize effectively during the think-aloud activity. The overall findings support the claim that the presented Math questions are capable of eliciting cognitively complex thinking in line with the question types' constructs from students with ADHD.

PARTICIPANT PERFORMANCE VIGNETTES

Participant performance vignettes (transcript excerpts) exhibiting highly successful (PL 1) outcomes in line with question types' constructs were obtained for fourteen of the fifteen Math questions, providing further evidence that the questions are capable of eliciting cognitively complex thinking from students with ADHD. The remaining question—question 6, a hard (PSB 7) multiple-choice Nonlinear Functions: Make Connections question outside of context—was answered correctly by only three participants, none of whom exhibited any expected behaviors. Transcript evidence indicates that many participants struggled with question 6 largely because the use of unknown constants made it difficult to identify an entry point—a challenge directly tied to the construct being measured.

PARTICIPANT PERCEPTIONS

Participants were, in the main, neutral or mixed in their attitudes toward the think-aloud activity (postexperience question 1), with smaller, roughly equal proportions of participants describing the experience as generally good or generally bad. A lack of appropriate or up-to-date content knowledge and a distaste for math itself were sometimes noted. Numerous participants called attention to the fact that the think-aloud methodology affected their test-taking process, although those who raised the issue were split over whether reading aloud helped them or was an impediment.

The most commonly mentioned general strategies for approaching the questions (postexperience question 2) were drawing on prior content knowledge, breaking down questions into more manageable parts, making (often heavy) use of a

calculator, and eliminating incorrect multiple-choice answer options. Several participants discussed using scratch paper productively. A few participants acknowledged having no real strategies at all.

Question ease (postexperience question 3) and difficulty (question 4) drew conflicting, often mirror-image responses. Some participants identified as notably easy those questions with graphs, questions set in context, questions involving the determination of slope, and questions for which a calculator was particularly useful, while other participants identified as notably hard those questions with graphs, questions set in context, questions with many variables and constants, geometry questions, and longer questions with more content and more to read.

When asked to discuss the impact of their ADHD symptoms on test taking (postexperience question 5), participants repeatedly raised the issue of having to deal effectively with distractions, and individual participants mentioned related challenges such as overcoming monotony, keeping a good pace, tracking important details, and ignoring extraneous information. Longer questions, questions set in context, and questions with higher information densities (factors that often co-occur) were also described as challenging due to participants' ADHD symptoms.

Although most participants didn't have any final, general comments (postexperience question 6), a few praised Bluebook for being "clean" and minimizing distractions and were thankful for the inclusion of a built-in graphing calculator.

General Discussion

Results from this cognitive lab study involving students with ADHD 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	19 of 24 (79%)	12 of 15 (80%)
Math	13 of 21 (62%)	12 of 15 (80%)

In terms of D_p, roughly four-fifths (79 percent) of Reading and Writing participants (n = 24) and just under two-thirds (62 percent) of Math participants (n = 21) 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, four-fifths (80 percent) of the Reading and Writing (n = 15) and 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).

Given the difficulty of many of the Reading and Writing and Math questions, the fact that all participants reported experiencing symptoms of ADHD to one degree or another (modally at the "moderate" level), and, perhaps most importantly, the inherent challenge of and relative or absolute unfamiliarity on the part of participants with thinking aloud while answering test questions, we judge these to be generally good results.

That said, the relatively low proportion of Math participants (62 percent) who met or exceeded the criterion for a good D_p is worth additional consideration. No firm causal explanation can be made, of course, from the sorts of data collected in this study, but these data do offer some hints.

First, the criterion-failing participants in Math were, with one notable exception, generally less successful in the think-aloud activity, in terms of raw question-answering performance, than were their criterion-meeting peers. Figure 3 displays the number of Math questions answered correctly (*#AC*) by D_p-criterion-passing and -failing participants (including the one participant who lacked a true differential).

Figure 3. Math: *#AC* by D_p Criterion-Passing and -Failing Participants.

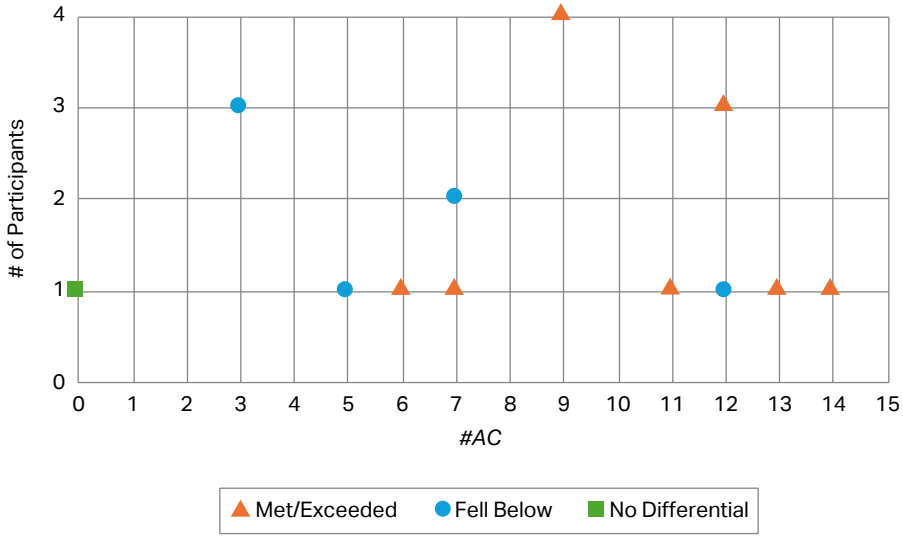


Figure 3 makes evident that participants who met or exceeded the criterion for a good D_p also tended to be more successful, in terms of *#AC*, than were their peers who failed to meet the criterion. This suggests that adequate content knowledge (or lack thereof) played an important role in whether the Math participants met/exceeded or failed to meet the criterion for a good D_p.

Second, and also favoring a content knowledge–centered explanation, responses to postexperience interview questions frequently suggested that at least some participants had a negative attitude toward math in general and that they either hadn't taken a math course in some time or couldn't recall previously learned math concepts and/or how to apply them appropriately. This lack of or uncertainty about adequate content knowledge likely also accounted for the postexperience interview responses received in which participants noted heavy or near-total reliance on a calculator during the activity.

This doesn't rule out the possibility that Math participants' ADHD symptoms negatively affected their activity performance and did so differently or more significantly than for Reading and Writing participants, a greater proportion of whom had criterion-passing D_ps. Indeed, Math participants identified several areas of challenge posed by their ADHD symptoms, including dealing with distractions, pressing through monotony, maintaining pace, processing details, and blocking out extraneous information. It seems likely, however, that these issues would also affect the Reading and Writing performance of test takers with ADHD—and, in fact, similar challenges were mentioned by some participants in the postexperience interview portion of that segment of the study. Given this, we speculate that the lower success of Math participants relative to Reading and Writing participants on the D_p criterion is attributable mainly to the former's struggles to marshal and apply adequate content knowledge.

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 from fourteen of the fifteen Math questions. The fact that all Reading and Writing and all but one of the Math questions analyzed for this study were able to elicit both correct answers and appropriate behaviors from students with ADHD—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 students with ADHD.

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:

- Participants tended to have a mixed or neutral impression of the think-aloud activity.
- The act of thinking aloud while answering test questions had an impact, 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.
- Strategy use was somewhat varied. Reading and Writing participants most commonly mentioned reading the question before reading the passage (something they could do normally but not during the think-aloud activity), rereading, highlighting, coming up with one's own answer before looking at provided answer choices, using content knowledge, eliminating incorrect answer choices, "plugging in" answer options in questions in the blank-completion format, and making use of a sense of "fit" between their answer choice and the stimulus material. Math participants most commonly cited using content knowledge, breaking down questions, making use of or outright relying on a calculator, eliminating incorrect answer choices, and using scratch paper. Math participants sometimes lamented that their content knowledge was incomplete or old.

- In terms of factors contributing to question ease and difficulty, Reading and Writing and Math participants were united in favoring short, compact questions over longer ones with more text and information. Reading and Writing participants also found the blank-completion format and the Rhetorical Synthesis question type relatively easy (the latter sometimes based on the erroneous assumption that it's unnecessary to read the student-generated notes), while questions with tables and graphs were relatively hard. Interestingly, about the same proportions of Math participants signaled that questions with graphs and questions set in context were easy and hard. Math participants also identified as hard those questions with many variables and constants, questions involving systems of equations, and questions concerning geometry (the last typically because it'd been some time since participants had studied or used geometry skills and knowledge).
- ADHD symptoms affected test taking negatively. Both Reading and Writing and Math participants mentioned challenges with maintaining focus and avoiding or minimizing distractions while testing. Longer questions and those with higher information densities were found by participants in both segments to be challenging.

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 = 24 for Reading and Writing; n = 21 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 students with ADHD 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 ADHD 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 somewhat 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 both a (greater or lesser) degree of artificiality and, in the case of some participants here, a certain psychological cost. 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, and, especially in the Math segment, this requirement, though known in advance, seems to have induced some level of anxiety and frustration among at least some participants, which potentially depressed performance. 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 have attention deficit hyperactivity disorder (ADHD) 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 students with ADHD 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 accommodations.

With regard to the first goal, the study's findings support the conclusion that students with ADHD 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 to ADHD students' performance residing in the test sections' designs or delivery method were identified.

It's important to note that the study's positive conclusions regarding students with ADHD are predicated on the assumption that these students have access as needed to appropriate testing accommodations. Nearly all participants (n = 45 across the Reading and Writing and Math segments) reported either having received or expecting to receive extended time and/or extended breaks as part of SAT Suite testing. This is necessary and desirable given that ADHD is, by definition, a condition affecting focus and persistence, and the provision of additional time helps level the playing field with respect to their peers without ADHD and the challenges it imposes.

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**.

American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed., text revision. American Psychiatric Association.

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/sslit.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 English Learners (ELs).* College Board. **https://satsuite.collegeboard.org/media/pdf/digital-sat-cognitive-lab-report-el.pdf**.

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 Pacic 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

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,

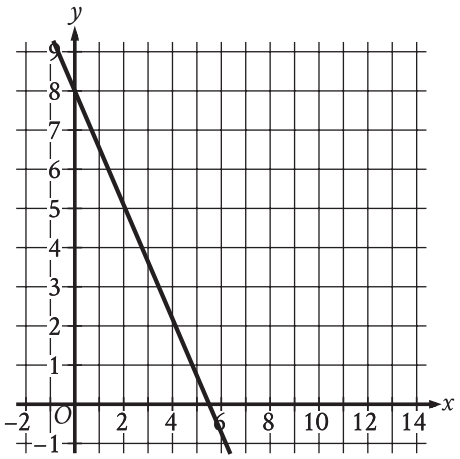
B) For example,

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

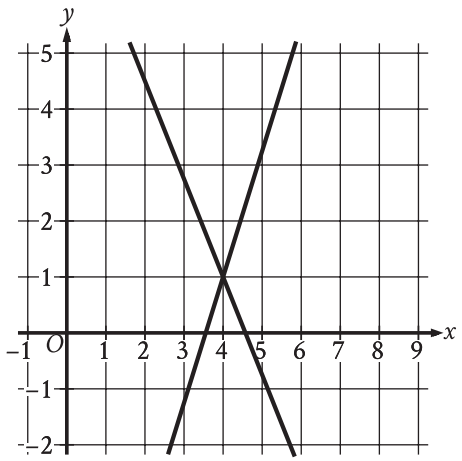
If $4x - 28 = -24$, what is the value of $x - 7$?

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 ?