

ETS GRE[®] Board Research Report

ETS GRE[®] – 14-05

ETS Research Report No. RR–14-28

Usability of Interactive Item Types and Tools Introduced in the New GRE[®] revised General Test

Wanda D. Swiggett

Laurie Kotloff

Chelsea Ezzo

Rachel Adler

Maria Elena Oliveri

December 2014

The report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board reports do not necessarily represent official Graduate Record Examinations Board position or policy.

The Graduate Record Examinations and ETS are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

GRE-ETS

PO Box 6000

Princeton, NJ 08541-6000

USA

To obtain more information about GRE programs and services, use one of the following:

Phone: 1-866-473-4373

(U.S., U.S. Territories*, and Canada)

1-609-771-7670

(all other locations)

Web site: www.gre.org

*America Samoa, Guam, Puerto Rico, and US Virgin Islands



RESEARCH REPORT

Usability of Interactive Item Types and Tools Introduced in the New GRE[®] revised General Test

Wanda D. Swiggett, Laurie Kotloff, Chelsea Ezzo, Rachel Adler, & Maria Elena Oliveri

Educational Testing Service, Princeton, NJ

The computer-based *Graduate Record Examinations*[®] (GRE[®]) revised General Test includes interactive item types and testing environment tools (e.g., test navigation, on-screen calculator, and help). How well do test takers understand these innovations? If test takers do not understand the new item types, these innovations may introduce construct-irrelevant variance, with test takers performing differently than they would with more familiar item types. Similarly, the navigational and other test environment tools are another potential source of variance, if some test takers understand how to use them and others do not.

In this study, we examined the reactions, engagement, and difficulties encountered as 20 potential test takers completed Verbal and Quantitative Reasoning sections of a practice GRE test. Participants were sophomores and juniors from colleges and universities in the local area. Their reactions were captured through the use of cognitive laboratory sessions that incorporated interviews that required test takers to think aloud, as well as researcher observations as test takers worked quietly. Results of the analysis of this data revealed that some participants needed time to figure out what was being asked of them when they encountered the new item types, although most were able to answer each item eventually. On the other hand, most participants stated that they did not even notice the test environment tools, and few were observed actually using the tools. Several participants provided suggestions about improving the usability of the new item types and test environment tools.

Keywords GRE; usability; interactivity; test environment tools; cognitive interviews

doi:10.1002/ets2.12028

The computer-based GRE revised General Test (hereafter, the GRE), which was introduced in August 2011, has several new item types and test environment tools. New item types in the Verbal Reasoning sections include text completion (with up to three blanks), sentence equivalence items, and items that require test takers to highlight a sentence in a passage to answer the question. A new item type in the Quantitative Reasoning section requires typing the answer into designated boxes (numeric entry). Both sections contain multiple-choice questions that may have more than a single correct response (multiple-answer multiple-choice). In addition to the new item types, test environment tools have been added to assist test takers as they navigate within each of the separately timed sections of the test. In both sections, the tools include previewing and reviewing capabilities, question tagging for review, and the ability to move back and edit answers before moving on to the next section. In the Quantitative Reasoning section, an on-screen calculator is also available.

The purpose of this study was to observe potential test takers' initial reactions to the new items types and tools. The study used cognitive interview methodologies to gather data on the perceptions and experiences of a sample of potential test takers as they completed verbal and quantitative items from computer-based practice test sections. Test takers also provided valuable feedback to improve the usability of these enhancements. The findings of this study can inform the development of future versions of GRE test preparation materials, including revisions of the visual layout of the computer test, such as the placement and wording of the informational and instructional text. Such improvements may better prepare test takers to understand the functionality of the test environment tools and the requirements for responding to the new item types.

Background

The GRE has been revised with new item types and tools that take advantage of the interactivity of using computers. These changes suggest the need for an evaluation of their effectiveness. These evaluative usability studies are also needed to understand whether there are any unanticipated effects on future test takers. Cognitive interview methodologies are

Corresponding author: W. D. Swiggett, E-mail: WSwiggett@ets.org

frequently employed in usability studies. In this study, we used a combination of those methodologies within the cognitive laboratory session.

Cognitive Interview Methods

We employed cognitive interviews for qualitative data collection in order to learn what our study participants were thinking as they interacted with the new item types and tools. Qualitative methodologies have been historically used to investigate the usability of test revisions. In this study, we used natural observations and observations using a think-aloud method, where we asked participants to verbalize their thoughts as they answered selected questions. We also used interview techniques involving open- and closed-ended retrospective questions. This combination of qualitative data collection techniques is described as a cognitive laboratory in this report.

In the early 1980s, software developers began conducting usability studies to test user interface and identify design features that potentially interfered with the software's intended use (Nielsen & Mack, 1994). Increasingly, these usability studies incorporated cognitive interviews to learn about users' thoughts and reactions as they interacted with the software. The methodology has evolved to include a combination of think-aloud processes, which asks users to vocalize their thought processes as they complete individual tasks as well as retrospective questions that ask about users' experience of the software program after they complete the interaction. For test developers, usability studies and cognitive interview methodologies have become important components of test development, as they provide a means of collecting evidence to help ensure that tests are fair and valid for intended test takers (e.g., Johnstone, Bottsford-Miller, & Thompson, 2006).

Almond et al. (2009) recommended including cognitive interviews during three phases of test development, each of which serves different purposes. They suggested conducting either exploratory or confirmatory cognitive interviews after items are developed, during pilot testing, during field testing, and potentially in conjunction with live testing or experimentally designed studies. Detailed individual reactions to items can provide helpful information on item difficulty, novel item types or item format issues, difficulties for specific groups of test takers, and test-taker preferences. Cognitive interviews can help determine whether test takers' incorrect answers were the result of construct-irrelevant features that led them astray or because they did not know the correct answer. Although the GRE had been administered on computers since the mid-1990s, the item types were identical to the item types that had been used on the paper-delivered test. When the GRE began exploring new item types that took better advantage of computer capabilities, usability studies were conducted at several points in the development process (e.g., Stone, King, & Laitusis, 2011).

The move to these enhanced items with computer-based testing was not only a change in the way test takers viewed and answered the question; the technology presented a more varied way for test takers to interact with the test in order to demonstrate their skills. With the introduction of new item types and tools for test takers, this usability study was warranted and the cognitive interviews were necessary. This study, however, was conducted after the test had been launched because of limitations in research resources. At this stage, the results of the study are most informative to the development of test preparation materials. They can also result in improvements in the instructions in future test forms. This study adds to the research that has already been conducted at the previously described phases of test development (e.g., McKinley, Mills, Reese, Schaeffer, & Steffen, 1993; Nissan & Schedl, 2012; Parshall, Harnes, Davey, & Pashley, 2010).

New Item Types

The GRE includes new item types and test environment tools in both the Quantitative and the Verbal Reasoning sections. In both sections, there are multiple-choice questions that are traditional, single-answer items. There are also multiple-choice items with one or more possible answers. Additionally, there are new item types that are unique to each of the two sections. (see Appendix A for examples of each item type.)

According to the GRE website, "in the Quantitative Reasoning section, more focus has been placed on data interpretation and real-life scenarios, with multiple-choice and numeric entry answers" (ETS, 2011, Test Content, section 3). In addition to data interpretation sets with single- or multiple-answer multiple-choice questions, the numeric entry items in the Quantitative Reasoning section require test takers to enter the exact numeric answer (unless rounding is specified). The response required can be a single number or a fraction, with separate entry boxes for the numerator and denominator. The section also includes quantitative comparison items that had been carried over from the paper test and from the previous computer-based test. Two quantities are provided and the test taker must determine whether they are equal,

whether one is greater than the other, or whether there is enough information provided to make that determination. Although this was not a new item type, we included it in this usability studies as its usability features had not previously been studied in this manner.

In the Verbal Reasoning section, the items are designed to “... test your ability to interpret, evaluate and reason from what you’ve read” (ETS, 2011, Test Content, section 3). In addition to the traditional reading comprehension sets with single- or multiple-answer multiple-choice questions, there are also select-in-passage questions where the test taker must highlight a portion of the passage in order to answer the question. There are also text completion questions where the test taker must complete text containing one to three blanks using the appropriate words for each blank (selecting from three to four options per blank). Additionally, there are sentence equivalence questions. These questions provide the test taker with one sentence containing one blank but require the test taker to select exactly two choices. When each choice is used to complete the sentence, both of the resulting sentences are expected to have the same meaning.

Test Environment Tools

The GRE testing interface provides tools that allow test takers to move efficiently within a test section (once the test takers leave the section, they cannot return to it). Therefore, using the tools can assist test takers before their time expires in each section. The mark and review tools are useful tools for checking one’s work before moving on to the next section of the test. After clicking the review icon, test takers are presented with a screen showing the item numbers within the test section. Next to each number is text stating that the item is either “Not Seen,” seen but “Not Answered,” or “Answered.” A check mark also appears next to the questions that have been marked using the mark tool. To use this feature, a test taker clicks the icon labeled “Mark” (available above every question in the section). A check mark appears on the icon and in the review screen for each question that was marked. (Screenshots of all of the test environment tools are provided in Appendix B. The review screen is shown in Figure B1.)

The help tool provides the test taker with a variety of information. The help screen contains multiple tabs relating to the test section and item type on which the test taker is currently working. The first tab provides directions specifically about the item type the test taker was viewing before clicking the help icon. It also illustrates a sample question and shows how it is answered (illustrated in Figure 1). The second tab displays the directions for the test section in which the test taker is working. The third tab provides general test-taking directions (that appear at the very beginning of the testing session). The fourth tab provides a description of all of the test environment tools available. When the test taker is in the Quantitative Reasoning section, a fifth tab provides a description of the on-screen calculator. (See Figures B2 and B3 for additional screenshots of the help screen.)

In the Quantitative Reasoning section, the interface provides an on-screen calculator. This tool benefits test takers as they demonstrate their quantitative skills and abilities. “Providing a calculator helps to assure that trivial computational errors are not interfering with assessment of the intended reasoning construct” (Bridgeman, Cline, & Levin, 2008, p. 1). The calculator is a basic, five-function calculator, including the square-root function. It also has memory capabilities, parentheses, and a button labeled “Transfer Display.” The Transfer Display button, when clicked, sends the calculated result from the calculator display to a numeric entry text box. The feature is available only when there is a single response required to answer a numeric entry question, not when two responses are required to complete a fraction. The on-screen calculator is the only calculator permissible during testing. (See Figure B3 for a screenshot of the calculator description from the help screen.)

Purpose and Rationale for the Study

The goal of this study was to learn from future test takers about the clarity of the format (i.e., visual layout) and directions of the new item types that are included in the GRE. We also sought to understand the ease of use and added value of the new computer-based tools. Using retrospective questions, observations, and the think-aloud method, the study focused on the following research questions:

1. How frequently do test takers use the optional tools (mark, review, help, and calculator), and what are their perceptions of the tools’ usefulness? Do they use some tools more than others? Are tools used more frequently for some questions or question types than others? What, if any, challenges do test takers encounter when using the tools? Are some tools particularly problematic or useful? How could tool displays and instructions be improved?

The screenshot shows the GRE Test Preview Tool interface. At the top, it says "IBT Client" and "GRE® Test Preview Tool Section 3 of 3". There are navigation buttons: "Quit Test", "Exit Section", "Review", "Mark", "Calc", "Help", "Back", and "Next". A "Break" button is also visible. The current question is "Question 1 of 11" and the time is "00 : 34 : 54".

The main content area displays a geometry problem. It shows a triangle PQR with a point S on the base PR . A line segment QS is drawn from vertex Q to point S . Below the diagram, it is stated that $PQ = PR$.

Below the diagram, two quantities are defined:

- Quantity A: PS
- Quantity B: SR

Four radio button options are provided:

- Quantity A is greater.
- Quantity B is greater.
- The two quantities are equal.
- The relationship cannot be determined from the information given.

At the bottom of the question area, it says "Click on your choice."

Figure 1 The help screen provides up to five screens of information (labeled with descriptive tabs). The fifth tab, describing the on-screen calculator, is only available in the Quantitative Reasoning section. In this screenshot, the first tab is illustrated. It provides directions and a sample question to instruct the test taker on how to complete the item type he or she reached before clicking the help icon. After using the help tool, the test taker clicks the "Return" icon to return to the test item.

2. If any, how severe are test takers' misunderstandings, confusion, errors, and difficulties with the new item types or the ways they are displayed (format)? Are some formats or item types particularly problematic? If so, how might they be fixed? How could item directions be clearer?
3. Are there test-taker characteristics that may or may not have an effect on how they interact with the new item types and tools? Demographic characteristics such as major, minor, academic honors, gender, and country of origin may affect the type of reactions and feedback we receive. Other background characteristics such as experiences with technology may also affect test takers' interactions with the test features. Unfortunately, the small number of participants limited the power of any subgroup analyses of results. Preliminary analyses suggested no clear relationships between the participating test takers' characteristics and their performance, so this research question is not pursued further in this report.

Method

Participants

Recruitment Procedures

Undergraduate sophomores or juniors who intended to take the GRE but had not yet begun preparing were recruited from the Central New Jersey and greater Philadelphia region. Participants were recruited in two ways. An announcement

Table 1 Gender Differences

Recruitment steps	Gender differences			
	Male		Female	
	<i>n</i>	%	<i>n</i>	%
1. Consent form (<i>n</i> = 35)	18	51%	17	49%
2. BIQ (<i>n</i> = 29)	16	51%	13	49%
3. Cognitive interview (<i>n</i> = 20)	13	65%	7	35%

was posted on the ETS internal website describing the study and the targeted participants. Also, the authors sent a similar description via e-mail to personal contacts that could connect them with potential participants. Individuals who expressed interest in the study completed an informed consent form and then an online background information questionnaire (BIQ). Participants who completed the BIQ were then contacted to participate in a cognitive laboratory session at the ETS campus in Princeton, NJ. Upon completion of the cognitive laboratory session, they were each given a \$75 Visa gift card to thank them for their participation in the study. The recruitment process concluded upon the completion of 20 cognitive laboratory sessions.

Response Rates and Background Characteristics

We recruited 35 students who completed the consent form (18 male and 17 female). Of those, 29 (83%) completed the BIQ. The final sample was the target size of 20 (57% of those who completed the consent form and 69% of those who completed the BIQ). However, the backgrounds of the 20 participants differed qualitatively from the backgrounds of those who were recruited but did not participate in the cognitive laboratory. The resulting sample became disproportionately male, more academically involved, and more academically accomplished. We used the data collected on the BIQ to make a comparison of those who were recruited and did not complete the cognitive laboratory versus those who completed the cognitive laboratory. We discovered group response rates differences by gender, GPA, academic involvement, and honors. The gender differences are shown in Table 1 along with the number of students who completed each step in the recruitment process.

The academic accomplishments of those who completed the study were more consistent with each other and stronger than the accomplishments of those who left the study. For example, the median grade-point average (GPA) of the nine participants who left the study after completing the BIQ was 3.02 (ranging from 2.30 to 3.90), whereas the median of those who completed the cognitive laboratory was 3.50 (ranging from 2.90 to 3.85). Also, only two of the nine who left the study listed participation in academic clubs, and two others listed earning academic honors or scholarships. This is in sharp contrast to those who completed the study. There were five participants who only listed participation in clubs, four who only listed earning academic honors or scholarships, and three who listed both.

We discovered other group differences when comparing the majors of the students. The majors of those completing the study were more uniform than the majors of those who left after completing the BIQ. Five of the participants who left the study were kinesiology majors; the other four majored in bioenvironmental engineering, psychology, early grades preparation, and real estate. Almost two-thirds of the final sample (13 of 20) majored in scientific or mathematical fields (e.g., engineering or finance). Additionally, three participants majored in music. The majors of the remaining participants were journalism, business administration, and government. One was undecided. Appendix C lists the majors and minors of every recruited student who completed the BIQ.

When indicating their ethnicity, 15 of the 29 respondents indicated that they were White, 4 were African American, 4 were Asian, 2 were Latino, 2 were multiracial, 1 was Egyptian, and 1 declined to respond. Although the proportion of minority participants remained the same, five of the nine who left the study after completing the BIQ were minority and were qualitatively different from the minority groups represented in the final sample. The three African Americans who did not participate each reported a higher GPA than the one African American who did participate. The other minorities who participated in the study reported higher GPAs and more accomplishments than those who left the study. The demographic descriptions of the 20 participants who completed the study are presented in the BIQ results tables (Appendix C).

Materials and Procedures

Informed Consent

Each participant signed an informed consent form that explained the purpose of the study and the conditions of the cognitive laboratory. The participants were informed that they would receive a \$75 Visa gift card when the cognitive laboratory was completed.

Background Information Questionnaire (BIQ)

The BIQ was administered to collect background information about the participants, such as their demographic characteristics (e.g., gender, race/ethnicity), major and minor fields of study, experience with and plans to take the GRE, and familiarity with different types of computer software. The survey was adaptive based on participants' answers. For example, upon selection of computer software programs with which they were familiar, they saw follow-up questions related only to those programs (such as how frequently each type of software is used).

POWERPREP® II Software

ETS created *POWERPREP*® II software, version 2.0 for the computer-based GRE so that potential test takers could prepare for the test by familiarizing themselves with its structure, content, and computer-based delivery. Students can download the free software and take multiple timed or untimed practice test forms on their personal computers. Because *POWERPREP* II mirrors the computer-based GRE in structure, format, layout, and content, it was a useful platform for gathering participant feedback about the revised test without having to use items that are currently operational. Each participant used an ETS-supplied laptop to complete two sections of *POWERPREP* II: a Verbal and a Quantitative Reasoning section.

Cognitive Laboratory Protocol and Interviewers

There were six interviewers involved in the study; four were female and are authors of this report. The remaining two interviewers were male and they did not participate in the development of the protocol. Interviewers were assigned based on a match between the interviewers' and the participants' schedules. As a result, two interviewers each completed one interview. The other interviewers completed two, three, five, and eight interviews respectively.

The interviewers followed a standard protocol for each research condition. The interviewers used a scripted protocol, and each interviewer was trained to read the script. The scripting was provided to eliminate the use of leading questions and the offering of unintended assistance to the participants. The protocol consisted of natural and think-aloud observations, as well as open- and closed-ended retrospective questions. When interviewers were being trained, they were provided with example phrases and prompts to say to participants. These phrases were used to encourage the participants to give more thorough answers while not providing leading information. During the laboratory session, interviewers recorded observations of participants' tool use and initial behaviors when encountering the new item types. In order to gather feedback on the full range of new item types included in the GRE, at least one example of each item type was included in the items selected for the untimed practice test (see the protocol outlines in Appendix D). All of the test environment tools (i.e., mark, review) were available throughout the entire cognitive laboratory session.

Generally, the protocol questions elicited participant feedback on a variety of features new to the GRE (i.e., the new item types and tools described above), specifically, the usability of the tools and visibility of various features relating to the items and the tools. Gathering feedback about the usability of the tools required questions regarding whether the participants understood the purpose of the tools and whether they considered the tools to be useful. The visibility of the various features included the layout of items on the screen, placement of informational text, and question directions. Participants were also asked for feedback on the content of section directions, content directions, and informational text.

Research Conditions

Study participants were assigned to one of two testing conditions based primarily on their major or minor field of study. Two protocols were used based on the research conditions. The protocol outlines are listed in Appendix D. For each

condition, participants completed both Quantitative and Verbal Reasoning sections. Participants first completed a short, timed practice test for one test section. At this stage, the interviewer primarily recorded observed tool use and, because this was a natural observation, participants were not interrupted. When participants completed all of the questions in this section, they were asked retrospective questions focused on their tool use.

Next, participants completed selected items in the untimed practice test (for a different test section), during which, the interviewers focused on participants' behaviors with the new item types, screen layout, directions, and other instructional text. While completing this section, participants were intermittently stopped and asked retrospective questions. They were also asked to verbalize their thoughts (think aloud) while completing some of the selected items.

- *VOQI condition* (verbal observation and quantitative interview): Participants first completed the Verbal Reasoning section of the timed practice test, which contained seven items. Participants then completed eight items from one of the Quantitative Reasoning sections of the untimed practice test. Nine participants were included in this condition.
- *QOVI condition* (quantitative observation and verbal interview): Participants first completed the Quantitative Reasoning section of the timed practice test, which contained 11 items. Participants then completed seven items from a Verbal Reasoning section of the untimed practice test. Eleven participants were included in this condition.

Most participants were assigned to conditions such that they took the timed practice test section (for the natural observation) in an area that was most closely related to their major or minor field of study. For example, if a participant was a mathematics major, he or she would be assigned to the QOVI condition, to be naturally observed completing the quantitative items without interruption. This method was intended to save time, as performance on the practice tests was not an integral part of the study (presumably a mathematics major would answer the quantitative questions more quickly than an English major). As recruitment was concluding, however, the goal of having roughly the same number of participants in each condition superseded this method of condition assignment.

Cognitive Interview Methodology

ETS researchers met with each study participant for an interview session that lasted approximately 1.5 hours. The participant was instructed to complete each item as if he or she were in normal GRE testing conditions (i.e., they could not ask questions, look up answers, etc.). Although they were encouraged to try their best, the participants were told that researchers were not focusing on their test performance. This was made explicit so that participants would not be nervous about whether they answered correctly or incorrectly. The participants were provided with sheets of blank paper, a pencil, and eraser in case they wanted to work out problems by hand (although the on-screen calculator was available in the Quantitative Reasoning section). The interviewers did not provide any instructions on how to take the test or any information to aid the participant (e.g., the presence of the on-screen calculator). Because the purpose of the study was to learn how potential test takers interacted with the interface, test-taking instructions from the interviewer would interfere with that goal.

Usability of Tools and Features

During the first part of the cognitive laboratory, participants completed the timed practice test in either the Quantitative or Verbal Reasoning section. As participants completed each item, the interviewers recorded observations about whether the new tools or features were used; they did not ask any interview questions. Although participants' answers to the test questions were not used in the analyses for the study, they were recorded and provided to the participant along with the answer key. If a participant asked a question related to the test, the interviewer instructed him or her to complete the test as if the interviewer was not present. Upon completion of the timed practice test section, the participant was asked a series of retrospective questions about each tool they used. If a tool was not used, participants were asked whether they noticed that the tool was available for them to use. If they saw that the tool was available for their use, they were asked to explain why they chose not to use it. Participants were also asked to provide feedback on ways to increase the usability of the available tools.

Usability of Verbal or Quantitative Reasoning Items

After the timed practice test was completed, and the retrospective questions were answered, participants were then given a description of the second half of the cognitive laboratory. They were told to stop after answering each question to be provided with instructions (either to skip to a different question, to answer retrospective questions, or to think aloud about the next question). The interviewer demonstrated thinking aloud on a question not used in the cognitive laboratory.

Once participants understood what was instructed, each person was directed to click “continue” and begin the test section. The first observation was whether they appeared to read the section directions. When participants proceeded to the first question, they were stopped and asked retrospective questions about the section directions. They were also provided with a screenshot of those directions to help them provide full responses.

Next, participants began completing selected items from an untimed practice test. The focus of this portion of the cognitive laboratory was on participants’ engagement with the item types, directions, layout, and content. When participants first encountered a new item type, researchers used a think-aloud method to learn of their initial reactions. Before moving to the question, participants were instructed to vocalize all of their thoughts as they answered the next question (as if they were talking to themselves, not the interviewer). For the remaining selected items, participants were instructed to simply complete the items as they would in a regular testing situation. Semistructured, retrospective questions were asked immediately after the non-think-aloud items were completed in order to obtain feedback on participants’ experiences with the item types and features that they encountered. Participants were also asked for impressions of and feedback about the content and layout of directions at various points. As participants completed the test questions, interviewers documented behaviors and comments that indicated participants’ ease, difficulty, confusion, or frustrations with the tools, items, directions, or layout of the screen. Interviewers were directed to ignore any observations relating to testing strategy or difficulty related to limited content knowledge, as neither was the focus of this study. The protocol outlines describing the portion of the cognitive laboratory focused on the item types are included in Appendix D.

Results

Background Information Questionnaire (BIQ) Results

The sample consisted of 20 undergraduate sophomores and juniors from eight local colleges and universities (13 male and 7 female). The participants tended to be academically strong, with more than half listing honors, scholarships, and participation in academic clubs. They also indicated familiarity with technology. For example, when asked which types of computer applications they used, all listed using at least 3 different kinds and one used as many as 11. All of the participants used word processing and most used collaborative websites, social or professional networking sites, and spreadsheets. Most indicated that they were comfortable exploring, as needed, the features of software applications. Further demographics and responses to other background and experience questions are shown in Appendix C. As stated earlier, none of the background variables showed systematic relationships with the participants’ observations and responses during the cognitive laboratory.

Usability of the Test Environment Tools

This section addresses the following research question: How frequently do test takers use the optional tools (mark, review, help, and calculator), and what are their perceptions of the tools’ usefulness? To address this question, we observed test takers as they worked quietly (on the quantitative section in the QOVI condition and the verbal section in the VOVI condition). We recorded whether the participants used each tool. After the observation, using retrospective questions, we directly asked the participants about their tool use (e.g., whether they noticed them, why they did not use them). If test takers do not notice or use a tool, it brings into question that tool’s usefulness for completing the GRE.

Table 2 shows data gathered from the observations and retrospective questions. Only about half the participants reported noticing the mark, review, and help tools. Of those who did notice those tools, only a small proportion actually used them. In the QOVI condition, fewer than half of the participants noticed the on-screen calculator, and of those who did notice it (8 of the 11 in that condition) only five actually used it.

Table 2 suggests that many participants who noticed the mark, review, and help tools did not use them. The most common reason participants gave for not using these tools was uncertainty about the functionality of the tool. The pressure

Table 2 Tool Visibility and Usability

	Tool			
	Mark <i>n</i> = 20	Review <i>n</i> = 20	Help <i>n</i> = 20	Calculator <i>n</i> = 11 ^a
Number of participants reporting that they ...				
Did not notice the tool	11	10	10	3
Noticed the tool, but did not use it	8	9	10	3
Used the tool	1	1	0	5
Number of participants who did not use a tool because ...				
They were uncertain as to the tool's function	5	5	4	0
They said that they did not need the tool	4	5	5	0

^aOnly the participants in the QOVI condition (*n* = 11) completed a quantitative reasoning section for the tool use focus of the cognitive laboratory.

of a timed test also influenced the decision of some participants who opted not to use the tools. One participant remarked that in a timed testing situation, he would not want to experiment with new and unfamiliar tools. Four participants stated that they did not need to use these tools; most said it was because they felt confident of their answers or because the practice test was relatively short. However, three of these participants indicated that the mark and review tools might be useful if they felt less confident about their answers or if the test had been longer. Five participants reported not needing the help tool because they felt that the format of the test was self-explanatory. Of the tools that were used, the participants stated that they were at least moderately easy to use. Only one person stated that the mark feature was “least useful” (when asked to rate it on a 3-point scale from least to most useful).

Eight of the eleven participants who took the Quantitative Reasoning section of the timed practice test (the QOVI condition) reported noticing the on-screen calculator. Of these, five used it during the test, found it easy to use, and in spite of some criticisms, reported that it was helpful. A participant reported that he would have found the calculator useful had he seen it; another did not use it because he forgot about it after he started the test (he read about it in the section instructions). Two participants commented that the calculator would have been more useful if, like more advanced calculators, it displayed all of the calculations entered (the entire equation rather than the result only), or allowed the user to do two different things at once like a graphing calculator. All of the participants who used the calculator also used paper and pencil.

Although the icon to select the calculator was next to the mark, review, and help icons, two of the participants who reported noticing the calculator stated that they did not notice some or all of the other icons. The retrospective interview also revealed that some participants who noticed the help tool mistakenly believed that it would only provide them with information that would not be useful (i.e., low-level information). None of the participants used the help tool. After being shown the feature and acknowledging that it provided more information than they anticipated, some stated that they still would not use it during an actual test. Instead, they believed that they would eventually “figure out” what they needed to know. (Figure 1 displays some of the information provided on the help screen.)

How to Make the Tools More Noticeable

The participants made several suggestions for making the tools more noticeable to participants and for making their functions better known. For example, the tools might be repositioned from the top of the page (where the participants stated that they have little reason to look) to vertically along the side of the screen, which is where the participant’s eye is more likely to fall as he or she is reading the item. To learn about the functionality of a tool, test takers must select and explore it. To do so is time-consuming, and some participants were unwilling to do it. Participants suggested that the test or section instructions include screenshots of the icons to accompany the text explanations of each test environment tool. They also suggested using text with brief descriptions that would appear when the mouse hovered over the icon or using pop-up reminders of the existence of the tools.

Usability of the New Item Types

These results address the research question: If any, how severe are test takers’ misunderstandings, confusion, errors, and difficulties with the new item types or the ways they are displayed (format)? During the second half of the cognitive

laboratory session, we used a combination of qualitative techniques and focused on participants' engagement with the new item types. We observed participants working on the items, silently or while they spoke their thoughts aloud. We also asked retrospective questions that were open- and closed-ended about their initial reactions to the item types. Finally, we asked participants if they had any feedback about the items.

Quantitative Reasoning Item Types

In addition to observing participants work on the items, we also observed whether they appeared to read the section directions, question directions, or use any of the available tools. Participants at this stage in the cognitive laboratory were fully aware of all of the tools available. Nine participants were observed and interviewed completing the Quantitative Reasoning item types in the untimed practice test (the VOQI condition): quantitative comparison, numeric entry, and data interpretation.

The quantitative comparison item type requires test takers to consider two quantities (labeled Quantity A and Quantity B) and determine which quantity is larger, whether they are the same, or whether not enough information is provided to make that determination. Figure 1 provides a description of this item type. All nine participants who saw this item type admitted to being confused by some aspect of the item. Seven said that it took them a while to determine what the question was asking them to solve because there was no explicit question or directions. Five participants reported initial uncertainty about what the word "quantity" referred to or meant. One participant said it would have been clearer if there were equal signs following "Quantity A" and "Quantity B" rather than having the quantities listed underneath the underlined titles. Another felt that it would have been clearer if the options were written as whole sentences, rather than fragments (i.e., "Quantity A is greater than Quantity B" or vice versa).

The numeric entry item type provides test takers with one or more text boxes within which to enter the computed answer to the question. Two text boxes could be present to represent the numerator and denominator of a fraction. The participants reported no problems with the requirement to enter their answer rather than selecting it from a set of answer options. Only two stated that this type of question was new for them. None of the participants who used the calculator used the "transfer to display" feature for this question type. This feature allows test takers to click an icon on the calculator so that the value in the calculator display can be transferred automatically to the text box (only available when a single text box was displayed). (See Figure B3 for a screenshot of the on-screen calculator description from the help screen.)

The data interpretation sets contain a chart, a table, or both (usually in the top half of the screen) and two or more questions associated with the data. (Figure 2 shows both a chart and table.) Test takers must be able to know what data from the chart and table are needed to answer the question, whether computation of the data is required, and how to correctly compute the answer. Five of the nine participants interviewed felt that the amount of data presented and the spatial layout of the tables made it difficult to identify the key elements needed to solve the problem. Some suggestions, however, may have been related to the limited data interpretation skills of the participants. There were comments that the tables should be separated more and that all of the important information should be grouped together.

Verbal Reasoning Item Types

Eleven participants were interviewed and observed answering the Verbal Reasoning questions in the untimed practice test (the QOVI condition): select-in-passage, text completion, and sentence equivalence. They were also observed, using the think-aloud method, and interviewed about reading the instructions for specified questions. (Appendix A provides screenshots with examples of each item type.)

These participants completed the select-in-passage (sentence highlight) items that require test takers to answer the test question by clicking on a passage in order to select a portion of the text. Once clicked, the entire portion of the passage that can be selected is highlighted. (See Figure A4 for an example with an answer chosen.) There is no need to click and drag in order to make a selection. While being observed thinking aloud, four participants attempted to highlight the sentence by dragging. Three displayed signs of confusion as to how to answer the question. One interviewer observed that a participant "was looking around the screen" as if trying to determine what to do. Two clicked randomly at the screen and appeared surprised when a sentence was highlighted.

When interviewed, several participants provided additional feedback about the select-in-passage item type. Two stated that they did not know what to do to answer the question. Two stated that they had no problems highlighting the sentence,

IBT Client

GRE® Untimed Practice Test Section 3 of 5

Question 15 of 20

Questions 14 to 16 are based on the following data.

In a survey, 100 travel agents each ranked Airlines *A*, *B*, and *C* in order of preference. Each of the 100 travel agents also rated the three airlines in five categories on a scale of 1 through 10, with 10 being the best rating.

DISTRIBUTION OF RANKINGS

Note: The notation *ACB* means *A* ranked 1st, *C* ranked 2nd, and *B* ranked 3rd.

AVERAGE RATING

Category	Airline		
	<i>A</i>	<i>B</i>	<i>C</i>
Convenience	5.1	8.0	4.3
Friendliness	5.0	5.5	5.4
Price	5.0	6.4	3.5
Promptness	6.5	6.9	4.1
Reliability	7.8	7.5	4.9

Airline *B*'s average rating for convenience was approximately what percent greater than Airline *A*'s average rating for convenience?

30%
 36%
 40%
 57%
 64%

Click on your choice.

Figure 2 Data interpretation set with both a chart and a table. Descriptive text describes the data and a line separates the data from the question to be answered by the test taker. Some items display the data to the left of the screen, with a line separating the data from the question.

and another participant stated that he was glad that it required a click to select instead of dragging to highlight the sentence. Another participant reported noticing the “click to select” instructions but because it was at the bottom of the screen, stated that it was not helpful. One suggestion was to spatially separate the sentences that were available for selection to make more obvious what the answer options were.

The text completion items require the test taker to click on the word(s) that completes the sentence(s) in the short passage. The test taker is presented with sentences with blank lines. The short passage could have up to three blanks, with three or four words available for each blank. The words available for each blank are presented in columns labeled with lowercase Roman numerals that correspond to the Roman numerals above each blank line. (See Figure 3 for an example.)

Observations of the participants’ initial reactions on the text completion items include one participant’s attempt to drag and drop the word from a column into the sentence. During the observation and the think-aloud portions, no other participants displayed visible difficulty with this item type, regardless of how many blanks there were in the question. Some commented, however that there was some initial confusion. One stated that he expected to select the words by clicking ovals (referring to the description of the single-answer multiple-choice items in the section directions), so the rectangular columns confused him initially. Another said that he was not sure what to do until he read the directions for the question but was confident that he would have eventually realized it on his own. Three participants stated that the text completion item type was new to them but only because there was more than one blank to fill. One participant commented positively about the use of columns and Roman numerals. Suggestions that involved more interaction with the technology included using drop-down menus instead of columns; another was for the selected word to automatically appear in the sentence.

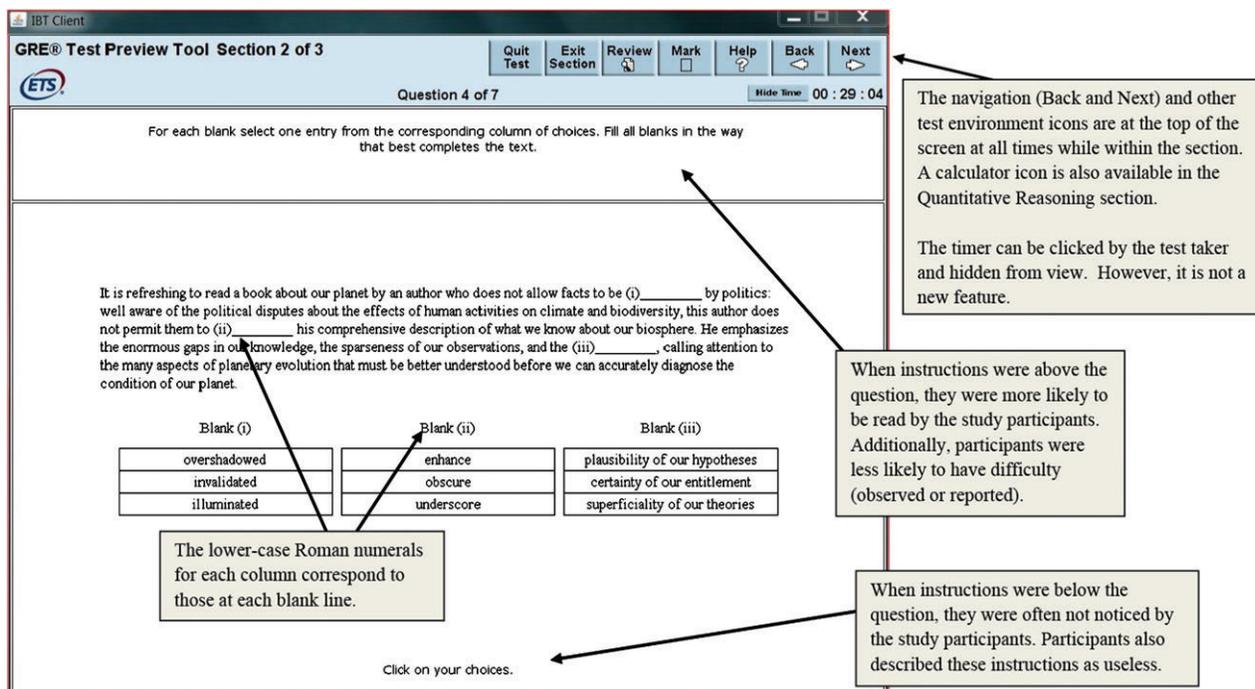


Figure 3 Layout of the test screen (sentence completion item type illustrated). Some features of the item type and the location of the test environment tools are illustrated. The different locations where instructions were frequently placed are also illustrated. (Not every item type has instructions in these locations.)

The sentence equivalence item type requires that test takers select exactly two words so that each would make the sentence have the same meaning, though the sentence has only one blank. There was some initial confusion or difficulty observed when five participants answered this item type. One of the participants selected three choices, instead of two. During the think-aloud portion, three read and reread the directions while verbally struggling with its meaning. One was very confused and stated that it was because the question had one blank but required two answers. Only 2 of the 11 stated that the sentence equivalence item type was not new.

When asked for feedback about the item, two stated that a single blank with two choices was confusing. One participant stated that he still was not certain whether the answers needed to be synonyms. Another believed that the test taker could choose more than one answer if he or she believed it fit in the sentence (instead of choosing exactly two answers). One participant admitted that she would have only selected one answer but because the word “two” was underlined in the directions, it caught her attention and made her read the full directions for the question. We also observed that if the computer system allowed only two options to be selected, some confusion (and incorrect responses) could be eliminated. Test takers are less likely to notice their error because more than two options can be selected and it looks like a multiple-answer multiple-choice question with square boxes.

Quantitative and Verbal Reasoning Sections

The multiple-answer multiple-choice item type was included in both sections of the test. This item type looks almost like a standard, single-answer multiple-choice item, except that the options have squares to click into instead of ovals. When clicked, an “X” appears in the square and it must be clicked again to be cleared. (See Figure A3 for an example.)

Fewer than half the participants from both conditions (7 of 20) stated that this type of item was new to them. None of these participants reported finding the format difficult, confusing, or hard to understand. However, when asked for feedback, one participant reported being so focused on the question and answer options that she did not see that the instructions were to select “all” that apply; although for her, this item type was not new.

Feedback About the Informational Text

The participants in both conditions were asked about the directions for the section and also for each of the questions. If the question had directions, the participants were also asked whether they noticed the directions as some directions were in two or more groups in different locations on the screen. (See Figure 3 for an example.) They were also asked whether they noticed any of the informational text (e.g., stating that a passage was to be used for a set of questions) as some were in smaller white font or placed in blue horizontal bars on the screen. Participants were also asked whether the directions or instructional text were clear or confusing in any way, in addition to having any other feedback.

Section Directions

Twelve of the twenty participants stated that they skimmed or read the directions before beginning either section of the untimed practice test. Of those who said they read it, only two admitted not reading it completely, and the remaining six read it completely. Every participant was provided a screenshot of the directions to refer to when asked for feedback. They all stated that the directions were clear and understandable. Nine, however, provided feedback for making the directions clearer. Some complained about the length of the directions or of certain paragraphs. One suggested adding more separations in the text. Others suggested more bolding; the participants who only skimmed the directions all read the bolded parts.

The paragraphs that were described as being the most challenging were those about the geometric figures not necessarily being drawn to scale. A participant stated that examples of what that meant would be helpful. Additionally, in the description of how single- versus multiple-answer multiple-choice questions were displayed, one student said it was unclear. He wondered whether the description of the presence of square boxes meant that there was definitely more than one answer, or if one answer was still possible. He said he felt that he was being tricked. Twelve other participants stated that the oval and square explanations (in bolded text) were very clear and straightforward. Some participants liked that the directions mentioned the tools, although they initially overlooked them while reading or skimming. One suggested providing screenshots in the directions that show where the icons are on the screen.

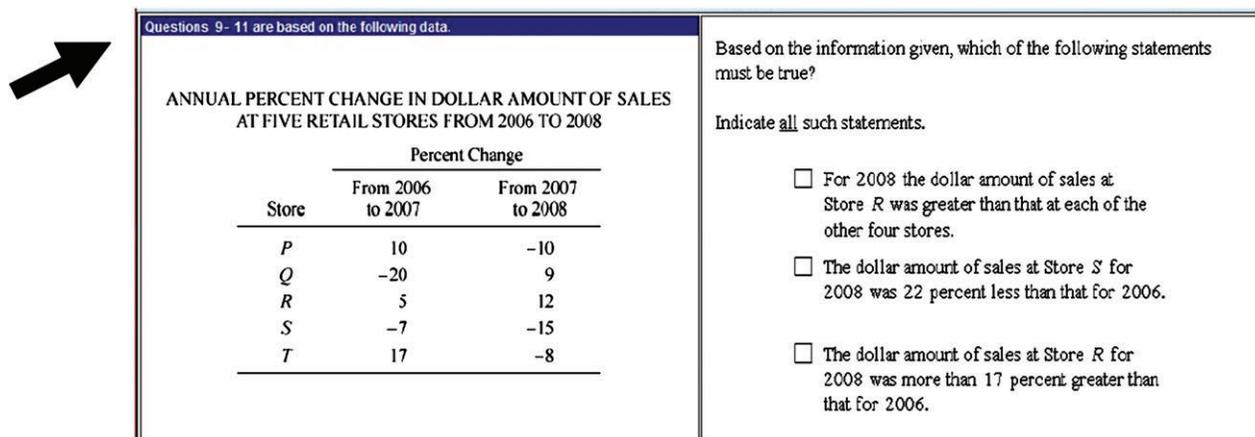
Placement of the Directions

When the instructions were placed above the question, participants noticed them regardless of the item type or test section (even if they did not read them). When the instructions were placed at the bottom of the screen, most participants did not notice them. One participant stated that the directions should be listed above the place where the test taker is expected to act. The directions at the bottom of the screen were generally described as being useless—either because the participants had already realized what to do by the time it was read, or because the material at the bottom of the screen restated the directions that were provided above the question. (See Figure 3 for an example.)

Information for Passages and Data

When a passage or data were used for a set of questions, a narrow blue bar with white text was present, stating which items corresponded with the passage or data (see Figure 4 for an example). Although the text was clearly stated, many participants did not notice the blue bar and realized that the passage or data were the same only when they advanced to the next question. Even when previously asked whether they noticed the blue bar, several participants still overlooked it on subsequent items. Participants commented that the smaller size of the text made it appear irrelevant and suggested that it be made to stand out more.

The feedback shared by the participants about the spatial layout, instructions, and informational text enrich our understanding of how future test takers will engage with the test. This insight offers a fuller description of how the new item types and tools can be perceived. In addition to providing new ways to assess quantitative or verbal reasoning, many of the innovations introduced in the test were provided as a benefit to test takers. Therefore, it is important that test takers understand how to properly respond to all of the item types, are aware of the presence and purpose of the test environment tools, and understand how to use the tools for their intended functions.



Questions 9-11 are based on the following data.

ANNUAL PERCENT CHANGE IN DOLLAR AMOUNT OF SALES
AT FIVE RETAIL STORES FROM 2006 TO 2008

Store	Percent Change	
	From 2006 to 2007	From 2007 to 2008
P	10	-10
Q	-20	9
R	5	12
S	-7	-15
T	17	-8

Based on the information given, which of the following statements must be true?

Indicate all such statements.

For 2008 the dollar amount of sales at Store R was greater than that at each of the other four stores.

The dollar amount of sales at Store S for 2008 was 22 percent less than that for 2006.

The dollar amount of sales at Store R for 2008 was more than 17 percent greater than that for 2006.

Figure 4 The blue bar contains informational text and is the only colored part of the screen. However, the bar is very thin and the font of the text is considerably smaller than that of the other text on the screen.

Discussion

The purpose of this study was to explore how test takers are likely to interact with the new item types and test environment tools on the computer-based GRE. Through observations, including the think-aloud method, and open- and closed-ended interviews, we learned a great deal about these innovations. In addition to learning how test takers interacted with these features, we also received feedback from our participants.

Test Environment Tools

The test environment tools that are available to test takers on the computer-based test were frequently unnoticed by the participants in this study. Often, the participants were unaware of what the functions of those tools were, and only a few were willing to explore them while taking the timed test. When the purpose and functions of the tools were explained, participants stated that they still may not have used some tools (e.g., help and review without mark), whereas others would have been useful (i.e., calculator and mark with review).

Although the test environment tools were available to assist test takers' navigation within each section, a few of the participants were not aware that they could move back and forth within a section. They only realized this after completing the section and reading the informational text informing them that they could go back and review their answers before continuing. Participants who were willing to move back and forth within the section mostly did so using the "back" and "next" icons, not the review tool. Some participants who marked items did so using the scrap paper and pencil that were available.

Item Types

For each of the item types, the participants eventually understood how to answer the question. There were some item types on which the participants hesitated before finally answering. Most noticeable was the quantitative comparison type of question. Several participants stated that they were not certain what the question was asking until reading the options. Additionally, some participants exhibited and later described confusion when they initially encountered the text selection items. However, once participants realized how to answer that type of question, they provided positive feedback about clicking, instead of dragging, to select the sentence that answers the question.

The multiple-answer multiple-choice items confused one participant only because he was not certain whether there was definitely more than one correct answer, or if only one correct answer was still a possibility. When initially encountering the question, the participants understood how to answer it because the answer choices displayed boxes instead of ovals—a distinction many were familiar with. Participants also had no difficulty answering the sentence equivalence questions where exactly two answers were required. The participants stated that if they forgot about the squares and ovals, the instructions clearly stated what to do.

The feedback concerning the format of the text completion items was positive as well. Participants generally liked that the options were separated in boxed columns and labeled with Roman numerals. The instructions and informational text (describing each table and figure) for the data interpretation sets also received positive feedback. Participants liked that there was a line separating the question instructions from the rest of the item. They also liked that there was text describing what was presented in the table or figure.

Instructional Text and Directions

In general, the participants stated that, when there were question directions, they were appropriate when the text was placed above the question or the options. The text at the bottom of the screen was frequently overlooked and described as not helpful, particularly because it was generally noticed after the participant had already answered the question. The text in the blue bar stating that passages or data sets corresponded to multiple questions was also frequently overlooked. Some participants stated that, whether they noticed it or not, it was not useful information because they would eventually figure it out. Those who did not notice the text stated that it should be larger; otherwise, it appears as if the text is not important information.

For most item types, whenever participants were uncertain how to respond, they read instructions and found them useful. Having words bolded or underlined in the instructions was especially important, particularly for the sentence equivalence items. The explanation of the boxes and ovals that contained bolding in the section directions also proved invaluable for the participants' understanding. The imprecise instructions for the select-in-passage items and lack of a stem or instructions in the quantitative comparison items, however, were problematic for many of the participants in the study. Although they eventually understood how to respond to those item types, their confusion (which was not content-related) could have cost them precious time during an actual test. Additionally, the instructions at the bottom of the screen and the blue bar containing a statement about the question set were described as "useless" by many participants.

Limitations and Future Research

A few limitations to this study should be noted. First, the sample was relatively small and not representative of the full GRE test-taking population. Perhaps due to the small sample size, there were few clear relationships between participants' background characteristics and their performance. Examining subgroup differences would better be addressed in a larger, more diverse sample. Furthermore, the students who chose to participate (those who completed both the BIQ and the study) were academically stronger than the remaining students who were recruited. Consequently, further research should be conducted to determine whether these results generalize to a larger, more diverse sample. Nevertheless, the observation that even academically talented individuals had some difficulties with the new item types suggests the importance of informative test preparation materials.

Second, the types of responses observed might have been limited because participants knew they were taking part in a study. Although we found similar results whether test takers talked aloud or were observed as they worked quietly, participants might not have tried as hard as they would for the real GRE. Test takers would also be more likely to familiarize themselves with the item types and the test environment tools before taking the real GRE. For example, some participants admitted that they did not use the test environment tools because they were taking a shortened practice test rather than being in a real testing situation. Further research might collect test taker response patterns as they complete a real GRE. For example, using technology that captures the active desktop and mouse movements, researchers could determine whether actual test takers are using the tools in a real testing situation. Combined with the background information already collected about test takers, potential relationships that exist could be determined. Researchers could also request that individuals from that testing complete an additional survey in order to learn about additional background characteristics (such as familiarity with various forms of technology) that may have a relationship to whether test takers notice and use the various tools.

Finally, specific hypotheses about improvements to the item types, test environment tools, or their instructions could be tested through more controlled experimentation or focused questioning rather than the relatively open-ended, qualitative methods used in this study. To gain general feedback regarding the format of the items, a questionnaire that included screenshots could be used to ask specific questions about the text size, the positioning of the instructions, and the clarity of the instructions. If a questionnaire with screenshots were provided at the conclusion of a POWERPREP practice test,

additional information could be gained relative to how noticeable the tools or informational text were. Researchers could also get general, self-reported feedback regarding each of the research questions included in the study.

Conclusion

The revisions to the GRE test include new item types that present a more varied way of measuring test takers' quantitative and verbal reasoning skills. However, the instructions for answering the questions appropriately must be clear to all test takers. This study provides insight regarding potential sources of construct-irrelevant variance that may cause test takers to lose time or answer questions partially and thus incorrectly. Providing instructions that are more noticeable and thorough may be all that is needed to alleviate confusion. To make the test environment tools more effective, however, test takers need to know, in advance of the timed test section, how to access the tools and their functions. The results of this study provide areas of improvement for the GRE revised General test. Once improvements are made to address those issues, the innovations to the new assessment will enhance the ways that all test takers respond to and are assessed by the test.

References

- Almond, P. J., Cameto, R., Johnstone, C. J., Laitusis, C., Lazarus, S., Nagle, K., ... Sato, E. (2009). *White paper: Cognitive interview methods in reading test design and development for alternate assessments based on modified academic achievement standards (AA-MAS)*. Dover, NH: Measured Progress.
- Bridgeman, B., Cline, F., & Levin, J. (2008). *Effects of calculator availability on GRE quantitative questions* (Research Report No. RR-08-31). Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (2011). *Frequently asked questions about the GRE revised General Test: Test content*. Retrieved from http://www.ets.org/gre/revised_general/faq/
- Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaluate test designs for student with disabilities and English language learners* (NCEO Technical Report No. 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- McKinley, R. L., Mills, C. N., Reese, C. M., Schaeffer, G. A., & Steffen, M. (1993). *Field test of a computer-based GRE General test* (GRE Board Professional Report No. 88-08P). Princeton, NJ: Educational Testing Service.
- Nielsen, J., & Mack, R. L. (1994). *Usability inspection methods*. New York, NY: Wiley & Sons.
- Nissan, S., & Schedl, M. A. (2012). Prototyping new item types. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 281–294). London, England: Routledge.
- Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. J. (2010). Innovative items for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 215–230). New York, NY: Springer.
- Stone, E., King, T., & Laitusis, C. C. (2011). *Examining the comparability of paper-based and computer-based administrations of novel item types: Verbal text completion and quantitative numeric entry forms* (Research Memorandum No. RM-11-03). Princeton, NJ: Educational Testing Service.

Appendix A

Examples of Item Types

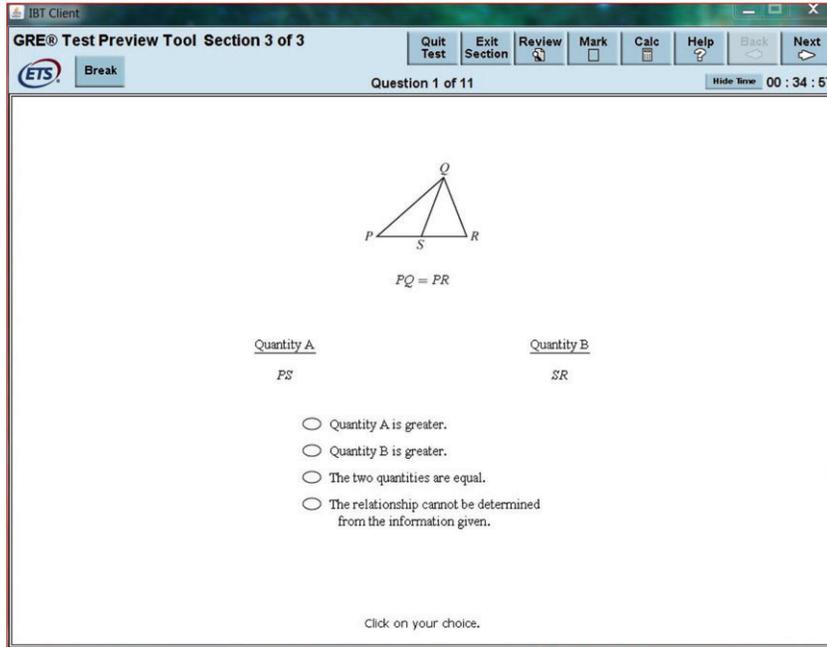


Figure A1 Quantitative comparison item type with single-answer multiple-choice (indicated by the ovals next to the answer choices).

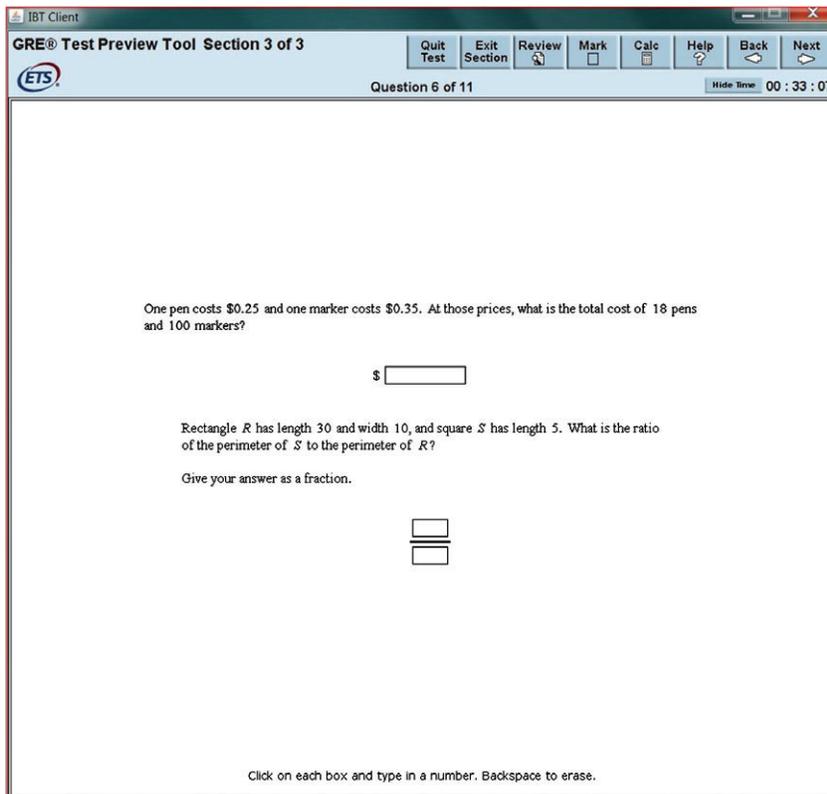


Figure A2 Numeric entry item type (two separate examples are shown together).

The screenshot shows the GRE Test Preview Tool interface for Section 3 of 3, Question 10 of 11. The interface includes a top navigation bar with buttons for 'Quit Test', 'Exit Section', 'Review', 'Mark', 'Calc', 'Help', 'Back', and 'Next'. A 'Break' button is also visible. The question text reads: 'Based on the information given, which of the following statements must be true? Indicate all such statements.' Below the text is a table titled 'ANNUAL PERCENT CHANGE IN DOLLAR AMOUNT OF SALES AT FIVE RETAIL STORES FROM 2006 TO 2008'. The table has three columns: 'Store', 'From 2006 to 2007', and 'From 2007 to 2008'. The data is as follows:

Store	From 2006 to 2007	From 2007 to 2008
P	10	-10
Q	-20	9
R	5	12
S	-7	-15
T	17	-8

Below the table are three multiple-choice options, each with a square checkbox:

- For 2008 the dollar amount of sales at Store R was greater than that at each of the other four stores.
- The dollar amount of sales at Store S for 2008 was 22 percent less than that for 2006.
- The dollar amount of sales at Store R for 2008 was more than 17 percent greater than that for 2006.

 At the bottom of the question area, it says 'Click on your choice(s)'.

Figure A3 Multiple-answer multiple-choice item in a data interpretation set. The square boxes next to the answer choices indicate that there are multiple options available. Otherwise, there would be ovals next to the answer choices.

The screenshot shows the GRE Test Preview Tool interface for Section 2 of 3, Question 3 of 7. The interface includes a top navigation bar with buttons for 'Quit Test', 'Exit Section', 'Review', 'Mark', 'Help', 'Back', and 'Next'. A 'Break' button is also visible. The question text reads: 'Select the sentence that distinguishes two ways of integrating rock and classical music.' Below the text is a passage about Philip Glass's music. A blue bar highlights a sentence in the passage: 'Popular elements do not appear out of place in Glass's classical music, which from its early days has shared certain harmonies and rhythms with rock music.' Below the passage is a single-choice option with a radio button: 'Click in the passage to highlight the sentence of your choice.'

Figure A4 Select-in-passage item type demonstrated with selected text. The blue bar of informational text is displayed for all sets of questions.

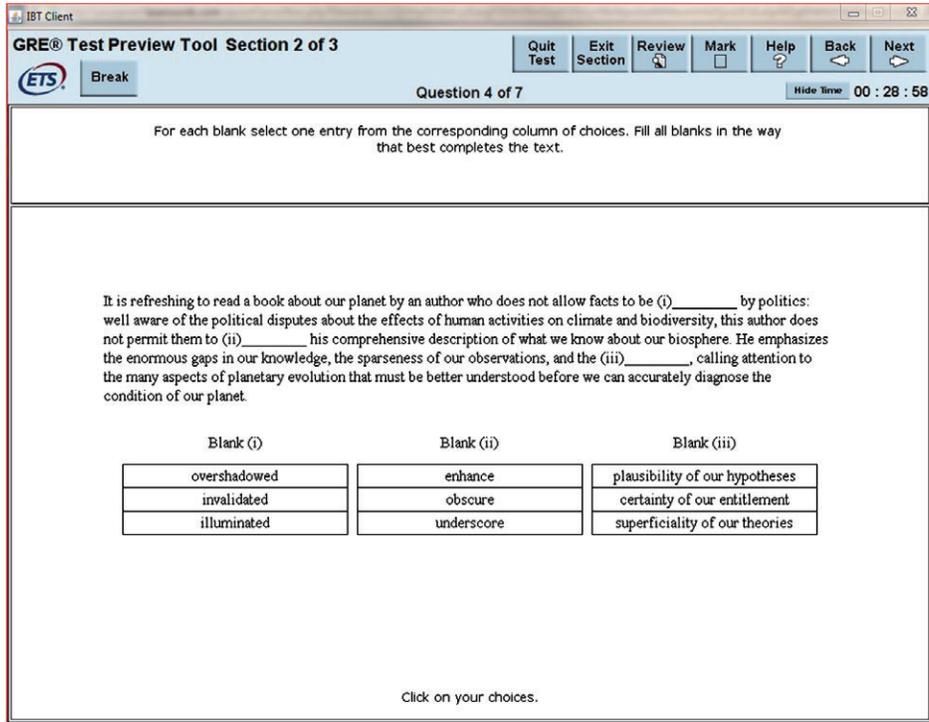


Figure A5 Text completion item type. This item type can have one, two, or three blanks with up to four options per blank.

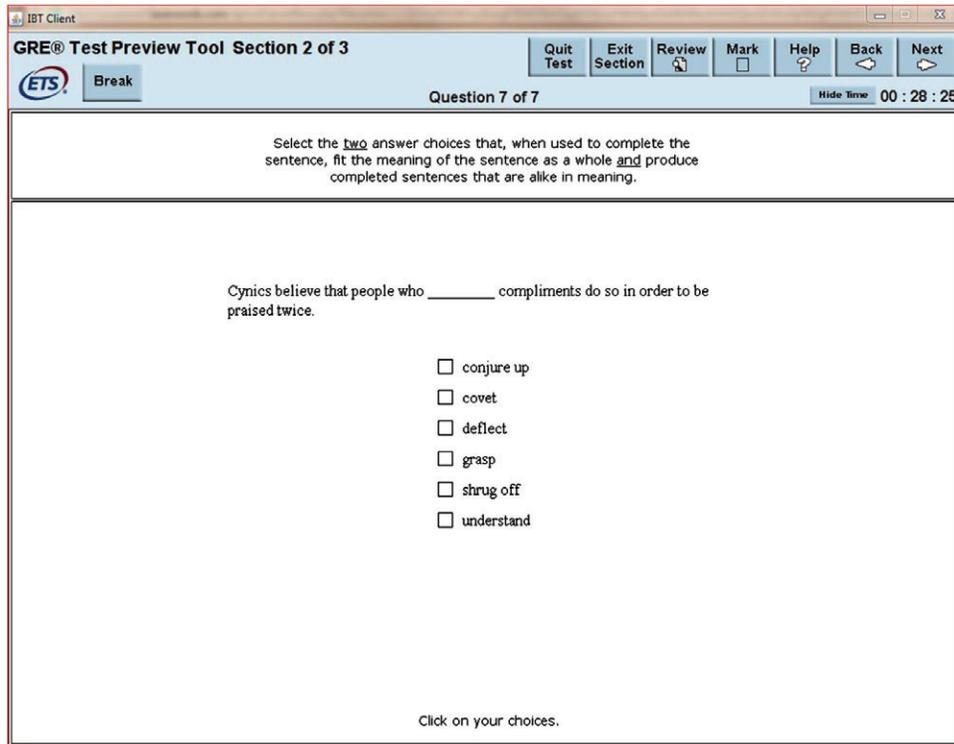


Figure A6 Sentence equivalence item type. There are squares to select because exactly two answers must be chosen.

Appendix B

Examples of Test Environment Tools

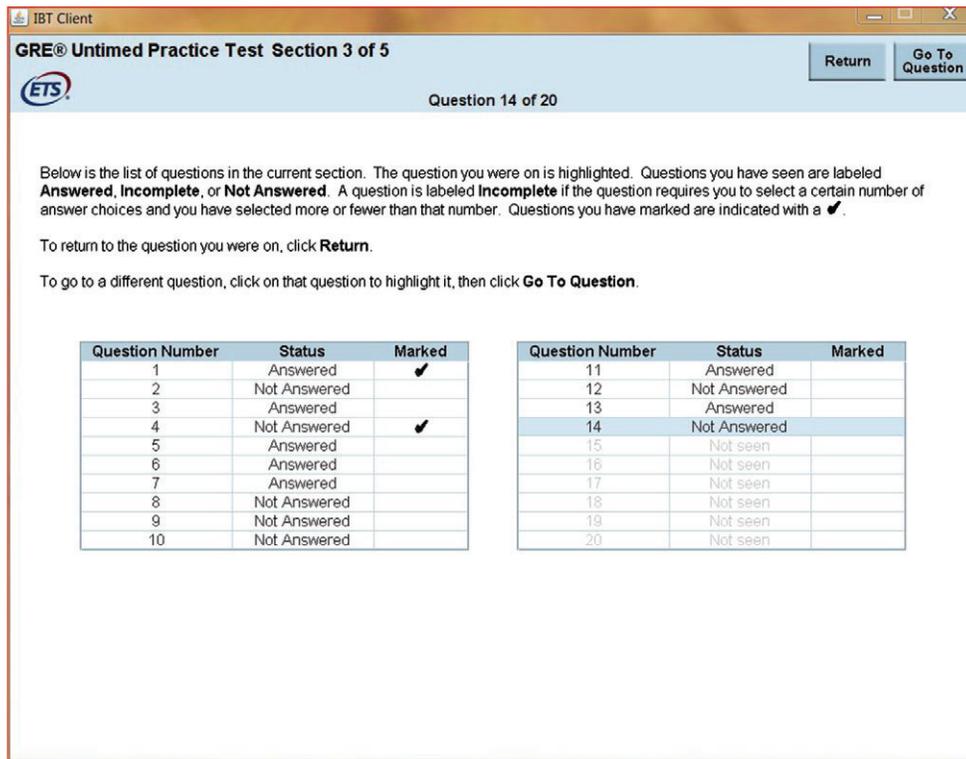


Figure B1 Mark and review screen. Questions 1 to 14 have been seen in this example. Some have been marked and/or answered.

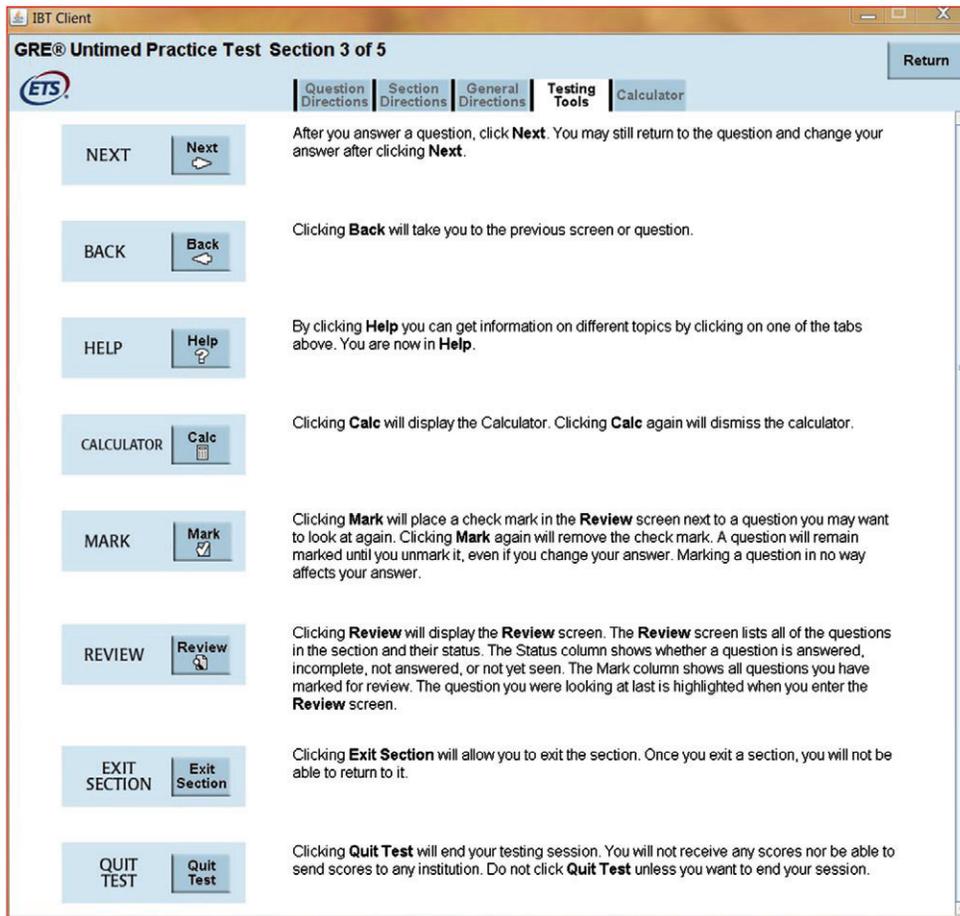


Figure B2 The test environment tools described within the help screen. The help screen offers tabs along the top on which test takers can click to access a variety of information. The first tab displays the direction for the item type that the test taker was looking at before clicking the help icon. The section directions (Quantitative or Verbal) can also be displayed. General test directions (e.g., what is and is not allowed during the testing session) can be read as well. In the Quantitative Reasoning section, a tab that describes the functionality of the on-screen calculator is also included.

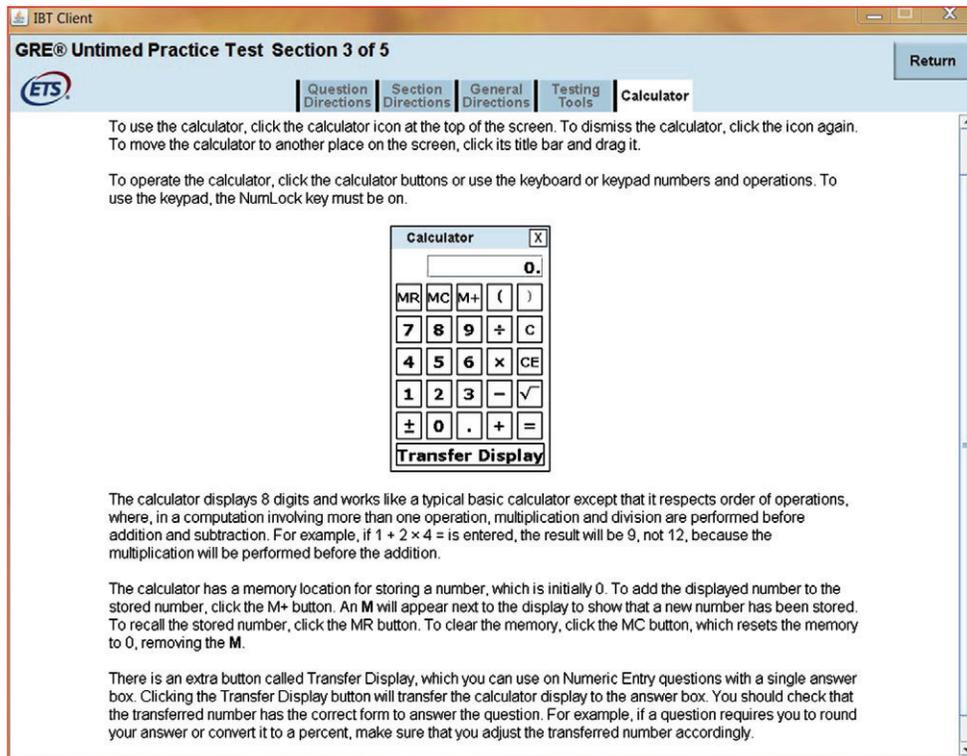


Figure B3 Calculator description and instructions from the help screen. When selected, the calculator appears and can be moved around the screen by the test taker.

Appendix C

Background Information Questionnaire With Results

GRE® Cognitive Lab Study Background Questionnaire

Introduction

Thank you for agreeing to participate in this study. Please complete this survey prior to us scheduling you for your participation in the study. The completeness of your answers is important.

This survey will take approximately 10 minutes to complete.

If you have any questions regarding the survey or the study, please contact Wanda Swiggett at WSwiggett@ets.org. Questions about the study may also be directed to Maria Elena Oliveri at MOliveri@ets.org.

ETS launched the GRE revised General Test on August 1, 2011 (**the new GRE**). From that date to now, it is the only available version. Before August 1, 2011, the only available GRE test was the previous version (**the old GRE**).

- **What is your most recent experience with the GRE test?**
You have: (*select one*)
 - (0) taken the old GRE
 - (0) taken the new GRE
 - (20) not taken any version of the GRE
- **What is your most recent experience studying for the GRE test?**
You have: (*select one*)
 - (19) not studied for any version of the GRE
 - (0) studied for the old GRE using test prep materials published by ETS (either free or purchased)
 - (1) studied for the new GRE using test prep materials published by ETS (either free or purchased)
- **Please describe your plans for taking the GRE test.**
You: (*select one*)
 - (1) will not take the GRE test
 - (11) plan to take the GRE test within the next two years
 - (0) plan to take the GRE test more than two years from now
 - (8) do not know when or if you will take the GRE test
- **How do you think you would perform on the test (with appropriate studying and preparation)?**
(*A percentile indicates your rank compared to other students. For example, a score in the 50th percentile means that 50% of the other students scored below you.*)

Please select the percentile within which you believe you would be placed.

Percentile Prediction Responses	Quantitative Reasoning	Verbal Reasoning
90 th – 99 th	5	3
80 th – 89 th	9	11
70 th – 79 th	1	1
60 th – 69 th	2	2
50 th – 59 th	3	1
30 th – 39 th		1
no answer		1

- How frequently do you use the following type(s) of computer application(s)?

Frequency, Median, and Mode of Computer Application Use

	Rarely	Occasionally	Often	Very often	Med (1–4)	Mod (1–4)
Word processing (e.g., Pages, Word)		2	5	13	4	4
Spreadsheets (e.g., Excel, Numbers)	1	5	3	6	3	4
Publishing (e.g., Adobe, Publisher)	2	2		3	2	4
Database (e.g., Access, FileMaker)	1	1	2	1	3	3
Bibliographic software (e.g., Endnote, RefWorks)		2	1		2	2
Collaborative websites (e.g., Blackboard, GoogleDocs)	1	1	7	7	3	—
Social or professional networking (e.g., Twitter, LinkedIn)		1	2	13	4	4
Interactive video games (multiple players over an Internet connection)		3	5	2	3	3
Video games (single player, Internet may not be needed)	1	2	3	2	3	3
Applications specific to your field of study (e.g., for research or design)			7	2	3	3
Other computer application (please list one):			3	1	3	3

- How do you use the features of the following type(s) of computer application(s)?

Frequency, Median, and Mode of Behavior With Computer Applications

	Learn the basics about the features I need	Master the features I need	Master the features I need & explore related features	Master the features I need & explore some advanced features I may not need	Master many of the features whether I need to use them or not	Med (1–5)	Mod (1–5)
Word processing		6	4	5	5	4	2
Spreadsheets	4	5	2	2	2	2	2
Publishing	1	2	2	2	2	3	—
Database	2		2	1		3	—
Bibliographic software		2	1			2	2
Collaborative websites	2	4	5	3	2	3	3
Social or professional networking		2	4	3	7	4	5
Interactive video games		1	1	4	4	4	—
Video games		1	1	3	3	4	—
Applications specific to your field of study			1	5	3	4	4
Other computer application (please list one):			2	2		4	—

- What do you do when features change with newer software versions?

Frequency, Median, and Mode of Behavior With New Versions of Computer Applications

	Relearn the basics about the features I need	Remaster the features I need	Remaster the features I need & explore related features	Remaster the features I need & explore a variety of other features	Master many of the features, whether I need to use them or not	Med (1–5)	Mod (1–5)
Word processing	1	6	5	4	4	3	2
Spreadsheets	5	6	1	2	1	2	2
Publishing		3	2	2		3	2
Database	1		3	1		3	3
Bibliographic software		2	1			2	2
Collaborative websites	1	7	5	1	2	3	2
Social or professional networking		5	3	2	6	4	5
Interactive video games		3		5	2	4	4
Video games		3		4	1	4	4
Applications specific to your field of study	1		1	4	3	4	4
Other computer application (please list one):			2	2		4	—

- Please indicate your gender.
 - (13) Male
 - (7) Female
- Please indicate your ethnicity.
 - (11) White/Caucasian
 - (1) Black/African-American
 - (1) Hispanic/Latino
 - (3) Asian
 - (0) Native American/Alaskan Native
 - (0) Pacific Islander
 - (2) Other/Multi-Racial: (a) Italian, Irish, and German; and (b) Japanese, German, and Italian
 - (1) Decline to Respond
- What is your native country?
 - United States ($n = 17$), Canada ($n = 1$), Guatemala ($n = 1$), and India ($n = 1$)

If you are not a U.S. native, how long have you lived in the U.S.? (Whole numbers only)

Time Lived in U.S.

Country	Years	Months
Canada	28	3
Guatemala	4	0
India	14	

- **What is your major?**
 () Please write your major(s): _____ () Currently undecided

Majors Listed

Accounting (<i>n</i> = 1)	Government (<i>n</i> = 1)
Broadcast Journalism (<i>n</i> = 1)	Kinesiology (<i>n</i> = 3)
Business Administration (<i>n</i> = 1)	Mechanical Engineering (<i>n</i> = 1)
Chemistry (<i>n</i> = 1)	Music Industry (<i>n</i> = 3)
Civil Engineering (<i>n</i> = 1)	Pre-medical (<i>n</i> = 1)
Electrical Engineering/Technology (<i>n</i> = 3)	Psychology (<i>n</i> = 1)
Finance (<i>n</i> = 1)	Undecided (<i>n</i> = 1)

- **What is your minor?**
 () Please write your minor(s): _____ () I don't currently have a minor

Minors Listed

Astronomy (<i>n</i> = 1)
Dance (<i>n</i> = 1)
International Relations ^a (<i>n</i> = 1)
Math (<i>n</i> = 2)
Middle Eastern Studies ^a (<i>n</i> = 1)
Physics (<i>n</i> = 1)
No current minor (<i>n</i> = 14)

^aOne participant listed both International Relations and Middle Eastern Studies.

- **GPA**
 What is your GPA?: _____
Is your GPA on a 4.0 or 5.0 scale?
 (20) 4.0
 (0) 5.0
 (0) Other: _____

Grade Point Averages Listed

GPA	<i>n</i>
2.90	2
3.00	3
3.10	1
3.46	1
3.50	5
3.56	1
3.68	2
3.72	1
3.74	1
3.76	1
3.80	1
3.85	1

Please list your academic clubs and honors. (Please do not include social or service organizations).

Clubs and Honors Listed

Academic clubs	<i>n</i>	Academic honors	<i>n</i>
Little Investment Bankers	1	Dean's List	6
Society of Women Engineers	1	Recipient of the STEM-ESP Scholarship	1
Science and Math Outreach Club	1	Honor's Director's List	1
Pre Physical Therapy Association	1	Phi Theta Kappa Honors Society	1
National Society of Colligate Society	1	Letters of recognition (one from the Dean)	1
Diamond Gems Dance Team*	1	Society of Women Engineers, American Society of Civil Engineers	1
Guitar Ensemble ^a	1	Little Black Book Publications	1
Rock Ensemble ^a	1	Undergraduate Black Wharton Association	1

Notes. Nonacademic clubs and honors as well as those earned in high school were omitted from this list and not included in the data analysis. Scholarships awarded in high school for college were included.

^aThese students' majors or minors were related to dance or music. Therefore, they were included.

Please list your name so that we may contact you. We will use the e-mail address you provided on the consent form.

(Please use the same name you entered on the consent form)

Name _____

Thank You!

Thank you for taking our survey. If you've completed the survey, we will contact you within 3 business days to schedule your visit to ETS. When you arrive, your role as a research participant will be further explained.

We will use the contact information provided on the consent form. If you have any updates to your contact information, please e-mail that information to wswiggett@ets.org or to Moliveri@ets.org.

Additional Background Information (Collected during Cognitive Laboratory)
College or University Attended

College or University	<i>n</i>
Camden Community College	1
Drexel University	5
New Jersey Institute of Technology	1
Raritan Community College	1
Rowan University	2
Rutgers University	5
Temple University	4
Wesleyan University	1

Appendix D

Protocol Outlines

These describe items selected for the cognitive lab section focused on the item types.

Table D1 Quantitative Observation and Verbal Interview (QOVI) Protocol Outline (Using the Quantitative Reasoning Section)

Question #	Format type	Think aloud	Retrospective question	Comment
All	Verbal section of test preview		Observational notes	Look for tool use, ease or difficulty with item types, appearing to read or skip instructions
	Section directions		Tool use questions	Observe if they appear to skip or read it.
1	Quantitative Comparison	Think aloud	Question format, format directions, etc.	1. Cover with paper and ask “Section Directions” questions. 2. After think aloud, complete retrospective questions.
10	Skip Questions 2–9 Single-answer multiple-choice (horizontal)			
11	Single-answer multiple-choice (vertical)			
12	Numeric entry	Think aloud	Question format, inc. horizontal versus vertical format directions, etc.	Directions, same passage, format change
14	Skip Question 13 Dataset numeric entry			
15	Dataset 1 single-answer multiple-choice		Question format, format directions, etc.	Directions, same passage, format change
18	Skip Questions 16–17 Numeric entry			
19	Multi-option multiple-choice Skip Question 20	Think aloud	Item format questions Open-ended feedback	

Table D2 Quantitative Observation and Verbal Interview (QOVI) Protocol Outline (Using the Verbal Reasoning Section)

Question #	Format type	Think aloud	Retrospective question	Comment
All	Quantitative section of test preview		Observational notes	Look for tool use, ease or difficulty with item types, appearing to read or skip instructions
	Section directions		Tool use questions	Observe if they appear to skip or read it.
	Skip Question 1			1. Cover with paper and ask “Section Directions” questions 2. Move on to Question 2
2	Verbal text completion 1 blank			
3	Verbal text completion 2 blanks			
4	Verbal text completion 3 blanks	Think aloud	Question format, format directions, etc.	
	Skip Questions 5–6			

Table D2 Continued

Question #	Format type	Think aloud	Retrospective question	Comment
7	Passage 1 single-answer multiple-choice			
8	Passage 1 highlight sentence	Think aloud	Question format, format directions, blue info bar, etc.	
10	Skip Question 9 Passage 2 multi-option multiple-choice	Think aloud	Question format, format directions, etc.	
13	Skip Questions 11 – 12 Verbal sentence equivalence	Think aloud	Question format, format directions, etc.	Observe if they read/understood the directions (only 2 answers).
	Skip remaining questions		Open-ended feedback	

Suggested citation:

Swiggett, W. D., Kotloff, L., Ezzo, C., Adler, R., & Oliveri, M. E. (2014). *A usability of interactive item types and tools introduced in the New GRE® revised General Test* (GRE Board Research Report No. 14-05, ETS Research Report No. RR-14-28). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12028

Action Editor: Brent Bridgeman

This report was reviewed by the GRE Technical Advisory Committee and the Research Committee and Diversity, Equity and Inclusion Committee of the GRE Board.

ETS, the ETS logo, GRADUATE RECORD EXAMINATIONS, GRE, LISTENING. LEARNING. LEADING., and POWERPREP are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at <http://search.ets.org/researcher/>