CollegeBoard

The Cognitively Complex Thinking Required by Select Digital SAT® Suite Questions

College Board
January 2024

The Cognitively Complex Thinking Required by Select Digital SAT Suite Questions

January 2024

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

Dona Carling, Michael Gosche, Juliette Cabrera | **Math section analysis**

Jay Happel | **Sample analysis**

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

Vidlet, Inc. | **Cognitive interviews**

Suggestion Citation:
College Board. 2024. *The Cognitively Complex Thinking Required by Select Digital SAT Suite Questions*. New York: 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

Executive Summary

This report presents and discusses the results of a 2023 cognitive interview study conducted by College Board, with the assistance of vendor Vidlet, Inc., involving select questions from the digital SAT® Suite of Assessments. Twenty-six high school juniors and seniors volunteered to think aloud as they answered a set of twenty Reading and Writing section questions, while another twenty-three students participated in thinking aloud through a set of twenty Math section questions. Questions from both sections were chosen to be broadly representative of the sections' designs, including key skill/knowledge elements, question difficulty levels, subject areas, question formats (for Math), and text complexity levels (for Reading and Writing). Each participant engaged in a one-on-one interview session conducted via Zoom, wherein students were briefed on the task by a trained interviewer, experienced modeling of thinking aloud by the interviewer, had one or more opportunities to practice thinking aloud themselves, and then conveyed as much as possible about their concurrent thoughts as they worked through and attempted to answer a set of digital SAT Suite test questions.

Transcripts were produced from these interview sessions and analyzed qualitatively and quantitatively by College Board assessment and subject matter experts.

Qualitatively, each student's response to each test question was coded against a set of required (Reading and Writing) or expected (Math) behaviors. These behaviors, predefined by the College Board research team, described the aspects of cognitively complex thinking various question types are intended to elicit. Each student participant was judged by the researchers to have or have not demonstrated each of these behaviors in their response to the questions, and their responses were coded correspondingly. Vignette candidates of students exhibiting these behaviors and, in the process, demonstrating exemplary (if not necessarily perfect) thinking through a given question were also identified during the coding stage.

In quantitative terms, the College Board researchers tabulated several statistics from the coding. The most important metric for each Reading and Writing and Math question is referred to in this report as the *differential*. This differential is the

arithmetic difference between (1) the number of students who answered a given test question correctly and (2) the number of students who both answered the question correctly and also demonstrated all required (Reading and Writing) or at least one expected (Math) behavior. A low differential—one of 5 or lower—was deemed evidence of a given test question having performed as intended, as the majority of students would have demonstrated requisite elements of cognitively complex thinking in line with the question type's intended construct (i.e., the academic concept the question type is trying to assess students' attainment of). A higher differential, by contrast, was suggestive that a given question wasn't performing as intended, though mitigating factors may have led the researchers to conclude that the question was still capable of eliciting aspects of cognitively complex thinking.

All examined Reading and Writing questions and the vast majority (85 percent) of examined Math questions performed as intended, with differentials from 0 to 5. Two Math questions had differentials greater than 5, but the qualitative evidence suggests that students were still exhibiting aspects of cognitively complex mathematical reasoning. A third Math question was answered correctly by no student, so although it technically had a differential of 0, it was considered an outlier. Vignettes of student performance associated with each of the forty questions supply additional evidence that the questions elicited cognitively complex thinking from student participants.

The key finding of this study is strong confirmation of the hypothesis that the digital SAT Suite assessments are capable of eliciting cognitively complex thinking from student test takers. This is important because, first, a large body of evidence supports the conclusion that students need to be able to engage in such thinking to be college and career ready (i.e., prepared to succeed in college or workforce training programs without remediation) and, second, because the U.S. Department of Education requires states using the digital-suite tests (or other off-the-shelf large-scale standardized assessments) as part of their education accountability systems to supply evidence that the tests are capable of eliciting such thinking. Based on the findings reported here, policymakers should have high confidence that the tests of the digital SAT Suite of Assessments satisfy these criteria. In addition, the results and the methodology laid out in this report may be useful to researchers interested in evaluating the cognitive demands of large-scale standardized assessments.

Section 1: Introduction

Several converging factors argue persuasively for the need for tests of college and career readiness to assess higher-order or cognitively complex thinking in students. First and foremost, evidence (e.g., College Board 2019) indicates that postsecondary educators expect incoming students to already be able to demonstrate such skills as drawing reasonable inferences, making appropriate connections, analyzing arguments, citing textual evidence, and applying information and problem-solving strategies to novel situations. Second, federal peer review guidelines for state educational accountability systems (U.S. Department of Education 2018) require evidence that assessments of college and career readiness measure cognitively complex thinking in students. Third, specifications documentation for various testing programs designed to measure college and career readiness, including the digital SAT Suite of Assessments (College Board 2023a), claim that their test items assess the complexity of thinking required of students to be college and career ready.

This report presents the results of a 2023 study conducted by College Board, with the assistance of vendor Vidlet, Inc., to ascertain whether select test questions of the digital SAT Suite, which comprises the SAT, PSAT/NMSQT®, PSAT™ 10, and PSAT™ 8/9 college and career readiness assessments, are capable of eliciting cognitively complex thinking from student test takers. A positive finding would be important because it would offer evidence that the digital SAT Suite tests (1) measure important college and career readiness prerequisites, (2) are appropriate for use as part of state educational accountability systems, and (3) conform to College Board's own claims for their tests, as laid out in specifications documentation (College Board 2023a).

The principal mechanism of this study, which closely follows the approach used in an earlier project involving the paper-based SAT Suite (College Board and HumRRO 2020), is the use of cognitive interviews with a sample of high school juniors and seniors. During these interviews, which were prepared jointly by College Board and Vidlet and conducted by trained Vidlet staff, participants were asked to think aloud—that is, verbalize any and all of their thoughts—as they worked through either a set of twenty digital SAT Suite Reading and Writing questions or a set of twenty digital SAT Suite Math questions. Twenty-six students

participated in the Reading and Writing component of the study, while twenty-three participated in the Math component. These participants' interviews were transcribed, and College Board content experts coded the transcripts against a set of required or expected cognitive behaviors, which varied by test section and question type. Whether a given participant had answered a given question correctly was also recorded.

Two related forms of analysis of the coded data are presented in this report. On the quantitative side, a derived statistic called the *differential* was calculated as a measure of the extent to which each sampled digital SAT Suite test question performed as expected. For a given test question, the differential was found using the formula $D = C - A$, where D is the differential, C is the number of participants answering a given question correctly, and A is the number of participants who both (1) answered correctly and (2) demonstrated all required (Reading and Writing) or at least one expected behavior (Math). The differential thus represents the arithmetic difference between the total number of participants answering a given question correctly and the number of those participants who also enacted the question type's construct by demonstrating particular behaviors.

A 0 to low differential (up to 5) would be interpreted as evidence that the question performed as intended, as the vast majority of students answering correctly would have also demonstrated the requisite behavior(s). A moderate to high differential, by contrast, would be interpreted as suggestive of a potential weakness in a question, as it would have proved to be possible for numerous students to have answered the question correctly without exhibiting the requisite behavior(s). A differential of 6 or higher wouldn't be proof that a question-level flaw existed, but it would prompt close scrutiny and discussion by the College Board researchers. In addition, a differential of 5 or lower might still be a concern if no or few participants answered the question correctly, as the metric presumes that a substantial proportion of participants were able to correctly answer.

On the qualitative side, vignettes of exemplary student performance were identified while the transcripts were being coded for demonstration of behaviors. These vignettes serve to concretize and illustrate the kinds of cognitively complex thinking participants were provoked into demonstrating by various digital SAT Suite test questions.

The results in both Reading and Writing and in Math offer strong evidence that the sampled digital SAT Suite test questions, which are broadly representative of the tests' designs, are capable of eliciting cognitively complex thinking from students. Zero to low differentials were associated with all questions in Reading and Writing and with the vast majority (85 percent) of questions in Math, and vignettes were found and are presented in this report to exemplify this thinking. The behavior of the three nonconforming Math questions—two with differentials above 5 and one with a 0 differential but no participants answering correctly—is also analyzed and presented in Section 5: Discussion, the conclusion being that these questions still elicited aspects of cognitively complex thinking.

Structure of This Report

Section 2: Literature Review provides an overview of the research literature in support of the use of cognitive interviews and think-aloud protocols as valid sources of data concerning otherwise typically unobservable mental processes. Section 3: Methodology addresses in more detail the approach used in this study, including an account of the process steps followed and an analysis of the student samples used. Section 4: Results details the study's findings, first in Reading and Writing and then in Math. This section includes quantitative analysis of the test questions in each subject area by type as well as one or more interview vignettes for each question. Section 5: Discussion evaluates the study's findings. Section 6: Implications addresses the significance of the findings for both policymakers and researchers. Following Section 7: Conclusion and a full list of references, the appendix provides documents relevant to the study, including recruitment materials, a sample protocol, and student practice items.

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, Johnson, and Stephens 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, Woodward, and Li (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 "notably different" ways (215) from 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, Imamoğlu, and Katmer Bayraklı (2017) also used verbal protocol analysis to examine students' approaches to math problem-solving, in this case involving 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 (139–140).

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 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, English learners and native speakers, and students who are neurodivergent as well as students who aren't.

Verbal Protocols as Data in Research of Large-Scale Assessment Designs

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 assessments. One such study is that of Johnstone, Bottsford-Miller, and Thompson (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 particular 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, Liu, Altman, and Thurlow (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 (233):

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."

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 long-term memory) 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, Johnson, and Stephens (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, Johnson, and Stephens 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 primarily 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, Bottsford-Miller, and Thompson (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 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 semistructured postexperience interviews with participants, as a secondary data source while paying heed to Taylor and Dionne's (2000) recommendations for limiting reactivity in questioning; however, as will be subsequently discussed, study timing limitations led to these interviews not being conducted in several cases, and the inconsistently available results of these questions haven't been analyzed. Third, the initial instructions given to participants for the concurrent verbal protocols were kept as simple and nondirective (in Taylor and Dionne's [2000, 415] words, as "infrequent and neutral") 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 digital 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, although a specific "silent" control group wasn't employed, the study does make use of several checks on the typicality of participants' responses. These checks include the selection of participant samples consistent with the demographic and achievement distribution of the typical digital SAT Suite test-taking population and, when necessary, comparison of participants' verbalizations with their actual recorded responses to the examined test questions.

Section 3: Methodology

Test Question Selection

College Board subject matter experts began the research process for this study by identifying sets of digital 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 the digital SAT Suite tests—the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9—are intentionally similar (College Board 2023a), 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. During this selection process, the subject matter experts decided to exclude 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.

Twenty Reading and Writing questions and twenty Math questions were ultimately selected for study. These questions were drawn from actual digital SAT Suite item pools rather than developed specifically for this study and were intended to be representative of questions students might encounter on test day. The individual test questions had previously undergone rigorous internal quality control checks to ensure their content soundness (accuracy) as well as their appropriateness for use with secondary-level students in a large-scale, high-stakes standardized assessment of their college and career readiness. Because this study was conducted prior to the domestic launch of the digital SAT Suite in the 2023–2024 academic year, some Reading and Writing test questions hadn't been previously pretested; in those cases, College Board test developers provisionally assigned them to difficulty levels (i.e., *performance score bands*, defined below) based on expert judgment. (This limitation is further discussed in Section 6: Implications.)

Collectively, the Reading and Writing and Math question sets were intended to represent a wide range of skill/knowledge testing points, subject areas, question difficulty levels, text complexities, and question formats consistent with the tests' designs, although, as discussed below, especially low-difficulty questions

were intentionally omitted from the study. It should be noted that while broadly representative, neither set was designed to be a full testing experience (i.e., a *test form*) such as students would encounter on test day for any of the digital-suite programs.

Table 1 breaks down the characteristics of the digital SAT Suite questions included in the study. Each question has several characteristics:

- **Test section.** Reading and Writing or Math
- **Q#.** Question number (1–20)
- **Content domain.** One of the major conceptual divisions within each of the two test sections. Represented Reading and Writing content domains were Craft and Structure, Information and Ideas, and Expression of Ideas, while a fourth area, Standard English Conventions, was excluded, as discussed above. All four Math content domains—Algebra, Advanced Math, Problem-Solving and Data Analysis, and Geometry and Trigonometry—were represented.
- **Skill/knowledge testing point.** The skill or knowledge element targeted by the question (e.g., Words in Context in Reading and Writing; Unit Rates in Math)
- **Subject area.** The content area sampled by the question. In Reading and Writing, these subject areas are literature, history/social studies, the humanities, and science. In Math, some questions are grounded in social studies or science contexts or in real-world topics, while others ("None") assess "pure" math absent context.
- **Performance score band (PSB).** A numerical rating of a question's statistical difficulty aligned to the test sections' scales. Questions in PSBs 1 to 3 are considered easy and are associated with Reading and Writing section scores from 200 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 set included questions ranging in PSB from 3 to 7; questions in PSBs 1 and 2 were excluded from consideration, as the research literature (e.g., Bainbridge and Sanderson 1995) suggests that such relatively cognitively simple tasks are unlikely to elicit much conscious thought at all.
- **Question format.** All Reading and Writing questions, both in the study and on the actual digital 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.

Table 1. Digital SAT Suite Reading and Writing and Math Questions Studied.

Test Section	Q #	Content Domain	Skill/Knowledge Testing Point	Subject Area	Performance Score Band (PSB)	Question Format
Reading and Writing	1	Craft and Structure	Words in Context	Science	5	MC
	2			History/social studies	7	MC
	3		Text Structure and Purpose	Literature	3	MC
	4			Science	4	MC
	5		Cross-Text Connections	Humanities	6	MC
	6			Humanities	7	MC
	7	Information and Ideas	Central Ideas and Details	Humanities	4	MC
	8			Literature	4	MC
	9		Command of Evidence: Textual	Literature	6	MC
	10			Science	3	MC
	11		Command of Evidence: Quantitative	History/social studies	4	MC
	12			Humanities	6	MC
	13			Science	5	MC
	14			History/social studies	4	MC
	15		Inferences	Science	6	MC
	16			History/social studies	4	MC
	17	Expression of Ideas	Rhetorical Synthesis	Science	4	MC
	18			Humanities	5	MC
	19		Transitions	History/social studies	4	MC
	20			History/social studies	3	MC
Math	1	Algebra	Linear Functions: Interpret	Science	3	MC
	2	Problem-Solving and Data Analysis	Fit a Model	None	4	MC
	3	Advanced Math	Nonlinear Functions	Real-world	4	MC
	4	Algebra	Linear Functions/ Inequalities in One Variable: Create and Use	Science	4	MC
	5	Algebra	Linear Functions/ Inequalities in One Variable: Create and Use	Real-world	4	SPR
	6	Advanced Math	Nonlinear Functions	Science	6	MC
	7	Problem-Solving and Data Analysis	Unit Rates	Science	5	SPR
	8	Geometry and Trigonometry	Equation of a Circle	None	5	MC
	9	Problem-Solving and Data Analysis	Probability	Real-world	5	SPR
	10	Geometry and Trigonometry	Special Right Triangles	None	7	MC
	11	Problem-Solving and Data Analysis	Sample Proportion	Social studies	6	MC
	12	Advanced Math	Make Connections	None	6	MC
	13	Geometry and Trigonometry	Volume	None	6	MC

(continued)

Test Section	Q #	Content Domain	Skill/Knowledge Testing Point	Subject Area	Performance Score Band (PSB)	Question Format
Math *(continued)*	14	Problem-Solving and Data Analysis	Derived Units	Science	7	MC
	15	Advanced Math	Determine Conditions	None	7	MC
	16	Algebra	Linear Equations in Two Variables: Make Connections	None	7	MC
	17	Advanced Math	Nonlinear Equations: Solve	None	6	SPR
	18	Problem-Solving and Data Analysis	Percentages	Real-world	7	MC
	19	Advanced Math	Rewrite	None	7	MC
	20	Algebra	Linear Systems: Determine Conditions	None	7	SPR

Note the different organization of Reading and Writing and Math sections above: while the study's Reading and Writing questions were grouped by content domain, its Math questions were grouped by difficulty level. These two approaches broadly reflect the presentation order in which actual test takers would be administered the questions in the two sections.

As a group, the twenty sampled Reading and Writing questions represented three of the section's four content domains (with Standard English Conventions being excluded), all major skill/knowledge categories within those three domains, all four subject areas sampled in the section, and a range 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 categories within those domains, all three subject areas sampled in the section as well as questions set outside of context, a range of difficulty from 3 (easy) to 7 (hard), and both question formats used in the section (multiple-choice and student-produced response). Comparatively, the Math questions skewed harder, on average, than did their Reading and Writing counterparts, a circumstance discussed in Section 6: Implications.

Question-Level Construct Definition

The same College Board subject matter experts who selected the questions for the study next identified *constructs* for the questions by skill/knowledge testing point. These constructs, in the form of lists of behaviors demonstrable by test takers, described the kinds of thinking students were expected to exhibit if they approached answering the questions as intended. For each Reading and Writing question type (e.g., Words in Context), staff developed a list of behaviors test takers were required to exhibit in order to answer as intended. Because many Math questions include, by design, multiple legitimate pathways for test takers to pursue in answering, these behaviors were defined as expected rather than required, and students needed only to exhibit at least one of them to be considered as having enacted the construct. Additionally, both Reading and Writing and Math staff identified common behaviors that skillful test takers may or may not exhibit; these optional behaviors were coded for (see below) but aren't analyzed in this report.

Protocol Development

The lead author of this study next developed closely parallel Reading and Writing and Math protocols for conducting the cognitive interviews in which students would participate (see exhibit 4 in the appendix for a sample). These protocols were designed as guides for the interviewers conducting sessions with students. The guides included instructions for conducting the sessions, scripts for

interviewers to follow, and suggested probes and prompts that interviewers could use during sessions should students lapse into silence while working through the test questions. Consistent with best practices (as discussed in Section 2: Literature Review), interviewers 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 students' responses. Interviewers 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, Inc., the vendor contracted by College Board to conduct the interviews on its behalf, was briefed and trained on the protocol and given multiple opportunities to provide feedback and suggest refinements.

Sample Recruitment, Selection, and Characteristics

SAMPLE RECRUITMENT AND SELECTION

The main goal of the College Board–developed recruitment approach for this study was to select student samples for both the Reading and Writing and Math cognitive interview activities that closely mimicked the typical digital SAT Suite test-taking population.

College Board staff developed the recruitment approach and materials used in the study. The recruitment method was a direct email (see exhibit 1 in the appendix) to eleventh- and twelfth-grade students who had previously taken the PSAT/NMSQT / PSAT 10 or SAT tests and had elected to receive emails from College Board. It was important that participating students had prior College Board test scores, as a key sample selection criterion was ensuring, to the extent possible, that the study's samples reflected the widest possible span of achievement on the tests and behaved like a typical digital SAT Suite test-taking population. Prior PSAT/NMSQT / PSAT 10 scores could be used as good proxies for SAT achievement levels, given that the tests are on the same vertical scale. (See College Board 2023a, section 2.2.8.2, for more details on the digital SAT Suite's vertical scale.)

The recruitment email described the study and its eligibility requirements, including a willingness and ability to participate in a roughly ninety-minute virtual session with an interviewer. A link in the email navigated interested students to an online form (see exhibit 3) that provided more information about the study, including the purpose for conducting the study, an indication of the voluntary nature of participation, and a description of the activity students would be asked to participate in. They were also informed about the study's incentive, which was a $150 gift card to be delivered on successful completion of the activity. (As noted in the headers for exhibit 1 and exhibit 2, College Board and Vidlet jointly determined before the interviews were conducted that the $100 compensation referred to in the recruitment materials should be increased to $150 to better reflect the time and effort required to successfully complete the activity.)

Interested students answered questions about their eligibility and availability and were also asked to provide their first and last names as well as their email

address. These student-provided data were matched to records in the College Board database; demographics and test score history were drawn from this database, and only students who provided data that could be matched to data in the College Board database were deemed eligible, as the study required evidence of their previous test scores. Students were informed that their names and email addresses would be shared with our subcontractor, Vidlet, for scheduling their session and obtaining consent forms. College Board staff then selected eligible candidates in accordance with the goal of having highly diverse samples as consistent as possible with the characteristics of the typical digital-suite test-taking population.

Upon receipt of the list of these eligible participants, Vidlet staff contacted the sampled students through targeted emails. After communication was established and participation confirmed, all eligible students were randomly assigned to participate in either the Reading and Writing or Math cognitive interviews. After initial confirmation, Vidlet requested and collected documentation from students, including a consent form (see exhibit 3) acknowledging voluntary participation in the study, and confirmation of the interview time, date, and location. Vidlet obtained parent or guardian consent for students under eighteen years of age.

In total, fifty thousand students who had previously taken the SAT, PSAT/NMSQT, or PSAT 10 and who had opted in to receive College Board communications were emailed about the study opportunity. Of those contacted, 198 (0.4 percent) completed the application process, 53 were accepted into the study, and 49 ultimately participated.

SAMPLE CHARACTERISTICS

This subsection presents demographic and score information for the twenty-six Reading and Writing and twenty-three Math cognitive interview participants. Demographic information is presented separately for the participants in each sample group (Reading and Writing; Math) as well as for all participants.

The forty-nine participants in the study were nearly evenly split in terms of the cohort year from which they came. Table 2 shows that twenty-four participants (thirteen in Reading and Writing and eleven in Math) were from the class of 2023 (i.e., were high school seniors at the time of the study), while the remaining twenty-five (thirteen in Reading and Writing and twelve in Math) were from the class of 2024 (i.e., were current high school juniors).

Table 2. Cognitive Interview Participants by Cohort Year.

	Reading and Writing		Math		Total	
Cohort	#	%	#	%	#	%
Class of 2023	13	50	11	48	24	49
Class of 2024	13	50	12	52	25	51
Total	26		23		49	

Table 3 shows that when it comes to gender, slightly more participants identifying as female (fifty-four) than as male (forty-six) took part in the Reading and Writing cognitive interviews, while slightly more participants identifying as male (fifty-two) than as female (forty-eight) took part in the Math interviews (as noted

above, student assignment to Reading and Writing or Math was random). Overall, participants were nearly evenly divided in terms of gender, with twenty-five identifying as female, twenty-four identifying as male, and none identifying as another gender. Table 3 also shows the percentages by gender of all SAT test takers nationwide (College Board 2022, 4). The total group of participants in these cognitive interviews very nearly matched the national SAT population ("National Percentage") when it comes to gender representation, but the Reading and Writing and Math groups each deviated from it slightly.

Table 3. Cognitive Interview Participants by Gender.

	Reading and Writing		Math		Total		National
Gender	#	%	#	%	#	%	Percentage*
Female	14	54	11	48	25	51	51
Male	12	46	12	52	24	49	48
Another	0	0	0	0	0	0	<1
Total	26		23		49		

*Source: College Board 2022, p. 4.

When it comes to racial/ethnic representation, the participants in the two sample groups differed from the national SAT test-taking population but in different ways.

Table 4 shows that compared to all SAT test takers nationwide, the percentages of Reading and Writing study participants identifying as Asian and White were higher, while that of participants identifying as two or more races/ethnicities was the same and those of participants identifying as Black/African American and Hispanic/Latino were lower. As for the Math participants, the percentages identifying as Asian, White, and from two or more races were lower than those in the national population, while the percentage identifying as Hispanic/Latino was greater than that of the national population. In both samples, the percentages identifying as American Indian/Alaska Native and Native Hawaiian/Other Pacific Islander were similar to the national percentages. In both groups, the percentage of participants opting not to report their race/ethnicity was greater than the corresponding national percentage, and proportionally more Math participants than Reading and Writing participants chose not to report.

Table 4. Cognitive Interview Participants by Race/Ethnicity.

	Reading and Writing		Math		Total		National
Race/ethnicity	#	%	#	%	#	%	Percentage*
American Indian/ Alaska Native	0	0	0	0	0	0	1
Asian	4	15	1	4	5	10	10
Black/African American	2	8	3	13	5	10	12
Hispanic/Latino	4	15	6	26	10	20	23
Native Hawaiian/Other Pacific Islander	0	0	0	0	0	0	0
White	12	46	8	35	20	41	42
Two or more races/ ethnicities	1	4	0	0	1	2	4
No response	3	12	5	22	8	16	8
Total	26		23		49		

*Source: College Board 2022, p. 4.

Participants were also asked to identify the first language(s) they'd learned (table 5). The percentage of Reading and Writing participants reporting English only (58 percent) was comparable to the national percentage (59 percent), while the percentage of Math participants (52 percent) was slightly less. The percentages of both Reading and Writing and Math participants reporting English and another language (15 and 17 percent, respectively) as their first languages were slightly less than the national percentage (19 percent). The percentage of Reading and Writing participants reporting a language other than English as their first language (4 percent) was less than the national percentage (10 percent), while the percentage of Math participants reporting this (13 percent) was slightly greater. In both groups, the percentage choosing not to report their first language(s) (23 and 17 percent, respectively) was greater than the national percentage (12 percent).

Table 5. Cognitive Interview Participants by First Language(s) Learned.

	Reading and Writing		Math		Total		
First Language	**#**	**%**	**#**	**%**	**#**	**%**	**National Percentage***
English only	15	58	12	52	27	55	59
English and another language	4	15	4	17	8	16	19
Another language (other than English)	1	4	3	13	4	8	10
No response	6	23	4	17	10	20	12
Total	26		23		49		

*Source: College Board 2022, p. 4.

Participants were also asked to report their current best language(s) (table 6). Over three-quarters of Reading and Writing participants (77 percent) and over half of Math participants (57 percent) reported English only. Twelve percent of Reading and Writing participants and 30 percent of Math participants said English and another language were their best languages. No participants said a language other than English was their best. Twelve percent of the Reading and Writing participants and 13 percent of the Math participants chose not to respond. (National percentages aren't available for comparison on this dimension.)

Table 6. Cognitive Interview Participants by Best Language.

	Reading and Writing		Math		Total	
Best Language	**#**	**%**	**#**	**%**	**#**	**%**
English only	20	77	13	57	33	67
English and another language	3	12	7	30	10	20
Another language (other than English)	0	0	0	0	0	0
No response	3	12	3	13	6	12
Total	26		23		49	

Sampling for the cognitive interview study was based in part on the previous SAT, PSAT/NMSQT, or PSAT 10 scores of those selected, with the goal of having the cognitive interview samples reflect a wide range of achievement as measured by the SAT and these PSAT-related assessments. Table 7 shows that for both

samples, over half (mostly from the class of 2024 cohort) had previous PSAT/NMSQT or PSAT 10 scores, and fewer than half had previous SAT scores.

Table 7. Cognitive Interview Participants by Digital SAT Suite Test Previously Taken.

	Reading and Writing		Math		Total	
Testing Program	**#**	**%**	**#**	**%**	**#**	**%**
PSAT/NMSQT / PSAT 10	15	58	13	57	28	57
SAT	11	42	10	43	21	43
Total	26		23		49	

Table 8 and table 9, respectively, show the means, standard deviations, minima, and maxima of the paper and pencil Evidence-Based Reading and Writing section scores of the Reading and Writing cognitive interview participants and of the paper and pencil Math section scores of the Math cognitive interview participants. The tables also show national mean scores. All SAT section scores, whether paper-based or digital, are on a scale from 200 to 800, while all PSAT/NMSQT and PSAT 10 section scores, regardless of mode, are on a scale from 160 to 760. (For details on how the performance score bands were determined, see College Board 2023b.)

Table 8. Cognitive Interview Participants: Reading and Writing—Prior Mean Section Scores.

Testing Program	#	Mean	SD	Min	Max	National Mean*
PSAT/NMSQT / PSAT 10	15	523	113	360	730	481
SAT	11	608	139	380	780	529
Total/Average	26 (total)	559 (average)	129 (average)			

* Sources: College Board 2022, p. 9 (PSAT/NMSQT / PSAT 10 mean); College Board 2022, p. 6 (SAT mean)

Table 9. Cognitive Interview Participants: Math—Prior Mean Section Scores.

Testing Program	#	Mean	SD	Min	Max	National Mean*
PSAT/NMSQT / PSAT 10	13	461	36	410	530	467
SAT	10	522	159	370**	730	521
Total/Average	23 (total)	487 (average)	110 (average)			

* Sources: College Board 2022, p. 9 (PSAT/NMSQT / PSAT 10 mean); College Board 2022, p. 6 (SAT mean)

** As discussed below, one participant with a prior SAT Math score of 200 (the section minimum) was mistakenly included in the sample. The score of the second-lowest-performing student is reported here as the "true" minimum.

Participants as a group in the Reading and Writing cognitive interviews were higher performing than was the class of 2022, which took the paper-based versions of the SAT Suite assessments. The mean PSAT/NMSQT / PSAT 10 paper and pencil Evidence-Based Reading and Writing section score of the fifteen students who reported those was 523, compared to the national mean PSAT/NMSQT / PSAT 10 section score of 481, and the minimum PSAT/NMSQT / PSAT 10 section score reported was 360. (Minimum national section scores for either the SAT or the

PSAT/NMSQT / PSAT 10 are typically the lowest reported score: 160 for PSAT/NMSQT / PSAT 10 and 200 for the SAT.) The mean paper and pencil SAT Evidence-Based Reading and Writing section score of the eleven students who reported those was 608, while the national mean was 529, and the minimum paper and pencil SAT section score reported was 380. Participants in the Math cognitive interviews, on the other hand, were similar in performance to or slightly lower performing than the class of 2022. The mean paper and pencil PSAT/NMSQT / PSAT 10 Math section score was 461 and the mean SAT Math section score was 522, compared to national means of 467 and 521, respectively.

Note that a student with a past SAT Math section score of 200—the minimum—was included in the sample. It's most likely that this score represents the student for some reason having not attempted the Math section when they took the SAT. This student's inclusion in the sample was an oversight, and future College Board studies in this vein will correct for that error. Table 9 thus reports the second-lowest SAT Math section score (370) as the "true" minimum.

Something else to note about the Math participants is the much smaller range of prior PSAT/NMSQT / PSAT 10 scores than SAT scores they represent. Excluding the participant whose prior SAT Math section score was 200, the nine participants with previous paper-based SAT scores had Math section scores spanning a broad expanse of the score range (370–730). The fifteen participants with prior paper-based PSAT/NMSQT / PSAT 10 scores, despite having a mean nearly equal to the national mean, spanned only thirteen scale score values: 410–530. This indicates that the Math sample tended to overrepresent average section scores and underrepresent higher- and lower-achieving students.

Finally, table 10 and table 11 show the distributions of Reading and Writing and Math cognitive interview participants, respectively, by the performance score bands (PSBs) their SAT or PSAT/NMSQT / PSAT 10 scores fall into. Note that the score ranges for the bands used below are the same as those used for the questions themselves as proxies of question difficulty.

For Reading and Writing, the most populous performance score band was 680–800, at 30.8 percent of participants, followed by 420–480 at 26.9 percent, 610–670 at 15.4 percent, and 490–540 also at 15.4 percent. Approximately 90 percent of all Reading and Writing participants fell into one of these bands. For Math, 26.1 percent of participants fell into score band 470–540, followed by 420–460 at 21.7 percent, 680–800 at 17.4 percent, and 550–600 at 13.0 percent. About 80 percent of Math participants fell into one of these bands. (National percentages aren't available for comparison on these dimensions.)

Table 10. Cognitive Interview Participants: Reading and Writing—Prior Achievement by Performance Score Band (PSB).

Performance Score Band (Section Score Range)	#	%
1 (<370)	1	3.8
2 (370–410)	1	3.8
3 (420–480)	7	26.9
4 (490–540)	4	15.4
5 (550–600)	1	3.8
6 (610–670)	4	15.4
7 (680–800)	8	30.8

Table 11. Cognitive Interview Participants: Math—Prior Achievement by Performance Score Band (PSB).

Performance Score Band (Section Score Range)	#	%
1 (<370)	2	8.7
2 (370–410)	1	4.3
3 (420–460)	5	21.7
4 (470–540)	6	26.1
5 (550–600)	3	13.0
6 (610–670)	2	8.7
7 (680–800)	4	17.4

In short, the Reading and Writing and Math samples in this cognitive interview study were largely like their respective national populations on several demographic variables, including gender, race/ethnicity, and self-reported first language(s), with the observed variance largely attributable to the small sample sizes and the voluntary nature of participating in this activity. In terms of prior SAT Suite test achievement, the Reading and Writing sample was somewhat more able than is true nationally, while the Math sample was comparably or slightly less able. Given the recruitment constraints of the study, the College Board researchers deemed the samples sufficiently comparable to their national counterparts to make them appropriate for analysis.

Cognitive Interviews

Once recruitment for the study was completed, College Board's vendor, Vidlet, began setting up interviews with each participant. After recruited students had confirmed their participation in the study, the Vidlet team collected consent forms and scheduled students for ninety-minute sessions. Sessions were conducted one-on-one between a Vidlet interviewer and a student; no College Board personnel were present except on one occasion, when the study lead sat in (with his camera off and with the participant's knowledge) for interviewer training purposes.

During each session with participants, the Vidlet interviewer executed the protocol that College Board had previously developed and that the vendor had been trained on. In brief, each interview consisted of the following elements:

- Welcome
- A briefing on the study, participants' role in it, and participants' right to decline to answer questions and/or stop participating at any time
- Modeling by the interviewer, who, reading from a script, demonstrated the think-aloud process on a sample question (not analyzed in the study)
- One or (at the interviewer's discretion) two chances on the participant's part to practice thinking aloud on sample questions (not analyzed in the study)
- A thinking-aloud period of approximately seventy minutes, during which the participant worked through as many as twenty Reading and Writing or Math questions
- A set of debriefing questions
- Wrap-up

After several interviews had been conducted, College Board and Vidlet jointly determined that participant noncompletion of the think-aloud portion of the activity due to time constraints was an emergent issue. College Board directed Vidlet to allow participants to continue thinking aloud into the time allotted for debriefing questions. Thus, while some participants were able to think aloud through all twenty questions and then answer the debriefing questions, others weren't. As they represented an incomplete sample, responses to the debriefing questions aren't analyzed in this report.

Interview sessions were conducted remotely via Zoom and were videorecorded with the participants' awareness and consent. Participants were asked to share their device screen with the interviewer so that the latter could follow along as a given participant worked through the Reading and Writing or Math questions.

Test questions themselves were presented to the participants via a Qualtrics survey, as technical limitations in place at the time precluded the use of the actual test delivery platform designed for the digital SAT Suite tests. The Qualtrics platform was set up to mimic as closely as possible the actual operational test delivery interface to minimize platform effects. Each participant was assigned a unique, non–personally identifying code (e.g., "student RW1") and given a unique link at which to connect to the survey. Each participant's survey included access to the test section directions, to which students could return at any time. The Math section interview also included access to the sheet of common mathematical formulas provided to actual test takers as well as connection to the Desmos® Graphing Calculator, an online tool that test takers can employ during operational testing; participants were also allowed to use their own handheld calculator provided that it complied with College Board's acceptable device policy (**https://satsuite.collegeboard.org/digital/what-to-bring-do/calculator-policy**). Each test question was presented to participants on a separate "page," and participants were able, if desired, to navigate freely among the questions (though they were encouraged to proceed linearly as much as possible). The Qualtrics platform allowed participants to click an answer choice (for multiple-choice questions) or to enter a self-generated answer (for Math student-produced response questions); these data were recorded and occasionally used by the researchers to clarify or confirm ambiguous verbal responses.

Vidlet conducted the interviews from April 12 to April 30, 2023. Vidlet subsequently produced verbatim transcripts of each interview and submitted them, along with the associated video recordings and certain ancillary data, to College Board. These ancillary data included reports of any irregularities during interviews (e.g., late arrival, connectivity issues) and records of participants' start and stop times on each question. These ancillary materials aren't analyzed in this report, but no major irregularities were reported.

Coding and Analysis

CODING

The lead College Board researcher uploaded the interview transcripts 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. Team members were also directed to code as "vignette candidates" any participant response that exhibited the required/expected behaviors and that served to illustrate well-reasoned responses without significant errors, omissions, or uncorrected missteps. As a supplement, the team concurrently recorded, in Microsoft Excel, whether each participant had exhibited the required/expected behaviors for each of the questions; these Excel spreadsheets served as the basis for tabulating the statistics presented in Section 4: Results. The coding process resulted in approximately four thousand codes being assigned to forty-nine participants' interactions with the forty studied questions.

ANALYSIS

The College Board researchers analyzed the coded data both qualitatively and quantitatively. Qualitative analysis consisted of assigning behavior codes to participants' transcribed responses and identifying and selecting illustrative examples (*vignettes*) of participants' cognitively complex thinking in accordance with the behaviors defined for each question type.

The most important aspect of the quantitative analysis, which builds on the preceding qualitative behavior coding, is a derived statistic developed for this and similar studies (College Board and HumRRO 2020) called the *differential*. This statistic is determined for each studied test question using the formula $D = C - A$, where D is the differential, C is the number of participants answering a given question correctly, and A is the number of participants who both (1) answered correctly and (2) demonstrated all required (Reading and Writing) or at least one expected behavior (Math).

This statistic, D, calculated for each test question in the study, is posited to express the extent to which a given question functions as intended. C in the above expression simply represents the count of the number of participants who answered a given question correctly, while A represents the number of correctly answering participants who, by exhibiting all required behaviors (for Reading and Writing section questions) or at least one expected behavior (for Math section questions), have enacted the question type's construct.

Note that because of differences in the disciplines of literacy and math, the lists of required Reading and Writing behaviors included providing the best answer, while the lists of expected Math behaviors don't. This makes no difference in how the differential was calculated between test sections, as the Math calculation still considers whether participants answered a given question correctly.

Zero or low differentials—those of 5 or lower—are associated with test questions that perform as expected in eliciting cognitively complex thinking, as most if not all correctly answering participants also approached the question as intended. Moderate or high differentials, by contrast, may be suggestive of shortcomings in test questions, as substantial numbers of participants were deemed to have found ways to answer correctly without enacting the cognitively complex behaviors associated by College Board with the question type. This might be the case if, for example, participants could bypass demonstrating passage comprehension—a behavior associated with all Reading and Writing questions—and still answer

correctly. A moderate to high differential wasn't considered conclusive proof that a given test question was unable to elicit cognitively complex behavior, but it did suggest that the question warranted close scrutiny.

Two Math questions had differentials of greater than 5, while an additional Math question with a technical differential of 0 had no participants answering correctly. These three questions are analyzed in detail in Section 5: Discussion. No Reading and Writing questions had a differential above 5 or were otherwise nonconforming.

Section 4: Results

This section reports the results of the study's examination of twenty Reading and Writing questions and twenty Math questions via the cognitive interview methodology. The analysis proceeds first by test section (Reading and Writing, and then Math), followed by content domain within each test section, and finally by question type within each domain. Quantitative results by question type are first summarized, and then vignettes of successful student performance on these questions follow. The total number of sampled students answering a given question is listed in the *n*-count in the associated tables; these numbers are inconsistent from question to question due to varying response rates, an issue touched on again in Section 6: Implications. In the tables and discussions that follow, "PSB" refers to *performance score band*, the question difficulty (and student achievement) measure discussed in Section 3: Methodology.

For each question, the percentage of students answering correctly shouldn't be taken as indicative of the measured difficulty of the question when given to a pretesting sample for potential inclusion in an operational test. Sample sizes used in this study are incomparably smaller than pretesting sample sizes and may yield percentage-correct data quite different from that yielded during standard pretesting.

In the following tables displaying student performance on one or more Reading and Writing or Math test questions, color coding is used to help differentiate results, with different shades used for each quartile. Blue shading is applied to behaviors, while purple shading highlights data used in calculating the differential. The lightest shades of blue and purple represent the bottom quartile, while the darkest shades represent the top quartile.

Reading and Writing

The following subsection details the results of the Reading and Writing test questions included in the study. Note that consistent with the test section's design, questions were presented to students in approximately the order they would receive them in an actual test: questions in the Craft and Structure content domain came first, followed by Information and Ideas questions and then Expression of Ideas questions, with like question types within each content domain grouped together to reduce task switching.

CRAFT AND STRUCTURE

Words in Context Questions

Words in Context questions on the digital SAT Suite tests assess whether students can effectively determine the meaning of or skillfully use high-utility academic words and phrases in context. *High-utility academic words and phrases*, sometimes known as tier two vocabulary (Beck, McKeown, and Kucan 2013), are those commonly encountered in texts but relatively seldom in conversation (which is composed mainly of tier one vocabulary) and are neither obscure nor exclusive to one domain of knowledge, such as science (i.e., tier three vocabulary). Examples of high-utility academic vocabulary include *trait*, *observe*, *portrayal*, and *critique*. College Board refers to these words and phrases as high utility because knowledge of such vocabulary is extremely useful to unlocking the meaning of texts, especially more complex texts, across a range of subject areas.

To answer Words in Context questions as intended, students are expected to demonstrate the following 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.

Students exhibiting both these behaviors provide evidence of engaging in cognitively complex thinking that involves, first, reading and comprehending an appropriately challenging text from one of several possible subject areas and, second, choosing the answer option that either represents the most logical and precise high-utility academic word or phrase to use in that context (when the passage contains a blank to be completed) or supplies the most appropriate definition of a high-utility academic word or phrase identified in the passage. Because blank-completion questions are new to the digital SAT Suite's Reading and Writing section and are much more prevalent in the test design, the two Words in Context questions studied used that approach.

Table 12 summarizes how students performed on the two Words in Context questions included in the study.

Table 12. Student Performance on Reading and Writing: Craft and Structure—Words in Context Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors			Answered Correctly	Demonstrated Both Behaviors and Answered Correctly	Differential
			1	2	Both			
1 *n* = 25	Science	5	16 (64%)	7 (28%)	6 (24%)	7 (28%)	6 (24%)	1
2 *n* = 25	History/ social studies	7	19 (76%)	18 (72%)	17 (68%)	18 (72%)	17 (68%)	1

For this and all tables in the Results section, color gradations indicate percentage quartiles, with darker shades denoting higher percentages. Purple columns highlight data used to calculate differentials.

Table 12 indicates that both Words in Context questions included in the study performed as expected, with differentials of 1. Although relatively few students answered question 1 correctly, those who did almost always demonstrated both behaviors, while a majority of students at least demonstrated some level

of passage comprehension. Results for question 2, which more students in the sample answered correctly, follow the same pattern, with nearly all the students who answered correctly also exhibiting passage comprehension.

Vignettes from students answering correctly and demonstrating both required behaviors illustrate the kinds of complex thinking elicited by Words in Context questions on the digital SAT Suite assessments.

Question 1

Question 1, a medium-difficulty (PSB 5) question set in a science context, asks students to best complete the text (i.e., fill in the blank) with the most logical and precise word or phrase.

> Some foraging models predict that the distance bees travel when foraging will decline as floral density increases, but biologists Shalene Jha and Claire Kremen showed that bees' behavior is inconsistent with this prediction if flowers in dense patches are _____: bees will forage beyond patches of low species richness to acquire multiple resource types.
>
> Which choice completes the text with the most logical and precise word or phrase?
>
> A) immature
> B) homogeneous
> C) depleted
> D) dispersed

To answer this question correctly, students must determine from the passage that bees will extend their foraging range beyond nearby dense flower patches if these patches have "low species richness" and therefore don't offer access to "multiple resource types."

Choice B, *homogenous*, is the best answer, as it clearly indicates that when the flowers in nearby dense patches are highly similar, bees will range beyond them to look for greater resource variety. Answering this question correctly requires more than prior knowledge of the meaning of the words *immature*, *homogeneous*, *depleted*, and *dispersed* because any of these options could be meaningfully "read into" the passage, while only *homogeneous* supplies the word that logically and precisely completes the thought expressed in the passage.

Student RW17 begins their successful approach to question 1 by paraphrasing the topic of the passage, thereby indicating some conceptual understanding of the task, and restating the researchers' claim.

> So this text talks about bees and the distance they travel when foraging. And it's asking me to fill in the blank. So the sentence that it wants me to fill in, or the phrase, is the biologists "show[ed] that bees' behavior is inconsistent with this prediction if flowers in dense patches are," blank. And it says, "bees will forage beyond patches of low species richness."

Methodology Notes: Vignettes

All vignettes in this report are as close to verbatim representations of students' transcribed responses as possible. Omissions (typically made to reduce repetition of passage and/or question content) are noted by ellipses ("…"). Text in [brackets] has been inserted for clarity, most commonly to unambiguously identify answer choices. Material in "quotation marks" (and possibly including bracketed text) represents verbatim quotations from a given test passage or question. Note that answer choice letters (A–D) are provided with the multiple-choice questions to increase clarity about which choice is being referred to; because of technical limitations, the Qualtrics survey didn't include such letters, although many students supplied them on their own. Note that the genderless third person pronoun "they" is consistently used in this section to refer to individual students.

Student RW17 then uses a "plug-in" strategy in an effort to obtain a clearer sense of how each answer choice would fit into the passage's blank. In doing so, the student offers up prior-knowledge definitions of *depleted* and *dispersed* but indicates an initial lack of understanding of *homogeneous*, which is the question's best answer.

> So for A, *immature*, I would plug it in. "Bees' behavior is inconsistent with this prediction if flowers in dense patches are immature: bees will forage beyond patches of low species [richness]—" So I'm not sure if that makes sense, so I'm just going to go to B, "showed that bees' behavior is inconsistent with this prediction if flowers in dense patches are homogeneous." I don't really know what *homogeneous* means, so I'm going to hold off of that one too. C, *depleted*. "Showed that bees' behavior is inconsistent with this prediction if flowers in dense patches are depleted." So I think this one does make sense because it says that bees will forage beyond patches of low species [richness]. And that's kind of what *depleted* means. And then *disperse*[*d*] means "more spread out." So "showed that bees' behavior is inconsistent with this prediction if flowers in dense patches are dispersed." So this one might also make sense just because it goes with the low species richness.

It's worth noting that student RW17's approach initially rules in *depleted* and *dispersed* as possible correct answer choices. By design, Words in Context questions pose incorrect answer choices (*distractors*) that are at least surface plausible in terms of meaning and that can be "read into" the context without awkwardness. To answer such questions correctly, then, students must use context clues and may use other techniques, such as calls to prior knowledge and division of words into meaningful, more recognizable parts (base words and affixes), to determine the best answer.

Seemingly realizing that prior vocabulary knowledge alone isn't enough to correctly answer question 1, student RW17 returns to the context, paraphrasing the passage's main idea and restating the researchers' claim.

> So I'm either between C [*depleted*] or D [*dispersed*], so I'm just going to plug them in again. . . . So from what I'm getting, it's saying that bees will keep going beyond patches of low species [richness] to acquire different types [of resources], so I'm going to say—let me see. That "bees' behavior is inconsistent with this prediction if flowers in dense patches are"—okay.

With a clearer sense of the meaning of the passage, student RW17 uses the context to select *homogeneous*, a basic definition to which the student has recalled or perhaps inferred, and concurrently to rule out *depleted* and *dispersed*.

> So I think C and D are actually wrong because I think *homogeneous* means "the same." So I think that one makes sense because it's saying that bees will go beyond same patches so that they could go for "multiple resource types." So they don't want to stay in homogeneous patches. So I'm going to go with B.

Question 2

Question 2, a hard (PSB 7) question set in a history/social studies context, requires a similar combination of passage-based reasoning and vocabulary understanding to answer correctly. To properly answer this question, students need to understand the complex content of the passage and then select the phrase that best completes the passage.

> While scholars believe many Mesoamerican cities influenced each other, direct evidence of such influence is difficult to ascertain. However, recent excavations in a sector of Tikal (Guatemala) unearthed a citadel that shows _____ Teotihuacan (Mexico) architecture—including a near replica of a famed Teotihuacan temple—providing tangible evidence of outside influence in portions of Tikal.
>
> Which choice completes the text with the most logical and precise word or phrase?
>
> A) commonalities with
> B) animosities toward
> C) refinements of
> D) precursors of

The passage asserts that new, tangible evidence of Mesoamerican cities' influence on each other has been uncovered in the form of striking similarities between a citadel in Tikal, Guatemala, and stylistic elements used in Teotihuacan architecture.

Choice A, *commonalities with*, is the best answer here, as it appropriately indicates that the citadel in Tikal was intentionally highly similar in design to Teotihuacan structures in Mexico. Choice B, *animosities toward*, makes little sense in context, as it suggests that the Tikal builders held some sort of grudge against Teotihuacan architecture. Choice C, *refinements of*, is blocked by the passage's phrase "near replica," which suggests that those in Tikal tried to emulate rather than improve on a preexisting Teotihuacan temple. That the direction of influence went from Mexico to Guatemala and not the other way around (thus, blocking choice D, *precursors of*) is also indicated by "near replica," which establishes that the designers or builders of the citadel in Tikal were familiar with "a famed Teotihuacan temple," as well as by the passage's phrase "outside influence in portions of Tikal," which further clarifies that Tikal was the recipient of influence.

Student RW22 begins their analysis by providing a text-based rationale for the best answer.

> *Commonalities with* could make sense because if there's possible tangible evidence of influence, that would make sense if there's some commonality there because influence would mean that there is.

The student next identifies choice B, *animosities toward*, as illogical in and therefore inappropriate for the context.

> *Animosities toward*, that's more of an emotion. I don't think you can tell through architecture so much.

After considering choice C, *refinements of*, student RW22 rules out choice D, *precursors of*, using the passage's clues that architectural influence spread from Mexico to Guatemala and not the reverse.

> *Precursors of*? Does that mean the Tikal—or, no, that would mean this Mexican architecture would have—oh, no, no, no. The Tikal architecture would have been made before the Mexican architecture, which means—in portions of Tikal. Wait, Tikal architecture would be before the new thing. So I don't think—that that would mean that the Mexican architecture wouldn't influence Tikal. I don't think that makes sense.

Student RW22 ultimately chooses the best answer mainly on the basis of the passage's reference to the later Tikal architecture including an almost exact copy of a Teotihuacan temple.

> Let's see. So it's either *commonalities* [*with*] or *refinements* [*of*]. . . . I think *commonalities* [*with*] makes the most sense because they include here a near replica. They add that sentence, which makes most sense if there's commonalities. I think that's it, though it could be *refinements* [*of*].

Although student RW22 never fully rules out *refinements of* as an option, their rationale for affirmatively selecting *commonalities with* exhibits a strong sense of the passage's content and the meaning of the phrase itself.

Text Structure and Purpose Questions

The digital SAT Suite's Text Structure and Purpose questions assess whether students can successfully analyze the structure of texts, primarily in terms of ascertaining the contribution that specific elements make to the whole, as well as discern the main rhetorical purpose of texts.

To answer Text Structure and Purpose questions as intended, students are expected to demonstrate the following behaviors, depending on whether the question asks about a part-to-whole relationship of an element to the entire passage or about the main purpose of the passage.

Part-whole relationship

1. Read and demonstrate comprehension of the passage.
2. Select the answer choice that best describes the main function of the underlined portion of the passage in the passage as a whole.

Main purpose

1. Read and demonstrate comprehension of the passage.
2. Select the answer choice that best states the main purpose of the passage.

Both question types are designed to elicit cognitively complex thinking from students. For questions about part-whole relationships, students are expected to develop a clear sense of the overall message and structure of an appropriately challenging passage in one of several subject areas in order to determine the main rhetorical role that a particular, substantive part of the passage (e.g., a clause, a sentence; designated by underlining) plays in the passage as a whole. For questions about main purpose, students are expected to use an understanding of the content of a given passage to ascertain its primary aim and to distinguish

that main purpose from subordinate purposes, mere details, and/or unsupported assertions made about the passage in one or more answer choices.

Table 13 summarizes how students performed on the two Text Structure and Purpose questions included in the study.

Table 13. Student Performance on Reading and Writing: Craft and Structure—Text Structure and Purpose Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors			Answered Correctly	Demonstrated Both Behaviors and Answered Correctly	Differential
			1	2	Both			
3 Main purpose *n* = 26	Literature	3	26 (100%)	25 (96%)	25 (96%)	25 (96%)	25 (96%)	0
4 Part-whole relationship *n* = 26	Science	4	25 (96%)	24 (92%)	24 (92%)	24 (92%)	24 (92%)	0

Table 13 indicates that both Text Structure and Purpose questions included in the study performed as expected, with differentials of 0. Each student who answered one or both of the questions correctly demonstrated both required behaviors.

Vignettes from students answering correctly and demonstrating both required behaviors illustrate the kinds of complex thinking elicited by Text Structure and Purpose questions on the digital SAT Suite assessments.

Question 3

Question 3, an easy (PSB 3) question set in a literature context, asks students to read a passage and then determine the passage's main purpose.

The following text is from Holly Goldberg Sloan's 2017 novel *Short*.

> More than two years ago my parents bought a piano from some people who were moving to Utah. Mom and Dad gave it to my brothers and me for Christmas. I had to act really happy because it was such a big present, but I pretty much hated the thing from the second it was carried into the hallway upstairs, which is right next to my bedroom. The piano glared at me. It was like a songbird in a cage. It wanted to be set free.
>
> ©2017 by Holly Goldberg Sloan

Which choice best states the main purpose of the text?

A) It establishes how the narrator feels about the piano.

B) It describes the event that led the narrator's parents to buy a piano.

C) It explains why the narrator always wanted a piano close to her bedroom.

D) It suggests that the narrator's brothers are talented piano players.

In working through the question, students are expected to realize that the passage's main purpose is to establish how the narrator feels about the piano (choice A), as every element included in the passage furthers this writerly aim. While a partial reason for why the parents bought the piano is included in

the passage, it's only a subordinate element of the passage, making choice B incorrect. Neither answer choice C nor D is supported by the passage, so each can be ruled out as a reasonable assertion of the passage's main purpose.

Student RW1 follows closely the analysis path described above, and their response is representative of many others'. They demonstrate their solid grasp of the passage by immediately identifying the best answer based on their own interpretation of the text, and then they provide passage-based rationales for ruling out each of the distractors.

> So I think immediately, I already know that the first option is the best answer because I feel like as I was reading [the passage], I saw that, um, it doesn't really explain—I don't think the purpose of the test—text was to explain why the narrator's parents bought the piano because it's just stating that—how they got the piano, but not, like, like, the actual events that led up to it. That was only in, like, the first sentence, and there's—the rest of the passage is there, and it does not talk about that. Um, she—and then this third option, I'd say it's not that because she clearly says how she "hated the thing from the second it was carried into the hallway," which is right next to her bedroom. She did not want it to be close to her bedroom because she literally hates the piano. And it's not the last option because nowhere does it say that her brothers are talented piano players. It just says that it was gifted to her and her brother[s] for Christmas, but it doesn't suggest anything about them being piano players. So I would say [choice] A because it just talks about how she was not happy about getting the piano for Christmas and how she felt like it was "a songbird in a cage" and "it wanted to be set free."

Question 4

Question 4, a medium-difficulty (PSB 4) question set in a science context, asks students to determine the main function of the underlined portion in relation to the passage as a whole. To answer this question successfully, students need to understand both the substance of the whole passage and the specific role that the underlined portion plays in the passage.

Part of the Atacama Desert in Peru has surprisingly rich plant life despite receiving almost no rainfall. Moisture from winter fog sustains plants once they're growing, but the soil's tough crust makes it hard for seeds to germinate in the first place. Local birds that dig nests in the ground seem to be of help: they churn the soil, exposing buried seeds to moisture and nutrients. Indeed, in 2016 Cristina Rengifo Faiffer found that mounds of soil dug up by birds were far more fertile and supported more seedlings than soil in undisturbed areas.

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

A) It identifies the reason particular bird species dig nests in Atacama Desert soil.

B) It explains how certain birds promote seed germination in Atacama Desert soil.

C) It describes the process by which seeds are deposited into Atacama Desert soil.

D) It elaborates on the idea that the top layer of Atacama Desert soil forms a tough crust.

The passage as a whole is mainly about how some birds in the Atacama Desert likely contribute to the flourishing of plant life in that arid environment by digging nests that disturb the soil. The underlined portion—in this case, a clause—serves to explain why the birds may promote "rich plant life" in the desert. The fact that the underlined portion follows a colon provides some clue to its function, as phrases and clauses following colons typically amplify, elaborate on, or explain the preceding content in the sentence.

Choice B is the best answer, as the underlined portion notes how nest-digging birds "churn the soil, exposing buried seeds to moisture and nutrients." The passage doesn't explain why some bird species dig nests in the soil, so choice A is incorrect; similarly, the passage doesn't describe how seeds are (initially) deposited into the desert soil, so choice C can be ruled out. While the fact that the Atacama Desert soil has a tough crust is mentioned in the passage, the underlined portion doesn't build on that idea specifically, which makes choice D attractive but incorrect.

Student RW11's analysis focuses on identifying the best answer (without direct consideration of the alternatives). In the process, the student demonstrates comprehension of the passage as a whole as well as a clearheaded sense of how the underlined portion contributes to the passage's message.

> Well, the under[lined] portion reads, "they [churn] the soil, exposing buried seeds to moisture and nutrients," "they" referring to local birds in the desert. And it says that the nests that they dig "in the ground seem to be of help" because "the soil's tough crust makes it hard for seeds to germinate," but the nests that they dig seem to be of help. And then it explains how they dig the nests in the underlined portion, that they churn the soil exposing buried seeds to moisture and nutrients.

Cross-Text Connections Questions

On the digital SAT Suite tests, Cross-Text Connections questions assess whether students can read and comprehend pairs of appropriately challenging passages on the same topic or closely related topics and then draw a reasonable, text-based conclusion about some aspect of the relationship between the two passages. This relationship can be as simple as opposing viewpoints on a topic, but the majority of passage pairings feature more subtle and complex interactions, such as a second passage expanding on a minor point found in or reframing an argument presented in a first passage.

To answer Cross-Text Connections questions as intended, students are expected to demonstrate the following 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 (which precedes the answer choices).

Because of the number of components involved, Cross-Text Connections questions are some of the most inherently complex found in the digital SAT Suite's Reading and Writing section. Students demonstrating all the above behaviors have exhibited understanding of the gist of each passage separately, properly conceptualized the intended relationship between the two passages as signaled by the question, and selected the answer choice that best reflects that relationship.

Table 14 summarizes how students performed on the two Cross-Text Connections questions included in the study.

Table 14. Student Performance on Reading and Writing: Craft and Structure—Cross-Text Connections Questions.

			Demonstrated Required Behaviors							
Question	**Subject Area**	**PSB**	**1**	**2**	**3**	**4**	**All**	**Answered Correctly**	**Demonstrated All Behaviors and Answered Correctly**	**Differential**
5 *n* = 26	Humanities	6	20 (77%)	24 (92%)	23 (88%)	18 (69%)	15 (58%)	18 (69%)	15 (58%)	3
6 *n* = 26	Humanities	7	24 (92%)	22 (85%)	20 (77%)	9 (35%)	9 (35%)	9 (35%)	9 (35%)	0

Table 14 indicates that both Cross-Text Connections questions included in the study performed as expected, with one having a differential of 3 and the other a differential of 0.

Vignettes from students answering correctly and demonstrating all required behaviors illustrate the kinds of complex thinking elicited by Cross-Text Connections questions on the digital SAT Suite assessments.

Question 5

Question 5, a hard (PSB 6) question set in a humanities context, requires students to read and understand two differing but overlapping perspectives on author Virginia Woolf's book *Orlando*. The first passage stresses how much of an outlier *Orlando* is within Woolf's oeuvre, while the second passage argues for the novel's importance despite its unusual characteristics.

Text 1
Virginia Woolf's 1928 novel *Orlando* is an oddity within her body of work. Her other major novels consist mainly of scenes of everyday life and describe their characters' interior states in great detail, whereas *Orlando* propels itself through a series of fantastical events and considers its characters' psychology more superficially. Woolf herself sometimes regarded the novel as a minor work, even admitting once that she "began it as a joke."

Text 2
Like Woolf's other great novels, *Orlando* portrays how people's memories inform their experience of the present. Like those works, it examines how people navigate social interactions shaped by gender and social class. Though it is lighter in tone—more entertaining, even—this literary "joke" nonetheless engages seriously with the themes that motivated the four or five other novels by Woolf that have achieved the status of literary classics.

Based on the texts, how would the author of Text 2 most likely respond to the assessment of *Orlando* presented in Text 1?

A) By concurring that the reputation of *Orlando* as a minor work has led readers to overlook this novel but maintaining that the reputation is unearned

B) By agreeing that *Orlando* is less impressive than certain other novels by Woolf but arguing that it should still be regarded as a classic

C) By conceding that Woolf's talents were best suited to serious novels but asserting that the humor in *Orlando* is often effective

D) By acknowledging that *Orlando* clearly differs from Woolf's other major novels but insisting on its centrality to her body of work nonetheless

Besides having to read and understand the two passages, students must make a two-part inference to select the best answer, choice D. The author of Text 2 shares with the author of Text 1 the view that *Orlando* is unusual among Woolf's works but, in contrast to the author of Text 1, contends that the novel is still important because, despite having a lighter tone, it covers the same general thematic content as Woolf's other significant fiction. To reach this understanding, students need a clear sense of both passages individually as well as where the two texts agree and differ in perspective. Choice A is incorrect because Text 2 never concurs that *Orlando*'s reputation has kept potential readers away. Choices B and C are incorrect because the author of Text 2 regards *Orlando* as among Woolf's "great novels."

Student RW13 begins working the question by summarizing in their own words the gist of Texts 1 and 2, in the process noting the differences in perspective represented by each passage.

> And then I will probably go back just to see if I completely understand the opinions being explained in Text 1 and Text 2. Text 1 is basically saying that "*Orlando* is an oddity." It's different from her other work. It's pretty much just a joke. It doesn't say anything about it actually tackling serious topics. And then Text 2 is the one that focuses on how *Orlando* is like Woolf's other great novels. So [the author of Text 2] clearly believe[s] that *Orlando* deserves to be considered a classic.

After reading choice A and deciding to come back to it later, student RW13 then rules out choices B and C, the question's other distractors.

> "By agreeing that *Orlando* is less impressive than certain other novels [by Woolf]." [The authors of Texts 1 and 2] definitely do not agree. [The author of Text 2 is] saying that *Orlando* is just as good even though it's "lighter in tone." So it's not [choice] B. "By conceding that Woolf's talents were best suited to serious novels but asserting that the humor in *Orlando* is often effective." I think [the author of Text 2] also would not agree that [Woolf's] talents were best suited to serious novels because what they're trying to say is that *Orlando*, even though it's not—even though it's lighter, it's just as good. So they're not saying that she had a worse sense of style or a worse writing style in this book.

Student RW13 next offers a text-based rationale for the best answer, observing not only that the authors of Texts 1 and 2 agree that *Orlando* is an outlier among Woolf's fiction but also that they disagree about how seriously to treat the work.

> [Choice] D, "By acknowledging that *Orlando* clearly differs from Woolf's other major novels but insisting on its centrality to her body of work nonetheless." I think that one makes the most sense to me right now because they're saying the style is different but its centrality is the same. It talks about the same psychological themes. So that's what I'm thinking [the best answer] is, but I'm going to go back and read [choice] A just to make sure since I didn't completely understand what it was [saying].

The student concludes by returning to and ruling out choice A on the grounds that the author of Text 2 wouldn't consider *Orlando* a minor work.

> So it says: "By concurring that the reputation of *Orlando* as a minor work has led readers to overlook this novel but maintaining that the reputation is unearned." The reputation that it's talking about, the reputation of *Orlando* as a minor work. So I guess that would be saying that both passages—I mean, Text 1 and Text 2—would both agree that it's a minor work. But I think there's actually nothing in Text 2 that says that it is a minor work. Text 1 kind of sees it as a minor work, but Text 2 does not. It's talking about how the other ones are—her other novels are considered classics and this one is like those works, so it should be considered a classic as well. I don't think they would agree. I guess I'll just stick with my answer of [choice] D, not go back on that.

Question 6

Question 6, a hard (PSB 7) question also set in a humanities context, requires students to read, comprehend, and integrate information and ideas from two passages on the nature, role, and value of credits in different creative fields.

Text 1

Films and television shows commonly include a long list of credits naming the people involved in a production. Credit sequences may not be exciting, but they generally ensure that everyone's contributions are duly acknowledged. Because they are highly standardized, film and television credits are also valuable to anyone researching the careers of pioneering cast and crew members who have worked in the mediums.

Text 2

Video game scholars face a major challenge in the industry's failure to consistently credit the artists, designers, and other contributors involved in making video games. Without a reliable record of which people worked on which games, questions about the medium's development can be difficult to answer, and the accomplishments of all but its best-known innovators can be difficult to trace.

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

A) By suggesting that the scholars mentioned in Text 2 rely more heavily on credits as a source of information than film and television researchers do

B) By pointing out that credits have a different intended purpose in film and television than in the medium addressed by the scholars mentioned in Text 2

C) By observing that a widespread practice in film and television largely prevents the kind of problem faced by the scholars mentioned in Text 2

D) By recommending that the scholars mentioned in Text 2 consider employing the methods regularly used by film and television researchers

The gist of Text 1 is that the film and television industries consistently and thoroughly document contributors, while the gist of Text 2 is that the video game industry doesn't. The key point of interaction between the passages concerns the value of credits as a research tool. Text 1 notes that the practice of having transparent and complete credits allows researchers to more easily study "the careers of pioneering cast and crew members" who've worked in films and television, while Text 2 observes that the comparative lack of crediting in video games can leave "questions about the medium's development . . . difficult to answer" and makes "the accomplishments of all but its best-known innovators . . . difficult to trace."

Choice C is the best answer. The "widespread practice" choice C refers to is the prevalence of crediting in the film and television industries, and the "problem faced

by the scholars" referred to in Text 2 is the absence of similarly detailed credits in the video game industry. Choice A is incorrect because video game scholars have less access to systematic credits than do scholars in the film and television industries, so the former can't and don't rely more heavily on credits as a source of information. Choice B is incorrect because the two texts present credits as having the same basic function in the video game and the film and television industries. Choice D, like choice A, is incorrect because the scholars mentioned in Text 2 can't use "the methods regularly used by film and television researchers" because those latter methods rely on systematic credits that don't exist in the video game industry.

After reading the two passages and before analyzing the answer choices, **student RW4** offers their own tentative conclusion about the relationship between Text 1 and Text 2.

> I think that author of Text 1 would probably say something along the lines of standardization could help or something along the lines—along the lines of—I don't know, because they agree that standardization and having this credit sequence is a good thing. So they would probably agree with that.

Student RW4 then uses text-based reasoning to block two of the distractors. To rule out choice A, the student cites the basic agreement between the two passages on the value of thorough credits to researchers of both the film and television and the video game industries. They also recall the basic contrast: the former have a longstanding practice of crediting all contributors, while the latter doesn't.

> Answer choice A, "By suggesting that the scholars mentioned in Text 2 rely more heavily on credits as a source of information than film [and] television researchers do." Well, I mean, that's suggesting that scholars mentioned in Text 2 rely more heavily on credit as a source. I don't feel like that's correct because they both rely on—or that the author is saying that in Text 2, they need the credits and they don't have them. And in Text 1, they're saying they have the credits and they're great. So I think they would all agree that credits are a great source to be relied upon, and to rely more heavily on the credits. Yeah. If anything, Text 1 can rely on the credits more because they're consistently there and they don't have to worry about them not being there.

The student uses similar reasoning to rule out a second distractor, choice B, observing that despite differences in the extent to which credits are available in the film and television and the video game industries, the role and value of credits are essentially the same.

> Choice B, "By pointing out that credits have a different intended purpose in film and television than in the medium addressed by the scholars mentioned in Text 2." Definitely not. Both say that it's to find the people who worked on it who were great. Just Text 1 says, ". . . ensure that everyone's contributions are duly acknowledged." Text 2 says, "Without a reliable record of which people worked on which games, questions about the medium's development can be difficult to answer, and [the

> accomplishments of] all but [its] best-known innovators can be difficult to trace," saying that . . . they need them for finding who did what well.

Student RW4's consideration of the question's best answer, choice C, illustrates that they fully understand the basic agreement in perspective across Texts 1 and 2.

> Answer choice C, "By observing that a widespread practice in film and television largely prevents the kind of problem faced by the scholars mentioned in Text 2." I think that answer choice definitely has potential as they are saying in Text 1 that the widely used practice of having a credits roll allows them to address the issue of not being able to find who did what part and how impactful they were. Well, not how impactful they were, but what they focused on, and then you could know what they did well and see if that's someone you need to talk to. So that's definitely an answer choice to keep in my back pocket.

When evaluating choice D, the final distractor, the student is prompted to reconsider their initial supposition about the relationship between the two passages.

> And then D, "By recommending that the scholars mentioned in Text 2 consider employing the methods regularly used by film and television researchers." Although this was how I first—yeah, although this is how I first summarized it, I think it's probably—actually, how would the author of Text 1 most likely respond? Text 1 would most likely respond by saying that Text 2 should employ methods regularly used [in the film and television industries].

In the process, student RW4 draws a subtle but necessary distinction between this answer choice and a more accurate statement about the relationship between the texts: the issue isn't that video game scholars *should* make more use of credits in their research but rather that they *can't* because these credits don't exist to the extent that they do in film and television.

> Although I guess they wouldn't do that because the emphasis would not be, "Here are our methods we have that work well for us" because as line one says of Text 2, they don't consistently credit the artist. The issue isn't that they don't know how to credit the artist. It's that they don't consistently do it. . . . All right. So answer choice [D] is saying "By recommending that the scholars mentioned in Text 2 consider employing [the] methods regularly used by film [and television researchers]." They already employ those methods. So it can't be that because they're already employing them. She isn't giving new methods.

The student concludes by reaffirming the correctness of the question's best answer and concurrently ruling out the above distractor.

> "By observing [that] a widespread practice in film and television . . . largely prevents the kind of problem faced [by the scholars mentioned in Text 2]." Yeah. Yeah, I think that can make more sense than this one . . . because she would observe that film and television don't have this issue because of the fact that the practice is widespread, instead of saying, "Here's a new way that you can solve this issue." Because Text 1 would

probably say something along the lines of, "You should do this better like us," which is observing the widespread practice and how that has fixed their issue, instead of saying, "Here's a new solution that we use that you guys don't," which would be more of an answer choice [like choice] D. So I'm going to go with [choice] C.

INFORMATION AND IDEAS

Central Ideas and Details Questions

As the name implies, Central Ideas and Details questions take two general forms on the digital SAT Suite tests: one in which students must determine the main idea of a passage, and the other in which they must make use of important details in a passage to answer a comprehension question.

To answer Central Ideas and Details questions as intended, students are expected to demonstrate the following behaviors:

1. Read and demonstrate comprehension of the passage.
2. Select the answer choice that best represents the passage's main idea or the most reasonable conclusion based on details presented in the passage.

Students exhibiting the above behaviors offer evidence of having engaged in cognitively complex thinking. For questions focused on the main idea of passages, students not only must demonstrate an accurate understanding of a given passage's content but also must be able to summarize the passage's central point and distinguish that point from both subordinate points made in the passage and mere details. Details-focused questions call on a similar level of text-based reasoning to determine what conclusion a given passage best supports.

Table 14 summarizes how students performed on the two Central Ideas and Details questions included in the study.

Table 15. Student Performance on Reading and Writing: Information and Ideas—Central Ideas and Details Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors			Answered Correctly	Demonstrated Both Behaviors and Answered Correctly	Differential
			1	2	Both			
7 *n* = 26	Humanities	4	22 (85%)	16 (62%)	16 (62%)	16 (62%)	16 (62%)	0
8 *n* = 26	Literature	4	23 (88%)	18 (69%)	17 (65%)	18 (69%)	17 (65%)	1

Table 15 indicates that both Central Ideas and Details questions included in the study performed as expected, with differentials of 0 and 1.

Vignettes from students answering correctly and demonstrating both required behaviors illustrate the kinds of cognitively complex thinking elicited by Central Ideas and Details questions on the digital SAT Suite assessments.

Question 7

For question 7, a medium-difficulty (PSB 4) question set in a humanities context, students need to ascertain that the passage's main idea is that Richard Hunt

uses abstraction rather than "extreme accuracy" to depict his artistic subjects, an assertion best represented by answer choice C. This idea is directly expressed in the passage's first sentence, while the rest of the passage provides supporting examples in the form of *Arachne* and *The Light of Truth*.

> In many of his sculptures, artist Richard Hunt uses broad forms rather than extreme accuracy to hint at specific people or ideas. In his first major work, *Arachne* (1956), Hunt constructed the mythical character Arachne, a weaver who was changed into a spider, by welding bits of steel together into something that, although vaguely human, is strange and machine-like. And his large bronze sculpture *The Light of Truth* (2021) commemorates activist and journalist Ida B. Wells using mainly flowing, curved pieces of metal that create stylized flame.
>
> Which choice best states the text's main idea about Hunt?
>
> A) He uses different kinds of materials depending on what kind of sculpture he plans to create.
> B) He tends to base his art on important historical figures rather than on fictional characters.
> C) He often depicts the subjects of his sculptures using an unrealistic style.
> D) He has altered his approach to sculpture over time, and his works have become increasingly abstract.

In ruling out the distractors, students should recognize that choice A represents, at best, a subordinate, rather than main, point made by the passage; that choice B is factually incorrect per the passage; and that choice D is unsupported by the passage.

Student RW14 offers some interpretive commentary on Hunt and his works, as depicted in the passage, while recounting the passage's key elements.

> Okay. So once again, before even looking at the . . . answer choices, Richard Hunt. That's the artist. What is the text stating about the artist himself? So, at the start, it just states his name. He "uses broad forms rather than extreme accuracy," so maybe stating he's more artistic because he uses something different than—he doesn't try to be as accurate. "To hint at specific people or ideas." Talked about his first major work, how he constructed the mythical character. So once again, a mythical character instead of some historical leader. That shows more creativity to me. "A weaver who was changed"—yeah. Literally, a weaver who turns into a spider. That shows way more creativity to me. "By welding bits of steel together into something that, although vaguely human, is strange and machine-like." So this is just screaming "creativity" to me. And then "his large bronze [sculpture] *The Light of Truth*"—so that one does credit a historical person, but it does state using creativity stuff, "stylized flame."

The student then describes their sense of the gist of the passage before reading the answer choices.

> So to me, I think they're just calling him creative and good at his work, so I'm looking for answer choices that kind of fit that.

As it turns out, this encapsulation doesn't directly embody the question's narrower and more precise best answer. Nonetheless, the student has demonstrated strong comprehension of the passage and called attention to the fact that Hunt created art representing both historical and mythical figures, a point that will subsequently rule out one of the question's distractors.

Student RW14 then correctly determines that choice A, one of the distractors, represents a subordinate rather than the main idea of the passage.

> [Choice A,] "He uses different kinds of materials depending on what kind of sculpture he plans to create." So I'm just going to go back to the materials used. It talks about "welding bits of steel together." And then this one talks about "flowing, curved pieces." The same thing. I swear it said bronze somewhere. [*mumbles while reading aloud*] "And his [large] bronze"—so he did use bronze for this other one, but he used steel for the spider. I don't think that's the main idea, so I'm going to probably cross that out. I don't think that's the main idea. I can just go ahead and move on.

Student RW14's prior summation of the passage content enables them to easily block another distractor, choice B.

> [Choice B,] "He tends to base his art on important historical figures [rather than on fictional characters]." Okay, I'm just going to go ahead and cut that there. Yeah. He made Ida B. Wells, but he also made Spider-Arachne Man. So, no, I'm going to go ahead and cross that out.

The student then uses passage-based reasoning to provisionally settle on the question's best answer, in doing so making the intended leap from the passage's reference to "broad forms" (and the two examples that follow) to choice C's use of "unrealistic style."

> [Choice C,] "He often depicts the subjects of his sculpture[s] using an unrealistic style." That kind of works because it kind of calls to—he doesn't use accuracy. He uses "broad forms," so he tries to be more creative. That kind of calls "creative" to me. So I'm going to put a little dash next to that to the side, and I'm going to move on to the third—or final [answer choice].

The passage doesn't support choice D's assertion that Hunt has "altered his approach to sculpture over time" or that "his works have become increasingly abstract," as both 1956's *Arachne* and 2021's *The Light of Truth* are essentially equally abstract in style. As student RW14 considers this distractor, they make a misstep in reasoning: they correctly reject the answer choice but do so on the errant basis that Hunt's works have grown *less* abstract over time.

> [Choice D,] "He has altered his approach to sculpture over time, and his works have become increasingly abstract." Okay. What I'm thinking here

> is even though his first sculpture was *Arachne*—the spider, whatever—and he welded bits of steel together to create something vaguely human, he just used bits of steel, and even though it was abstract—I'd say that was pretty abstract. And Ida B. Wells, using some interesting bits of what? It's still metal and still Ida B. Wells, so I wouldn't call that abstract. So I'm not really loving this one.

The student later clarifies that their main reason for ruling out this choice was because they "don't think [it] says enough about that here in the passage" to make it the best answer. While the student's rationale remains incomplete and imperfect, they nonetheless evince conceptual awareness that part of the task posed in this question is to differentiate the main idea from subordinate ideas and details.

Ultimately, the student reaffirms the best answer choice, C, tying their selection back to their initial assessment of the passage's message.

> "He often depicts the subjects [of his sculptures using] an unrealistic style." I like that one because it kind of talks about creativity, which is what I was saying before.

Question 8

Question 8, a medium-difficulty (PSB 4) question set in a literature context, focuses on the passage's key analogy.

The following text is from Ezra Pound's 1909 poem "Hymn III," based on the work of Marcantonio Flaminio.

As a fragile and lovely flower unfolds its gleaming
foliage on the breast of the fostering earth, if
the dew and the rain draw it forth;
So doth my tender mind flourish, if it be fed with the
sweet dew of the fostering spirit,
Lacking this, it beginneth straightway to languish,
even as a floweret born upon dry earth, if the
dew and the rain tend it not.

Based on the text, in what way is the human mind like a flower?

A) It becomes increasingly vigorous with the passage of time.
B) It requires proper nourishment in order to thrive.
C) It draws strength from changes in the weather.
D) It perseveres despite challenging circumstances.

According to the passage, just as flowers need moisture in the forms of dew and rain to survive, the human mind needs its own form of "food" to thrive. Selecting the best answer, choice B, requires students to parse this analogy and its component parts—a common activity in English language arts classes. Choices A and C aren't supported by the passage. Choice D draws its appeal from echoing the speaker's sense of possible calamity, but it doesn't accurately describe the speaker's analogy.

After reading the passage and the answer choices, **student RW9** reiterates the question being posed.

> So, basically, what the question is asking [is] how the human mind is like a flower.

The student then rereads the opening lines of the passage and offers their own summation of the analogy.

> So, in the poem, "As a fragile and lovely flower unfolds its gleaming foliage on the breast of the fostering earth, if the dew and the rain draw it forth; / So doth my tender mind flourish, if it be fed with the sweet dew of the fostering spirit." For right there, when it says, "if it be fed [with] the sweet dew of the fostering spirit," I feel like it's talking about how the mind can flourish with nutrients, which kind of relates to . . . how the human mind, given proper nutrients, . . . could thrive like a flower. From right here, it says, "[if] [the] dew and the rain draw it forth; / So doth my tender mind flourish," which I feel like it's trying to relate right there.

Student RW9 then offers text-based rationales for ruling out each of the three distractors in the question, finding little evidentiary support for these choices.

> But for answer choice A, "It becomes increasingly vigorous with the passage of time." I don't really see how time can be related in this poem because they didn't really talk about time at all. So I wouldn't feel like that's a right answer choice.
>
> But [choice] C, it says, "It draws strength from changes in the weather," which I could see how this could be a possible correct answer choice because it does talk about how "the dew and the rain draw it forth" twice, . . . which would make sense about the weather part. But I just didn't feel like how the changes in the weather would relate to the human mind because that didn't make sense.
>
> But answer choice D, it says, "It perseveres despite challenging circumstances." . . . For answer choice D, I didn't see it as the correct answer because it just didn't make sense how it perseveres despite challenging circumstances. . . . I don't really know how to explain it. But just [choice D] didn't make sense because I didn't feel like persevering was the main focus of the poem and how it related to the mind.

The student then provides a similarly text-based rationale for the question's best answer, choice B.

> But I did feel like answer choice B made the most sense because [the passage] does talk about how the flower's flourishing, how it's gleaming on the fostering earth in the rain and the dew or the rain because it's getting the right nutrients. So as the mind, if it gets the right nutrients, it can flourish like the plant is, or the flower.

Command of Evidence: Textual Questions

Command of Evidence questions on the digital SAT Suite tests take one of two main forms. The first, discussed in this subsection, consists of questions focusing

on the skillful use of textual evidence to support a specified claim or point. The second is made up of questions about quantitative data displayed graphically; these are discussed in the next subsection.

Textual evidence questions on the digital-suite tests may appear in differing formats, though the emphasis of each is on assessing whether students can use provided evidence to substantiate a claim or point made in the question itself. In some cases, as with question 9 in this study, students may be asked to identify the line(s) from a literature text that best support an interpretive claim or point about the larger work; in other cases, as with question 10, students may be asked to consider how best to support a claim or point (such as a scientific hypothesis) when given possible, hypothetical research findings.

To answer Command of Evidence: Textual questions as intended, students are expected to demonstrate the following behaviors:

1. Read and demonstrate comprehension of the passage (or, in cases lacking a standard passage, such as question 9 below, comprehension of the answer choices).
2. Demonstrate an understanding of the relationship between the criterion set forth in the question's stem (which precedes the answer choices) and the passage.
3. Select the answer choice that best meets the criterion set forth in the question's stem.

Students offering evidence of the above behaviors have demonstrated cognitively complex thinking. When answering, students must show a clear understanding of the passage (or, as in question 9, the answer choices), the claim or point set forth in the question, and how the passage, the claim/point, and the best answer choice work together. In microcosm, students are thereby demonstrating the ability to reason carefully to support argumentative claims (including literary interpretations) and informational points with textual evidence.

Table 16 summarizes how students performed on the two Command of Evidence: Textual questions included in the study.

Table 16. Student Performance on Reading and Writing: Information and Ideas—Command of Evidence: Textual Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors				Answered Correctly	Demonstrated All Behaviors and Answered Correctly	Differential
			1	2	3	All			
9 *n* = 26	Literature	6	18 (69%)	21 (81%)	17 (65%)	13 (50%)	17 (65%)	13 (50%)	4
10 *n* = 25	Science	3	23 (92%)	25 (100%)	25 (100%)	23 (92%)	25 (100%)	23 (92%)	2

Table 16 indicates that both Command of Evidence: Textual questions included in the study performed as expected, with differentials of 4 and 2.

Vignettes from students answering correctly and demonstrating all required behaviors illustrate the kinds of cognitively complex thinking elicited by Command of Evidence: Textual questions on the digital SAT Suite assessments.

Question 9

Question 9 is a hard (PSB 6) question set in a literature context. This difficult question doesn't have a traditional test passage preceding the question; rather, it embeds what would typically be passage content into the four answer choices. The question lays out the interpretive claim to be supported by one of the four answer choices. Each of the quotations in questions such as this is an accurate representation of the original text, and such quotations are carefully selected to ensure that students, irrespective of background knowledge (including whether they've previously read the sampled work of literature), are able to make sense of them.

> "On Virtue" is a 1766 poem by Phillis Wheatley. Wheatley addresses the poem directly to the quality of virtue, imploring it to assist her in reaching a future goal: _____
>
> Which quotation from "On Virtue" most effectively illustrates the claim?
>
> A) "O thou bright jewel in my aim I strive / To comprehend thee. Thine own words declare / Wisdom is higher than a fool can reach."
>
> B) "Attend me, *Virtue*, thro' my youthful years! / O leave me not to the false joys of time! / But guide my steps to endless life and bliss."
>
> C) "But, O my soul, sink not into despair, / *Virtue* is near thee, and with gentle hand / Would now embrace thee, hovers o'er thine head."
>
> D) "I cease to wonder, and no more attempt / Thine height t'explore, or fathom thy profound."

The question establishes that the interpretive claim to be supported is that "Wheatley addresses the poem directly to the quality of virtue, imploring it to assist her in reaching a future goal." The best answer, choice B, strongly supports this claim: Wheatley directly addresses her plea to a personified Virtue, whom she asks to "guide [her] steps to endless life and bliss." Choices A and D are incorrect because they suggest the speaker has simply given up on understanding the nature of virtue. Choice C is incorrect because the speaker directly addresses the soul, not virtue, and because Virtue is presented more as a generalized source of comfort in distress than as an aid in meeting a goal.

After reading the question, **student RW15** evaluates the various answer choices, in the process demonstrating a grasp of the meaning and implication of each.

> So [choice] A, "O thou bright jewel in my aim I strive / To comprehend thee. Thine own words declare / Wisdom is higher than a fool can reach." So it's [not] talking about how—it's [not] talking about virtue and assisting her in reaching her future goal. So I would say that [choice] A does not really cover that right away, but we can still keep it in the loop. So let's look at [choice] B, "Attend me, *Virtue*, thro' my youthful years! / O leave me not to the false joys of time! / But guide my steps to endless

> life and bliss." So I would say [choice] B automatically has a little bit of a better context with talking about "endless life and bliss," guiding on the steps of her journey to reach that future goal. So I would say that definitely is a little bit better than the first one. But for [choice] C, "But, O my soul, sink not into despair, / *Virtue* is near thee, and with gentle hand / Would now embrace thee, hovers o'er thine head." That one is a little, again, broad. It doesn't really cover assisting her in a future goal. And then [choice] D, "I cease to wonder, and no more attempt / Thine height t'explore, or fathom thy profound." So again, very broad. Doesn't really cover assisting in a future goal.

The student concludes by reaffirming the question's best answer.

> The only one is [choice] B. B is talking about time. It's talking about "guide my steps to endless life and bliss." And it's talking directly to the quality of virtue and imploring it to assist her in reaching a future goal. And that's the only one that really talks about the future.

Question 10

Question 10 is an easy (PSB 3) question set in a science context. Unlike question 9, question 10 has a traditional reading passage preceding the question and answer choices.

> Scientists have long believed that giraffes are mostly silent and communicate only visually with one another. But biologist Angela Stöger and her team analyzed hundreds of hours of recordings of giraffes in three European zoos and found that giraffes make a very low-pitched humming sound. The researchers claim that the giraffes use these sounds to communicate when it's not possible for them to signal one another visually.
>
> Which finding, if true, would most directly support Stöger and her team's claim?
>
> A) Wild giraffes have never been recorded making humming sounds.
> B) The giraffes only produced the humming sounds at night when they couldn't see one another.
> C) Giraffes have an excellent sense of vision and can see in color.
> D) Researchers observed other animals in European zoos humming.

To answer this question effectively, students need to understand the passage content, the criterion set forth in the question (providing the best support for Stöger and her team's claim), and the impact that each of the answer choices would have on the researchers' claim. Note that the phrase "if true" is included in the question to reassure students that they don't need to consider the factual accuracy of the answer choices; they need only to assess their (hypothetical) value in supporting the researchers' claim. This claim, according to the passage, is that "giraffes use [humming] sounds to communicate when it's not possible for them to signal one another visually." Given that, choice B turns out to be the best answer, as it indicates that giraffes seem to generate the sounds only "at night when they couldn't see one another." Choice A, which suggests limits to scientists'

observations of giraffes humming, is incorrect because it's either irrelevant to the specific claim in the question or actually serves to weaken it. Choice C, which describes giraffes' vision, is also incorrect because it provides information irrelevant to the claim. Choice D is, again, not strictly germane to the researchers' specific claim and is therefore also incorrect.

Student RW7 demonstrates a conceptual understanding of the task, beginning their analysis of the question with a distillation of the claim being asserted.

> So basically, they're asking which finding would basically mostly support Stöger and her team's claim. Okay. So basically, what you have to ask yourself first is—or what we need to establish first is what is her claim, what was the research team's claim. So basically, their claim—yeah, because their claim is that giraffes use very low-pitched humming sounds when they're unable to communicate with each other visually. So they're saying, basically, they use it to communicate with each other when they can't do so visually. So they're asking me what directly supports that claim.

The student then exhibits text-based reasoning to assess the impact that each presented finding, if true, would have on the claim. Although the student's exclusion of choice A, one of the incorrect answer choices, is imprecise, their reasoning on the best answer (choice B) and the other two distractors (choices C and D) is clear and explicit.

> So choice A says, "Wild giraffes have never been recorded making humming sounds." I don't think that makes sense. Choice B, "The giraffes only produced the humming sounds at night when they couldn't see one another." Okay. Choice B actually makes a lot of sense because they're saying that they—they say that the giraffes only produce the humming sounds at night when they couldn't see one another. And her claim says that they use them when they're not able to signal to one another visually, and at night you can't see things, so it would make sense that they would use the humming sounds to communicate. Choice C says, "Giraffes have an excellent sense of vision and can see in color." Okay. Choice C don't make sense because they're saying that basically they can—they're saying that their vision is good and that they can see in color, but that's not what she's saying. [Choice D,] "Researchers observed other animals in European zoos humming." They're telling you about giraffes. I would say choice B is the best choice.

Command of Evidence: Quantitative Questions

Quantitative evidence questions on the tests of the digital SAT Suite assess the ability of students to locate, analyze, and make strategic use of data from tables, bar graphs, and line graphs to support claims and make informational points as specified in the questions themselves.

To answer Command of Evidence: Quantitative questions as intended, students are expected to demonstrate the following behaviors, with some variation by type of informational graphic (table, graph).

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, *or* 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 table *or* 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.

Students whose interviews offer evidence of carrying out these behaviors have demonstrated cognitively complex thinking. To successfully answer Command of Evidence: Quantitative questions involving tables and graphs, students must read and interpret the passage, which provides crucial context for understanding the included table or figure. Students must also understand the substance of the associated informational graphic, including what the graphic as a whole represents as well as the nature of its components (i.e., tabular data, bars, lines). They must additionally have a clear grasp of the criterion established by the question, which indicates what argumentative claim or informational point is meant to be supported by data from the table or figure. Finally, students must synthesize elements of the passage, informational graphic, and question to arrive at the best answer among the provided choices.

Table 17 summarizes how students performed on the four Command of Evidence: Quantitative questions included in the study.

Table 17. Student Performance on Reading and Writing: Information and Ideas—Command of Evidence: Quantitative Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors					Answered Correctly	Demonstrated All Behaviors and Answered Correctly	Differential
			1	2	3	4	All			
11 *n* = 24	History/ social studies	4	18 (75%)	24 (100%)	18 (75%)	21 (88%)	16 (67%)	21 (88%)	16 (67%)	5
12 *n* = 22	Humanities	6	15 (68%)	15 (68%)	15 (68%)	11 (50%)	10 (45%)	11 (50%)	10 (45%)	1
13 *n* = 23	Science	5	17 (74%)	21 (91%)	18 (78%)	17 (74%)	14 (61%)	17 (74%)	14 (61%)	3
14 *n* = 21	History/ social studies	4	18 (86%)	19 (90%)	19 (90%)	19 (90%)	16 (76%)	19 (90%)	16 (76%)	3

Table 17 indicates that all four Command of Evidence: Quantitative questions included in the study performed as expected, with differentials ranging from 1 to 5.

Vignettes from students answering correctly and demonstrating all required behaviors illustrate the kinds of cognitively complex thinking elicited by Command of Evidence: Quantitative questions on the digital SAT Suite assessments.

Question 11

Question 11 is a medium-difficulty (PSB 4) table-based question set in a history/ social studies context. Students are presented with a table of four locations in the Navajo Nation and those locations' average high and low temperatures, in degrees Fahrenheit, in July. The accompanying passage indicates that the large expanse of

and diversity of elevations within the Navajo Nation have an impact on the climate residents at various locations experience.

Average Temperatures in July in Four Locations in the Navajo Nation

Location	Average highest temperature (Fahrenheit)	Average lowest temperature (Fahrenheit)
Teec Nos Pos	94°	65°
Cameron	99°	65°
Ramah	83°	50°
Tuba City	83°	50°

The Navajo Nation has the largest land area of any tribal nation in the United States: over 27,000 square miles in the Southwest. Because this area is so huge and its communities are located at various elevations, the people of the Navajo Nation can experience different climate conditions depending on where they live. For example, in July, _____

Which choice most effectively uses data from the table to support the claim?

A) the lowest temperature for both Cameron and Teec Nos Pos was 65°.

B) the lowest temperature for both Ramah and Tuba City was 50°.

C) Tuba City's average highest temperature was 94°, while Teec Nos Pos's was 93°.

D) Ramah's average highest temperature was 83°, while Cameron's was 99°.

To answer correctly, students must understand that the question is asking for the example from among the provided choices that best supports the passage's claim that "the people of the Navajo Nation can experience different climate conditions depending on where they live." This claim is best supported by an answer choice that establishes, using data from the table, a wide divergence in average highest and lowest temperatures in July.

Choice D best accomplishes this goal, as it accurately uses data from the table to establish such a divergence: Ramah's and Cameron's average highest temperatures in July are sixteen degrees different (83 degrees versus 99 degrees). Choices A and B are incorrect on two grounds. First, each refers to the "lowest temperature" rather than the *average* lowest temperature at two locations and thus mispresents the table. Second, both pairs of temperatures cited are the same and therefore don't establish a striking dissimilarity, as sought by the question's criterion. Choice C incorrectly represents data from the table: Tuba City's average highest temperature in July was 83 degrees, not 94 degrees, and Teec Nos Pas's average highest temperature in July was 94 degrees, not 93 degrees.

Student RW20 begins their approach to the question by reading and summarizing the data contained in the table. This includes reading the table's title, making sense of the table's rows and columns, and showing understanding of the data contained in the table's cells.

> "Average temperatures in July in four locations in the Navajo Nation." So there's four locations: Teec Nos Pos, Cameron, Ramah, and Tuba City. And there's two columns for each location: "average highest temperature" and the "average lowest temperature." So I'm just scanning through the data, and it looks like—before I go into the question—it looks like the average highest temperature is obviously bigger than the average lowest temperature because it's in the definition. And other than that, the numbers look pretty similar. The first two locations seem to be larger—by 10 degrees, compared to the last two. So knowing that information, I'm going to be going into the question now.

After reading the question itself, student RW20 describes the question's criterion and what the nature of the best answer would be.

> So the question prompt is pretty straightforward. And they just, they're just, they have a "For example, in July." So they want us to show that "the people of the Navajo Nation can experience different climate conditions depending on where they live." So they just want to establish a contrast here. So I'm going to look for the option that's the most contrasting and different.

Student RW20 then processes the various answer choices, checking each for both accuracy relative to what's reported in the table and appropriateness for supporting the passage's claim—in the latter case, for the answer choice that establishes the clearest and widest contrast.

> So option A, "The lowest temperature for both Cameron and Teec Nos Pos was 65[°]." And then looking at both of them, and that is correct, actually. But it doesn't establish a contrast, which is what we're trying to get. So I'm ruling out option A. Option B, "the lowest temperature for both Ramah and Tuba City was 50°." And although this is correct [per the table], [it's wrong for] the same reason as option A, because we want to—I want to establish a contrast, and that's what, like, the question is trying to say because that's why they have "for example": they want to show that people in different parts of the Navajo Nation can experience different climate conditions because the area is so big and because it's so different in different places. So that leaves me with option[s] C and D. Option C, "Tuba City's average highest temperature was 94°, while Teec Nos Pos's was 93°." And this actually is factually incorrect, so I can rule it out right now. And that leaves me with option D. Let me check and make sure that it is correct. "Ramah's average highest temperature was 83°, while Cameron's was 99[°]." Compared to all the other answers, it establishes the most contrast. And I'm just scanning through the table once again. It looks like it's also factually correct. So I'm going to click on option D.

It's worth noting that student RW20 doesn't observe that choices A and B, about the "lowest temperature" in two locations, are incorrect also because they

misstate the table's information, which concerns the highest and lowest *average* temperatures in July in four locations. Nonetheless, the student recognizes that neither choice, taken on its own terms, provides the contrast called for in the question's stem.

Question 12

Question 12 is a hard (PSB 6) question set in a humanities context. In the table, students are given information about four individuals: their years active in the film industry and their known professional contributions. The passage accompanying the table contextualizes the tabular data by noting that "counts of those four figures' output should be taken as bare minimums rather than totals" because "so many films and associated records for this era have been lost."

Credited Film Output of James Young Deer, Dark Cloud, Edwin Carewe, and Lillian St. Cyr

Individual	Years active	Number of films known and commonly credited
James Young Deer	1909–1924	33 (actor), 35 (director), 10 (writer)
Dark Cloud	1910–1920	35 (actor), 1 (writer)
Edwin Carewe	1912–1934	47 (actor), 58 (director), 20 (producer), 4 (writer)
Lillian St. Cyr (Red Wing)	1908–1921	66 (actor)

Some researchers studying Indigenous actors and filmmakers in the United States have turned their attention to the early days of cinema, particularly the 1910s and 1920s, when people like James Young Deer, Dark Cloud, Edwin Carewe, and Lillian St. Cyr (known professionally as Red Wing) were involved in one way or another with numerous films. In fact, so many films and associated records for this era have been lost that counts of those four figures' output should be taken as bare minimums rather than totals; it's entirely possible, for example, that _____

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

A) Lillian St. Cyr acted in far more than 66 films and Edwin Carewe directed more than 58.

B) James Young Deer actually directed 33 films and acted in only 10.

C) Dark Cloud acted in significantly fewer films than did Lillian St. Cyr, who is credited with 66 performances.

D) Edwin Carewe's 47 credited acting roles include only films made after 1934.

The "example" referred to in the passage's last sentence, which students are expected to complete via their answer selection, is intended to illustrate the

passage's claim that the film outputs of the four individuals presented in the table should be viewed as minimums and not as exhaustive totals. The best answer, choice A, does this by indicating that the number of St. Cyr's and Carewe's actual credits may be much higher than the "number of films known and commonly credited" to each person, noted in the table's rightmost column. Choice B is incorrect because it contends that Young Deer's actual output was lower, not higher, than what's presented in the table. Choice C contains accurate information about both Dark Cloud and St. Cyr, but the choice doesn't effectively support the passage's claim or complete the passage's example because it simply compares the known output of two of the individuals included in the table instead of suggesting that the table's numbers are likely undercounts of the individuals' true output. Choice D is at odds with the table's information, which establishes that Carewe's forty-seven credited acting roles took place between 1912 and 1934, the years that, according to the table, Carewe was active.

Like student RW20 for question 11, **student RW16** starts their approach to question 12 by reading through and summarizing the table.

> "Credited Film Output of James Young Deer, Dark Cloud, Edwin Carewe, and Lillian St. Cyr." Um, so this, this, this chart is specifically talking about the, uh, film and credits of certain individuals.

After reading through the table, passage, and question, the student then offers their own encapsulation of the passage's gist.

> So the passage is saying that, um, the, uh, known, uh, credits of these, uh, filmmakers are minimums. They—they're—the known is just the bare minimum for which they actually did. So, even though we know that these—they're accredited for, uh, so, so many films, they probably produced, acted [in], or directed so many more. Um, so it's asking, you know, uh, essentially who—well, let me look at the answer choices first.

Student RW16 proceeds to work through each of the answer choices (in reverse order), settling on the question's best answer, choice A.

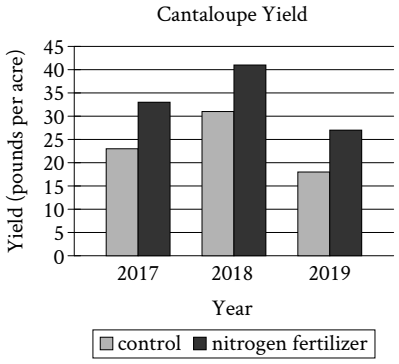
> Um, so [choice] D, it says, "after 1934," even though he was active up, up, up until 1934. So I don't think D's correct just because it doesn't support the data. Um, [choice] C is comparing two of the filmmakers. Um, and it's saying that "Dark Cloud acted in significantly fewer films than did Lillian St. Cyr, who [is] credited with 66 [performances]." Um, that is true, but, um, we are—the, the passage is not talking about comparing these, uh, these filmmakers. We're—it's talking about, um, how they were probably in way more films than, uh, than is previously known. So I don't think [this] is correct either. Um, [choice B,] "James Young Deer actually directed 33 films and acted in only 10." Um, that's true. That is directly using the data. Um, so that is true, but also I don't think that supports the passage. Um, it's just saying—it's just stating facts. Um, but the passage is not talking about facts. The passage is actually talking about—uh, it's hypothesizing what could be possible. And then [choice] A, "Lillian St. Cyr acted in far more than 66 films and Edwin Carewe directed more than 58." So, um, it's saying that the known number is much lower than

what they actually did, um, which is directly supporting the passage and makes A the only possible correct answer.

Note that the student's rationale for ruling out choice B is somewhat opaque but ultimately correct. Strictly speaking, choice B can't be "true" per the table, as the student claims, because the table itself lists higher numbers of film credits. However, the student rightly observes that choice B can't effectively complete the passage's example because it posits lower, not higher, outputs for Young Deer than he's commonly credited with.

Question 13

Question 13, the first of the study's two Command of Evidence: Quantitative questions incorporating a graph, is a medium-difficulty (PSB 5) question set in a science context. To answer this question correctly, students must read and understand the graph, noting that its bars refer to cantaloupe yields in three years under two different conditions (the experimental condition, in which nitrogen fertilizer was used, and a control condition, in which a fertilizer without nitrogen was used); determine from the passage that the claim to be supported is that "nitrogen fertilizer increases cantaloupe yield"; and then determine which answer choice provides data from the graph that best support this claim.



To test the effects of a nitrogen fertilizer on cantaloupe production, researchers grew cantaloupe plants and harvested their fruit over three years. In each year, half the plants were grown using a nitrogen fertilizer, and the other half were grown using a control fertilizer that contained no nitrogen. The researchers concluded that the nitrogen fertilizer increases cantaloupe yield.

Which choice best describes data in the graph that support the researchers' conclusion?

A) The yield for plants treated with the nitrogen fertilizer increased from 2017 to 2018.

B) In every year of the experiment, plants treated with the nitrogen fertilizer had a greater yield than did plants treated with the control fertilizer.

C) The 2018 yield for plants treated with the control fertilizer was greater than was the 2019 yield for plants treated with the nitrogen fertilizer.

D) In every year of the experiment, plants treated with the nitrogen fertilizer had a yield of at least 30 pounds per acre.

Choice B is the best answer. The key comparison derivable from the graph, which this answer choice aptly summarizes, is that the experimental condition involving the application of nitrogen fertilizer outyielded the control condition in each of the three years depicted in the figure. Choice A is factually correct but irrelevant to responding appropriately to the question, as the question's logic demands a cross-condition comparison, not a within-condition one. Choice C is also accurate per the graph but is similarly irrelevant as evidence since it merely compares the cantaloupe yield in one year under the control condition to the yield in a different year under the experimental condition. Choice D is factually incorrect—2019's experimental condition yielded fewer than thirty pounds of cantaloupe per acre—and is in any case immaterial, as the appropriate choice here draws a comparison across time between the two conditions in the experiment.

Student RW17, like many others quoted so far, starts their successful approach to answering by paraphrasing for themselves the content being presented.

> So the question is asking, which best supports the data? And [in] the data, we see the plants that were grown with the control fertilizer and then the other ones were grown with the nitrogen. And we see that the nitrogen fertilizer is greater than the control fertilizer.

The student then evaluates the answer choices, demonstrating in the process the understanding that the best answer, as in their summary of the data, must accurately capture the intended comparison between the experimental and control conditions.

> [Choice] A, "The yield for plants treated with the nitrogen fertilizer increased from 2017 to 2018." So we do see that's true, but I'm going to hold off on that one because I feel like the researcher's conclusion is comparing how the nitrogen fertilizer was better than the control fertilizer. So I wouldn't really compare it to itself. [Choice] B says, "In every year of the experiment, plants treated with the nitrogen fertilizer had a greater yield than did plants treated with the control fertilizer." So this I'm going to say is a possible answer because it's true; we do see that the nitrogen fertilizer had a greater yield than the control one. And the researchers, that's what they're trying to test—that the nitrogen fertilizer did help the cantaloupe production. So I'm going to say [this] might be the answer. [Choice] C, "The 2018 yield for plants treated with the control fertilizer was greater than was the 2019 yield for plants treated with the nitrogen fertilizer." So, again, I'm not going to say it's this one just because I feel like comparing them or comparing it to itself isn't really what the researchers were trying to do. And then [choice] D says, "In every year of the experiment, plants treated with the nitrogen fertilizer had a yield of at least 30 pounds per acre." So this one I'm not going to say either just because thirty pounds per acre, we don't really have anything to compare this number to, so we don't really know what it means.

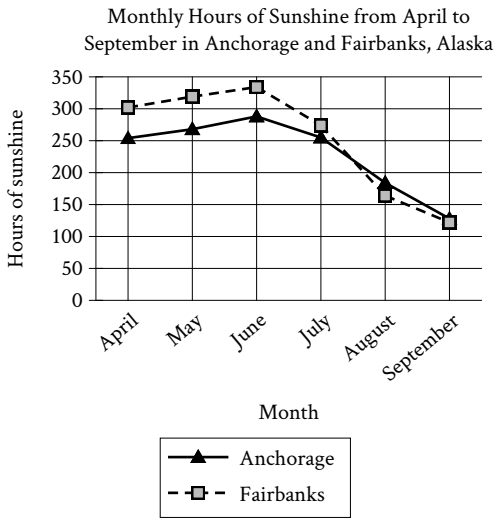
Although the student doesn't recognize or at least point out that this last choice is factually incorrect per the graph, they do, critically, understand that even if this option were true, it wouldn't supply evidence that would support the researchers' conclusion.

Having considered all the choices, student RW17 wraps up by affirming that the second option they considered was the best answer to the question.

> So for [choice] B, "In every year of the experiment, plants treated with the nitrogen fertilizer had a greater yield than did plants treated with the control [fertilizer]." So I'm going to go with B as my answer because that's what the researchers were trying to find. And they concluded that the nitrogen fertilizer increases cantaloupe yield, so I'm going to go with B.

Question 14

The last of the studied Command of Evidence: Quantitative questions is a medium-difficulty (PSB 4) question with a line graph set in a history/social studies context. The two lines represented in the graph identify the monthly hours of sunshine from April to September in two locations in Alaska. To answer this question correctly, students must determine from both the graph and passage that the two cities "show a similar pattern in the monthly hours of sunshine from April to September."



A student is researching monthly hours of sunshine in different cities in Alaska. When comparing trends in Anchorage and Fairbanks, the student concludes that the two cities show a similar pattern in the monthly hours of sunshine from April to September.

Which choice best describes data from the graph that support the student's conclusion?

A) The monthly hours of sunshine in both Anchorage and Fairbanks increase from April to June and then decrease from June to September.

B) Anchorage and Fairbanks both have less than 200 monthly hours of sunshine from April to September.

C) Anchorage and Fairbanks both have more than 300 monthly hours of sunshine from April to June and less than 200 hours from July to September.

D) The monthly hours of sunshine in both Anchorage and Fairbanks hold steady in June and July before beginning to decline in August.

The passage notes that the student researcher has identified a "similar pattern" in the hours of sunshine Anchorage and Fairbanks experience from April to September. The deliberate use of "pattern" here strongly implies that the best answer will accurately describe a trend or trends common to both cities and evident in the graph.

Choice A is the best answer here because it captures the basic similarity in the hours of sunshine experienced by the two cities over time: they increase from April to June and then decrease from June to September. Choice B is factually erroneous given the information in the graph. While choice C tries to represent a trend across the studied time span, it does so with inaccurate information about monthly hours of sunshine in the two cities. Choice D is incorrect because it's factually inaccurate—the hours of sunshine don't "hold steady in June and July" in the two cities but rather begin decreasing from June to July—and because it fails to accurately represent the overall trend in the two cities from April to September.

In successfully approaching this question, **student RW26** starts off by connecting the passage's reference to a "similar pattern" to what they find in the graph.

> Two cities show a similar pattern in the monthly hours of sunshine. So that's the conclusion. So I guess the pattern being seen is that they're both pretty much the same throughout the late spring, early summer. And then they declined towards the end of the summer.

Having correctly discerned the intended sort of answer, the student rules in the best response and rules out the distractors using largely the same sort of rationale provided above.

> So [choice A,] "The monthly hours of sunshine in both Anchorage and Fairbanks increase from April to June and then decrease from June to September." So April to June, that is true—it is increasing in both places and decreasing after that. So this one is true, and it supports the student's conclusions. So I feel like this is a good answer choice. [Choice B,] "Anchorage and Fairbanks both have less than 200 monthly hours of sunshine from April to September." This one, it's not talking about trends and the things. [The passage] talks about [a] similar pattern of the monthly hours. [This answer choice is] just talking about one sort of comparison between the two that they have less than 200 monthly hours. So that's not the best answer choice. [Choice C,] "Anchorage and Fairbanks both have more than 300 monthly hours of sunshine from April to June and less than 200 hours from July to September." Again, as long as we talk about the pattern throughout the years. [The passage is] talking about the pattern from April to September, mainly just talking about—and also [this answer choice is] not even accurate, looking at the graph, and so also the [third] answer is inaccurate, looking at the graph. So, I mean, that one's not a good answer choice either. [Choice D,] "The monthly hours of sunshine in both Anchorage and Fairbanks hold steady in June and July before beginning to decline in August." "In June and July." No, it is declining in July because it's in the graph. In August, it also declined. It starts declining earlier. So that one's wrong. So the first answer choice is probably the best answer choice.

Student RW26's reasoning for ruling out choice B is imperfect—this answer option does posit a trend but one that's not supported by the graph—but in every other respect their approach to the question is exemplary and illustrative of cognitively complex thinking.

Inferences Questions

On the digital SAT Suite tests, Inferences questions assess students' ability to reach reasonable, text-supported conclusions based on what passages say explicitly and strongly imply. Inferences questions include a blank, which students must "fill in" with the most logical option among the provided answer choices.

To answer Inferences questions as intended, students are expected to demonstrate the following behaviors:

1. Read and demonstrate comprehension of the passage.
2. Select the answer choice that most logically completes the passage.

Students offering evidence of the above behaviors have demonstrated cognitively complex thinking. In addition to reading and comprehending the overall content of a passage associated with an Inferences question, students must determine which of the proffered answer choices completes the passage most logically. To accomplish the latter, they must carefully evaluate what the passage says directly as well as what it strongly implies to reach a supportable conclusion.

Table 18 summarizes how students performed on the two Inferences questions included in the study.

Table 18. Student Performance on Reading and Writing: Information and Ideas—Inferences Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors			Answered Correctly	Demonstrated Both Behaviors and Answered Correctly	Differential
			1	2	Both			
15 *n* = 21	Science	6	17 (81%)	14 (67%)	11 (52%)	14 (67%)	11 (52%)	3
16 *n* = 21	History/ social studies	4	21 (100%)	17 (81%)	17 (81%)	17 (81%)	17 (81%)	0

Table 18 indicates that both Inferences questions included in the study performed as expected, with differentials of 3 and 0.

Vignettes from students answering correctly and demonstrating both required behaviors illustrate the kinds of cognitively complex thinking elicited by Inferences questions on the digital SAT Suite assessments.

Question 15

Question 15 is a difficult (PSB 6) Inferences question set in a science context. Like other questions of the type, this question requires students to read and understand an appropriately challenging passage and then determine which choice among the answer options most logically follows from what's been presented in the text—in this case, teasing out what suggestion is implied by the passage's information about mosses growing in the desert.

Mosses can struggle in harsh desert conditions because these plants require enough sunlight for photosynthesis but not so much that they risk drying out. Researchers Jenna Ekwealor and Kirsten M. Fisher found several species of *Syntrichia caninervis*, a type of desert moss, growing under quartz crystals in California's Mojave Desert. To evaluate whether these semitransparent rocks benefited the moss, the researchers compared the shoot tissue, a measure of plant growth, of *S. caninervis* when growing on the soil surface versus when the moss was growing under the quartz rocks. They found that the shoot tissue was 62% longer for moss growing under the quartz as compared to moss on the soil surface, suggesting that _____

Which choice most logically completes the text?

A) quartz crystals do not transmit the necessary sunlight for photosynthesis in *S. caninervis*.

B) quartz crystals are capable of supporting *S. caninervis* growth if the crystals are not too thin.

C) *S. caninervis* growing under quartz crystals experience lower light intensity and are thus able to retain more moisture.

D) *S. caninervis* is one of the few types of moss that can survive under semitransparent rocks.

To answer this question correctly, students must trace the line of reasoning presented in the passage and determine which of the provided conclusions is logically entailed, or strongly indicated, by that reasoning. Schematically, students might break this passage down as follows:

1. Mosses can dry out in harsh, bright desert conditions but still need enough sunlight for photosynthesis.
2. Scientists found a type of desert moss growing under quartz crystals.
3. The scientists wondered whether growing under quartz crystals benefited the moss.
4. Shoot tissue, a measure of plant growth, was sixty-two percent longer for moss growing under the quartz than for moss growing on the soil surface.
5. This finding suggests that . . .

Choice C is the best supported, most logical suggestion here. *S. caninervis* needs enough sunlight for photosynthesis but not so much that it dries out. The *S. caninervis* samples under quartz crystals showed evidence of greater growth (i.e., more shoot tissue) than did mosses growing on the soil surface. This implies that *S. caninervis* under the crystals experienced more supportive growing conditions. Given that the passage indicates that *S. caninervis* needs just the right amount of sunlight to thrive, it's reasonable that *S. caninervis* growing under quartz crystals experience lower light intensity and are thus able to retain more moisture. Choice A is illogical because the passage observes that *S. caninervis* thrives under the quartz crystals, so it must be experiencing photosynthesis. Neither choice B nor choice D is supported by the passage.

After reading through both the question and answer choices, **student RW11** paraphrases the gist of the passage.

> So I'm going to read back through what it would be suggesting to see what needs to complete the text, and then I'll read back through the context. "They found that the shoot tissue was 62% longer for moss growing under the quartz as compared to moss on the soil surface, suggesting that"—so it would be suggesting something [like] that the quartz produced more growth [than did moss] in the soil surface. So I'm going to look in the passage to see if that is mentioned anywhere. They're evaluating whether the semitransparent rocks benefited the moss. And so they compared the shoot tissue when [the moss was] growing on the soil surface versus when the moss was growing under the quartz rocks. So it would indicate that the quartz support[s] the growth.

Despite their apt summary of the passage, student RW11 momentarily gravitates toward choice B, one of the question's distractors. The appeal of choice B to the student appears to be that it suggests the same sort of balancing act as that introduced in the passage's first sentence: desert mosses need some sunlight to survive, but not too much.

> So I would say choice B, "quartz crystals are capable of supporting *S. caninervis* growth if the crystals are not too thin," because these plants require enough sunlight for photosynthesis but not so much that they risk drying out.

As they work through the answer choices more fully, however, student RW11 recognizes their earlier mistake and rules out choice B as well as the other two distractors.

> It's not choice A. They do transmit the necessary sunlight for photosynthesis. I said [the best answer was choice] B, but the passage doesn't mention anything about the crystals being too thin. I don't think it says anything about choice D, it being "one of the few types of moss that can survive under semitransparent rocks," because it says—

Student RW11 then provides a text-supported rationale for the best answer.

> But it does say that they "struggle in harsh desert conditions" just because they require enough sunlight. Choice C, they grow under—"[*S. caninervis*] growing under quartz crystals experience lower light intensity and are thus able to retain more moisture." [The passage] does reference needing enough sunlight but not so much that they risk drying out. So because it's under the crystal, it can retain more moisture because—but it still has enough for photosynthesis. But it can retain more moisture because it's a lower intensity, and they will be less likely to dry out.

Question 16

Question 16 is a medium-difficulty (PSB 4) Inferences question set in a history/social studies context. In this question, students must determine the logical consequence that follows from the results of a study on interruptions in the workplace.

Some businesses believe that when employees are interrupted while doing their work, they experience a decrease in energy and productivity. However, a team led by Harshad Puranik, who studies management, has found that interruptions by colleagues can have a social component that increases employees' sense of belonging, resulting in greater job satisfaction that benefits employees and employers. Therefore, businesses should recognize that _____

Which choice most logically completes the text?

A) most employees avoid interrupting colleagues because they don't appreciate being interrupted themselves.

B) in order to maximize productivity, employers should be willing to interrupt employees frequently throughout the day.

C) in order to cultivate an ideal workplace environment, interruptions of work should be discouraged.

D) the interpersonal benefits of some interruptions in the workplace may offset the perceived negative effects.

The underlying structure of the passage can be broken down in a manner similar to that for question 15's text.

1. Some businesses believe interruptions reduce employees' energy and productivity.
2. A research team, however, found that interruptions by colleagues can actually improve employees' engagement in work.
3. Therefore, businesses should recognize that . . .

Choice D most logically completes the text because it acknowledges that whether interruptions are good or bad for employees depends on the circumstances: while some interruptions could have negative effects, others, specifically those involving colleagues interrupting coworkers, could have positive social effects. Choice A is unsupported by the passage and is therefore incorrect. Choice B is wrong because it ignores the contingent nature of interruptions, which can be either good or bad depending on the circumstances. Choice C is incorrect because it erroneously posits that all interruptions are problematic.

After reading the passage, **student RW27** provides a succinct summary of the gist, albeit one that lacks some of the passage's nuance.

> So they [Puranik and team] agree with interruptions.

Using that frame, student RW27 then evaluates the answer choices. Although the student neglects the qualified nature of the researchers' endorsement of workplace interruptions when summarizing the passage, they clearly exhibit textual comprehension as they work through the choices, observing that some such interruptions were found to be beneficial. In their selection of the best answer, choice D, student RW27 uses vocabulary knowledge and context clues to ascertain that "offset" essentially means "take away."

[Choice A,] "most employees avoid interrupting colleagues because they don't appreciate being interrupted themselves." No. That doesn't go with the passage. [Choice B,] "in order to maximize productivity, employees—" or [rather] "employers should be willing to interrupt employees frequently throughout the day." Maybe not frequently. They just said some interruptions increased sense of belonging. So I don't totally agree with this one. Here, [choice C,] "in order to cultivate an ideal workplace environment, interruptions of work should be discouraged." Okay, they're not talking about discouraging because they said that it can show a sense of belonging. So greater job satisfaction. So don't agree with that one. [Choice D,] "the interrupted—," oh no, "the interpersonal benefits of some interruptions in the workplace may offset the perceived negative effects." In this sentence, I think they mean "offset." And then in "the interpersonal benefits of some [interruptions] in the workplace may offset the perceived negative effects." So maybe take away the perceived negative effects. If that's what that means, I agree. So I'm going to go with this option. Because they do.

EXPRESSION OF IDEAS

Rhetorical Synthesis Questions

The digital SAT Suite's Rhetorical Synthesis questions assess students' ability to combine information and ideas in ways aligned to specified writerly goals. Each Rhetorical Synthesis question includes three elements:

- A set of "notes," in bulleted-list form, gathered by a hypothetical student researching a given topic
- A writerly goal that the student seeks to accomplish using relevant information from the notes
- A series of answer choices presenting differing arrangements of select portions of the notes, with one of these arrangements representing the most effective approach to meeting the student's goal

To answer Rhetorical Synthesis questions as intended, students are expected to demonstrate the following behaviors:

1. Read and demonstrate comprehension of the notes.
2. Demonstrate an understanding of the relationship between the notes and the criterion (i.e., the hypothetical student's goal) set forth in the question.
3. Select the answer choice that best meets the criterion set forth in the question.

Students demonstrating the above behaviors have exhibited several aspects of cognitively complex thinking. First, they've gained a clear sense of the topic under discussion by reading the notes, which contain factually accurate information, reasonable interpretations, logical conclusions, and the like, and mentally supplied the cohesive links that plausibly connect the information and ideas they're presented with. Second, they've grasped the writerly goal identified in the question itself. Third, they've applied the understanding that the question's best answer should include accurate representations of only the information and ideas from the notes most relevant to meeting the writerly goal and that differing syntactical arrangements of words, phrases, and clauses—as reflected in the question's answer choices—produce varying rhetorical effects, such as adding emphasis or establishing a contrast.

Table 19 summarizes how students performed on the two Rhetorical Synthesis questions included in the study.

Table 19. Student Performance on Reading and Writing: Expression of Ideas—Rhetorical Synthesis Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors				Answered Correctly	Demonstrated All Behaviors and Answered Correctly	Differential
			1	2	3	All			
17 *n* = 20	Science	4	18 (90%)	19 (95%)	16 (80%)	15 (75%)	16 (80%)	15 (75%)	1
18 *n* = 20	Humanities	5	17 (85%)	17 (85%)	17 (85%)	15 (75%)	17 (85%)	15 (75%)	2

Table 19 indicates that both Rhetorical Synthesis questions included in the study performed as expected, with differentials of 1 and 2.

Vignettes from students answering correctly and demonstrating all required behaviors illustrate the kinds of cognitively complex thinking elicited by the digital SAT Suite's Rhetorical Synthesis questions.

Question 17

Question 17, a medium-difficulty (PSB 4) question set in a science context, asks students to selectively use the provided notes to establish an advantage of a new type of platinum catalyst.

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

- Platinum is a rare and expensive metal.
- It is used as a catalyst for chemical reactions.
- Platinum catalysts typically require a large amount of platinum to be effective.
- Researcher Jianbo Tang and his colleagues created a platinum catalyst that combines platinum with liquid gallium.
- Their catalyst was highly effective and required only trace amounts of platinum (0.0001% of the atoms in the mixture).

The student wants to explain an advantage of the new platinum catalyst developed by Jianbo Tang and his colleagues. Which choice most effectively uses relevant information from the notes to accomplish this goal?

A) While still highly effective, the new platinum catalyst requires far less of the rare and expensive metal than do other platinum catalysts.

B) Platinum is a rare and expensive metal that is used as a catalyst for chemical reactions; however, platinum catalysts typically require a large amount of platinum to be effective.

C) Researcher Jianbo Tang and his colleagues created a platinum catalyst that combines platinum, a rare and expensive metal, with liquid gallium.

D) Like other platinum catalysts, the new platinum catalyst requires a particular amount of the metal to be effective.

Choice A is the best answer, as this option most effectively uses relevant information from the notes to explain that the new catalyst remains effective despite using much less of the rare and costly metal platinum than do other catalysts. Choice B is incorrect because it merely describes platinum and platinum catalysts in general without specific reference to the new catalyst discussed in the notes. Choice C is incorrect because it just describes the platinum catalyst Tang and colleagues produced without explaining why it's superior to other catalysts. Choice D is incorrect because it establishes a general similarity between the new catalyst and other catalysts and thereby underplays the key benefit of the new catalyst.

After observing that the topic in the question is "very cool," **student RW22** evaluates the answer choices. They first note that the best answer, choice A, accurately reflects the notes and expresses an advantage of the new platinum catalyst.

> "While still highly effective, the new platinum catalyst requires far less of the rare and expensive metal than do other platinum catalysts." That's true. It does compare platinum catalysts usually requiring a large amount versus the new one [that] only needed, like, one ten-thousandth, or .0001 percent. So yes, that would make sense based on the notes.

Student RW22 then rules out the other answer choices on the intended basis: none of them is as successful as choice A in identifying an advantage of the new catalyst.

> [Choice B,] "Platinum is a rare and expensive metal that is used as a catalyst for chemical reactions; however, platinum catalysts [typically] require a large amount of platinum to be effective." Okay. That's true, but it doesn't explain the advantage of the new one that uses less platinum, which I think is what the question is asking for. [Choice C,] "Researcher Jianbo Tang and his colleagues created a platinum catalyst that combines platinum, a rare and expensive metal, with liquid gallium." Again, that doesn't really explain the advantage of the new one. It explains what it is, but not why it's better. [Choice D,] "Like other platinum catalysts, the new platinum catalyst requires a particular amount of [the] metal to be effective." That, again, doesn't really say anything because it's saying there's a similarity between the two catalysts, but it's not saying that there's any reason to pick one or the other. I think the first one is the only one that does that, that makes a compelling point.

Question 18

Question 18 is another medium-difficulty (PSB 5) Rhetorical Synthesis question, this time set in a humanities context and with the goal of describing a particular work of art in the exhibition "Labor of Love."

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

- Isabel Toledo was a Cuban-American fashion designer and artist.
- In 2019, Isabel and her husband, Ruben Toledo, created an exhibition called "Labor of Love" at the Detroit Institute of Arts.
- Many of the Toledos' works in "Labor of Love" honored other works in the museum's collection.
- In a collage called "The Choreography of Labor," Toledo blended her dress designs with images of Detroit laborers.
- The images of the laborers were from murals by Mexican artist Diego Rivera.

The student wants to describe a work from the exhibition "Labor of Love." Which choice most effectively uses relevant information from the notes to accomplish this goal?

A) "The Choreography of Labor" was a collage that blended dress designs from its creator, Isabel Toledo, with images of laborers from Diego Rivera's murals.

B) Many of the works in "Labor of Love," including "The Choreography of Labor," honored other works in the museum's collection.

C) In 2019, two artists collaborated on an exhibition, "Labor of Love," displaying their art in a Detroit museum that also featured murals by Diego Rivera.

D) "Labor of Love," at the Detroit Institute of Arts, was a 2019 exhibition of works by Cuban-American artists Isabel and Ruben Toledo.

Choice A is the best answer, as it most effectively uses relevant information from the notes to describe "The Choreography of Labor," a work by Isabel Toledo that was shown as part of the "Labor of Love" exhibition and that "blended dress designs from its creator . . . with images of laborers from Diego Rivera's murals." Choice B is incorrect because although it mentions a specific work that was part of the "Labor of Love" exhibition, it doesn't describe "The Choreography of Labor" in any detail but instead simply names a feature shared by many of the works in the exhibition. Choices C and D are incorrect because they only offer information about the "Labor of Love" exhibition in general.

Student RW13 begins their successful analysis by quickly eliminating choices C and D on the same grounds as above.

> Okay. First of all, [choices] C and D don't really explain anything about the exhibition. They're just saying that there was one. So they're kind of the same thing. Don't really say anything about it. And what [the student is] trying to accomplish is to describe [Isabel Toledo's] work, which [choices C and D] don't do. So we can rule those out.

The student next observes that choice B is closer to being viable than are choices C and D but that B, too, doesn't successfully meet the goal set forth in the question.

> [Choice B] is a little bit more descriptive. It talks about "The Choreography of Labor" and how it "honored other works in the museum's collection," but it's not really detailed.

They then settle on choice A, the best answer, as the most descriptive of an individual work in the exhibition.

> The best option is A because it's talking about . . . "The Choreography of Labor," which is part of the collection and it's "a collage that blended dress designs from its creator . . . with images of laborers." So that is a way that—they could use that to describe her work. They're like, "This is a work from 'Labor of Love.'" And then you understand what it is, how it looks, and—yeah. So that's the best option.

Transitions

The final Reading and Writing question type examined in this study focuses on assessing students' ability to skillfully use transition words and phrases to enhance the logic and cohesion of texts.

To answer Transitions questions as intended, students are expected to demonstrate the following behaviors:

1. Read and demonstrate comprehension of the passage.
2. Select the answer choice that completes the passage with the most logical transition.

Students demonstrating both these behaviors offer evidence of having engaged in cognitively complex thinking. They must first show an understanding of the gist of an appropriately challenging passage, and then they must use that understanding as well as knowledge of various transitional relationships, such as contrast and exemplification, to complete the passage in a way that makes the text most logical and cohesive.

Table 20 summarizes how students performed on the two Transitions questions included in the study.

Table 20. Student Performance on Reading and Writing: Expression of Ideas—Transitions Questions.

Question	Subject Area	PSB	Demonstrated Required Behaviors			Answered Correctly	Demonstrated Both Behaviors and Answered Correctly	Differential
			1	2	Both			
19 *n* = 20	History/ social studies	4	19 (95%)	17 (85%)	16 (80%)	17 (85%)	16 (80%)	1
20 *n* = 19	History/ social studies	3	17 (89%)	17 (89%)	16 (84%)	17 (89%)	16 (84%)	1

Table 20 indicates that both Transitions questions included in the study performed as expected, with differentials of 1.

Vignettes from students answering correctly and demonstrating both required behaviors illustrate the kinds of cognitively complex thinking elicited by the digital SAT Suite's Transitions questions.

Question 19

Question 19, a medium-difficulty (PSB 4) question set in a history/social studies context, requires students to determine that a cause-effect transition is necessary to make the passage most logical.

> Before the 1847 introduction of the US postage stamp, the cost of postage was usually paid by the recipient of a letter rather than the sender, and recipients were not always able or willing to pay promptly. _____ collecting this fee could be slow and arduous, and heaps of unpaid-for, undeliverable mail piled up in post offices.
>
> Which choice completes the text with the most logical transition?
>
> A) On the contrary,
> B) Consequently,
> C) Regardless,
> D) For example,

Choice B is the best answer, as the blank in the passage should be filled in with the cause-effect transition "consequently": two results of expecting mail recipients to pay the cost of postage were that fees were collected in a "slow and arduous" way and that "heaps" of abandoned mail "piled up in post offices." Choices A, C, and D are incorrect because each fails to complete the passage with a logical transition indicating a cause-effect relationship. "On the contrary" means "just the opposite"; "regardless" means "despite everything"; and "for example" signals exemplification.

In successfully answering this question, **student RW23** relies mainly on a substitution tactic, putting three of the four answer choices directly into the blank and reading behind and ahead of the blank to check the fit of the options. The student's approach isn't particularly precise, as they somewhat effortfully try to explain the intended meaning of "on the contrary" and "regardless." That said, student RW23 makes it clear that they understand that a causal transition is needed in the blank to signal that slow fee collection and stacks of undeliverable mail were results of the typical practice of charging recipients rather than senders for postage.

> I'll read the sentence before and after so I'll know what to put there. So it says, "The cost of . . . postage was usually paid by the recipient of [a] letter rather than [the] sender, [and] recipients were not always able or willing to pay promptly." Blank. Maybe, [choice A,] "On the contrary, collecting this fee could be slow and arduous, and heaps of unpaid-for, undeliverable mail piled up in post offices." "On the contrary." I don't think it fits. It doesn't sound like the right transition. It's not the right kind of transition to use. It does not feel right. It's not really, "on this hand, it's like this." It's not describing something like that. So the next choice would be [choice B,] "consequently." Some "recipients were not always [able or] willing to pay promptly. Consequently, collecting this fee could be slow and arduous, and heaps of unpaid-for, undeliverable mail piled up in [post] office[s]." That could be correct because since they're not always

willing or able to pay, this happened, which is a way to explain, so that could be the correct transition. Third choice [choice C] is "regardless." ". . . were not always [able or] willing to pay promptly. Regardless, collecting this fee could be slow—" It doesn't make any sense because they're not saying, "Well, regardless of this, it's still like this." It's not explaining that kind of way. So it shouldn't be that one. And as for [choice D,] "for example," it's not giving an example of something. It's a cause-and-effect kind of thing. So it would be "consequently."

Question 20

Question 20 is an easy (PSB 3) question also set in a history/social studies context. To answer this question correctly, students need to recognize that an adversative transition (i.e., one signaling opposition or contradiction) is needed in the blank.

> It has long been thought that humans first crossed a land bridge into the Americas approximately 13,000 years ago. _____ based on radiocarbon dating of samples uncovered in Mexico, a research team recently suggested that humans may have arrived more than 30,000 years ago—much earlier than previously thought.
>
> Which choice completes the text with the most logical transition?
>
> A) Similarly,
> B) In conclusion,
> C) As a result,
> D) However,

Choice D is the best answer because "however" logically signals the simple contrast set up between the two sentences in the passage: what was once thought to be true is no longer believed to be so. Choices A, B, and C are incorrect because each fails to complete the passage with a logical transition indicating an adversative relationship. "Similarly" indicates likeness; "in conclusion" indicates a summation; and "as a result" indicates consequence.

Somewhat like student RW23 for question 19, **student RW6** provides a clear and explicit basis for selecting the best answer, noting that a word or phrase signaling contradiction is required, though they're less precise about the reasoning for ruling out choices A and B.

Option A, "similarly." "Similarly, based on radiocarbon dating—" Okay, that doesn't really make sense. [Choice B,] "in conclusion." It could be "in conclusion," but it's not. [Choice C,] "as a result." It cannot be "as a result" because it's contradicting, so it can't be a result. [Choice D,] "however." "However" makes the most sense because it's showing—okay, so humans thought that people came into the Americas 13,000 years ago. However, they arrived more than 30,000 years [ago]. So "however" makes more sense because it's showing that "No, it's not 13,000; it was actually 30,000." So that makes more sense.

Math

This subsection presents the results of the twenty Math questions analyzed in the study. Note that unlike in Reading and Writing, for which questions were grouped by content domain, questions for Math were scrambled, meaning that, for example, an Algebra question could be followed by an Advanced Math question, and so on. This is because in actual Math section testing, questions are ordered by difficulty from easiest to hardest, which has the practical effect of mixing up the order of presentation, whereas Reading and Writing questions are presented in a fixed order by content domain. In the analysis that follows, questions have been regrouped into their respective content domains—Algebra, Advanced Math, Problem-Solving and Data Analysis, and Geometry and Trigonometry—for easier discussion. In the study itself, the presented Math questions moved generally from easiest to hardest (as measured by PSB), with some slight deviations.

In addition, two key differences between the Reading and Writing and Math analyses are worth noting here. First, the Math behaviors listed for each question type are *expected* rather than *required*, meaning that students needed only to demonstrate one of the behaviors to have been considered enacting the question's intended construct. This is because, by design, Math questions tend to have multiple solving pathways, only one of which a given student is expected to pursue. Second, answering correctly isn't one of the listed behaviors, as it is for Reading and Writing. However, the differential is calculated in a similar way and represents the difference between the number of students answering a given question correctly and the number of students who both answered correctly and exhibited at least one of the expected behaviors.

ALGEBRA

Questions in the Algebra content domain of the digital SAT Suite tests align most closely with topics covered in a typical rigorous first-year secondary algebra course, including assessing the skills and knowledge associated with working with linear expressions, linear equations in one and two variables, linear functions, systems of linear equations, and linear inequalities. Test questions cover such skills and knowledge as creating and using a linear equation; identifying an expression or equation that represents a situation; interpreting parts of a linear equation in context; making connections between linear equations, graphs, tables, and contexts; determining the number of solutions and the conditions that lead to different numbers of solutions; and calculating and solving. The test questions aligned to algebra skill/knowledge elements range in difficulty from relatively easy to relatively complex and challenging. The test questions require students to demonstrate skill in generalization, abstraction, and symbolization, with a strong emphasis on equivalence and using structure. Many of the test questions are constructed to allow for more than one solving strategy.

Five Algebra questions were included in this cognitive interview study: one Linear Functions: Interpret question, two Linear Functions/Inequalities in One Variable: Create and Use questions, one Linear Equations in Two Variables: Make Connections question, and one Linear Systems: Determine Conditions question.

Linear Functions: Interpret

To answer Linear Functions: Interpret questions as intended, students are expected to demonstrate at least one of the following behaviors:

1. Read and demonstrate comprehension of the context described.
2. Interpret an element of a linear equation based on the context described.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves, first, reading and comprehending the description of a context and, second, deciding on the proper interpretation of an element of the given linear equation based on that context. Once students have made sense of the context and determined their interpretation of the element, they must select the phrase, clause, or sentence from the four provided answer choices that best matches their interpretation of the element.

Table 21 summarizes how students performed on the Linear Functions: Interpret question included in the study.

Table 21. Student Performance on Math: Algebra—Linear Functions: Interpret Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors			Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
			1	2	One or Both			
1 MC *n* = 23	Science	3	14 (61%)	19 (83%)	21 (91%)	18 (78%)	18 (78%)	0

MC = multiple-choice

Table 21 indicates that the Linear Functions: Interpret question included in the study performed as intended, with a differential of 0. Every student who demonstrated one or both expected behaviors answered the question correctly. To show evidence of properly interpreting an element of a linear equation, students needed to demonstrate a correct interpretation of a number that represents a rate of change in the given equation. While students needed to comprehend the described context (expected behavior 1) to make an accurate interpretation of the element (expected behavior 2), not all students directly demonstrated their comprehension through their explanations captured during the interview process; thus, the number of students who demonstrated expected behavior 2 is greater than the number of students who demonstrated expected behavior 1.

Vignettes from students answering correctly and demonstrating one or both expected behaviors illustrate the kinds of complex thinking elicited by Algebra—Linear Functions: Interpret questions on the digital SAT Suite assessments.

Question 1

Question 1 is a relatively easy (PSB 3) multiple-choice question set in a science context. The question asks students to read and understand the context and to identify the best interpretation of the number 16 in the equation in terms of that context.

$$d = 16t$$

The given equation represents the distance d, in inches, where t represents the number of seconds since an object started moving. Which of the following is the best interpretation of 16 in this context?

A) The object moved a total of 16 inches.

B) The object moved a total of $16t$ inches.

C) The object is moving at a rate of 16 inches per second.

D) The object is moving at a rate of $\frac{1}{16}$ inches per second.

Choice C is the correct answer. To answer this question correctly, students are expected to determine from the passage that the variable d on the left-hand side of the equation represents distance, in inches, and that the variable t on the right-hand side of the equation represents time, in seconds. Then students must understand that to compute the distance in inches, the time in seconds must be multiplied by a rate in inches per second. Therefore, 16 is the rate at which the object is moving in inches per second.

Student M1 begins their successful approach to question 1 by stating their understanding of the context and what the variables in the equation represent. Student M1 then uses the units for the variables to figure out the units that the number 16 must have, which helps them determine the correct interpretation of 16 in this context.

> When I read the question, I think, well, *d* is the distance and *d* is in inches, but they're multiplying 16 seconds by the time. So 16 has to do something with time. So my initial thought would be the object is moving at a rate of 16. So I want to say 16 is the starting of the rate of seconds. I don't know if the way I said that makes sense. But that's the way I interpreted it. So the object is moving at a rate of 16 inches per second. So I want to say that's the initial time, but it's not saying initial. It's just asking my best interpretation. So I would say the object is moving at a rate of 16 inches per second.

Student M1 seems to want an initial value for time, which isn't relevant for this context but doesn't impede identifying the correct answer.

Student M23 also shows some comprehension of the context by identifying that the variable *t* represents time, though they don't provide a clear explanation for the meaning of the variable *d*. However, that understanding may be implied by the student's first statement.

> Um, since it's distance, distance is always positive. And, um, it's telling us it's moving [in] seconds, and *t* represent[s] a number of seconds. So, so it's—the object is moving at a rate of—no. Yeah. The object is moving at a rate of 16 inches per second.

Student M23 does correctly identify the best interpretation of the number 16 in the context but doesn't provide clear verbal detail about how they came to that interpretation.

Linear Functions/Inequalities in One Variable: Create and Use Questions

To answer the Linear Functions/Inequalities in One Variable: Create and Use question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Write a linear equation or inequality in one variable.
2. Solve a linear equation or inequality in one variable.
3. Apply algebraic reasoning to a context that represents a linear equation or inequality.
4. Solve by guessing and checking.
5. Choose appropriately whether to round up or down.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that typically involves, first, writing a linear equation or inequality based on the description of the context and, second, solving the linear equation or inequality. Alternatively, students could use algebraic reasoning to solve the problem, a behavior that mimics the actions a student would take to solve an equation or inequality had they themselves written it. Then students must either choose the answer option that best matches their answer to the question or input their answer into the space provided.

Table 22 summarizes how students performed on the Linear Functions/Inequalities in One Variable: Create and Use questions included in the study.

Table 22. Student Performance on Math: Algebra—Linear Functions/Inequalities in One Variable: Create and Use Questions.

Question	Subject Area	PSB	Demonstrated Expected Behaviors						Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
			1	2	3	4	5	One or More			
4 MC *n* = 23	Science	4	1 (4%)	1 (4%)	18 (78%)	n/a	n/a	18 (78%)	16 (70%)	16 (70%)	0
5 SPR *n* = 23	Real-world	4	7 (30%)	7 (30%)	13 (57%)	5 (22%)	14 (61%)	23 (100%)	18 (78%)	18 (78%)	0

MC = multiple-choice; SPR = student-produced response

Table 22 indicates that the Linear Functions/Inequalities in One Variable: Create and Use questions included in the study performed as intended, each with a differential of 0. Every student who demonstrated one or more of the five expected behaviors answered the question correctly. For question 4, students were expected to write and solve a linear equation in one variable. Most of the students in the sample took the alternative approach mentioned above and used the information in the context to reason algebraically to solve for the unknown quantity. For question 5, students were expected to write and solve a linear inequality in one variable and then decide whether to round the initial solution up or down to match the context. While all students had to decide whether to round up or down, not all students gave evidence of their decision-making during the interview. Notably, for question 5 many students used a guess-and-check method

to find a solution. Students then chose the correct answer from the four answer choices provided or entered their answer in the space provided.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of complex thinking elicited by Linear Functions/Inequalities in One Variable: Create and Use questions on the digital SAT Suite assessments.

Question 4

Question 4 is a medium-difficulty (PSB 4) multiple-choice question set in a science context. This question asks students to determine an unknown value based on a given context. This context represents a common linear pattern in which a starting amount changes evenly over time.

> A candle is made of 17 ounces of wax. When the candle is burning, the amount of wax in the candle decreases by 1 ounce every 4 hours. If 6 ounces of wax remain in this candle, for how many hours has it been burning?
>
> A) 3
>
> B) 6
>
> C) 24
>
> D) 44

Choice D is the correct answer. This context represents a common linear pattern in which a starting amount (in this case, 17 ounces of wax) experiences a constant rate of change (here, a decrease by 1 ounce every 4 hours), meaning the change is the same amount for even increments of time. To answer this question correctly, students can either write an equation in one variable that represents the situation in the given context and solve for the number of hours the candle has been burning or use algebraic reasoning to determine the number of hours the candle has been burning.

In order to find this number of hours, students could write the equation $17 - \frac{1}{4}t = 6$, where 17 is the amount of wax, in ounces, the candle has before it begins to burn; $-\frac{1}{4}$ is the constant rate of change that represents that the amount of wax in the candle decreases by 1 ounce every 4 hours; t is the amount of time the candle has been burning, in hours; and 6 is the amount of wax, in ounces, remaining in the candle. To solve this equation, students would likely subtract 17 from both sides of the equation, which gives $-\frac{1}{4}t = -11$. Then students would multiply both sides of this equation by -4, which results in $t = 44$. Alternatively, students may choose to use algebraic reasoning that follows a pattern similar to that involved in solving the equation, such as first observing that the candle starts with 17 ounces of wax and decreases by 1 ounce every 4 hours and then reasoning that 4 hours after the candle has been burning, 16 ounces of wax remain; 8 hours after the candle has been burning, 15 ounces of wax remain; and so on, until the number of ounces of wax remaining reaches 6 ounces.

Student M7 begins their successful approach to question 4 by constructing an equation with a constant and assigning variables that represent the given situation.

> On my paper, I'm going to first note the total amount, 17 [ounces], from the start. When the candle is burning, the amount of wax is subtracting 1 [ounce] every four 4 hours. So 1 every 4 hours, so that would be the amount over the time. So $-\frac{1}{4}t$ or x. I like using x. I believe that's how we begin, but I'll check. $17-\frac{1}{4}t[x]$, and I want to say that equals 6.

Student M7 next solves the equation for the time variable, which they referred to as both t and x. Student M7 first eliminates the fraction by multiplying both sides of the equation by the denominator in the fraction (4) and then isolating the variable to solve for its value.

> And I'll solve for x or what would be time in this example. So to simplify and answer this, I'm going to multiply everything by 4. 17×4 is 68, $[68]-x=24$. Let me subtract 24 from both sides and then add x. So $68-24$, that's 44, and that equals x. So that says 44 hours.

Student M7 uses algebraic reasoning to check their answer against the information given to make sure the answer makes sense.

> However, let me make sure this is correct. So every hour, it's losing 1 ounce. So it's going to leave 11 ounces within the 4 hours. No, within 44 hours, it'll lose 11 ounces. It starts with 17, ends with 6. If I lose 11 from the 17, that's 6, so 44 hours would be the correct answer.

Student M21 follows the alternative path to solve the question. They begin by writing down what was given in the question and trying to determine what the context means algebraically.

> So first I'll write down all the values that it gives me. So 17 ounces of wax initially. And when the candle's burning, the amount of wax in the candle decreases by 1 ounce every 4 hours. So I'll also write that down. Decreases 1 ounce every 4 hours. All right, so 6 ounces of wax remain in this candle for how many hours? So I'll write down the end amount, 6 ounces of wax at the end. So I'll just set it equal to 6 for the ounces . . .

Student M21 seems unsure how to use the given rate or write an appropriate equation from the information provided in the question, but they do understand how to find the number of hours.

> . . . and $17-1$. So I'll just go maybe 6. 6 ounces of wax remain in this candle. 1 ounce every 4 hours. So it's lost 11 ounces, $11\times4=44$. So I'll go with 44 hours.

Student M9 begins their successful approach to question 4 by using algebraic reasoning skills, but their reasoning is different than student M21's. Student M9 first figures out by how much wax the candle had decreased, then uses the rate to get the amount of wax used in 4 hours, then again uses the rate to get the amount used in 8 hours, and then uses the constant rate of increase pattern to find the solution.

Because we know that the candle is made out of 17 ounces and if it's left with 6, I can do 17 – 6, which will give me 11 ounces. And this is when the candle is burning, the amount of wax will decrease by 1 ounce every 4 hours. So every 4 hours it decreases 1 ounce. So 1 equals 1 ounce those 4 hours. So that means that 2 ounces would equal 8 hours. So it's just been multiplied. So what I can do is just do 11 times 4. And it has been run for around 44 [hours]. [*Student double-checks result using different values.*] Yeah, I believe that's it. Okay.

Question 5

Question 5 is a medium-difficulty (PSB 4) student-produced response question set in a real-world context. This question asks students to find an unknown value based on the context described. Question 5's context exhibits a linear pattern similar to that in the context in question 4: a starting amount changes at a constant rate by an incremental amount over each unit of time. While question 4 represents a linear equation in one variable, question 5 represents a linear inequality in one variable.

> An event planner is planning a party. It costs the event planner a onetime fee of $35 to rent the venue and $10.25 per attendee. The event planner has a budget of $200. What is the greatest number of attendees possible without exceeding the budget?

The correct answer for this question is 16. To answer this question correctly, students should either write an inequality in one variable that represents the situation described, use algebraic reasoning, or use a guess-and-check method to find the number of attendees. Then students need to determine whether their decimal solution requires rounding in order to meet the requirements in the question.

To find the greatest number of attendees possible without exceeding the budget, students could write the inequality $35 + 10.25a \leq 200$, where 200 is the dollar amount of the budget that the event planner can't exceed; 35 is the cost of the onetime fee, in dollars; 10.25 is the cost per attendee, in dollars; and a is the number of attendees. Alternatively, students may choose to solve for the number of attendees using algebraic reasoning that follows a pattern similar to writing and solving an inequality but doesn't use variables. Students may also choose to use guess-and-check by trying different values for the number of attendees until they narrow in on the largest number that keeps the total cost under $200. Once students find an answer through one of these paths, they must decide whether an answer that includes a decimal part should be rounded up or down so as not to exceed the budget. Since the value of a represents the number of attendees, it must be a whole number. Therefore, any decimal value has to be rounded in the direction that keeps the cost under the budget—that is, rounded down to the nearest whole number.

Student M5 begins their successful approach to question 5 by making sense of the context.

> It costs the event planner a onetime fee of $35 to rent a venue, so that's how much is the venue. So 35 is always going to be constant. 35. And 10.25 per attendee. The event planner has a budget of $200. Okay.

Student M5 then uses a guess-and-check method to get from the fixed cost (the onetime fee) to the greatest number of attendees that the budget could support.

> So what I would do is I would do 200 divided—no, I would do 10.25 on my calculator until I get as close to 200 as I can. Minus 35. So I'm going to do, first, 10×6, which is 60, so that's too low. So $10.25 \times$ maybe 15. We have 153. I'm going to do $10.25 \times$ maybe 17. We have $174 + 35$ is 209, so that's a little bit too up. Um, 10.25×16, $164 + 35 = 199$. That's literally perfect. So she can have, uh, 16 attendees without exceeding the budget.

Student M5 doesn't mention rounding their answer, but the description of their thinking shows an understanding of the need for a whole number: 15 is not enough attendees and 17 is too many, but 16 is the correct number.

Student M20 begins their successful approach to question 5 by writing an inequality that represents the situation in the question and then using the correct operations to solve for the unknown value. In the process, student M20 expertly explains the need to round the initial solution down to obtain the correct answer of 16.

> In order to do this, I set up an inequality. So it's $10.25x$, where x is the amount of people, $+35$ because that's the onetime fee, and that'll be no greater than 200. Subtract 35 from both sides, $10.25x$ is no greater than 165. Divide both sides by 10.25. The answer is 16.09, but since I can't have a decimal amount of people, the answer would be 16. That's the greatest amount of people.

Student M25 begins their successful approach to question 5 by subtracting the onetime fee from the budget and then determining how many times the per-attendee fee could be paid from the remaining budget amount. Then student M25 accurately describes why the answer must be rounded.

> So you have $200 for this party, and then you subtract 35 because of the onetime fee. And so that would be $165 left over. And then per attendee, it is $10.25. So you'd do $165 divided by 10.25, $10.25 per attendee, and see. You get 16. And you can't have 0.9 other person, so the max people that could attend would be 16 people without exceeding the budget.

Note that the reasoning used by student M25 closely mirrors the process used by student M20 to solve the inequality they wrote.

Linear Equations in Two Variables: Make Connections

To answer the Linear Equations in Two Variables: Make Connections question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Make connections between an algebraic representation and a graph of a linear equation in two variables not set in context.
2. Identify the linear equation for a translated graph.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves, first, examining a graph and identifying elements of the equation for that graph and then, second, identifying the equation that represents the graph before it was translated.

Table 23 summarizes how students performed on the Linear Equations in Two Variables: Make Connections question included in the study.

Table 23. Student Performance on Math: Algebra—Linear Equations in Two Variables: Make Connections Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors			Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
			1	2	One or Both			
16 MC *n* = 20	None	7	11 (55%)	6 (30%)	11 (55%)	7 (35%)	6 (30%)	1

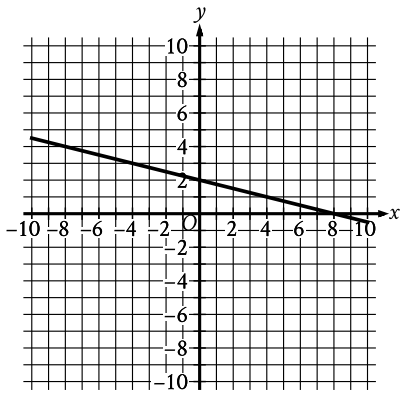
MC = multiple-choice

Table 23 indicates that the Linear Equations in Two Variables: Make Connections question performed as intended, with a differential of 1.

Vignettes from students answering correctly and demonstrating one or both expected behaviors illustrate the kinds of complex thinking elicited by Linear Equations in Two Variables: Make Connections questions on the digital SAT Suite assessments.

Question 16

Question 16 is a hard (PSB 7) multiple-choice question without a context. The question shows a linear graph and informs students that the graph represents a translation of the graph of the function *f*. Students are asked to identify which equation from the four answer options defines the function *f* prior to the translation.



The graph of $y = f(x) + 14$ is shown. Which equation defines function f?

A) $f(x) = -\frac{1}{4}x - 12$

B) $f(x) = -\frac{1}{4}x + 16$

C) $f(x) = -\frac{1}{4}x + 2$

D) $f(x) = -\frac{1}{4}x - 14$

Choice A is the correct answer. To answer this question correctly, students must make a connection between the given equation and the given graph. If the graph shown is $y = f(x) + 14$, then students need to determine the relationship between the graph and the equation for $f(x)$. Successful students will realize that $y = f(x) + 14$ represents a vertical shift up of 14 units to the graph of $y = f(x)$. Therefore, the equation that defines function f is an equation of the graph shown with a translation down 14 units. Since the graph shown has a slope of $-\frac{1}{4}$ and a y-intercept of (0, 2), it follows that an equation that defines function f is $f(x) = \left(-\frac{1}{4}x + 2\right) - 14$, or $f(x) = -\frac{1}{4}x - 12$. Therefore, function f, represented in the graph of $y = f(x) + 14$, is defined by $f(x) = -\frac{1}{4}x - 12$.

Student M13 begins their successful approach to answering question 16 by determining the slope of the given graph by counting the gridlines and then recognizing that the slope must be negative.

> Let's see. I'm going to find a good point. When it goes 1 down, it goes 1, 2, 3, 4 [over]. So $\frac{1}{4}$ is a slope. And it's negative. Okay.

Student M13 next rereads the question. They then count the distance in the graph from the y-axis to the line to determine the y-coordinate of the y-intercept (2) of the line graphed.

> The graph of $y = f(x) + 14$ is shown. Which of the following? Which equation defines function f? So $f(x)$ is in this function? $f(x)$, so 1, 2, 1, 2.

It's not clear whether student M13 is using numbers from the options or has silently reasoned that to get to the 2 in the graph after a shift of 14 to $f(x)$, the function $f(x)$ had –12 as the y-coordinate of the y-intercept.

> So 14. $-12 + 14$ is 2. Yep. $-12 + 14$. It's $+2$. So I would go with that.

Student M8 begins their successful approach to question 16 by making the connection quickly between the function graphed, the function they're seeking, and the equation. Then they're able to identify the correct equation. Student M8 first identifies the y-coordinate of the y-intercept of the given graph and then connects that value (referred to as the b-value) to what it would have to be for the graph of $f(x)$.

> So I know that for this one, the—my intercept is 2. So, looking at this, I just know that a first something plus 14 would have to equal 2 for the $+b$-value. So, looking at this, I just know that it would be that because the—all the slopes are constant, so I'm just looking at the b-value, and since $f(x)$ is just -12, if I added 14, it would just be $+2$. And that's what that is, so.

Linear Systems: Determine Conditions

To answer the Linear Systems: Determine Conditions question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Understand the conditions for the number of solutions for a linear system of equations.
2. Find the value of a constant in a linear system of equations.
3. Use a graph to determine the solution(s) to a linear system of equations.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves first examining the given system of linear equations, which aren't written in standard form ($Ax + By = C$) and include an unknown constant, and then recognizing the conditions that lead to a system of linear equations that has no solutions. Then students need to solve for the unknown constant based on the condition that leads to no solution. Alternatively, a student may use a graph to identify the value of the unknown constant that leads to the condition of no solution for the system of linear equations.

Table 24 summarizes how students performed on the Linear Systems: Determine Conditions question included in the study.

Table 24. Student Performance on Math: Algebra—Linear Systems: Determine Conditions Question.

			Demonstrated Expected Behaviors						
Question	Subject Area	PSB	1	2	3	One or More	Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
20 SPR n = 19	None	7	4 (21%)	3 (16%)	1 (5%)	4 (21%)	3 (16%)	3 (16%)	0

SPR = student-produced response

Table 24 indicates that the Linear Systems: Determine Conditions question included in the study performed as intended, with a differential of 0. Every student who answered correctly demonstrated one or more of the three expected behaviors.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of complex thinking elicited by Linear Systems: Determine Conditions questions on the digital SAT Suite assessments.

Question 20

Question 20 is a hard (PSB 7) student-produced response question without a context. The question gives a system of two linear equations in two variables, x and y, with an unknown constant, r. Students are asked to identify the value of r if the system has no solution. This is a very challenging question because the system of linear equations isn't written in a standard form, it includes fractions and negative values for elements of the given equations, and one equation includes an unknown constant.

$$48x - 72y = 30y + 24$$

$$ry = \frac{1}{6} - 16x$$

In the given system of equations, r is a constant. If the system has no solution, what is the value of r?

The correct answer to this question is −34. To answer this question correctly, students must articulate what it means for a system of two linear equations in two variables to have no solution. This means that the two lines are nonintersecting, or parallel (i.e., they have the same slope), and don't coincide (i.e., they don't have

the same y-intercept). Students can determine this by solving algebraically, or they can determine this graphically. To find the value of r algebraically, students will likely rewrite the given equations in either standard form $Ax + By = C$ or slope-intercept form $y = mx + b$. Once the two equations are both in the same format using either of these typical forms, students can find the value for r that leads to the system of linear equations having no solution. Alternatively, students may choose to solve graphically. They might use the Desmos Graphing Calculator embedded in the test delivery system or their personal graphing calculator. Since there is an unknown constant in the equation, students would need to investigate what value of r leads to a system of parallel lines that aren't coincident.

Student M7 begins solving the problem by speaking somewhat inelegantly about how one can know there's no solution. Student M7 doesn't use mathematical language precisely or consistently, but one can follow their solution path and realize that the math steps used to solve for the value of the constant r lead to the correct solution.

> So no solution means they do not equal each other, or if they do equal each other, that's all real solutions. So can't be with each other.

"Means they do not equal each other" can reasonably be presumed to mean that when solving the system of equations algebraically, the student has discovered that the interim solution when solving for one variable leads to a false statement, such as $0 = 5$, which means there's no value for that variable and thus no solution to the system. Student M7 correctly states that "or if they do equal each other, that's all real solutions," which can reasonably be presumed to mean that if solving the system algebraically, the student has discovered that the interim result when solving for one variable yields an equation that is always true no matter the value of the variable, such as $5 = 5$, and that the value of that variable is "all real numbers," and thus the system has infinitely many solutions.

Next, student M7 rewrites the equations in standard form. They wisely note that careful work is required to avoid making errors when rewriting. They then rewrite both equations in the order presented.

> They have to be two different numbers when they first simplify, meaning $48x - 72y = 30y + 24$. My original thought is this is going to be a lot of manipulation of the equation, which can make annoying mistakes. So I got to make sure to be careful with it. I add 72 to both sides. Or I'll subtract 30. I'll keep it on the same side. $-102y$. [inaudible] Yeah. $-102y = 24$. $r \times y$ equals—I'm going to add $16x$. So $ry + 16x = \frac{1}{6}$. Okay. So now that I've [inaudible] just a little bit, let me fix this so it's a little better. $16x + ry$. Okay, so it's $48x$, $16x + ry$. What is the value r? Okay.

At this point, student M7 has rewritten the two equations in standard form, with x- and y-terms on one side of the equation and the constant term on the other:

$$48x - 102y = 24$$
$$16x + ry = \frac{1}{6}$$

Student M7 talks through what needs to happen next in order for the two equations to represent parallel lines. They correctly observe that the x-term in the

second equation needs to be multiplied by −3, and while they don't describe this clearly, they do correctly multiply the whole second equation by −3. This allows the *x*-term to be eliminated when the two equations are added together. When solving a system of equations that represents parallel lines using the elimination method, both variables will be eliminated and result in an equation where the two sides of the equation aren't equivalent. Student M7 refers to this as "cancel out." They then reason about what value of *r* would eliminate the variable *y* when the manipulated equations are added together. Note that student M7 refers to "102," the coefficient on the *y*-term from the first equation, a couple of times, which they recognize would need to be equivalent to $-3r$ for the sum of the *y*-terms to be 0 and thereby also be eliminated.

> So in order for this not to work, I know I need to have *x* and *y* cancel each other out. However, I know I ha[ve] to multiply 16 by 3. So it has to be $102 - r$, but it has to be $-3r$. So I know it would have to be some—*r* would have to be multiplied by 3 and so when added to 102. Okay. So $-102 - 3r = 0$. $-3r = 102$. Subtracted [the] 2 [equations], 102 divided by 3, or $-3 \ldots -34$. So it's -34×3. -102. If I subtract the two, it cancel[s] out. So I know it is -34.

Student M7 next shows some concern about whether they've gotten the sign correct and checks their work to be sure. They seem to believe a negative answer isn't an acceptable answer but then realize the test section's directions say they can enter a negative response. (In the final paper-based version of the SAT Suite, negative answers couldn't be provided for student-produced response questions in Math, nor would questions in that version call for such answers.)

> Oh. All right. Is it just positive 34? [inaudible] Let me see. Make sure all my signs are correct. $48x - 72y = 30y + 24$. And say *r* is 34, $34y = \frac{1}{6} - 16x$. Okay. Okay. [inaudible] Would it have to be positive or negative? Well, negative is not an option. So I believe it's 34. I'll double-check for the sake of making sure. [inaudible] 24 and 34. $16x + 34y =$ [inaudible]. I don't think that it can be the right answer because I thought it would have to be negative. I don't believe negative is an option. So I think. Make sure. It says I could put a negative, but whenever I try to type negative, it doesn't work. Okay. So I can put a negative. Oh, [inaudible] at the beginning. Okay. Yeah. So -34. That would make sense.

Student M29 also has a successful solution to question 20, this time using a graphical approach. Student M29 starts with an algebraic solving method similar to student M7's but decides that there should be an easier way and eventually graphs the system of equations.

> Right, yes, I'm just looking at this and seeing maybe if I put in—for the first equation, if I just put *x* to one side and then plug in that value to the second equation to just get *r*, which is a constant, and only have one value. But there should be an easier way to do this. I am still thinking that—I'm going to try doing what I first did, and I'm assuming I just probably made a algebra error, where I'm trying to have the equations be equal to each other. So $48x$ [inaudible]. So $-102y + 48x - 24$. And then $ry + 16x - \frac{1}{6}$. Let me just see. 18. Yeah. It will not work out that way,

> unfortunately. So let me see. Oh, I can graph them? Let me see. I'm going to try graphing them and then see if that will help me. 48*x*. Well, never mind.

Student M29 doesn't describe how they graphed the equations, but one can reasonably infer from what student M29 says next that both equations were graphed.

> So in order for it to have no solutions, that cannot cross. Right now, they look to be perpendicular to each other.

"That cannot cross" is likely a reference to the fact that if there is no solution to the system, the two lines cannot intersect.

Student M29's assertion that they're "just plugging in random values" can reasonably be presumed to refer to entering random values for *r*. Note that while narrating their work, student M29 first mistakenly says "to make sure it's perpendicular" but later corrects to "parallel to each other."

> Let me see. I'm just plugging in random values. I don't really know what [inaudible]. Let me try plugging negative values. Oh, interesting. So plugging in negative values, which is what I thought earlier, does make it move away from each other. I just have to make sure it's perpendicular to each other so—parallel to each other so it never crosses, because 33 seems to cross over here. So let me just see what value would work. 34. Oh, I lost my graph. So 34 does not seem to cross. I'm just going to make sure, plugging in 35. 35 definitely crosses. All right. So I'm just going to put 34.

At the end, student M29 seems to be making an observation comparing use of the Desmos Graphing Calculator, which is embedded in the test delivery software, and a handheld graphing calculator. For many of the latter, the equation would need to have been rewritten in a form of "$y =$ ____" in order to have been entered.

> I do think we can't easily plug this into the calculator. So we would have to set *y* equal to a certain function first. So with decimals, this question is much easier than if we were doing the actual SAT with only the calculator.

ADVANCED MATH

The Advanced Math topics assessed on the digital SAT Suite tests extend those covered in the Algebra content domain into the realm of nonlinear equations and functions and align most closely with topics mastered in a typical rigorous second-year secondary algebra course and sometimes beyond. Since these Advanced Math test questions build on skills and knowledge first mastered with linear expressions and equations, it follows that these topics should also be well represented on college and career readiness assessments such as those of the digital SAT Suite. As a result, skill/knowledge elements in Advanced Math are represented on the digital SAT Suite tests in relatively high proportions. The Advanced Math content domain assesses skills and knowledge associated with working with quadratic, exponential, polynomial, rational, radical, absolute value, and conic section equations and functions. Similar to Algebra questions, questions in the Advanced Math domain cover skill/knowledge applications similar to those in the Algebra domain but with different types of equations, including creating and

using a nonlinear equation; identifying an expression or equation that represents a situation; interpreting parts of an equation in context; making connections between equations, graphs, tables, and contexts; determining the number of solutions and the conditions that lead to different numbers of solutions; and evaluating and solving using nonlinear equations as well as systems that include a nonlinear equation. The test questions in the Advanced Math domain range in difficulty from relatively easy to relatively complex and challenging. Many of the test questions represent challenging, authentic problems in context for which students can draw on strategies developed during their coursework to solve.

Six Advanced Math questions were included in this cognitive interview study: two Nonlinear Functions questions, one Make Connections question, one Determine Conditions question, one Nonlinear Equations: Solve question, and one Rewrite question.

Nonlinear Functions

To answer the Nonlinear Functions questions as intended, students are expected to demonstrate at least one of the following behaviors:

1. Interpret an element based on a context.
2. Identify a nonlinear model.
3. Make connections between the description of a context and ordered pairs in that description.
4. Find the value of a constant in a nonlinear equation.
5. Guess-and-check using the provided answer choices.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves first interpreting a context; then either (a) identifying an exponential equation that models the description of the context or (b) creating an exponential equation (or a part of such an equation) that models the description of the context; and then using the exponential equation to answer a question about the context.

Table 25 summarizes how students performed on the Nonlinear Functions questions included in the study.

Table 25. Student Performance on Math: Advanced Math—Nonlinear Functions Questions.

			Demonstrated Expected Behaviors								
Question	**Subject Area**	**PSB**	**1**	**2**	**3**	**4**	**5**	**One or More**	**Answered Correctly**	**Demonstrated One or More Behaviors and Answered Correctly**	**Differential**
3 MC *n* = 23	Real-world	4	13 (57%)	11 (48%)	1 (4%)	n/a	8 (35%)	20 (87%)	19 (83%)	18 (78%)	1
6 MC *n* = 23	Science	6	n/a	n/a	8 (35%)	9 (39%)	6 (26%)	14 (61%)	13 (57%)	12 (52%)	1

MC = multiple-choice

Table 25 indicates that the Nonlinear Functions questions included in the study performed as intended, with differentials of 1. In both cases, all but one student demonstrating one or more of the expected behaviors answered each question correctly.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of cognitively complex thinking elicited by Nonlinear Functions questions on the digital SAT Suite assessments.

Question 3

Question 3 is a medium-difficulty (PSB 4) multiple-choice question set in a real-world context. The question asks students to read and understand a financial context about an account balance and create a nonlinear equation that could define this balance.

> A company opens an account with an initial balance of \$36,100.00. The account earns interest, and no additional deposits or withdrawals are made. The account balance is given by an exponential function A, where $A(t)$ is the account balance, in dollars, t years after the account is opened. The account balance after 13 years is \$68,071.93. Which equation could define A?
>
> A) $A(t) = 36{,}100.00(1.05)^t$
>
> B) $A(t) = 31{,}971.93(1.05)^t$
>
> C) $A(t) = 31{,}971.93(0.05)^t$
>
> D) $A(t) = 36{,}100.00(0.05)^t$

Choice A is the correct answer. To answer this question correctly, students need to understand that exponential growth is modeled by an equation of the form $A(t) = A_0\left(1 + \frac{r}{100}\right)^t$, where $A(t)$ is the current account balance, A_0 is the initial account balance, r is the growth (or interest) rate, and t is the time, in years. Since it's given that the initial balance of the account is \$36,100.00, it follows that $A_0 = 36{,}100.00$. Substituting 36,100.00 for A_0, it follows that $A(t) = 36{,}100.00\left(1 + \frac{r}{100}\right)^t$. Students must then make a connection between the contextual explanation of "the account balance after 13 years is \$68,071.93" and translate that into an ordered pair, $(13, 68{,}071.93)$. Students would substitute 13 for t and 68,071.93 for $A(t)$ into the equation and solve for r, which yields approximately 5. Substituting 5 for r in the equation yields $A(t) = 36{,}100.00\left(1 + \frac{5}{100}\right)^t$, or $A(t) = 36{,}100.00(1.05)^t$. Alternatively, students can evaluate each of the given options by substituting 13 for t, where the correct answer would then be the answer choice that would lead to $A(13) = 68{,}071.93$.

Student M16 begins their successful approach to question 3 by stating their understanding of the context and acknowledging that this question models exponential growth.

> Okay. A company opens up—okay, so this looks like it's a, um, rate of change question, an exponential growth question. So, um, a company opens an account with an initial balance of 36,100, the account earns interest, and, uh, no additional deposits or withdrawals are made. The account balance is given by an exponential equation, um, on function A where $A(t)$, um, is the account balance in dollars t years after the account is open. The account balance after 13 years is 68,071.93. Which equation could define A?

After rereading the question, student M16 references the positive rate of change, which allows them to identify the two answer choices that include the correct growth rate (1.05), and although they don't verbalize it, comes to their conclusion by referencing the initial balance given in the question, allowing them to identify the correct answer choice.

> Um, okay, so it looks like the rate of change is 5%, so it's either [choice] A or B. Um, yeah, it's [choice] A because we know that the initial balance is 36,100.

Student M9 begins their successful approach to question 3 by stating their understanding of the context and acknowledging that substituting 13 for t in the answer choices should result in the account balance, $A(t)$, being equal to 68,071.93. Note that student M9 uses imprecise language when saying "$A \times t$ is account balance," though they do seem to understand that $A(t)$ is the function value and not a multiplication expression.

> Okay. So, first, I want to see that A is the exponential function. So I'll write that down so I can remember it. And then $A \times t$ is account balance. Okay. So this right here would be probably the account balance for that initial. And t is the years after the account is open. So I know we have 13 years. So that's going to be our t. And a total of 68,000. And I'm trying to find A. Okay. Okay, in order to find this. And we try to find something that's going to give me this number. So I would probably plug into my calculator just to be sure. The first one, and then put t as 13. And it's going to give me 68,071 93. So I could have that as an option.

To be certain, student M9 checks another answer choice to gain confidence in their initial selection.

> And I would probably check another number just in case—the bottom one, $36{,}100.00(0.05)^{13}$. And that does not give me the answer. So I believe it's this one.

Question 6

Question 6 is a hard (PSB 6) multiple-choice question set in a science context. The question asks students to read and understand a context with two sets of values and then create an exponential equation in the given form to identify a constant within this equation.

> A scientist initially measures 12,000 bacteria in a growth medium. 4 hours later, the scientist measures 24,000 bacteria. Assuming exponential growth, the formula $P = C(2)^{rt}$ gives the number of bacteria in the growth medium, where r and C are constants and P is the number of bacteria t hours after the initial measurement. What is the value of r?
>
> A) $\frac{1}{12{,}000}$
>
> B) $\frac{1}{4}$
>
> C) 4
>
> D) 12,000

Choice B is the correct answer. To answer this question correctly, students must understand that the initial number of bacteria, 12,000, is the value of C in the given equation, $P = C(2)^{rt}$. Once this connection is made, students then need to understand that when $t = 4$, $P = 24{,}000$. Students could substitute these values into the given equation, which yields $24{,}000 = 12{,}000(2)^{4r}$. Dividing both sides of this equation by 12,000 yields $2 = (2)^{4r}$. It follows that $4r = 1$, or $r = \frac{1}{4}$. Alternatively, after substituting the values for P, C, and t as identified above in the given formula, students could then evaluate the answer choices by substituting the value for r from each answer option to determine the key.

Student M15 begins their successful approach to question 6 by connecting the given information to the variables given in the equation. Student M15 represents the given information in terms of the variables t, C, and P, indicating that $t = 4$ and that C represents the initial measurement of 12,000 bacteria.

> If there is 24,000 and C are constants—does that mean—so $t = 4$. r is what we're trying to find. C is the initial value, I believe, so that equals 12,000.

Student M15 interprets from the context that when $t = 4$, $P = 24{,}000$, and substitutes these values in the given equation, $P = 12{,}000(2)^{rt}$. They erroneously divide the time by 100, thinking the rate is a percentage. Then they explore a couple of ideas about how to solve for r.

> And, yeah, we're trying to find r. We have C and t, and $P = 24{,}000$. We have $24{,}000 = 12{,}000(2)^{r0.04}$ because for exponential growth, you always have to take it and divide it by a hundred, so it's 0.04, or because, yeah, it's always be like that for hours, I believe too. So if I were to do 24,000 divided by 12,000—well, actually, it's 12,000 divided by 2, so it's just put to the 5th power.

Student M15 then plugs the values of t and C that they'd previously identified into the answer options. They then check to see whether the equation then yields the corresponding value of P, 24,000, that they'd previously identified and, thus, an equation that represents the situation.

> Let me just plug in the answers to see which one I could get that could equal it. $12{,}000 \times 2$. I was never good at these exponential problems. But the constant value of r—P is the number of bacteria after the initial measurement. . . . Gives 24,000. Okay. So I've put in, plugged in $12{,}000(2)^{0.25(4)}$. Yeah, and the 0.25 is representative of 1 over 4. And I did get 12,000—or 24,000 bacteria, so I'm just going to put that in, and then I'd come back and try to plug more things in, but since I got that as an answer, I'm probably not going to double-check everything else just because I'm already fairly okay on time.

Student M8 begins their successful approach to question 6 by offering their understanding of the context along with some observations about the given equation, such as the initial number of bacteria in the growth medium and the rate of growth. Student M8 doesn't explicitly say why they believe C is the initial value, but they appear to come to this conclusion while reading the question.

> So r would be, um, the rate. Yeah. So C would be 12,000, 12,000.

Student M8 then proceeds to evaluate the answer choices with the knowledge that when $t = 4$, $P = 24{,}000$. After performing this calculation, student M8 arrives at the answer.

> t would be—um, I, yeah. I think it would be—let me test this out by just plugging it in, so. Um, 4, so then—yeah, it would be $\frac{1}{4}$. Because after setting up the equation $12{,}000 \times 2$, uh, and then $\frac{1}{4} \times 4$, I was left with 24,000. And I think that's the most reasonable choice.

Make Connections

To answer the Make Connections question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Make connections between a graph and solutions: polynomial.
2. Use the solutions in a calculation.
3. Understand and use function notation.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves making a connection between the solutions to a polynomial and its x-intercepts and then using those solutions to calculate a value. Students can approach such questions graphically or by correctly understanding and using function notation.

Table 26 summarizes how students performed on the Make Connections question included in the study.

Table 26. Student Performance on Math: Advanced Math—Make Connections Question.

			Demonstrated Expected Behaviors						
Question	Subject Area	PSB	1	2	3	One or More	Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
12 MC n = 22	None	6	7 (32%)	6 (27%)	5 (23%)	7 (32%)	9 (41%)	6 (27%)	3

MC = multiple-choice

Table 26 indicates that the Make Connections question included in the study performed as expected, with a low differential of 3. Six of the nine students who answered the question correctly demonstrated one or more of the expected behaviors.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of complex thinking elicited by Make Connections questions on the digital SAT Suite assessments.

Question 12

Question 12 is a hard (PSB 6) multiple-choice question without a context. The question expects students to determine the x-intercepts of a function involving a transformation and then calculate the sum of the x-coordinates of the x-intercepts.

Function f is defined by $f(x)=(x+6)(x+5)(x+1)$. Function g is defined by $g(x)=f(x-1)$. The graph of $y=g(x)$ in the xy-plane has x-intercepts at $(a, 0)$, $(b, 0)$, and $(c, 0)$, where a, b, and c are distinct constants. What is the value of $a+b+c$?

A) -15

B) -9

C) 11

D) 15

Choice B is the correct answer. To answer this question correctly, students could use an understanding that function g can be obtained by substituting $x-1$ for x in function f: $f(x-1)=(x-1+6)(x-1+5)(x-1+1)$, which simplifies to $f(x-1)=(x+5)(x+4)(x)$. Next, setting each factor equal to 0 and solving for x leads to finding the x-intercepts of $y=g(x)$, which correspond to the values of a, b, and c. The values of a, b, and c, then, are -5, -4, and 0. Students would then find the sum of those values, which is -9. Alternatively, students could answer this question by graphing function f and observing that function g is obtained by shifting function f one unit to the right. They could then observe that the x-intercepts of function g are $(-5, 0)$, $(-4, 0)$, and $(0, 0)$.

Student M8 begins their successful approach to question 12 by correctly using function notation when substituting $x-1$ for x in function f to obtain function g.

> Okay. Um, we are finding—okay. $g(x)$ is equal to $f(x-1)$. So I guess you could just end up plugging this in, so I would plug in $x-1$ for each x-value. So $(x-1+6)(x-1+5)(x-1+1)$. So that would give me $(x+5)(x+4)x$.

Student M8 next demonstrates command of how to find the x-intercepts of function g while recognizing that the x-coordinates of the x-intercepts are the values of a, b, and c. Finally, student M8 finds the sum of those values.

> And for the x-intercept, so it's—see where all that equals to 0 and that would give me -5, -4, and 0. So if I wanted to add all those values, I would have -9.

Student M20 follows a successful solution path similar to that used by student M8.

> So in order to get $g(x)$, I'm going to plug in $x-1$ every time I see an x. So $(x-1+6)(x-1+5)(x-1+1)$. So getting $(x+5)(x+4)x$. So that means the x-intercepts are $(-5, 0)$, $(-4, 0)$, and $(0, 0)$. So sum of these x-values is -9.

Determine Conditions

To answer the Determine Conditions question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Evaluate values of a constant given in the answer choices to determine the number of solutions for a quadratic equation.

2. Use the discriminant to determine the number of solutions for a quadratic equation.
3. Use a graph to determine the solutions to a quadratic equation.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves determining the conditions for which a given quadratic with an unknown constant has more than one real solution. By evaluating the answer choices, students could substitute values for the constant into the given quadratic equation and then find the solution(s) for the equation by solving it or by graphing it and then making the connection that the solution(s) are the *x*-intercept(s) of the graph. Students could also calculate the value of the discriminant and then draw a conclusion about the range of values for the constant that would lead to the desired number of solutions for the quadratic equation.

Table 27 summarizes how students performed on the Determine Conditions question included in the study.

Table 27. Student Performance on Math: Advanced Math—Determine Conditions Question.

			Demonstrated Expected Behaviors						
Question	Subject Area	PSB	1	2	3	One or More	Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
15 MC *n* = 20	None	7	8 (40%)	2 (10%)	3 (15%)	8 (40%)	9 (45%)	6 (30%)	3

MC = multiple-choice

Table 27 indicates that the Determine Conditions question included in the study performed as expected, with a differential of 3. Six of the nine students who answered the question correctly demonstrated one or more of the expected behaviors.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of complex thinking elicited by Determine Conditions questions on the digital SAT Suite assessments.

Question 15

Question 15 is a hard (PSB 7) multiple-choice question without a context. The question asks the students to determine the value of an unknown constant in a quadratic equation given that the equation will have more than one real solution.

$$64x^2 + bx + 25 = 0$$

In the given equation, b is a constant. For which of the following values of b will the equation have more than one real solution?

A) -91

B) -80

C) 5

D) 40

Choice A is the correct answer. To answer this question correctly, students could substitute the values of b from the answer choices into the given quadratic equation. The wording of the question makes this a likely solution path for students. When solving numerically like this, after substituting in the value of b from a given answer choice, students still need to determine whether that choice yields two real solutions, which can be done by either graphing or solving the equation. When solving graphically, the option that yields two x-intercepts will be the correct answer. Alternatively, students could calculate the discriminant, $b^2 - 4(64)(25)$, and apply an understanding that for the equation to have more than one real solution, $b^2 - 4(64)(25) > 0$, or $b^2 > 6{,}400$. Students could evaluate each number in the answer choices to determine which satisfies this condition, or recognize that for this to be true, $|b| > 80$.

Student M9 begins their successful approach to question 15 by acknowledging that the answer can be determined by evaluating the given answer choices. Note that it's unclear what's meant by "lower minimal solutions," as it isn't referenced again in their transcript for this solution.

> Okay. So we have $64x^2 + xb + 25 = 0$. So we don't know what x is. But that's fine. What I can probably do is have x be 1 for now. So I can just have—and I'll just probably apply the numbers that were given. And then see which one has the lower minimal solutions.

As student M9 substitutes the options for b from the answer choices into the given equation, they aren't sure what to do next. Student M9 starts with answer choice C (5), squares 64, and then realizes that squaring 64 isn't correct and doesn't mention it again. Next, they try answer choice B (–80) but quickly move on to choice D (40).

> So $64^2 + 5$, for example. So that value doesn't make it as an answer because we're—also it's probably one of these negative numbers. So 64^2 is 4,096. And then negative 80. One new solution. $64x^2 + 40x + 25$.

Student M9 attempts to factor the quadratic that results from substituting 40 (answer choice D) in for the constant b in hopes of finding the solution(s) but doesn't seem to finish that approach. Next, student M9 identifies the values of A, B, and C from the standard form of a quadratic equation ($Ax^2 + Bx + C$).

> So when I factor this, what I'm going to get is. So we know that A is going to be 64. B is going to be 40, and then C is going to be 25. But there's really no numbers that we can factor this. Maybe if I change it to 5. Or maybe that's going to give it to me. 64, 5, and 25. But that's not going to give it to me. So I can try the negative numbers. And A, it's going to be 64. B is going to be -80. Maybe that one's going to give it to me. 1,264. -80. That's not going to give me either.

At this point, student M9 acknowledges that -80 gives one solution, 0.625, and proceeds to substitute -91 for b in the quadratic equation. Unlike in their other attempts to solve by factoring, student M9 now uses technology (either their own calculator or the Desmos Graphing Calculator) to solve the equation by graphing to determine the number of solutions. Student M9 doesn't clearly articulate what they're doing with the technology but does say what the two values of x are when

the value of the constant b is -91. Note that the actual solutions are approximately 1.0498 and 0.3721. Student M9 may have substituted 91 for b and thus gotten the two solutions for $64x^2 + 91x + 25 = 0$, or they may have identified the solutions as negative when they are actually positive.

> So we're left with 91. I think I'm going to err. Let's see 4. Okay. So I got $-.625$. And then 64. So assuming that it's probably 80. But I'm just going to double-check with the 91 solution. 64. -91 and 25. But that also gives us—but they're looking for different solutions. So I'm going to have to go with -91 because it gave me -1.04 and $-.37$.

Student M8 begins their successful approach to question 15 by deciding that a formula is needed to determine the correct answer. Student M8 starts their approach with answer choice D. Their reference to a "common denominator" isn't pertinent to solving the question, so presumably student M8 means "common factor."

> Okay, so in the given equation, b is a constant. Which of the values would be—? Because you have more than one real solution. So I guess I would just look for something that's—factorable. Uh, uh, $64x^2$ where, uh, uh, minus—$+40x+25$. So I guess I would just look for a common denominator. Thinking of the formula that determines whether or not the equation has one known, uh, real solution. So I'm trying to recall it. I think it's $-\frac{b}{c}$. But again, I'm not 100 percent sure. Um, yeah. Um, oh. So I'm not too sure. 'Cause I know that you have to use a formula for this problem.

At this point, being unsure of the formula, student M8 changes their approach and substitutes the values in the answer choices in for b in the given quadratic equation. Next, student M8 graphs each equation and determines that of the options, the graph has two x-intercepts only when $b = -91$.

> I guess I would, in the worst case, just graph this out and then see where there are, um, maybe solutions, as I'm not too sure about the formula. Okay. Um, sorry, I'm just gonna graph that out. Um, so it seems that, um, 40 would not work out and—um, yeah, I'm just trying out and testing the different values of b to see if there would be any solutions. Okay. And, um, looking at the graph, it seems that -91 would be the only one to cross the x-axis two times, so I think it would be -91.

Student M7 begins their successful approach to question 15 by demonstrating an understanding of the discriminant and what conditions allow for a quadratic equation to have more than one real solution.

> Okay. So I know that in order to determine how many solutions there are, it's just $-b$ or it's not. Square root of $b^2 - 4ac\left[\sqrt{b^2 - 4ac}\right]$, that has to be positive and not 0.

Student M7 then evaluates each answer choice. After trying 5 (answer choice C) and 40 (choice D), student M7 makes an astute observation in evaluating the "biggest number" since squaring a negative number results in a positive number.

> So what would give me that? 5 would be $25 - 4ac$. That is definitely going to be negative, 100 percent going to be negative. It's 40. That's $160 - 4 \times 64 \times 25$, which is 6,400, most likely negative as well. I'm going to choose the biggest number, probably -91. -90 or 91×91 because it's—although it's negative, it's squared. So $91 \times 91 - 4 \times$—let me put on the right parentheses. $-4(64)(25)$. That gives me a positive number. So it's definitely 91. Or I think it is 91, at least, because taking the square root of that is plus or minus, and it will have more than one real solution.

Even though student M7 feels as though they've identified the correct answer, they continue their process and evaluate the remaining option. Student M7 again shows command of what the discriminant provides in stating that when the value of the discriminant is 0, there's going to be only one solution.

> Let me just check -80. So that's 80×80. $6,400 - 4 \times 64 \times 25$. And that gets me a 0. So I know that's only going to be one solution. Therefore, this is going to be -91.

Nonlinear Equations: Solve

To answer the Nonlinear Equations: Solve question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Find the value of a constant in a nonlinear equation.
2. Solve a nonlinear equation.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves solving a nonlinear equation with a radical and an unknown constant while being given the greatest solution.

Table 28 summarizes how students performed on the Nonlinear Equations: Solve question included in the study.

Table 28. Student Performance on Math: Advanced Math—Nonlinear Equations: Solve Question.

			Demonstrated Expected Behaviors					
Question	Subject Area	PSB	1	2	One or Both	Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
17 SPR *n* = 20	None	6	1 (5%)	0 (0%)	1 (5%)	3 (15%)	1 (5%)	2

SPR = student-produced response

Table 28 indicates that the Nonlinear Equations: Solve question included in the study, which has a differential of 2, was exceptionally challenging for students. One of the three students who answered the question correctly demonstrated one of the expected behaviors.

A vignette from the student who answered correctly and demonstrated one of the expected behaviors illustrate the kinds of complex thinking elicited by Nonlinear Equations: Solve questions on the digital SAT Suite assessments.

Question 17

Question 17 is a hard (PSB 6) student-produced response question without a context. The question asks students to find the value of an unknown constant in a nonlinear equation given the equation's greatest solution.

> $\sqrt{5(x-k)} = x-k$
>
> In the given equation, k is a positive constant. The greatest solution to the equation is 12. What is the value of k?

The correct answer for this question is 7. To answer this question correctly, students could recognize the structure of this equation and let $y = x-k$. Substituting y for $x-k$ in the given equation yields $\sqrt{5y} = y$. Squaring both sides of this equation yields the quadratic equation $5y = y^2$. Subtracting $5y$ from both sides of this equation gives $0 = y^2 - 5y$. Since both terms on the right-hand side of this equation have a common factor of y, the equation can be rewritten as $0 = y(y-5)$. Therefore, $y = 0$ or $y - 5 = 0$. Since $y = x-k$, substituting $x-k$ for y in these two equations yields $x-k = 0$ or $x-k-5 = 0$. Therefore, $x = k$ and $x = k+5$ are the solutions to the given equation. It's given that k is a positive constant, so $k+5$ is greater than k. It's also given that the greatest solution to the equation is 12; therefore, $5+k = 12$, or $k = 7$.

Student M16 begins their successful approach to question 17 by substituting the given greatest solution for x into the given equation. While this is likely not the most common solution path, the student does demonstrate the basic understanding that to find the key, the equation can't have two unknown values.

> Oh, oh, I'm gonna start by plugging in the greatest solution. If there—if 12 is a solution [then] we can get k. Hmm. So $12 - k = 12 - k$.

Student M16 then attempts to solve the equation by squaring both sides. There are multiple ways to solve this resulting quadratic equation; student M16 elects to expand the squared binomial and combine like terms.

> So based off that, I can just square to the right side. $5(12-k) = (12-k)^2$. So foiling that out [i.e., using the FOIL method of multiplying two binomials], um, I get $144 - 12k - 12k + k^2$. So $k^2 - 24k + 144$. So $5(12-k) = k^2 - 24k + 144$. Um, $5 \times 12 - k$ is, um, $60 - 5k$. $60 - 5k = k^2 - 24k + 144$. Um, I think I can just move that to the other side and then solve, um, that for a quadratic. Um, $-24 + 5$, um, then $144 - 60$. So $0 = k^2 - 84k$, oh, uh, $+84$. Okay. So $k^2 - 19k = 84$.

At this point, student M16 determines that the solutions are 7 and 12 but doesn't describe how they got from the last equation articulated to that solution. Student M16 then appears to check these two values of k in the given equation to arrive at a key of 7. Student M16 doesn't verbalize how they came to this conclusion, but since it was given that the greatest solution to the equation is 12, it's possible that they chose 7 as it is less than 12.

> Um, based off of those two solutions, what is the value of k? I guess I can test—okay, so the only solution that seems to work is 7, I think, 'cause, um, $12 - 12$ would result in 5×0. So the $\sqrt{0} = 12 - 12$, 0. Oh, that would work. Well, 7 [*corrects self*] $12 - 7$, 5. So $5 \times 5 =$ [*corrects self*] $\sqrt{5 \times 5} = 5$ Yeah. That was two answers. So I think I'm just gonna go with 7.

Rewrite

To answer the Rewrite question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Determine conditions for which a value must be an integer.
2. Recognize that when multiplying two polynomials together, the constant terms multiplied together are a constant.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves rewriting a given expression with unknown integer constants and understanding the conditions for which a value must be an integer as it relates to a different given expression. Students must be able to recognize that when two integer constants are multiplied together, the product is a constant. This is a conceptually sophisticated question in which students can quickly become overwhelmed with the given information and in trying to determine what is and isn't needed to obtain the correct answer.

Table 29 summarizes how students performed on the Rewrite question included in the study.

Table 29. Student Performance on Math: Advanced Math—Rewrite Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors			Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
			1	2	One or Both			
19 MC *n* = 19	None	7	1 (5%)	2 (11%)	2 (11%)	6 (32%)	0 (0%)	6

MC = multiple-choice

Table 29 indicates that the Rewrite question included in the study didn't perform as expected, with a differential of 6. None of the students who answered the question correctly demonstrated any of the expected behaviors.

Vignettes from students attempting to answer the question illustrate some aspects of cognitively complex thinking elicited by Rewrite questions on the digital SAT Suite assessments, even though none of the expected behaviors were demonstrated.

Question 19

Question 19 is a hard (PSB 7) multiple-choice question without a context. The question gives two equivalent expressions written in different forms, both with unknown constants, some of which are specified as integer constants. The answer choices are expressions with unknown constants, and using the given information about those constants, students are to decide which choice must be an integer in this situation.

The expression $4x^2 + bx - 45$, where b is a constant, can be rewritten as $(hx + k)(x + j)$, where h, k, and j are integer constants. Which of the following must be an integer?

A) $\frac{b}{h}$

B) $\frac{b}{k}$

C) $\frac{45}{h}$

D) $\frac{45}{k}$

Choice D is the correct answer. To answer this question correctly, students could multiply the binomial factors in the second expression, which results in $hx^2 + (hj + k)x + jk$. It's given that this expression is equivalent to $4x^2 + bx - 45$; therefore, $jk = -45$, or $j = -\frac{45}{k}$ (since $k \neq 0$). It's given that k and j are integers, so of the given answer choices, $\frac{45}{k}$ must be an integer.

Student M29 begins their approach to question 19 by gathering their thoughts and attempting to process the given information.

> Okay. So this is just vocab, "integer" is just—yeah, I think it is just a value. It cannot be a fraction. So let's see. I'm just rereading to see what all of these constants are, can be rewritten.

In their attempt to find h and j, student M29 recognizes that when the binomial factors are multiplied, the coefficient of this x^2 term will equal the coefficient of the x^2 term in the expression $4x^2 + bx - 45$, in this case 4.

> So I am going to try to rewrite it myself to figure out what h and j is equal to. So $4x^2 + bx - 45$, and $hx + k$. So h should be 4, so it's $4x$; I know that. And then in the other bracket, it has to be x, so then when it multiplies, it's $4x^2$.

As student M29 continues, they make the connection that -45 is the product of two constants, although without specifically identifying the constants as j and k. Student M29 then focuses heavily on determining the value of b even though this information isn't necessary to select the correct answer.

> And then b is just what you add up, and it has to—when you multiplied it to numbers, it has to be -45. A integer, so b just disappears. That makes it harder. So 45 could just be -9×5, but it could also be -5×9. I'm assuming it should be $+9 - 5$, as b is positive. But since we're not given b, I cannot be 100 percent sure. The following must be an integer. So b [is] 45.

Student M29 makes an error in thinking the value of b is the sum of j and k, forgetting that the value of h is 4, not 1. In doing so, they conclude that $b = 4$, $h = 4$, $k = 9$, and $j = -5$ and proceed to evaluate the answer choices using those values to determine which must be an integer. Both choice A and choice D result in integers, and the student incorrectly selects choice A.

> Okay. So I'm assuming, so b should be when you add 9 and -5, so that should be 4. b is equal to 4, and then h is equal to 4. k is equal to 9. j is equal to -5, although k and j are interchangeable. So, for the first option, $b \div h$ would be $4 \div 4$, which would be 1. That's an integer. I'm just going to continue checking. $b \div k$ would be $4 \div 9$ or -5, which is not an integer. $45 \div 4$, that's not an integer either. And $45 \div 9$—actually, 45 divided by—that is an integer. So it's either the first or last option. I'm just going to check. b, h. As I'm not sure about the value of k, I'm just going to go with the first option, which is b over h.

While this was a failed attempt, student M29 showed signs of the kinds of cognitively complex thinking meant to be elicited by the question.

As mentioned, none of the students who answered this question correctly demonstrated either of the behaviors expected; however, the six students who answered correctly demonstrated some consistent logic in doing so. Most notably, these students dismissed answer choices A and B, containing *b*, as the best answer due to *b* being defined as a constant, whereas the question is asking which choice must be an integer. Additionally, some students noted that *h* was likely 4, and since it was associated with x^2, choice C was likely incorrect. To gain further confidence that the key was choice D, some students noted that 45 would come at the end, in the last term of the expression, without a variable in that last term, which would be where *k* would be in the equivalent form.

PROBLEM-SOLVING AND DATA ANALYSIS

A wealth of evidence supports the conclusion that students need to leave secondary school prepared to work with data and armed with statistical thinking skills and data acumen. Additionally, students need to grasp concepts from the study of probability to understand the importance of randomness in statistics. Two foundational topics that flow through the math curriculum, typically introduced in grade 6 and returned to throughout high school, are developing an understanding of proportional reasoning and applying proportional relationships to solve single-step and multistep problems. Proportional reasoning is an important skill when solving percent-based problems, including those involving discounts, tips, sales tax, interest, unit rates, and percent increase and decrease, and thus it's assessed, at appropriately challenging levels, throughout the digital SAT Suite, including on the SAT.

The Problem-Solving and Data Analysis content domain assesses knowledge and skills in using ratios, rates, proportional relationships, unit analysis, percentages, probability and conditional probability, one- and two-variable data, scatterplots, models, inference from sample statistics, and evaluating statistical claims. Unlike topics covered in the Algebra and Advanced Math content domains, the topics addressed by the digital SAT Suite in Problem-Solving and Data Analysis aren't aligned to those covered in a specific secondary-level math course. State education systems include the topics covered in this domain in a variety of courses, starting with middle school/junior high school math and continuing through high school coursework. The test questions in the Problem-Solving and Data Analysis domain range in difficulty from relatively easy to relatively complex and challenging and test a wide range of reasoning skills.

Six Problem-Solving and Data Analysis questions were included in the cognitive interview study: one question each in the categories Fit a Model, Unit Rates, Probability, Sample Proportion, Derived Units, and Percentages.

Fit a Model

To answer the Fit a Model question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Use slope to identify an algebraic model that fits data represented in a scatterplot.
2. Use the *y*-intercept to identify an algebraic model that fits data represented in a scatterplot.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves analyzing a scatterplot and then finding elements of an equation for a line of best fit for the data. Then students use the results of their analysis to choose the equation from the answer choices that best represents an equation of a line of best fit for the data. Students must recognize the linear trend throughout the data and identify characteristics, such as slope and *y*-intercept, of a possible line of best fit. Often an equation in the form $y = a + bx$, where a represents the *y*-coordinate of the *y*-intercept and b represents the slope of the line, is the form of the equation used to fit linear data.

Table 30 summarizes how students performed on the Fit a Model question included in the study.

Table 30. Student Performance on Math: Problem-Solving and Data Analysis—Fit a Model Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors			Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
			1	2	One or Both			
2 MC *n* = 23	None	4	17 (74%)	13 (57%)	19 (83%)	17 (74%)	17 (74%)	0

MC = multiple-choice

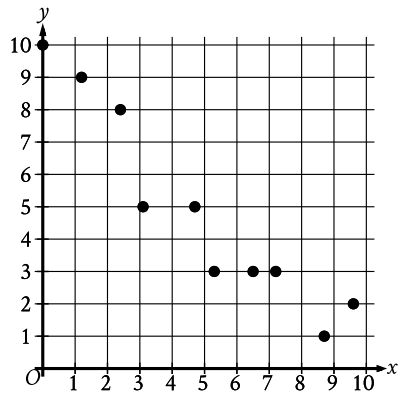
Table 30 indicates that the Fit a Model question included in the study performed as intended, with a differential of 0. Every student who demonstrated one or both expected behaviors answered the question correctly.

Vignettes from students answering correctly and demonstrating one or both expected behaviors illustrate the kinds of cognitively complex thinking elicited by Fit a Model questions on the digital SAT Suite assessments.

Question 2

Question 2 is a medium-difficulty (PSB 4) multiple-choice question without a context. The question gives a scatterplot with ten data points in a downward trend and doesn't show a line of best fit. The answer choices are distinguished by their varying values for the *y*-coordinate of the *y*-intercepts of and slopes for the line.

The scatterplot shows the relationship between two variables, x and y.



Which of the following equations is the most appropriate linear model for the data shown?

A) $y = 0.9 + 9.4x$

B) $y = 0.9 - 9.4x$

C) $y = 9.4 + 0.9x$

D) $y = 9.4 - 0.9x$

Choice D is the correct answer. To answer this question correctly, students could conceptualize a line of best fit and approximate two points on that line to compute the slope and *y*-intercept. Alternatively, given the answer choices, students could recognize that the *y*-intercept is close to (0, 10), thereby allowing them to eliminate choices A and B, and then recognize that the slope is negative, which results in choice D being the best answer.

Student M21 begins their successful approach to question 2 by acknowledging that because of the downward trend of the data, the slope of a possible line of best fit must be negative, thereby eliminating answer choices A and C. Additionally, student M21 recognizes that a possible line of best fit would intersect the *y*-axis at a value greater than $y = 1$, thereby eliminating choice B. Student M21 then concludes that the most appropriate linear model for the data shown is represented by choice D.

> So I can see that the graph is on a downward slope. So already I can eliminate $0.9 + 9.4x$. I can already eliminate these two. And judging by the starting point, I can eliminate the second option, $0.9 - 9.4x$. So I'm left with the last option, $9.4 - 0.9x$. And that is the answer I'm going to go with.

Student M3 begins their successful approach to question 2 by demonstrating understanding that a line of best fit follows the overall trend of the data in the scatterplot by identifying the data point where the scatterplot intersects the *y*-axis, meaning that based on the trend of the data, a *y*-intercept of the line of best fit would be near that data point. Then student M3 notices the direction of the trend of the data, going downward from left to right, yielding a negative slope.

After identifying these two pieces of necessary information about the features of the equation of a line of best fit—a *y*-intercept of about $(0, 9.52)$ and a negative slope—student M3 compares the answer options to these features, selecting the correct answer.

> Um, this question is asking which of the following equations is the most appropriate linear model for the data shown. And then we have a scatterplot here. So let's see. So the scatterplot starts at (0, 10). And then it goes until—let's see. That's about 9.52. Um, so then for the equation, it would be y equals—trying to find maybe, like, a slope first. Well, since it's going from—is that—left to right, it's gonna be a negative slope, um, since it is going down, so that's a clue. Um, so let's see. It would be—trying to see if I can find any more clues that might give me the answer. I can't find any more, so I'm going to say—I'm gonna take the information that I have right now and say that I would get $y = 9.4 - 0.9x$ because it looks like $0.9x$ is the closest to the slope as we'll get, so that's kind of what I thought.

Unit Rates

To answer the Unit Rates question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Use unit rates to solve a problem.
2. Compare values when using two unit rates.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves understanding a context, identifying unit rates from that context, using the two unit rates to make computations, and comparing the relative size of two phenomena based on the unit rates and a period of time.

Table 31 summarizes how students performed on the Unit Rates question included in the study.

Table 31. Student Performance on Math: Problem-Solving and Data Analysis—Unit Rates Question.

			Demonstrated Expected Behaviors					
Question	Subject Area	PSB	1	2	One or Both	Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
7 SPR *n* = 23	Science	5	19 (83%)	21 (91%)	23 (100%)	17 (74%)	17 (74%)	0

SPR = student-produced response

Table 31 indicates that the Unit Rates question included in the study performed as intended, with a differential of 0. Every student that demonstrated one or both expected behaviors answered the question correctly.

Vignettes from students answering correctly and demonstrating one or both expected behaviors illustrate the kinds of cognitively complex thinking elicited by Unit Rates questions on the digital SAT Suite assessments.

Question 7

Question 7 is a medium-difficulty (PSB 5) student-produced response question set in a science context. The question gives a unit rate for each of two natural phenomena and asks students to find the difference in the amount of time it takes for each phenomenon to happen a specified number of times.

> One of a planet's moons orbits the planet every 252 days. A second moon orbits the planet every 287 days. How many more days does it take the second moon to orbit the planet 29 times than it takes the first moon to orbit the planet 29 times?

The correct answer to this question is 1,015 (which students would enter without a comma in the provided field). To answer this question correctly, students must recognize that they need to find the number of days it takes each moon to orbit the planet 29 times and then subtract the two values to find the positive difference. To do this, students should multiply each of the given rates (the number of days it takes each moon to orbit the planet) by the number of times the moons orbit the planet (29): (252)(29) and (287)(29). Once each of these values is found, students can find the difference in the number of days it took each moon to orbit the planet 29 times: 8,323 − 7,308, or 1,015.

Student M21 begins their successful approach to question 7 by confidently talking through their clear intended solution path.

> So right away, just off of reading the question, I can tell that I'm going to have to multiply each planet's orbit by 29 and then just subtract the difference.

Student M21 then performs the necessary calculations and arrives at the correct answer.

> So I'll just go ahead and do 252×29 to find out the amount of days it takes to orbit that moon. So, 1 second. So 252×29. 7,308 for the first moon, so I'll just go ahead and write that down. A second one orbits the planet every 287 days. And now for the second moon, I will do 287×29 to figure out how long it takes to orbit. So 287×29 is 8,323. So I'll just subtract the two. And I'm left with $7{,}308 - 8{,}323$. 1,015. . . . I'll go with 1,015.

It's worth noting that student M21 verbalizes the required subtraction in the wrong order but has the presence of mind to enter the correct, positive value. Student M21 also says "orbit that moon" even though the question says each moon is orbiting the planet.

Student M28 begins their successful approach to question 7 by acknowledging they need to find the difference between the orbit durations.

> So first I need to find a difference. So $287 - 2$—let me just make sure. Yeah, 252. 35 days is the difference.

Next, student M28 recognizes that they need to multiply this difference by 29, the number of times each moon orbits the planet.

> So just 35×29, how do I pull up—um, one thousand a hundr—1,015 days would be the initial difference.

Probability

To answer the Probability question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Calculate, express, or interpret the probability of an event.
2. Determine an unknown number using probability and a description of a situation.
3. Set up and solve a proportion.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves understanding a context, identifying the possible outcomes and then determining the probability of each outcome, and using the probability to find the expected value requested.

Table 32 summarizes how students performed on the Probability question included in the study.

Table 32. Student Performance on Math: Problem-Solving and Data Analysis—Probability Question.

			Demonstrated Expected Behaviors						
Question	Subject Area	PSB	1	2	3	One or More	Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
9 SPR *n* = 23	Real-world	5	21 (91%)	17 (74%)	1 (4%)	21 (91%)	14 (61%)	14 (61%)	0

SPR = student-produced response

Table 32 indicates that the Probability question included in the study performed as intended, with a differential of 0. Every student that demonstrated one or more expected behaviors answered the question correctly.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of cognitively complex thinking elicited by Probability questions on the digital SAT Suite assessments.

Question 9

Question 9 is a medium-difficulty (PSB 5) student-produced response question set in a real-world context. The question indicates the total number of attendees at a conference and states that each attendee is assigned to one of three groups. The probability of selecting an attendee at random assigned to two of these groups is given. Students are to determine how many attendees are assigned to the third group.

> At a conference, there are a total of 275 attendees. Each attendee is assigned to either group A, group B, or group C. If one of these attendees is selected at random, the probability of selecting an attendee who is assigned to group A is 0.44 and the probability of selecting an attendee who is assigned to group B is 0.24. How many attendees are assigned to group C?

The correct answer to this question is 88. To answer this question correctly, students could determine the number of attendees assigned to groups A and B by multiplying their given probabilities by the total number of attendees. The number of attendees assigned to group A is 0.44(275), or 121. The number of attendees assigned to group B is 0.24(275), or 66. Since the total number of attendees at the conference is 275, all of whom have been assigned to groups A, B, or C, the number of attendees assigned to group C can be found by subtracting 121 and 66 from 275, resulting in 88. Alternatively, students could first add the given probabilities, 0.44 and 0.24, which results in 0.68, the probability of selecting a random attendee who's been assigned to either group A or B. Since all 275 attendees belong to exactly one group, the total probability of belonging to any group is 1. Therefore, $1 - 0.68$, or 0.32, is the probability of selecting a random attendee who's been assigned to group C. To determine the number of attendees assigned to group C, students could then multiply 0.32 and 275 to arrive at the correct answer, 88.

Student M24 begins their successful approach to question 9 by first adding the given probabilities and then using the complement of the event, in the process demonstrating a clear command of the skill assessed by the question.

> Well, this is a fill-in-the-blank. What I would do is do $275 \times$ the—actually, what I would do is $0.44 + 0.24$. And then 1 minus that amount. So minus 0.68 to find the percent of attendees that are assigned to group C.

Next, student M24 multiplies the total number of attendees by the probability of selecting an attendee in group C.

> And then I would do 275×0.32. And so there are 88 attendees in group C.

Student M30 begins their successful approach to question 9 by separately calculating the number of attendees in groups A and B.

> So let's see. I think I'm gonna do $275 \div 3$. Well, no. $275 \times .44$ is gonna be 121. So for group A, there's 121 attendees. And then for group B, I'm gonna do $275 \times .24$, which gives me 66.

Student M30 then subtracts the number of attendees in groups A and B from the total number of attendees at the conference, which results in the number of attendees assigned to group C.

> So you do $121 + 66$ is 187. $275 - 187$ is 88.

Finally, student M30 checks their answer via the alternative solution path discussed above by multiplying the sum of the given probabilities by the total number of attendees. However, the student only verbalizes how they arrived at the number of attendees *not* in group C and doesn't verbally include a step for arriving at a value of 88.

> And then just to double-check, I'm gonna do $.44 + .24$, which is .68. $275 \times .68$ is 187, which would mean that the first choice was right. So I'm gonna leave it at, um, 88 attendees for group C.

Sample Proportion

To answer the Sample Proportion question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Compute an estimated population mean from a sample proportion.
2. Understand that a sample proportion within a margin of error is a range.
3. Select a plausible value of the population mean using the population estimate and the margin of error.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves first reading and comprehending the context, then computing the estimated population mean, and then demonstrating an understanding of how to use the margin of error to determine the plausible values for this population mean.

Table 33 summarizes how students performed on the Sample Proportion question included in the study.

Table 33. Student Performance on Math: Problem-Solving and Data Analysis—Sample Proportion Question.

			Demonstrated Expected Behaviors						
Question	**Subject Area**	**PSB**	**1**	**2**	**3**	**One or More**	**Answered Correctly**	**Demonstrated One or More Behaviors and Answered Correctly**	**Differential**
11 MC *n* = 22	Social studies	6	8 (36%)	3 (14%)	10 (45%)	12 (55%)	15 (68%)	10 (45%)	5

MC = multiple-choice

Table 33 indicates that the Sample Proportion question included in the study performed as intended, with a differential of 5. Every student who answered correctly and demonstrated one or more expected behaviors showed evidence of having engaged in cognitively complex thinking. Some students demonstrated one or more expected behaviors but ultimately didn't answer the question correctly. Five of the students answered the question correctly but didn't demonstrate an expected behavior.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of cognitively complex thinking elicited by Sample Proportion questions on the digital SAT Suite assessments.

Question 11

Question 11 is a hard (PSB 6) multiple-choice question set in a social studies context. The question describes a situation in which a sample of 1,000 people was chosen at random from a population of 50,000 and surveyed about support for a proposed piece of legislation. The question then gives an estimated proportion of the sampled population that supports the legislation and an associated margin of error. Students are then asked to identify which of a given set of numbers is a plausible value for the total number of people in the population who support the proposed legislation.

From a population of 50,000 people, 1,000 were chosen at random and surveyed about a proposed piece of legislation. Based on the survey, it is estimated that 35% of people in the population support the legislation, with an associated margin of error of 3%. Based on these results, which of the following is a plausible value for the total number of people in the population who support the proposed legislation?

A) 350

B) 650

C) 16,750

D) 31,750

Choice C is the correct answer. To answer this question correctly, students need to comprehend the context and apply the understanding that when a statistic for an estimated proportion is given with a margin of error, it means that there is a range of plausible values for the true population value. Students could determine a plausible value for the total number of people in the population who support the legislation by using the sample proportion: 35% of 50,000, or 17,500. Then students could use the margin of error percentage to find the range of plausible values: ±3% of 50,000, or ±1,500, resulting in a range of 16,000 to 19,000. Alternatively, students could use the margin of error to identify the range of the proportion of the population that supports the proposed legislation. Students could find the range by adding 3% to and subtracting 3% from 35%: 32% to 38%. Then students could apply the least proportion of the range and the greatest proportion of the range to the population total to find the range of plausible values for the number of people from the population who support the proposed legislation: 32% of 50,000 is 16,000, and 38% of 50,000 is 19,000.

Student M26 begins their successful approach to answering question 11 by sorting through the information given, in the process demonstrating understanding of the difference between the population and the sample. They misstate that 35% of the sample would be 3,500 but later correct the error.

> So that's a lot to digest. I'm going to break it down. First and foremost, I have a population of 50,000, right? So that's a given. 50,000 people. Out of those 50,000, 1,000 were chosen at random, right? So the target group is 1,000, which is a straight number, and that's easy to calculate. Based on the survey, it estimated that 35% of people in the population, so it would be 3,500 out of the 1,000 here. And then it would be 35% of 50,000, right? They support the legislation. And there is an associated margin of error of 3%. Okay. Based on the results, they want the plausible value for the total number of people in the population who support the proposed legislation. Okay, so they're speaking population, which would be 50,000. If it were the 1,000 that were chosen, it would be 350, but it's not. So I would just do 35% of 50,000.

Student M26 next talks through the concept of margin of error and how to use that in solving the problem. They find 35% of the population and then find the lower bound of the plausible range, or 32% of the population. Student M26 then

examines the answer options and compares them to the population estimate and the lower bound of the range of the population estimate, selecting the correct answer, which is closest to these two values.

> When it comes to that margin error, I know I have to use it in the problem. But let me see what 35% of 50,000 is. So I would do 50,000, and I multiply that—excuse me, 50,000—I multiply that by 0.35. That gives me 17,500, but that's not necessarily an answer choice. That's where that margin error comes in. I am going to subtract by 3 on 35, which gives me 32. That's 16,000. Now, I know off bat that I can eliminate answer choice A. I can definitely eliminate answer choice B and of course D. D would be above the net 35%. I'm going mostly towards the answer choice C, the reason being it's the closest to the two numbers that I calculated as far as 17,500. And then if I subtract the 3, it would give me 16. So with that—excuse me—with that margin error in mind of 3% and calculating the population of 50,000, answer choice C would be my final answer.

Student M30 begins their successful and efficient approach by first finding the estimated population mean and then reasoning that a small margin of error means the plausible value would be close to the estimated population value.

> So it's asking you for the total number of people in the population who support the proposed legislation. So the sample size is 50,000. Well, the, the total population. And then the sample is 1,000. Um, okay. So $50{,}000 \times .35$, since that's what it's saying is the, like, number of people who were estimated to support the legislation. So $50{,}000 \times .35$ is 17,500. And there can be an estimated margin of error of 3%. So 350 would be way off. 650 would be way off. 31,000 would be way off. So the closest answer choice is 16,750, so I'm gonna go with that.

Derived Units

To answer the Derived Units question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Use a derived unit of measurement to find a missing value.
2. Use the volume of a cube to find the side length of the cube.
3. Set up and solve a proportion.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that typically involves first reading and comprehending the context and then using a derived unit that arises from a quotient to solve a new problem.

Table 34 summarizes how students performed on the Derived Units question included in the study.

Table 34. Student Performance on Math: Problem-Solving and Data Analysis—Derived Units Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors				Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
			1	2	3	One or More			
14 MC $n = 21$	Science	7	3 (14%)	3 (14%)	3 (14%)	3 (14%)	7 (33%)	3 (14%)	4

MC = multiple-choice

Table 34 indicates that the Derived Units question included in the study performed as intended, with a differential of 4. All three students who demonstrated one or more of the expected behaviors answered the question correctly. An additional four students answered the question correctly but didn't demonstrate any of the expected behaviors.

Question 14

Question 14 is a hard (PSB 7) multiple-choice question set in a science context. This question is challenging because it requires multiple steps to solve, involves using derived units to solve a problem, and requires some geometry skills in order to find a side length from the volume of a cube. The question gives the density of a certain type of wood and the mass of a cube-shaped sample of this wood. Students are then asked to find the length of one edge of the sample.

> The density of a certain type of wood is 353 kilograms per cubic meter. A sample of this type of wood is in the shape of a cube and has a mass of 345 kilograms. To the nearest hundredth of a meter, what is the length of one edge of this sample?
>
> A) 0.98
>
> B) 0.99
>
> C) 1.01
>
> D) 1.02

Choice B is the correct answer. To answer this question correctly, students should use the density of the wood and the mass of the sample, in their corresponding units, to determine the unknown volume of the sample. Students may know the formula for density $\left(\text{density} = \frac{\text{mass}}{\text{volume}}\right)$, or they may use the derived unit for the density value to write a proportion. Students can represent the unknown volume of the sample with a variable (such as V) and then write a proportion to represent the density relationships from the given information: $\frac{353\text{ kg}}{1\text{ m}^3} = \frac{345\text{ kg}}{V\text{ m}^3}$. Therefore, the volume of the sample is $\frac{345}{353}\text{m}^3$. Students would then use the fact that the sample is cube shaped and the volume formula for a cube to find the length of one edge of that cube. The volume formula for a cube is $V = s^3$. It follows that $\frac{345}{353} = s^3$. Taking the cube root of both sides of this equation yields $s = \sqrt[3]{\frac{345}{353}}$, or $s \approx 0.99$ m.

Student M16 successfully solves this problem by first identifying the relationship between mass and volume with respect to density.

> Um, the density of a certain type of wood—d [density] equals mass over volume—a sample of this wood is in the shape of a cube. Oh, okay. What is the length of one edge of this sample? Okay. So based off of that—I mean, I'm not really sure, um, right now, but I'm—I guess I'll just plug it into the $d = \frac{\text{mass}}{\text{volume}}$ formula just to get information.

Student M16 next makes a connection between the given density for the wood and the density of the sample and sets up a proportion. The solution process isn't articulated clearly, but student M16 nonetheless performs the steps correctly.

> So $353 = \frac{345}{\text{volume}}$. Okay, I can swap that. $345 \div 353$. So the volume is 0.977. What is the length of—oh, so I think based off of that volume. Wait, that doesn't make any sense. Density equals [mass over volume]. Yeah, I did that correctly.

After finding the volume for the sample, student M16 falters a bit before finding their way to the next step and then uses the value for the volume to find the side length of the cube.

> Am I supposed to find the area of this [stuff or that?]? Oh, okay. So, so one of the answer choices is cubed, I think. Pretty sure it's this [0.99]. Wait, I—I'm just gonna do the cube root of—the cube root of 0.977. So 0.99—yeah, yeah, that's it.

Student M19 also successfully answered question 14. Student M19 remembers the formula for density and applies it to the given information but lacks some confidence in this solution path.

> Okay. Density. Okay. If I remember correctly, density is—density equals volume over—or wait. It's mass divided by volume. Like, that's the formula. And we're given that the density is 353, uh, cubic meters, and then over, um, because we don't know—wait. No, we know what m is. We don't know what volume is. Then we have $\frac{345}{m}$, so then we would just have to solve for m. Um, okay. Then you would just—let me see real quick. 353—I think it would [be] divide 345 by 353. Okay. Then—and then I just need to see if that actually is how I would do it. Okay. Okay. So you would not divide 345 by 353 because when I put that answer back into m—wait. I mean V—it did not give the same thing, so. Mass is int—um, how would I get rid of—okay. Wait. Okay. I'm gonna do—okay. You would divide 345 by 353. So then you would get—the volume would be, like, .9777337, which is what I got from doing that because I do not remember how to do basic, um, algebra, so I just—so when I did the plug it into the formula, uh, I just, um—I was just—I just did 353 divided by 345, and I got 1-point—ah, 0-point-something, and I just plugged it into the equation to see if it would equal 353 and it didn't, so I was like, "Okay. That's not what you do." So then I did $345 \div 353$, then I got, like, .97. I put that back into, like, the equation I got from before, and I did get 353 from that, so I was like, "Okay. That's how you do it."

Student M19 then uses the formula for the volume of a cube to find the cube's side length.

So, so .97 is the volume of the cube, but it's asking for the length of one edge. So the length of one edge is s^3, so then you would do .97—or .97—. . . . It's actually .977337 $= s^3$, so then you would just take the cubed root of .97. So I don't know if that's even a fun—a function on this calculator. It is a function actually. Well, that's, that's nice. Okay. Wait. Oops. I have to clear that off. Okay. Okay. So 3 and then 0.977337. Okay. I got .99, and that's the second option.

Percentages

To answer the Percentages question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Write an expression to compute an increase by a percentage greater than 100.
2. Write an expression to compute a decrease by a percentage less than 100.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves understanding the relationship between a given value of an object and the value of that object after an increase by a percentage greater than 100 and then a decrease by a percentage less than 100.

Table 35 summarizes how students performed on the Percentages question included in the study.

Table 35. Student Performance on Math: Problem-Solving and Data Analysis—Percentages Question.

			Demonstrated Expected Behaviors					
Question	Subject Area	PSB	1	2	One or Both	Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
18 MC *n* = 19	Real-world	7	0 (0%)	5 (26%)	5 (26%)	0 (0%)	0 (0%)	0

MC = multiple-choice

Table 35 indicates that the Percentages question included in this study didn't perform as intended. None of the students who exhibited either behavior were able to answer the question correctly. The five students who exhibited expected behavior 2 understood that a decrease in value by *r*% is computed using a factor of $\left(1 - \frac{r}{100}\right)$. None of the nineteen students answering the question used the correct factor of $\left(1 + \frac{r}{100}\right)$ to compute an increase in a value of *r*% where *r* is greater than 100 (expected behavior 1). Many students gave evidence of holding the misconception that the factor to use to compute an increase in a value of *r*% where *r* is greater than 100 is just $\left(\frac{r}{100}\right)$ rather than $\left(1 + \frac{r}{100}\right)$. Thirteen of the students answering this question gave evidence of holding a second common misconception: a percent increase followed by a percent decrease can be computed by subtracting the percentages first. These students all used the incorrect factor for the percent increase.

The included student vignettes for this question illustrate the misconceptions that students held while answering.

Question 18

Question 18 is a hard (PSB 7) multiple-choice question set in a real-world context. The question asks for the net percentage increase in the value of an item after a percent increase of greater than 100 and then a percent decrease of less than 100.

> The value of a collectible comic book increased by 167% from the end of 2011 to the end of 2012 and then decreased by 16% from the end of 2012 to the end of 2013. What was the net percentage increase in the value of the collectible comic book from the end of 2011 to the end of 2013?
>
> A) 124.28%
>
> B) 140.28%
>
> C) 151.00%
>
> D) 209.72%

Choice A is the correct answer. The question's difficulty stems from two factors: one, the question requires multiple steps to solve and, two, the question is prone to the application of two common misconceptions. To answer this question, students should first assign a variable to the value of the collectible comic book at the end of 2011, say x. Then they should write an expression that represents the comic book's value at the end of 2012: $\left(1+\frac{167}{100}\right)x$, or $2.67x$. Then students should write an expression that represents a decrease of 16% from the value at the end of 2012 to the end of 2013. This would be written as $\left(1-\frac{16}{100}\right)(2.67x)$, or $(0.84)(2.67x)$. This is equivalent to $2.2428x$. This means that the comic book's value increased by a factor of $(1+1.2428)x$, or $\left(1+\frac{124.28}{100}\right)x$. This means that the net percentage increase is 124.28%.

Student M16 demonstrates the misconception related to using the factor of $\left(\frac{r}{100}\right)$ rather than the correct factor of $\left(1+\frac{r}{100}\right)$ for the increase in the comic book's value. Student M16 does, however, correctly use the factor $\left(1-\frac{r}{100}\right)$ for the decrease in value. Instead of assigning a variable to the value of the collectible comic book at the end of 2011, student M16 starts by assigning a value of \$1 to the comic book. This can be a very effective strategy for simplifying the computation in this type of question, even though student M16 was ultimately not successful in solving the question.

> So, uh, if we just assume that our initial value is 1, then we multiply that by 1.67, and then it gets—it gets decreased by 16%. So, um, multiply that by 0.84. So what was the net percent increase? So based off of that, I got 1.4028, and since my initial value was 1, I can assume that to just be a percentage. So I move that over to the right two decimal places. That's 140.28%.

Student M19 gives evidence of holding both the misconception demonstrated by student M16 of not adding 1 to the rate as well as the misconception that a percent increase followed by a percent decrease can correctly be computed by

subtracting the second factor (factor of decrease) from the first factor (factor of increase).

> Okay. So from 2011 to 2012, increase—[the comic book's value] increased by 167%, and then from 2012 to 2013 it decreased by 16%. So what was the net percentage increase in the value of the collectible from 2011 to 2013? So I would just do 167 – 16. Is it 16? It is 16. And then I would just—okay. So I got 151, which is this percentage, so I'm just gonna say it's [this].

GEOMETRY AND TRIGONOMETRY

Geometry is all about modeling the world around us, and knowledge of geometry helps lay the foundation for further achievement in math. Skills, knowledge, and concepts learned in the study of geometry are included in questions in the Geometry and Trigonometry content domain (for the PSAT 8/9 only, the Geometry domain) but are also woven into questions in the Algebra and Advanced Math domains, where geometric objects are sometimes used as contexts for building functions or modeling real-world scenarios. Geometry content on the digital SAT Suite is covered in secondary-level courses from grade 6 through high school. Trigonometry skills and knowledge are tested only on the SAT, PSAT/NMSQT, and PSAT 10, as these are typically taught and learned only in more advanced high school courses.

Test questions in the Geometry and Trigonometry content domain involve applying skills and knowledge in finding areas, perimeters, volumes, and surface areas; using concepts and theorems related to lines, angles, and triangles (PSAT 8/9 includes triangle angle sum theorem only); solving problems using right triangles (PSAT 8/9 includes Pythagorean theorem only); solving problems using special right triangles and right triangle trigonometry (SAT, PSAT/NMSQT, and PSAT 10 only); calculating using sine, cosine, and tangent (SAT only); solving problems using radian measure and trigonometric ratios in the unit circle (SAT only); and using definitions, properties, and theorems relating to circles (SAT only). These test questions vary in difficulty from easy to very hard and allow students to demonstrate problem-solving skills and knowledge using a variety of solving strategies.

Three Geometry and Trigonometry questions were included in the cognitive interview study, one question each in the categories Equation of a Circle, Special Right Triangles, and Volume.

Equation of a Circle

To answer the Equation of a Circle question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Use the coordinates of the center of a circle to write the equation for the circle.
2. Solve for the radius of a circle.
3. Identify the correct equation of a circle.
4. Identify the correct equation for a circle by substituting a given point in the answer choices.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves first determining the relationship

between the coordinates of the center of a circle and an equation of the circle in standard form and then solving for the length of the radius by either using the distance formula with the center point and the given point on the circle or using the Pythagorean theorem. Once the coordinates for the center and the length of the radius are known, students can identify the correct equation of the circle. Alternatively, students could choose to substitute the given point on the circle into one or more equations in the answer choices to determine the correct equation for the circle.

Table 36 summarizes how students performed on the Equation of a Circle question included in the study.

Table 36. Student Performance on Math: Geometry and Trigonometry—Equation of a Circle Question.

			Demonstrated Expected Behaviors							
Question	Subject Area	PSB	1	2	3	4	One or More	Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
8 MC *n* = 23	None	5	9 (39%)	2 (9%)	8 (35%)	2 (9%)	12 (52%)	10 (43%)	9 (39%)	1

MC = multiple-choice

Table 36 indicates that the Equation of a Circle question included in the study performed as intended, with a differential of 1. Nine of the ten students who answered the question correctly demonstrated one or more expected behaviors. Not all students articulated solving for the value of the radius (expected behavior 2) through their explanations captured during the interview; rather, they used the numbers in the answer choices to assume the length of the radius of the circle. All students who answered correctly needed to identify the correct equation of the circle, but not all these students successfully articulated the reasons for their choice. Two of the students used an alternative strategy of substituting the given point into the equations in the answer choices to determine which equation represented the correct answer.

Vignettes from students answering correctly and demonstrating one or more expected behaviors illustrate the kinds of cognitively complex thinking elicited by Equation of a Circle questions on the digital SAT Suite assessments.

Question 8

Question 8 is a medium-difficulty (PSB 5) multiple-choice question without a context. The question asks students to identify the correct equation of a circle in the *xy*-plane when given the coordinates for the center of the circle and a point on the circle.

> A circle in the *xy*-plane has its center at $(-4, 5)$ and the point $(-8, 8)$ lies on the circle. Which equation represents this circle?
>
> A) $(x-4)^2+(y+5)^2=5$
>
> B) $(x+4)^2+(y-5)^2=5$
>
> C) $(x-4)^2+(y+5)^2=25$
>
> D) $(x+4)^2+(y-5)^2=25$

Choice D is the best answer. The equation of a circle in standard form, where (h, k) is the center of the circle and r is the length of the radius of the circle, is $(x-h)^2+(y-k)^2=r^2$. Students could immediately determine the left side of the equation by substituting the coordinates of the center of the circle into the equation: $(x+4)^2+(y-5)^2=r^2$. To find the value of r, students could substitute -8 for x and 8 for y in the previous equation, which yields $(-8+4)^2+(8-5)^2=r^2$, or $(-4)^2+(3)^2=r^2$. Simplified, this gives $16+9=r^2$, or $25=r^2$. It follows that the equation for the circle is $(x+4)^2+(y-5)^2=25$. Alternatively, after obtaining the equation $(x+4)^2+(y-5)^2=r^2$, students could solve for the radius, r, by finding the distance between the center of the circle, $(-4, 5)$, and the given point, $(-8, 8)$, using either the distance formula or the Pythagorean theorem. If a student uses the Pythagorean theorem, they may create a right triangle using the two given points and a third point, $(-8, 5)$, that forms a right triangle. It follows that the horizontal distance between $(-4, 5)$ and $(-8, 5)$ is $|-8-(-4)|$, or 4, and that the vertical distance between $(-8, 5)$ and $(-8, 8)$ is $|8-5|$, or 3. Thus, the two legs of the right triangle are 3 and 4. A student may either recognize that this is a 3-4-5 triangle and that, thus, the hypotenuse (which is the radius of the circle) is 5, or a student may use the Pythagorean theorem to get a length of 5.

Student M20 begins their successful approach to question 8 by identifying how the coordinates of the center of the circle are situated in the equation.

> Well, first, the left side of the equation has to be $(x+4)^2+(y-5)^2$ because that point's the center of the circle.

Student M20 then talks through finding the length of the radius of the circle. They sketch a graph and then note that they'll use the distance formula. They recognize that the vertical distance and the horizontal distance between the two given points form the legs of a right triangle, identify a common Pythagorean triple in the form of the lengths of the two legs of the right triangle, and thereby determine the length of the hypotenuse, which is also the radius of the circle.

> And then I would . . . just draw it out so I can visualize the circle a little bit. Putting the center here, $(-4, 5)$ and $(-8, 8)$, [*counts out gridlines*] 2, 3, 4, 5, 6, 7, 8. And then I'd use distance formula to get the distance between these two. So the bottom leg, its distance is 4 because $-4-(-8)=4$. And then the vertical leg would be 3 because 8 is 3 away from 5. And then since this is a 3-4-5 triangle, the distance between the radius and the outside point on the circle is 5, which means that it's this [choice D] because r is squared . . . on the right side of the equation and 5^2 is 25.

Student M26 takes a circuitous route to the correct answer. Student M26 first takes stock of the given information and thinks about a possible formula for finding the equation.

> Okay, so I'm given two points. I'm going to write these down for reference. Writing down −4 and then 5. And I'm actually going to put a little bit of information as far as what these points represent. So the circle in the xy-plane has its center at $(-4, 5)$. So that's the center, right? Then

I have a point which lies on the circle, and the point is $(-8, 8)$. So let me write down that point, $(-8, 8)$, that is a point on the circle, which lies on it. So now I have to find an equation which represents the actual circle. Let's see. Now, I know there's a formula [*student consults reference sheet*]. So there's a formula with these two points that I'm given in the center and then the actual point which lies on there to where I can find the equation of what represents the circle in a way. Let's see if that formula may be of use. They're asking me for an equation. It wouldn't be this. This would be a circumference of pi. Wouldn't be this either. Nothing triangle related.

After realizing no appropriate formula is provided in the reference sheet, student M26 looks at the answer choices for a clue regarding how to properly answer the question. They note that the coordinates for the center of the circle are found in the answer choices.

Okay. Since it's a xy-plane and I'm seeing here with these answer choices, they're taking the points and they're actually adding them together, right? So $(x-4)^2+(y+5)^2$, and that equals 5. What I don't see, actually, is the other point, $(-8, 8)$. All the answer choices, they stem from -4 and 5. And then that would equal a certain value.

Student M26 then thinks about how a calculator might be helpful and eventually decides to use the given point to evaluate which answer choice contains the correct equation for this circle. Note that in their first try of substituting the point $(-8, 8)$ in answer choice D, student M26 makes an error by saying that 3^2 would be 6 but later recognizes and corrects the error.

Okay. Now, with this one, I'm not too entirely sure what course of action I would take, but what I'm going to do is I'm going to take a look at what I have here with the center. And then I'm going to see which of these answer choices makes the most sense as far as it being truthful. With D, for example: $(x+4)^2+(y-5)^2=25$. Obviously if I put that in the calculator, I wouldn't get the actual answer that it equals 25. I have to know what x is and, therefore, what y is.

Now, I'm also thinking that perhaps the x and the y would be taken from the point like $(-8, +8)$, but then if I did—yeah, if I did 8, right, for the y in this case, if I did $8-5$, that would be 3. 3^2 would be 6. And if I did $-8+4$, that would be $(-4)^2$; that's 16. 16 plus 6, that's now, what, 22. I'm going to try that. I'm going to try that and go through each answer choice, work backwards a little bit, and see what I'm able to find. So I've established that D wouldn't work because -8, like I said, x plus—if -8 was x, then that would be -4. $(-4)^2$, 4 times 4, that would be 16. And then 16 plus—oh, it was 9, actually. Yeah, $16+9$ is 25. Yeah. Okay, so it might be D. That's an answer choice. Let me try the same thing here. So if I had $-8-4$, right, that would give me -12. And then $(-12)^2$, I believe that's, like, 144 plus—yeah, that would be a really big answer choice, so I wouldn't go with that. So far, I'm liking D, and I'll write it out here in the calculator. It was $-8+4$, which gives me -4. $(-4)^2$ would be 16. I had the other thing—let me write that down. I had the other variable,

which was—it was $8 - 5$. That gave me 3. 3^2, 3×3 would be 9. So I miscalculated earlier. I put $16 + 9$. That gives me 25. I'm going to go with answer choice D.

Special Right Triangles

To answer the Special Right Triangles question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Write and use an equation for perimeter.
2. Find a side length in a special right triangle.

Students exhibiting one or both of these behaviors provide evidence of engaging in cognitively complex thinking that involves understanding the relationship between the side lengths of a special right triangle that's isosceles. Students need to recognize the triangle described is a right triangle and that the expression given for the perimeter of that triangle is found by adding three side lengths: two that are the same (the legs of the right triangle) and one that is longer (the hypotenuse of the right triangle). Students also need to recognize that the triangle described is a special right triangle and that the lengths of the sides are therefore in a ratio of $1{:}1{:}\sqrt{2}$. Students then need to write an expression for the perimeter of the triangle using a variable for the leg length. Setting the expression equal to the given perimeter allows the equation to be solved for the variable representing the leg length.

Table 37 summarizes how students performed on the Special Right Triangles question included in the study.

Table 37. Student Performance on Math: Geometry and Trigonometry—Special Right Triangles Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors			Answered Correctly	Demonstrated One or Both Behaviors and Answered Correctly	Differential
			1	2	One or Both			
10 MC *n* = 23	None	7	0 (0%)	2 (9%)	2 (9%)	13 (57%)	2 (9%)	11

MC = multiple-choice

Table 37 indicates that the Special Right Triangles question included in the study didn't perform as intended. Many of the students who answered the question correctly didn't demonstrate either of the expected behaviors. Both of the students who got this question correct and who also exhibited one of the behaviors ultimately found the correct answer by checking the given perimeter against the side lengths provided in the answer choices and then using their knowledge of the ratio of the lengths of the sides in a special right triangle. The majority of students reasoned their way to the correct answer by first realizing that choices C and D were too large and then deciding they needed a radical in their answer (choice B) because there was one in the given perimeter value.

Vignettes from students answering correctly and demonstrating one or both of the expected behaviors illustrate the kinds of cognitively complex thinking elicited by Special Right Triangles questions on the digital SAT Suite assessments. Additional vignettes illustrate the creative reasoning used by some students to successfully guess the correct answer.

Question 10

Question 10 is a hard (PSB 7) multiple-choice question without a context. The question gives the perimeter of an isosceles right triangle and asks for the length of one leg of the triangle.

> An isosceles right triangle has a perimeter of $94+94\sqrt{2}$ inches. What is the length, in inches, of one leg of this triangle?
>
> A) 47
>
> B) $47\sqrt{2}$
>
> C) 94
>
> D) $94\sqrt{2}$

Choice B is the correct answer. To answer this question correctly, students are expected to recognize that an isosceles triangle is a triangle with two sides of the same length. An isosceles right triangle is also known as a 45°-45°-90° right triangle, since the angle measures of all isosceles right triangles are 45°, 45°, and 90°. To successfully solve this question using the expected behaviors, students must recognize the pattern of the lengths of the sides of a 45°-45°-90° right triangle. If one leg has a length of x units, the other leg also has a length of x units, and the hypotenuse has a length of $x\sqrt{2}$ units. Students would then write an equation for the perimeter of the triangle and set it equal to the given perimeter: $x+x+x\sqrt{2}=94+94\sqrt{2}$. This equation can be simplified on the left-hand side, giving $2x+x\sqrt{2}=94+94\sqrt{2}$. Factoring out a common factor on the left-hand side gives $x(2+\sqrt{2})=94+94\sqrt{2}$. Next, dividing both sides of the equation by $2+\sqrt{2}$ gives $x=\frac{94+94\sqrt{2}}{2+\sqrt{2}}$. This value can be rewritten by multiplying the right-hand side of the equation by $\frac{2-\sqrt{2}}{2-\sqrt{2}}$, a technique used to remove the radical expression from the denominator, which gives $x=\frac{94+94\sqrt{2}}{2+\sqrt{2}}\times\frac{2-\sqrt{2}}{2-\sqrt{2}}$. This simplifies to $x=\frac{188+188\sqrt{2}-94\sqrt{2}-188}{4+2\sqrt{2}-2\sqrt{2}-2}$, or $x=\frac{94\sqrt{2}}{2}$, which is equivalent to $x=47\sqrt{2}$. Therefore, the length of each leg of the isosceles right triangle is $47\sqrt{2}$.

Student M7 begins their successful approach to question 10 by talking through what they know about isosceles triangles, albeit using imprecise language: student M7 says "all of them correct" when they likely mean "all of them congruent," for example.

> I can't remember my triangles. Isosceles, I'm not sure if that's with all of them correct. I'm not sure. I'm going to try different numbers. This references trig. Isosceles? It's not equilateral. It's not scalene. Wait, unless that's not—is the isosceles with two? I'm going to just assume it's with two—if not, whatever.

Student M7 then tries to make sense of the value of the perimeter. They seem to rely on some knowledge of the special right triangle but don't get it quite right, which they later recognize.

> Has a perimeter of $94+94\sqrt{2}$. That would make sense because $\sqrt{2}$ is the hypotenuse, and then x is just x times $\sqrt{2}$. So it's probably 94 on two legs. $94\sqrt{2}$ is the hypotenuse. What is the length of one of the legs? I believe it's 94 unless that's the perimeter. It would be 94. Oh, no. No, no, no, no.

Student M7 then continues effortfully to figure out how the perimeter is related to the side lengths. Their response exemplifies the confusion that many students in the sample had in solving this question: the perimeter value shows $\sqrt{2}$ but then doesn't match the pattern of $x+x+x\sqrt{2}$ in an obvious way because the $\sqrt{2}$ seems to be in the wrong place.

> It's going to be 47 because when finding the perimeter, you add all sides together. And the only way to have $94+94\sqrt{2}$ is by—the 94 could be a combination of the two sides. But then why is $94\sqrt{2}$ if x is actually 47? $47\sqrt{2}$, unless it's just my wrong memory. However, $94+94\sqrt{2}$. Two are the same. Hypotenuse is, I believe, just one side times $\sqrt{2}$. Therefore, it's that, what should be $94\sqrt{2}+94\sqrt{2}$. Oh, actually, I'll come back to this one. But I'm going to say it's 47. I don't know if it's as simple as it would just be the 94 because if I wanted to add the perimeter, it'd be 94 twice. I think I'll just do with 47 and come back to it and hope maybe another question has a triangle that gives me the correct dimensions.

After finishing the other questions, student M7 returns to question 10 and continues to try to solve the question. Student M7 finds a formula on the reference sheet that shows the side lengths of a 45°-45°-90° right triangle to be s, s, and $s\sqrt{2}$.

> That's so weird because the s had to be 94 in order for this to make any sense. But if it's two 94s—it says an isosceles right triangle has a perimeter of 94. What is the length in inches of one leg of this triangle? Okay. . . . What would happen if I do $47\sqrt{2}$ plus $47\sqrt{2}$? Yeah, that's the main thing I would expect. Isosceles right triangle. It'd [be] $2s+s\sqrt{2}$. . . . Oh, I'm done. It's $47\sqrt{2}$. It has to be because then I know it's $47\sqrt{2}+47\sqrt{2}$ or each multiplied by that 2, that by 2, which is $94\sqrt{2}$. And then if you were to just multiply that by $\sqrt{2}$, which is the hypotenuse, then that ends up getting 94. And that's why it's $94+94\sqrt{2}$. I thought $94+\sqrt{2}$ was the hypotenuse. That's actually the two side lines added together.

Student M2 begins their successful approach to question 10 by reasoning about the answer choices, as did several other students, rather than relying on the geometry involved.

> I'm looking at the square root and maybe we divide—$94\div 2$, and that will be 47. And so we look at the answer choices with 47 and so—I'm assuming it would be $47\sqrt{2}$ 'cause if the right triangle has a perimeter of $94+94\sqrt{2}$ inches, then I'm assuming it should look similar to this. All right.

Volume

To answer the Volume question as intended, students are expected to demonstrate at least one of the following behaviors:

1. Make connections between two figures in a composite figure.
2. Find the volume of a cube or sphere.
3. Find the volume of a composite figure.

Students exhibiting one or more of these behaviors provide evidence of engaging in cognitively complex thinking that involves, first, making a connection between the description of a composite figure and the two solids that are part of it; second, finding the volumes of the two solids; and, third, using those values to find the volume of a space in the composite figure.

Table 38 summarizes how students performed on the Volume question included in the study.

Table 38. Student Performance on Math: Geometry and Trigonometry—Volume Question.

Question	Subject Area	PSB	Demonstrated Expected Behaviors				Answered Correctly	Demonstrated One or More Behaviors and Answered Correctly	Differential
			1	2	3	One or More			
13 MC *n* = 21	None	6	9 (43%)	12 (57%)	9 (43%)	16 (76%)	12 (57%)	9 (43%)	3

MC = multiple-choice

Table 38 indicates that the Volume question included in the study performed as intended, with a differential of 3. Nine of the twelve students who answered the question correctly demonstrated one or more of the expected behaviors.

Question 13

Question 13 is a hard (PSB 6) multiple-choice question without a context. This question is challenging because it requires multiple steps to solve. The question describes a three-dimensional figure composed of a cube with a solid sphere inside of it that touches the center of each cube face. The question asks students to find the volume of the space in the cube *not* taken up by the sphere.

> A cube has an edge length of 68 inches. A solid sphere with a radius of 34 inches is inside the cube, such that the sphere touches the center of each face of the cube. To the nearest cubic inch, what is the volume of the space in the cube not taken up by the sphere?
>
> A) 149,796
>
> B) 164,500
>
> C) 190,955
>
> D) 310,800

Choice A is the correct answer. To answer this question correctly, students need to imagine the figure described, making the connection that the sphere is inside the cube in such a way that the diameter of the sphere is the same

length as the width of the cube. Students should find the volumes of the cube and the sphere and then subtract the volume of the sphere from the volume of the cube, rounding the result to the nearest cubic inch. The formula for the volume of a cube is $V = s^3$, where s is the length of the edge of the cube. For this cube, the volume is $(68 \text{ in})^3 = 314{,}432 \text{ in}^3$. The formula for the volume of a sphere is $V = \frac{4}{3}\pi r^3$, where r is the length of the radius of the sphere. For this sphere, the volume is $\frac{4}{3}\pi(34 \text{ in})^3 \approx 164{,}636 \text{ in}^3$. Therefore, the volume, to the nearest cubic inch, of the space in the cube *not* taken up by the sphere is $314{,}432 \text{ in}^3 - 164{,}636 \text{ in}^3 = 149{,}796 \text{ in}^3$.

Student M25 answers this question correctly by first finding the volume of the cube using the formula $V = lwh$, which for a cube is equivalent to $V = s^3$. Note that the student records the answer of $68 \times 68 \times 68$ incorrectly as 314,422. The actual value of the volume of the cube is 314,432.

> So a cube has an edge length of 68 inches, so it would be—I am just drawing a diagram here so it would be clear. It would be 68 because we're looking for the volume. So then the edge length of the cube is just—they're all symmetrical sides, or same-length sides. So then the volume would be 68 length × width × height, which would be 68^3, or $68 \times 68 \times 68$, which is—let me just calculate that—which would be this big number. Wait, 314,422.

Student M25 then finds the volume of the sphere. They initially give the volume formula as $\frac{4}{3}\pi r^2$, with the radius squared, but the radius should actually be cubed. From student M25's response, it's clear that they correctly cubed the radius to get 39,304 in an interim calculation.

> A solid sphere with a radius of 34 inches inside the cube—so [reference] sheet, so the volume, as you could see right here, is $\frac{4}{3}\pi r^2$. And the radius is 34, so it would be $\frac{4}{3}\pi(34)^3$. So it's the same thing we did, define the volume of the cube right there. It would be 34×34, which would be 39,304, is our radius. And so we would multiply that by π, as in the function up here [reference sheet], which would be $39{,}304 \times \pi$, which would be this number [39304]. Now actually multiply it by π. But that would give us this number, which is 123,477. And I'll round to the nearest decimal, so that should be 0.16. And then we need to still multiply that by $\frac{4}{3}$. All right, and that would give us 164,636.21. And it's to the nearest cubic inch, so we could just round this. And that is the volume of the sphere . . .

Student M25 next explains that subtraction is needed to find the space not taken up by the sphere.

> . . . but we're trying to find the space that is not taken up. So you would subtract this number and the number we got earlier. So it would be 314,422 −, which is the volume of the cube, 164,636, which would be 149,796, which is right here [option A].

Student M29 uses a similar strategy to answer this question. They start by identifying the needed formulas and then talk through a plan for finding the volumes.

So I know I have a cube, and I'm just going to check the formulas. Make sure, yeah. So I'm going to put volume $= lwh$. And we're given $l = 68$, solid sphere. Okay. And then the volume of a sphere is equal to $\frac{4}{3}\pi r^3$. And we're given that $r = 34$. And that's touching the face. Okay. So I think I'm assuming that I'm just going to find the volume that is taken up by the sphere and then just seeing if I can subtract that. I'm just using my calculator to solve it. But we're given everything we need to find the volume of the sphere. So I'm just plugging into my calculator formula.

Student M29 then finds the volume for the sphere using the formula correctly, demonstrating skill at thinking about the reasonableness of the answer.

So the volume should be—it seems a little high, but 164,636. Oh, looking at the answers, it does make sense that it is pretty high.

Student M29 next rereads the question and reasons that the correct answer can't be greater than the volume of the sphere. Then they proceed to find the volume of the cube. Finally, they subtract the volume of the sphere from the volume of the cube to find the proper answer.

And then to the nearest cube, what is the amount of the space in the cube not taken up by the sphere? So just looking at the answers, I know it cannot be higher than the volume of the sphere itself. Actually, never mind. Okay. Let me see if I can figure out the volume of the cube because then I could easily just subtract the two. Cube has an edge length. Oh. Okay. So it kind of slipped my mind, but now looking back on it, a cube, all of the sides have to be the same length. So if one of them is 68 inches, I know it's basically just $68 \times 68 \times 68$, and let me put that into a calculator. It would be 314,432. And then I'm just going to subtract these, the cube and the sphere. All right. And I get exactly 149,796, which is option A.

Section 5: Discussion

The main goal for the cognitive interview study presented in this report was to collect evidence in support of the hypothesis that select questions on the digital SAT Suite assessments are capable of eliciting cognitively complex thinking from student test takers. The primary source of such evidence resides in the transcribed responses of samples of Reading and Writing and Math test takers thinking aloud as they worked through and answered a series of digital-suite questions during one-on-one interview sessions. The analysis of the data presented in this report provides strong evidence of the truth of that hypothesis.

As discussed in Section 3: Methodology, the study's main quantitative metric is the *differential*, the arithmetic difference between (1) the number of students who answered a given question correctly and (2) the number of students who both answered the question correctly and demonstrated all required behaviors (Reading and Writing) or at least one expected behavior (Math). For the cognitive interview data to strongly support the study's hypothesis, differentials for examined test questions should have been 5 or lower, as this suggests that in order to answer the question correctly, the vast majority of correctly answering students needed (or elected) to enact the question type's intended construct. A higher differential, by contrast, could indicate that test takers found a shortcut through that construct, enabling them to answer correctly without engaging in the requisite cognitively complex behaviors. Such might happen if, for example, one or more of the studied questions contained clues or other flaws that suggested the right answer to students. A high differential signals the possibility that a given question may be flawed, but mitigating circumstances, such as the fact that distractors test common misconceptions, may still allow for the question to elicit aspects of cognitively complex thinking.

Reading and Writing

Table 39 summarizes the quantitative results for the Reading and Writing section's questions.

Table 39. Reading and Writing Section Differentials: Summary.

Question	Content Domain	Question Type	Subject Area	PSB (Difficulty)	Differential
1	Craft and Structure	Words in Context	Science	5	1
2			History/social studies	7	1
3		Text Structure and Purpose	Literature	3	0
4			Science	4	0
5		Cross-Text Connections	Humanities	6	3
6			Humanities	7	0
7	Information and Ideas	Central Ideas and Details	Humanities	4	0
8			Literature	4	1
9		Command of Evidence—Textual	Literature	6	4
10			Science	3	2
11		Command of Evidence—Quantitative	History/social studies	4	5
12			Humanities	6	1
13			Science	5	3
14			History/social studies	4	3
15		Inferences	Science	6	3
16			History/social studies	4	0
17	Expression of Ideas	Rhetorical Synthesis	Science	4	1
18			Humanities	5	2
19		Transitions	History/social studies	4	1
20			History/social studies	3	1

As table 39 indicates, all examined Reading and Writing questions had differentials from 0 to 5, indicating that the questions performed as intended per the methodology established for this study. Vignettes presented in this report for each of the questions further substantiate the claim that the questions were able to elicit cognitively complex thinking in line with the questions' intended constructs.

Math

Table 40 summarizes the quantitative results for the Math section's questions.

Table 40. Math Section Differentials: Summary.

Question	Content Domain	Question Type	Question Format	Subject Area	PSB (Difficulty)	Differential
1	Algebra	Linear Functions: Interpret	MC	Science	3	0
4		Linear Functions/Inequalities in One Variable: Create and Use	MC	Science	4	0
5			SPR	Real-world	4	0
16		Linear Equations in Two Variables: Make Connections	MC	None	7	1
20		Linear Systems: Determine Conditions	SPR	None	7	0

Question	Content Domain	Question Type	Question Format	Subject Area	PSB (Difficulty)	Differential
3	Advanced Math	Nonlinear Functions	MC	Real-world	4	1
6			MC	Science	6	1
12		Make Connections	MC	None	6	3
15		Determine Conditions	MC	None	7	3
17		Nonlinear Equations: Solve	SPR	None	6	2
19		Rewrite	MC	None	7	6
2	Problem-Solving and Data Analysis	Fit a Model	MC	None	4	0
7		Unit Rates	SPR	Science	5	0
9		Probability	SPR	Real-world	5	0
11		Sample Proportion	MC	Social studies	6	5
14		Derived Units	MC	Science	7	4
18		Percentages	MC	Real-world	7	0
8	Geometry and Trigonometry	Equation of a Circle	MC	None	5	1
10		Special Right Triangles	MC	None	7	11
13		Volume	MC	None	6	3

Table 40 indicates that eighteen of the twenty examined Math questions performed as intended per the study's methodology, with differentials from 0 to 5, although one additional question, discussed below, poses a special case. Vignettes presented in this report for each of these questions further substantiate the claim that the questions were able to elicit cognitively complex thinking in line with the questions' intended constructs.

Two questions—question 19, Advanced Math: Rewrite, and question 10, Geometry and Trigonometry: Special Right Triangles—had differentials in excess of 5, the threshold established by this report's methodology for acceptable ("low") differentials. Possible issues with these two questions are discussed below. A third question—question 18, Problem-Solving and Data Analysis: Percentages—also bears discussion, as no student in the sample was able to answer the question correctly.

NONCONFORMING MATH QUESTIONS

Question 19—Advanced Math: Rewrite

Question 19 is a hard (PSB 7) multiple-choice question without a context and yielded a differential of 6, just above the methodology's cutoff for questions performing as intended. Only six of the sampled nineteen students answered the question correctly, and only two students—both of whom answered the question incorrectly—exhibited one or both of the expected behaviors. The question's high difficulty, both in terms of its performance score band and the number of students in this study answering correctly, may have posed some challenges to participants exhibiting the expected behaviors, as the sample as a whole struggled with the question itself, but the main reason for the high differential likely lies in the correctly answering students' use of question-answering tactics—and a broader understanding of math—to eliminate distractors down to the correct answer.

Question 19 presents an expression and a rewriting of that expression. The term *b* is defined as a constant, and the terms *h*, *k*, and *j* are defined as integer constants. Students are asked to determine which of four fractions must represent an integer. As noted in the corresponding portion of Section 4: Results, some students offered evidence of being able to rule out two of the answer choices, $\frac{b}{h}$ and $\frac{b}{k}$, because each contained *b*, the term defined by the question as a constant. In addition, some students gave evidence of being able to rule out the third distractor, $\frac{45}{k}$, on the correct supposition that because *h* was likely 4, the choice couldn't result in an integer.

Although this question didn't behave as intended per the study's methodology, students answering correctly did nonetheless exhibit aspects of cognitively complex thinking, as evidenced in the accompanying vignettes. Furthermore, as observed above, students who reached the correct answer by distractor elimination tended to show a fundamentally clear understanding of how the rewriting process should work, ruling out distractors on the basis of an understanding of the nature of constants and a correct assessment of the likely value of one of the variables. Finally, it's worth noting that the incorrect answer choices included in the question represent surface-reasonable misinterpretations of what must be an integer based on the rewriting of the expression, an argument against this question having a flaw that students exploited.

Question 10—Geometry and Trigonometry: Special Right Triangles

Question 10 is a hard (PSB 7) multiple-choice question without a context. Students' responses to this question yielded a differential of 11, much higher than the threshold set by the methodology for questions performing as intended. Thirteen of twenty-three students answered this question correctly; of those, only two demonstrated one or both of the expected behaviors.

Much like question 19, question 10 proved amenable to students applying mathematical reasoning without showing evidence of either having written and used an equation for perimeter and/or finding a side length of a special right triangle. Many students who answered correctly were able to deduce that the side lengths for one leg of the triangle proposed in answer choices C and D—both distractors—were too large and that the only remaining viable choice was the one that included a radical (choice B) because a radical was included in the perimeter value provided in the question itself.

Three factors again militate against the conclusion that question 10 had significant flaws that precluded students from demonstrating cognitively complex thinking. First, the vignettes associated with the students who correctly answered the question and demonstrated one or both expected behaviors indicate that the question is capable of eliciting cognitively complex thinking in accordance with the question's intended construct. Second, those students who answered question 10 correctly but failed to demonstrate one or both of the expected behaviors nonetheless demonstrated aspects of cognitively complex thinking and mathematical understanding—critically, a strong sense of reasonable values in the given scenario, whereby two distractors were deemed implausible and the correct answer was deemed necessary because it included a radical. Third, as in question 19, question 10's answer choices represent a range of at least surface-plausible options, and only either enacting the question's intended construct or

applying deduction and mathematical reasoning enables a clear path to answering the question correctly.

Question 18—Problem-Solving and Data Analysis: Percentages

Technically, question 18—a hard (PSB 7) multiple-choice question set in a real-world context—had a differential of 0 in the study, but this was a result of the fact that no participating student answered the question correctly, although five students did exhibit one of the expected behaviors (write an expression to compute a decrease by a percentage less than 100). As analyzed in Section 4: Results, students attempting to answer question 18 struggled to find the correct answer, which first involved calculating the over-100-percent increase in value of a collectible comic book over a one-year period and then a double-digit percent decrease in value over the subsequent one-year period. This is a very challenging task that's prone to the application of the kinds of common misconceptions students gave evidence of holding in the study. Rather than being a flawed question, the fact that no participating student answered the question correctly is a product of the question's intended challenge level and the lack of mathematical understanding that participating students brought to the task. This question may still serve a valuable role in helping ascertain the ceiling of high-achieving students' ability.

SUBSECTION SUMMARY

To reiterate, seventeen of the twenty Math questions included in the study performed as expected, with differentials of 5 or lower and vignettes supportive of the claim that these questions elicited cognitively complex thinking. Two of the questions that didn't perform as expected (question 19 in Advanced Math and question 10 in Geometry and Trigonometry) were deemed to lack significant flaws, while the third (question 18 in Problem-Solving and Data Analysis) seems simply to have been too difficult for the participating students, at least under the pressure of simultaneously solving and thinking aloud, though the question is sound and still has assessment value in terms of helping measure the achievement of the highest-performing digital SAT Suite test takers.

Section 6: Implications

This study has important implications for both policymakers and researchers, including for College Board as it plans future cognitive interview studies on the digital SAT Suite.

Policymakers

The results and discussion sections of this report provide a strong basis for the conclusion that the digital SAT Suite's Reading and Writing and Math sections include numerous questions that elicit cognitively complex thinking from students in accordance with both the requirements of college and career readiness in general and the U.S. Department of Education's expectations for large-scale standardized assessments used as part of state educational accountability systems. All examined Reading and Writing questions and the vast majority (85 percent) of examined Math questions performed as expected and in line with intended question-level constructs designed to elicit cognitively complex behavior. The two examined Math questions whose student responses exceeded the differential threshold of 5 were shown to lack significant flaws while still being able to elicit aspects of cognitively complex thinking, while a third question simply proved too difficult for the sampled students to answer during the study.

Researchers

The methodology employed in this and a prior study (College Board and HumRRO 2020) is proposed as a reasonable, vetted, albeit time- and effort-intensive way to ascertain whether a given assessment's (or assessment system's) questions are capable of eliciting cognitively complex thinking from test takers. It builds on a robust research base supporting the validity of using verbalizations obtained from concurrent think-aloud studies to surface and analyze aspects of cognition that would otherwise be difficult if not impossible to recover. It also establishes and provides a rationale for a derived metric—the *differential*—that lends a useful quantitative complement to an otherwise strictly qualitative analysis of student responses. As Section 3: Methodology and Section 4: Results make clear, this metric, created and abstracting from rigorous qualitative coding and analysis, provides a useful way to identify successful question functioning in relation to

defined constructs and to point to questions that yield some evidence of not performing as intended and are therefore worth further examination.

Undertaking this study has also identified methodological refinements that seem likely to enhance the soundness and quality of future results.

First, the number of test questions included in the study—twenty for each subject area—resulted in some students rushing to complete the activity or failing to complete it in the allotted time. In addition, some students were able to finish—or finish additional questions—only because early in the interview process, College Board and its vendor, Vidlet, jointly determined that priority should be given to allowing students as much of the study time as possible for answering test questions, meaning that postexperience interview questions were often not asked (and haven't been analyzed for this study). Variance in *n*-counts for student responses by question are largely a product of some students running out of time; some additional variance resulted from the College Board researchers concluding that a small number of student responses to individual questions couldn't be coded due to ambiguous transcripts.

Because ninety minutes is likely near the upper limit of the time that student volunteers would be willing and able to engage productively in this activity, the College Board research team has concluded that the question sets used in subsequent studies should be pared down to approximately fifteen or sixteen questions to ensure that all students can give their best effort and pay full attention throughout the activity.

Second, the proportion of hard questions in the Math sample bears further examination. Twelve of the twenty Math questions examined in this study came from performance score bands 6 and 7, the two highest; by contrast, only six of the twenty Reading and Writing questions came from these bands. High-difficulty Math questions were disproportionately selected for the study because they were deemed most likely to elicit cognitively complex behavior; however, students often struggled to answer these questions correctly at all, and question 18—a PSB 7 student-produced response question in the Problem-Solving and Data Analysis content domain—elicited no correct responses from the sampled students. It's likely that just as cognitively simple or routine tasks are too "easy" to elicit cognitively complex thought—or, indeed, much conscious reflection at all—questions that are highly cognitively demanding risk flummoxing the majority of student respondents, at least under the think-aloud conditions of the study. When the set of Math questions is reduced in number for use in subsequent cognitive interview studies, College Board will closely consider whether some of the highest-difficulty questions (including question 18) should be eliminated from the study. This wouldn't be done to "improve" (bias) the results of subsequent studies, as the overall results of this study have shown that hard questions can elicit cognitively complex thought; rather, it would be to carefully limit the range of presented questions to more closely mimic what a typical test-taking population would likely be able to solve while under the added pressure of thinking aloud to an interviewer.

Third, due mainly to the fact that College Board hadn't yet completed the transition to the digital-suite tests when this study was being conceptualized and conducted, some of the Reading and Writing test questions examined here had not yet been

pretested. As a result, the performance score band (PSB) designations for some questions were assigned to their bands judgmentally by College Board staff. While expert, such judgment is less preferable than actual pretest statistics. Future cognitive interview studies involving the digital-suite tests will only examine questions that have been pretested and for which appropriate (and "passing") statistics exist.

Fourth, regarding sample selection, a volunteering student who had a previous SAT Math section score of 200—the lowest possible—was included in the study. It's most likely that this student didn't attempt the Math section during their previous testing. Their inclusion was an oversight, and future College Board studies will avoid such inclusions.

Fifth, the authors of this study made the assumption that participants were highly unlikely to have been previously exposed to the test questions sampled in the study. This assumption seemed warranted because despite these questions being available publicly as samples or as part of College Board–supported test preparation, the first domestic testing using the digital SAT Suite tests was roughly half a year away. This assumption may not hold, however, for subsequent studies in this vein. Given that, College Board will carefully consider whether the questions (or subsets of the questions) examined in this study can still safely be used for this purpose.

Finally, in conducting subsequent studies using this methodology, College Board will seek to refine its recruitment materials and interview protocols, copies of which are included in the appendix). One area of improvement would be making potential participants more aware that they may be asked to answer Math questions during the activity. For this study, students were randomly assigned to conditions and weren't briefed on which simulated test section they might be asked to think aloud through. This resulted in some students not being fully aware that they would either need to have their own calculator available or be able to use the Desmos Graphing Calculator built into the testing platform without assistance from the interviewers, even though this information was provided. This relative lack of preparedness on the part of some participants to answer Math questions may have depressed achievement overall, though, again, the vast majority of Math questions performed as intended and in line with their question-level constructs. Another area of improvement would be in clarifying for students that their participation in the study would in no way impact their past or future SAT Suite scores. Although no student indicated this was a concern, it's a matter that should've been more fully spelled out in recruitment.

Section 7: Conclusion

This report provided and discussed the results of College Board's 2023 examination of select digital SAT Suite test questions using a cognitive interview/think-aloud methodology involving student test-taking volunteers. Responses from student interviews were transcribed and analyzed both qualitatively and quantitatively. Results indicate that all Reading and Writing and the vast majority (85 percent) of Math questions included in the study performed as expected, with few students answering the questions correctly without exhibiting all required (Reading and Writing) or at least one expected (Math) behavior. Based on the methodology outlined in this report, these questions proved capable of eliciting cognitively complex behavior from students in accordance with both college and career readiness requirements and the U.S. Department of Education's expectations for large-scale standardized assessments used as part of state educational accountability systems. This report concludes that policymakers can elect to adopt the digital SAT Suite tests as part of their statewide educational assessment systems with high confidence that the tests elicit the kinds and degree of cognitively complex thinking required of students to be college and career ready by no later than the end of high school. Researchers interested in studying standardized tests or test-taking patterns may also benefit from the methodology described and exemplified in this report, as it has proved over repeated studies to be an effective tool in analyzing question types against their purported constructs.

References

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, 37–60. Amsterdam: 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. Philadelphia: Taylor and Francis.

Beck, Isabel L., Margaret G. McKeown, and Linda Kucan. 2013. *Bringing Words to Life: Robust Vocabulary Instruction*, 2nd ed. New York: Guilford.

Bettman, James R., and Park, C. Whan. 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, no. 3 (December): 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, no. 1 (Spring): 234–55. **https://doi.org/10.2307/2490945**.

Bolton, Ruth N. 1993. "Pretesting Questionnaires: Content Analyses of Respondents' Concurrent Verbal Protocols." *Marketing Science* 12, no. 3 (Summer): 280–303. **https://www.jstor.org/stable/184025**.

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, no. 2 (Summer): 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*. Edmonton: 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, no. 2 (April/May/June): 197–221. **https://www.jstor.org/stable/26622508**.

College Board. 2019. *College Board National Curriculum Survey Report 2019.* New York: College Board. **https://satsuite.collegeboard.org/media/pdf/national-curriculum-survey-report.pdf**.

College Board. 2022. *2022 SAT Suite of Assessments Annual Report: Total Group.* New York: College Board. **https://reports.collegeboard.org/media/pdf/2022-total-group-sat-suite-of-assessments-annual-report.pdf**.

College Board. 2023a. *Assessment Framework for the Digital SAT Suite*, version 2.0 (August 2023). New York: College Board. **https://satsuite.collegeboard.org/media/pdf/assessment-framework-for-digital-sat-suite.pdf**.

College Board. 2023b. *Skills Insight for the Digital SAT Suite.* New York: College Board. **https://satsuite.collegeboard.org/media/pdf/skills-insight-digital-sat-suite.pdf**.

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

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

Goos, Merrilyn, and Peter Galbraith. 1996. "Do It This Way! Metacognitive Strategies in Collaborative Mathematical Problem Solving." *Educational Studies in Mathematics* 30, no. 3 (April): 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, no. 4 (December): 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. Minneapolis: 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. Minneapolis: 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, no. 1 (January): 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, no. 1 (Winter): 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, no. 1 (March): 35–57. **https://doi.org/10.1017/S0272263104026129**.

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, no. 1 (Winter): 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, 133–52. Canberra: 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, no. 3 (May): 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**.

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

Russo, J. Edward, Eric J. Johnson, and Debra L. Stephens. 1989. "The Validity of Verbal Protocols." *Memory and Cognition* 17, no. 6 (November): 759–69. **https://doi.org/10.3758/BF03202637**.

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, 89–111. Thousand Oaks, CA: 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, no. 2 (April): 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, no. 3 (September): 413–25. **https://doi.org/10.1037/0022-0663.92.3.413**.

U.S. Department of Education. 2018. *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process.* Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. **https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf**.

Vessey, Iris. 1986. "Expertise in Debugging Computer Programs: An Analysis of the Content of Verbal Protocols." *IEEE Transactions on Systems, Man, and Cybernetics* 16, no. 5 (September): 621–37. **https://doi.org/10.1109/TSMC.1986.289308**.

Appendix

Recruitment Materials

EXHIBIT 1

The following recruitment email was sent out to eleventh- and twelfth-grade students who had previously taken the PSAT/NMSQT / PSAT 10 or SAT exams and had elected to receive emails from College Board. Underlining indicates the placement of (now deactivated) hyperlinks. Note that after this form was used with prospective participants but before interviews were conducted, College Board and Vidlet, Inc., agreed that $150 rather than $100 would be a more appropriate compensation given the time and effort participants were asked to put in.

Dear [student],

College Board regularly conducts research to evaluate our assessments.

You're eligible to earn a $100 digital gift card for participating in an online research study that will take no more than an hour and a half. Your input will help us ensure the quality of our assessments for future students.

Learn More

This study will be conducted in early April. Learn about the study, and sign up to participate.

On the day of the study, you must have 90 continuous minutes to participate, as well as:

- Access to a quiet, uninterrupted workspace.
- Use of a desktop or laptop computer, Chromebook, or tablet.
- Access to a reliable internet connection with sufficient capacity for video and screensharing via Zoom.

This study will be conducted entirely online and consists of an interview in which you'll be asked to describe your approach to answering a series of SAT® questions. On successful completion of the activity, you'll receive an email with a $100 digital gift card that can be used at a variety of retailers.

How Do I Learn More and Sign Up?

Complete this form by **Wednesday, March 29**, to review the details of what you'll be asked to do and to sign up to participate. There's limited space in this research study, and you'll be informed on April 3 if you've been selected to participate. If you aren't selected, we'll notify you if other opportunities arise in the future.

Sincerely,
College Board

EXHIBIT 2

Prospective participants received access to an electronic version of the following interest form via a hyperlink from the email in exhibit 1. Note that the document's formatting has been modified slightly here for readability and that the name of the gift card vendor has been redacted. Note also that after this form was used with prospective participants but before interviews were conducted, College Board and Vidlet, Inc., agreed that $150 rather than $100 would be a more appropriate compensation given the time and effort participants were asked to put in.

Cog Lab Research Study Interest Form - March 2023

Q1 Thank you for your interest in this study. College Board regularly conducts research to evaluate our assessments. You are eligible to earn a $100 digital gift card for participating in an online research study that will take no more than an hour and a half. 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.

This study will be conducted entirely online and consists of a 90 minute interview where you'll be asked to describe how you respond to SAT questions. Upon successful completion of the activity, you will receive an email with a link to redeem a $100 digital gift card with a retailer of your choice.

Sign-up will take less than 10 minutes.

- ◯ Yes, I would like to sign up to participate. (1)
- ◯ No, I cannot meet the above. (2)

Q2 Are you sure you want to close this form without signing up?

- ◯ Yes, close this form. (1) [terminate]
- ◯ No, I want to sign up. (2)

Q3 If selected to participate, your name and email address will be shared with our contractor, Vidlet, who will send a link to schedule your interview and a consent form which must be signed by your parent or guardian, or you if you are over 18. You will be notified on April 3 if you have been selected.

Daytime and evening interview sessions will be held between April 6 and 16 and are available on a first-come, first-served basis. Once you schedule your session and return your signed consent form, you will receive a Zoom link to join at the time of your interview.

To complete your interview and earn a $100 digital gift card, you must have 90 continuous minutes to work as well as:

- access to a quiet, uninterrupted workspace.
- use of a desktop or laptop computer, Chromebook, or tablet.
- access to reliable internet connection with sufficient capacity for video and screensharing via Zoom.

During the interview, you will meet over Zoom one-on-one with a researcher from Vidlet. The interviewer will send a link to you for you to access the sample SAT questions. You will need to share your screen, and the interviewer will ask you to describe how you would approach answering the questions. The sessions will be recorded.

After you successfully complete the interview, Vidlet will process your gift card through digital payment platform [redacted]. You will receive a link from [redacted] to the email address provided, which you can use to redeem your payment in the form of a bank transfer, PayPal deposit, or a gift card of your choice – [redacted] has over 300 gift card options for you to choose from.

If you are selected to participate, College Board reserves the option to cancel your participation in its sole discretion if your participation is no longer needed. In such case, you will not receive a gift card.

Please confirm you agree to these requirements.

- ◯ I agree (1)
- ◯ I do not agree (2) [terminate]

Q4 Great. Let's confirm your information. The information you provide here will be used to confirm your eligibility for the study, and if selected your name and email address will be shared with Vidlet. Please note that ALL fields are required.

Q5 First Name

Q6 Last Name

Q7 Email

Q8 Grade Level

- ◯ 12th (1)
- ◯ 11th (2)

Q9 If selected to participate, you will receive an email from Vidlet on April 3 with the link to schedule your activity and consent form. Interviews will be held April 6-16.

During the interview, you will meet over Zoom one-on-one with a researcher from Vidlet. The interviewer will send a link to you for you to access the sample SAT questions. You will need to share your screen, and the interviewer will ask you to describe how you would approach answering the questions. The sessions will be recorded.

If you successfully complete the interview, you will receive a link to redeem a $100 digital gift card through [redacted] at your designated email address.

If you are selected to participate, College Board reserves the option to cancel your participation in its sole discretion if your participation is no longer needed. In such case, you will not receive a gift card.

By clicking submit, you are signing up to participate in this research study, your information will be used to confirm your eligibility to participate, and if selected will be shared with Vidlet.

- ◯ Yes I want to participate. (1)

End of Survey Page:

We thank you for your time spent taking this survey. Your response has been recorded.

Termination Page:

Thank you for your interest. We're sorry to hear that you cannot participate. Please watch your email for other opportunities.

EXHIBIT 3

The following is a copy of the consent form signed by participating students and their parents/guardians. As before, the name of the gift card vendor has been redacted.

Student Research Group Agreement

By signing this agreement, the student identified below ("**Student**"), with consent of their parent/guardian ("**Parent/Guardian**"), agree 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, [redacted]. Student will receive a link from [redacted] 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 – [redacted] 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 (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

Sample Protocol

EXHIBIT 4

The following is a Reading and Writing sample of the interview protocols used in this study. The Math version (not included) is closely parallel, with only subject area–specific differences.

Digital SAT Suite Reading and Writing Section
Cognitive Interview Script

General Instructions to Interviewer

> All roman-font text in this script indicates directions for the interviewer. These should *not* be read aloud to the student.
>
> All *italicized*-font text should be read word-for-word by the interviewer to the student.

This research activity involves a single interviewer leading a student through a "think-aloud" exercise in which the student verbalizes their thoughts as they engage with a series of test questions. This activity takes place virtually via Zoom or other videoconferencing tool. The test questions are presented to the student via Qualtrics in a survey format, with the student sharing their screen as they work. The activity is scheduled for 90 minutes, with time allocated in the following way:

A. Welcome and activity introduction; student setup (5 min)
B. Think-aloud modeling and practice (5 min)
C. Think-aloud activity (70 min)
D. Postexperience interview questions (10 min)
E. Wrap-up and thank-you

If, at any time, deviations from this script occur, the interviewer should document them on the **timing and irregularity form** (a copy of which is attached as appendix B). This form should be labeled with the student's identification code and the date and time of the interview.

> Each student should be assigned a unique identifier for use in reporting on the study.

[Additional procedural details to be established by Vidlet. These should include how to prepare for and set up virtual interviews and ensure that appropriate paperwork is completed.]

Late-Arriving Students

If the student arrives **15 or fewer minutes late** to their scheduled interview, the interviewer should still conduct the interview. Omit the postexperience interview questions (section D) if necessary to allow the student the full 70 minutes for the think-aloud activity (section C).

If a student is **more than 15 minutes late**, the interviewer should seek to reschedule the interview, as there will not be enough time to obtain an unhurried think-aloud experience.

Late arrivals should be noted on the timing and irregularity form.

A. Welcome and activity introduction; student setup (5 min)

[Additional procedural details to be established by Vidlet. These should include how to begin the virtual interviews and how participants will access the test content and share their screen.]

During the interview, sit within camera view. Provide the student with the following overview:

> *Thank you for taking time to participate in this research study today. Before I explain the activity, I want to give you some background. The purpose of this research study is to help College Board, the makers of the SAT, learn more about how students like you approach questions on the test—specifically, Reading and Writing test questions. I'll be reading a lot from this document today. This is to help ensure that all students participating in this research activity have as similar an experience as possible.*
>
> *This research is to evaluate the test questions, not you, so don't be concerned about whether you answer a particular question correctly.*
>
> *During our time together, please keep your computer's camera on and the microphone unmuted as much as possible. Please silence your phone, try to avoid distractions and interruptions, and don't allow others to join you in this activity. Please also close all computer applications except the ones being used in this activity so that you can better focus on this task.*
>
> *At any time during this study, you're welcome to take a break, use the restroom, or choose to stop participating. All you need to do is let me know. All the information we collect today will be used only for research purposes, and you will* not *be identified by name or other personally identifying information in our final report. After successfully completing all steps in the study, you'll receive a $100 gift card.*
>
> *Any questions or concerns so far?*

After addressing any questions or concerns, continue to section B.

B. Think-Aloud Modeling and Practice (5 min)

Overview

Read the following text to the student:

> *In today's session, you'll participate in what's known as a "think-aloud" activity. In this activity, you'll respond both aloud and onscreen to several Reading and Writing test questions we present you with. Once you've responded to all the test questions, I'll ask you a few follow-up questions about your experience today.*
>
> *In this activity, you'll think aloud as you work through each test question. You'll verbally share any and all thoughts you have about each question as you read and answer it. In doing so, you'll describe all the steps you take to obtain your answer as well as any other thoughts about the question that occur to you.*
>
> *Your goal today is to think aloud as fully, honestly, and freely as possible as you work through each question. Remember: We're evaluating the questions, not*

you. There are no right or wrong thoughts, and you won't be penalized in any way for anything you say. We also won't be scoring your answers today.

Interviewer Modeling Activity

Read the following to the student:

I realize you may not have participated in a think-aloud study before, so let's consider a couple of examples. First, I'll demonstrate thinking aloud using a sample test question. Then I'll give you a sample question so that you can practice thinking aloud before you begin answering the rest of the questions.

Direct the student to the interviewer practice question ("IP") in the Qualtrics survey so the student can follow along.

I'll start by reading the test directions, passage, and question aloud and then narrate what I'm thinking as I answer the question.

Demonstrate the think-aloud method using the interviewer practice question:

Directions: *The questions in this section address a number of important reading and writing skills. Each question includes one or more passages, which may include a table or graph. Read each passage and question carefully, and then choose the best answer to the question based on the passage(s). All questions in this section are multiple-choice with four answer choices. Each question has a single best answer.*

Passage and question: *The following text is from F. Scott Fitzgerald's 1925 novel* The Great Gatsby*.*

[Jay Gatsby] was balancing himself on the dashboard of his car with that resourcefulness of movement that is so peculiarly American—that comes, I suppose, with the absence of lifting work in youth and, even more, with the formless grace of our nervous, sporadic games. This quality was continually breaking through his punctilious manner in the shape of restlessness.

As used in the text, what does the word "quality" most nearly mean?

A) Characteristic

B) Standard

C) Prestige

D) Accomplishment

Reading this passage and question, it looks like I'm being asked to figure out how the word "quality" is used in the text. "Quality" appears in the last sentence of the passage: "This quality was continually breaking through his punctilious manner in the shape of restlessness." I have no idea what "punctilious" means, but I think I can still answer the question about "quality" without knowing that. Going back through the passage, I realize that "this quality" refers to Gatsby's "resourcefulness of movement."

I'm now looking at the answer choices and trying to figure out which one is the best answer here. "Characteristic," choice A, makes sense, because the passage is describing something that Gatsby regularly shows, like a trait. The passage tells us that Gatsby's restlessness is "continually breaking through

his punctilious manner." So it follows that this "quality" of Gatsby's is a characteristic, or something that he's known for.

Choice B, "standard," isn't as good an answer to the question. "Standard" to me suggests that Gatsby's being judged or held to a standard by someone, which doesn't make sense here. The narrator is just describing what Gatsby is like. I can rule out choice C, "prestige," pretty easily because it doesn't really make any sense here. A "prestige" isn't something someone can have or show. Choice D, "accomplishment," makes a little more sense, but it's also not as good an answer as "characteristic." The narrator isn't really saying that Gatsby has accomplished or achieved something by being restless. It's just the way he is.

So I'm going to mark choice A is my answer.

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?

After addressing any questions or concerns, continue to the first student practice question.

Student Practice Activity 1

Direct the student to the first student practice question, "SP1."

Now I'd like you to practice thinking aloud using this practice question.

Remember: Try to say everything that goes through your mind as you read and answer the question. Please begin by reading the question and answer choices out loud. Continue by thinking aloud as you answer the question.

The first student practice question ("SP1") is reprinted in appendix A for your information but should be presented to the student onscreen.

(Optional) Student Practice Activity 2

It's important that the student feel comfortable thinking aloud before beginning to tackle the actual study questions. If the student thinks aloud successfully through the first practice question, skip over the second practice question ("SP2," reproduced in appendix A for your convenience) and continue to section C with question 1. If, however, you think the student could benefit from additional practice, have the student continue to practice thinking aloud with SP2.

C. Think-Aloud Activity (70 minutes)

Read the following to the student:

For this activity, you'll read several passages and respond to 20 questions. Each passage or pair of passages below is followed by a single question. After reading each passage or pair of passages, choose the best answer to each question based on what is stated or implied in the passage or passages and in any accompanying graphics (such as a table or graph).

You'll have 70 minutes to complete the Reading and Writing test questions. While you're working, I'll be using a timer to keep track of how long you take to answer each question. This is just so that we have a better sense of how long each question is taking students to answer.

Answer as many questions as fully and completely as you can. Answer each question to the best of your ability, and then move on to the next. You should have enough time for all 20 questions, but don't worry if you don't make it all the way through the full set. If you finish early, you may review your answers if you wish.

Remember to verbalize any and everything that comes to mind as you work through each question. If you stop talking for a bit, I'll prompt you to keep thinking aloud.

Do you have any questions or concerns before we begin?

After addressing any questions or concerns, direct the student to advance to the first actual test question ("1") in the Qualtrics survey.

Take out the timing and irregularity form (a copy of which appears in appendix B) to track the student's time on each question, and commence the activity by reading the following to the student:

After you finish the activity, I'll ask you a few questions about your experience today.

I'll turn on the timer as soon as you begin reading the first test question aloud.

I'll begin recording our session now.

START RECORDING

Vidlet to add any necessary details about the recordings

My name is [interviewer], and I'm interviewing [student's name and identification code] on [date] at [time].

This is a Reading and Writing think-aloud session. It consists of 20 questions.

Don't forget to read each passage, question, and answer choice out loud.

You have 70 minutes to complete these 20 questions.

Please begin.

The interviewer should watch and listen attentively to the student as they work through the questions.

Probes and prompts

- The interviewer should prompt the student to talk if it's obvious they're making progress but not verbalizing. Such **prompts should be minimal and nondirective**—for example, "Please remember to say out loud what you're thinking."
- If the student goes silent and appears stuck on a question, allow them approximately 15 seconds of silence before probing. Then say "Remember to keep talking," "Please continue," "Go on," or the like.

As students work on the questions, the interviewer should **maintain the timing and irregularity form**.

- Record the start time, to the nearest second, for each question.
- Record the stop time, to the nearest second, for each question.
- When students complete their first attempt at answering a question, enter the start and stop times from the timer in the "Attempt 1" column. If the student returns to a question to complete it or to check their work, use the "Attempt 2" and "Attempt 3" columns, as needed, to track the time.
- If a student makes more than one attempt on a particular question, flag this on the form so that College Board can easily see that the student's efforts aren't confined to the first attempt.

When the student finishes, even if they finish early, continue to section D.

Note that students may ask for input on how well they did or for the answer to one or more questions. Inform them that they're not being scored on their answers and that their verbal responses are what we're interested in.

If the student doesn't finish within the 70 minutes, say:

> *Time's up. Please stop working. Don't worry about any questions you didn't answer.*

Continue to section D.

D. Postexperience Interview Questions (10 min)

Continue the interview by saying the following:

> *You've finished the part of the interview dedicated to answering the study's Reading and Writing test questions.*
>
> *Now I'd like to ask you a few questions about your experience today.*

Ask the following retrospective questions, in order. If time runs short, omit later questions as needed to allow the session to end at the approximately 90-minute mark.

Don't call attention to particular test questions or particular verbalizations you heard. Allow the student to answer the follow-up questions in their own way. If asked, you may rephrase the questions below for clarification, but don't model possible responses.

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 easy to answer? If so, which one and why?*
4. *Was there a particular type of question that you found hard to answer? If so, which one and why?*
5. *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?*

STOP RECORDING

Continue to section E.

E. Wrap-Up and Thank-You

Say:

> *That concludes our interview. Thank you for participating. Your input regarding these Reading and Writing questions is valuable to us.*

Gift Card
[Vidlet to describe the gift card procedure]

Conclude Session
Thank the student again for their participation.

Ensure that the videoconference session has ended.

Complete Paperwork
Ensure that the timing and irregularity form is complete and appropriately stored.

> If something nonstandard happened during the session, this should be recorded as an Irregularity on the form. The interviewer should note the circumstance (e.g., student was late; power went out) and its impact on the interview.

Finish Up
[Vidlet to provide directions to the interviewer for what to do next]

* * *

Appendix A: Student Practice Items

SP1

Some studies have suggested that posture can influence cognition, but we should not overstate this phenomenon. A case in point: In a 2014 study, Megan O'Brien and Alaa Ahmed had subjects stand or sit while making risky simulated economic decisions. Standing is more physically unstable and cognitively demanding than sitting; accordingly, O'Brien and Ahmed hypothesized that standing subjects would display more risk aversion during the decision-making tasks than sitting subjects did, since they would want to avoid further feelings of discomfort and complicated risk evaluations. But O'Brien and Ahmed actually found no difference in the groups' performance.

Which choice best states the main purpose of the text?

A) It presents the study by O'Brien and Ahmed to critique the methods and results reported in previous studies of the effects of posture on cognition.

B) It argues that research findings about the effects of posture on cognition are often misunderstood, as in the case of O'Brien and Ahmed's study.

C) It explains a significant problem in the emerging understanding of posture's effects on cognition and how O'Brien and Ahmed tried to solve that problem.

D) It discusses the study by O'Brien and Ahmed to illustrate why caution is needed when making claims about the effects of posture on cognition.

(Key: D)

SP2

Many animals, including humans, must sleep, and sleep is known to have a role in everything from healing injuries to encoding information in long-term memory. But some scientists claim that, from an evolutionary standpoint, deep sleep for hours at a time leaves an animal so vulnerable that the known benefits of sleeping seem insufficient to explain why it became so widespread in the animal kingdom. These scientists therefore imply that _____

Which choice most logically completes the text?

A) it is more important to understand how widespread prolonged deep sleep is than to understand its function.

B) prolonged deep sleep is likely advantageous in ways that have yet to be discovered.

C) many traits that provide significant benefits for an animal also likely pose risks to that animal.

D) most traits perform functions that are hard to understand from an evolutionary standpoint.

(Key: B)

Appendix B: Timing and Irregularity Form (Sample)

Timing and Irregularity Form: Reading and Writing

Student Identifier: _____

Date and Time of Interview: _____

Directions:

Timing: Using a timer, record in the "Attempt 1" column the start and stop times, to the nearest second, associated with the student working on a given test item. If the student returns to an item, also list the start/stop times for "Attempt 2" and, if necessary, "Attempt 3" in the same manner. Flag those rows for easy identification by College Board.

Irregularities: Record any deviations from the interview script in the irregularities section. Indicate what impact, if any, such irregularities had on the session.

	Attempt 1		Attempt 2		Attempt 3	
Item	Start	Stop	Start	Stop	Start	Stop
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Irregularities (e.g., late arrival, power loss):