Spring 2018

Evaluation of Paper Test Strategies on Computer-Administered Tests: The Impact of Item Response Marking on Test Performance

Erik Ekberg
Central Washington University, erik.ekberg@cwu.edu

EVALUATION OF PAPER TEST STRATEGIES ON COMPUTER-ADMINISTERED TESTS: THE IMPACT OF ITEM RESPONSE MARKING ON TEST PERFORMANCE

A Thesis

Presented to

The Graduate Faculty

Central Washington University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

Experimental Psychology

---

by

Erik Allan Ekberg

May 2018

CENTRAL WASHINGTON UNIVERSITY

Graduate Studies

We hereby approve the thesis of

Erik Allan Ekberg

Candidate for the degree of Master of Science

APPROVED FOR THE GRADUATE FACULTY

______________ ______________________________
Dr. Tonya Buchanan, Committee Chair

______________ ______________________________
Dr. Susan Lonborg

______________ ______________________________
Dr. Kara Gabriel

______________ ______________________________
Dean of Graduate Studies

ABSTRACT

EVALUATION OF PAPER TEST STRATEGIES ON COMPUTER-ADMINISTERED TESTS: THE IMPACT OF ITEM RESPONSE MARKING ON TEST PERFORMANCE

by

Erik Allan Ekberg

May 2018

Marking on paper tests (e.g., crossing-out incorrect answers) has been associated with improved test performance, especially for students with higher test anxiety. Despite these benefits, marking has not been implemented or evaluated on computer-administered tests. After measuring test anxiety in participants using an adapted version of the Test Anxiety Inventory Short-form (TAI-5C), we randomly assigned participants to either the required-mark (n = 85) or control (n = 88) condition and measured test performance on 18 practice Scholastic Aptitude Test (SAT) math questions. A multiple regression was performed to model test performance from test anxiety, condition group, and their interaction, $R^2 = .08$, $F(3, 169) = 5.17$, $p < .01$. We found that only test anxiety impacted test performance significantly, β = -0.06, $p < .01$. We discuss these findings, the limitations of our study, and suggest alternatives which may enhance the test taker experience during computer-administered tests.

Keywords: Marking, marking-functionality, computer-administered test, test anxiety, test strategies

ACKNOWLEDGMENTS

Thank you Khan Academy for granting me permission to use your materials for non-commercial purposes to conduct my research. Also, thank you Taylor and Deane (2002) for permitting me to use your modified version of the Test Anxiety Inventory to evaluate test anxiety in students.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

Figure Page

CHAPTER I

INTRODUCTION

At least 10% of college students have high enough test anxiety that it negatively impacts their performance on tests (Szafranski, Barrera, & Norton, 2012). To help these students perform equivalently to their lower anxiety peers, researchers have examined different testing strategies and their impact on performance. These researchers found that writing comments next to questions (Calvin, McGuigan, & Sullivan, 1957; McKeachie, Pollie, & Speisman, 1955; Smith & Rockett, 1958), eliminating possible responses with marks (Herman, 1996), and taking one minute to look over all questions before starting a test (Mavilidi, Hoogerheide, & Pass, 2014) are all associated with improved performance in students. Furthermore, these strategies appear to be most useful to students with high test anxiety (Calvin et al., 1957; Smith & Rockett, 1958). But despite the benefits of these strategies, the rise of computer-administered testing in the classroom and on standardized tests have made these traditional strategies unusable (Macedo-Rouet, Ney, Charles, & Lallich-Boidin, 2009).

These traditional strategies have been researched, developed, and evaluated exclusively on paper tests and require students to interact with tests outside of strictly answering questions. For example, McKeachie et al. (1955), Calvin et al. (1957), and Smith and Rockett (1958) found that allocating space on classroom exams for students to write comments improved test performance. Nield and Wintre (1986) followed up this research and found that students preferred multiple-choice questions with the option to elaborate (i.e., write comments) about their responses. Research that clearly demonstrates the importance of students' ability to interact with tests.

Building further on this research, Herman (1996) showed that being able to mark test item responses on multiple-choice tests (e.g., crossing out incorrect answers) also improved performance in students in comparison to students who were not afforded the option to mark. However, these strategies again require students to physically write on the test, a feature of paper tests not currently permitted on computer-administered tests. Consequently, many of these beneficial testing strategies are unemployable on current computer-administer testing software (CTS; Macedo-Rouet et al., 2009). Therefore, the current study evaluated the impact of increased interactivity between students and CTS through item response marking (IRM) as defined by Herman (1996).

CHAPTER II

LITERATURE REVIEW

Test Anxiety and Test Performance

Test anxiety is associated with reduced test performance (Ashcraft, 1988; Calvin et al., 1957; Hong, Sas, & Sas, 2006; Mavilidi et al., 2014; Osterhouse, 1975; Sarason, Mandler, & Craighill, 1952; Szafranski et al., 2012), particularly when there are consequences linked with that performance (Mavilidi et al., 2014; Sarason et al., 1952). When bad or good outcomes are associated with poor performance, students find themselves more prone to intrusive cognitions, like low self-confidence (Mavilidi et al., 2014) and negative social comparisons (Zatz & Chassin, 1985), which increase their cognitive load and inhibit their ability to focus on test content (Ashcraft, 1988; Mavilidi et al., 2014; Szafranski et al., 2012). A model conceptually supported by researchers demonstrating that high test anxiety students need more time to finish exams and need to exert more effort to achieve similar performance to their low test anxiety peers (Mavilidi et al., 2014). Furthermore, Sarason et al. (1952) found that test anxiety does not negatively impact performance as much in non-instructional (low stakes) tests when compared to instructional (high stakes) tests. Nevertheless, given the impact of test anxiety on test performance, test anxiety levels need to be accounted for when evaluating the impact of IRM on CTS.

Traditional Strategies and Test Anxiety Levels

Strategies used by students are abstract, however, they produce artifacts on paper tests called marks (Herman, 1996; Kim & Goetz, 1993). Marks usually represent option elimination, answers selection (Herman, 1996; Kim & Goetz, 1993), and elaboration in the form of figures, equations, notes (Herman, 1996; Hong et al., 2006; Kim & Goetz, 1993), and comments on test pages (Calvin et al., 1957; Herman, 1996; McKeachie et al., 1955; Smith & Rockett, 1958). Marking behavior is thought to serve as an outlet for test anxiety (McKeachie et al., 1955) and a way for students to think more critically about test content (Herman, 1996). Although no explicit model or theory has been evaluated in the literature, researchers have shown that marking behavior does not benefit all students.

Marking behavior has been demonstrated to help high test anxiety students the most (Calvin et al., 1957; Smith & Rockett, 1958). McKeachie et al. (1955) found that forcing students to elaborate on their responses did not improve performance in general and actually decreased performance in certain conditions; particularly when low test anxiety students are forced to elaborate in low stakes testing situations (Smith & Rockett, 1958). But when taking test anxiety into account, Calvin et al. (1957) and Smith and Rockett (1958) found that students with high test anxiety benefited the most from elaboration. Furthermore, Herman (1996) found that the type of marking behavior, operationally defined as option elimination, answer selection, or elaboration, was not predictive of test performance, and rather that the absence or presence of marking behavior was more predictive of performance differences. Consequently, it is important to measure the amount of test anxiety a student experiences to properly evaluate the impact of IRM on performance.

Current Study

Marking behavior is associated with improved test performance on paper tests (Calvin et al., 1957; Herman, 1996; McKeachie et al., 1955). However, current CTS does not permit marking behavior. With the rise of CTS in the classroom and on standardized tests (Macedo-Rouet et al., 2009), marking may be a beneficial feature adopted onto CTS to improve the student experience. This increased interactivity may be especially beneficial for high test anxiety students who benefit the most from marking behavior (Calvin et al., 1957; Smith & Rockett, 1958) while other traditional strategies become unusable on CTS (Macedo-Rouet et al., 2009). Therefore, the purpose of this study was to evaluate the impact of forced marking behavior on student test performance with respect to test anxiety on CTS during a low stakes test.

Hypothesis 1. Given that high test anxiety negatively impacts student performance (Ashcraft, 1988; Calvin et al., 1957; Hong et al., 2006; Mavilidi et al., 2014; Osterhouse, 1975; Sarason et al., 1952; Szafranski et al., 2012) we hypothesized that as test anxiety increased performance would decrease for students who cannot mark the CTS.

Hypothesis 2. We further notice that marking behavior is beneficial to high test anxiety students (Calvin et al., 1957; Smith & Rockett, 1958) while decreasing performance in low test anxiety students during low stake tests (Smith & Rockett, 1958). Given this interaction, we hypothesized that if students were forced to mark the CTS then as test anxiety increased performance would increase.

CHAPTER III

METHODS

Design

This study evaluated the impact of forced marking behavior and test anxiety on test performance using a between subjects design. In this study, marking behavior was defined by usage of the item response marking (IRM) functionality provided by the CTS. The first factor of this study was the presence and usage of IRM on the CTS which had two levels: the control and required-mark condition. In the control condition, IRM was not present on any test questions, so all test questions were displayed to participants in a similar manner to current CTS. In the required-mark condition, IRM was present and participants were required to make at least two marks on each of the designated questions within the study (forced IRM). The second factor was test anxiety experienced by participants and the dependent variable was test performance defined by the number of correct responses on an objective multiple-choice test (OMCT).

Participants

A convenience sample of 204 undergraduate psychology students from Central Washington University (CWU) were recruited through the CWU Department of Psychology SONA system and were compensated with 1.5 extra credit points in one of their psychology courses in which they were enrolled. Because we wanted to examine marking behavior on computer-administer tests, we restricted eligibility to students who (a) had access to a computer device with access to the Internet, and (b) were at least 18 years of age. Also, participants who failed to completed the test anxiety measure used in the study were removed from the dataset prior to analysis. Additionally, all participants who did not at least view the demographic section of this study, because they had closed their internet browser either before or during the OMCT, were removed from the dataset prior to analysis. Participants who failed any of these checks (n=31) were excluded from any further analyses.

Of the remaining 173 participants (124 females, 46 males, 1 other, and 2 undisclosed), participants aged 60 years to 18 year (M=20.75, SD=5.49) described themselves predominantly as Caucasian or White (69%), then Hispanic or Latino (11%), Asian or Pacific Islander (8%), or Other (8%), and lastly as Black or African American (3%), or chose not to disclose their ethnicity (1%). Also, participants disclosed varied levels of math education in pre-algebra or below (8%), introductory algebra (12%) intermediate algebra (36%), pre-calculus or higher (45%), or chose not to disclose their math education (1%).

Materials

Computer Testing Software. Since no current CTS provides IRM, a CTS was developed using the Ruby on Rails framework using MathJax, JQuery, and HTML to render content on participants' computer devices. To make the CTS accessible to participants, the CTS was hosted on Amazon Web Services (AWS) using their Elastic Beanstalk platform on at least two Elastic Computer 2 instances to prevent down time in case of single instance failure. Furthermore, participant data were collected and stored through the CTS using HTTPS to an encrypted and password protected PostgreSQL database also hosted on AWS.

CTS Item Format. The layout and styling of all CTS questions in this study followed a similar layout and style to the Scholastic Aptitude Test (SAT) CTS (Heimbach, 2015). By default the CTS provides a radio-button per response, meaning that only a single final response may be submitted by a participant (see Figure 1). However, for open-ended questions a text-box appeared instead (see Figure 2).

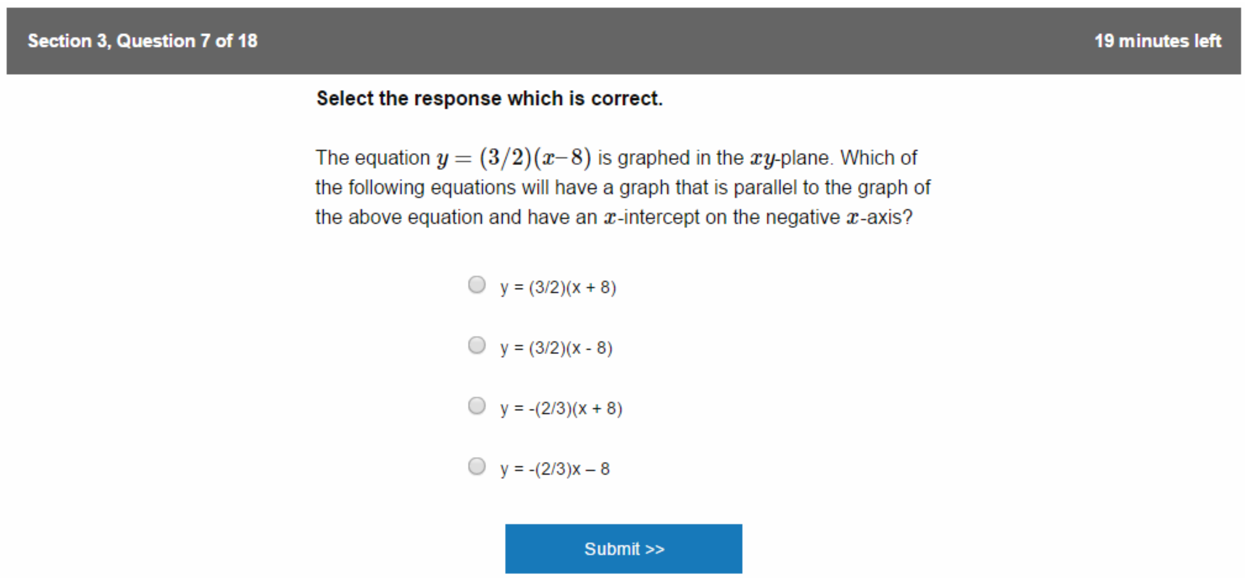


Figure 1. Example of an OMCT question and how the CTS displayed radio-button questions in the control condition.

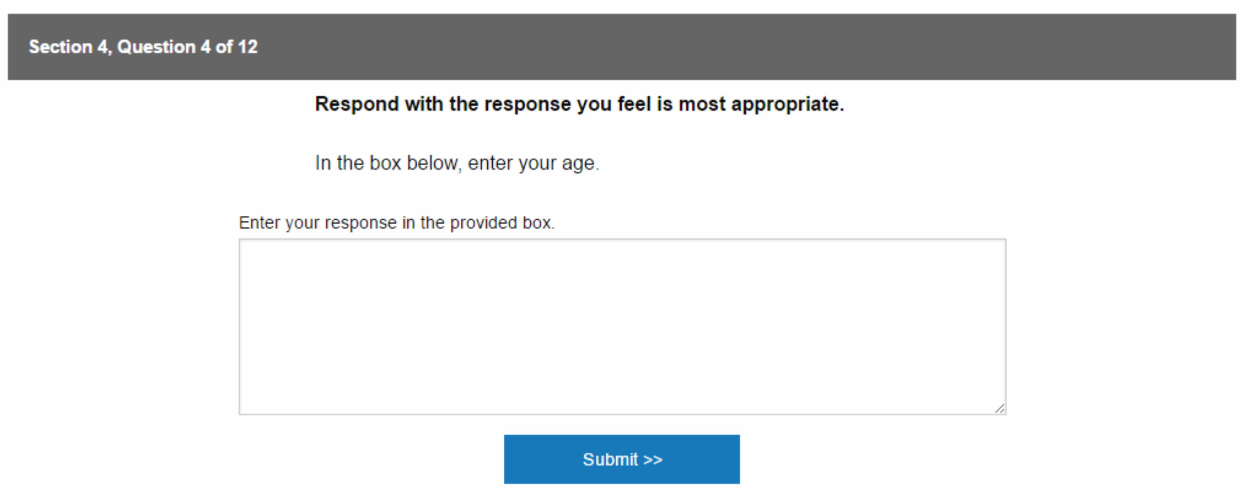


Figure 2. Example of a DQ question and how the CTS displayed text-box questions.

CTS IRM Functionality. During designated questions in the study, participants in the required-mark condition were presented with the IRM functionality provided by the CTS. The IRM functionality in this study provided a green "mark" button to the left of each radio-button which when hovered by a participant would turn light-blue (see Figure 3). When clicked, it would toggle the display settings of the adjustment radio-button text, turning the text gray and putting a line through it (see Figure 4). In the study, toggling the display of the radio-button text was considered a mark. In the required-mark condition, forced IRM required participants make at least two marks. If a participant did not make at least two marks or failed to select a final response, then the CTS produced an error and would not let the participant continue in the study until they made the minimum number of marks (see Figure 5).

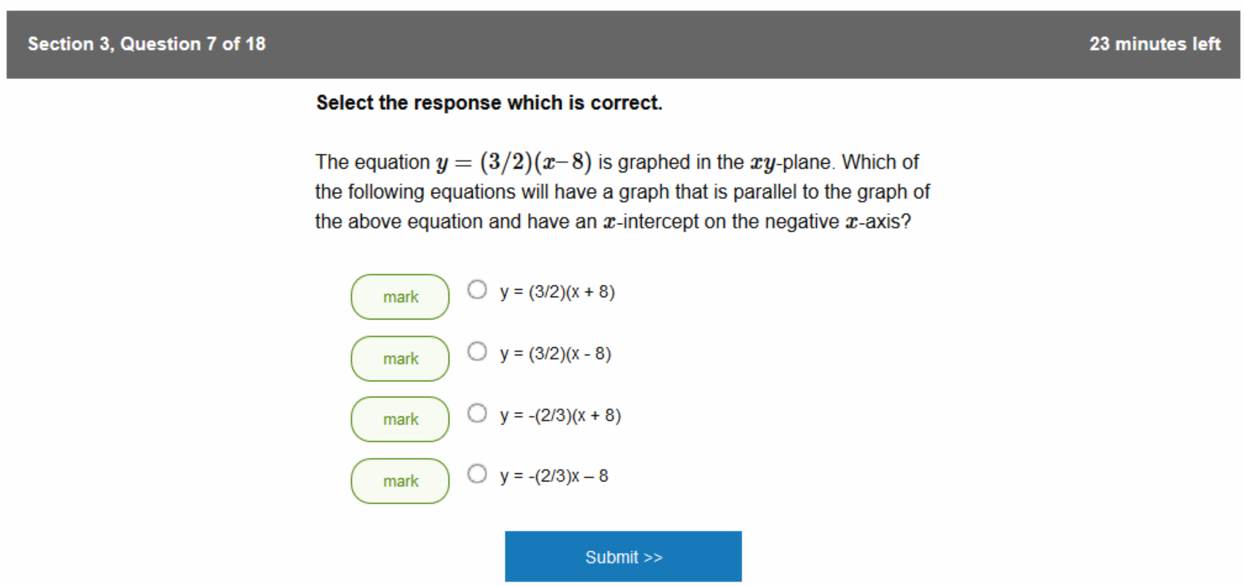


Figure 3. Example of an OMCT and IRM on the CTS in the required-mark condition.

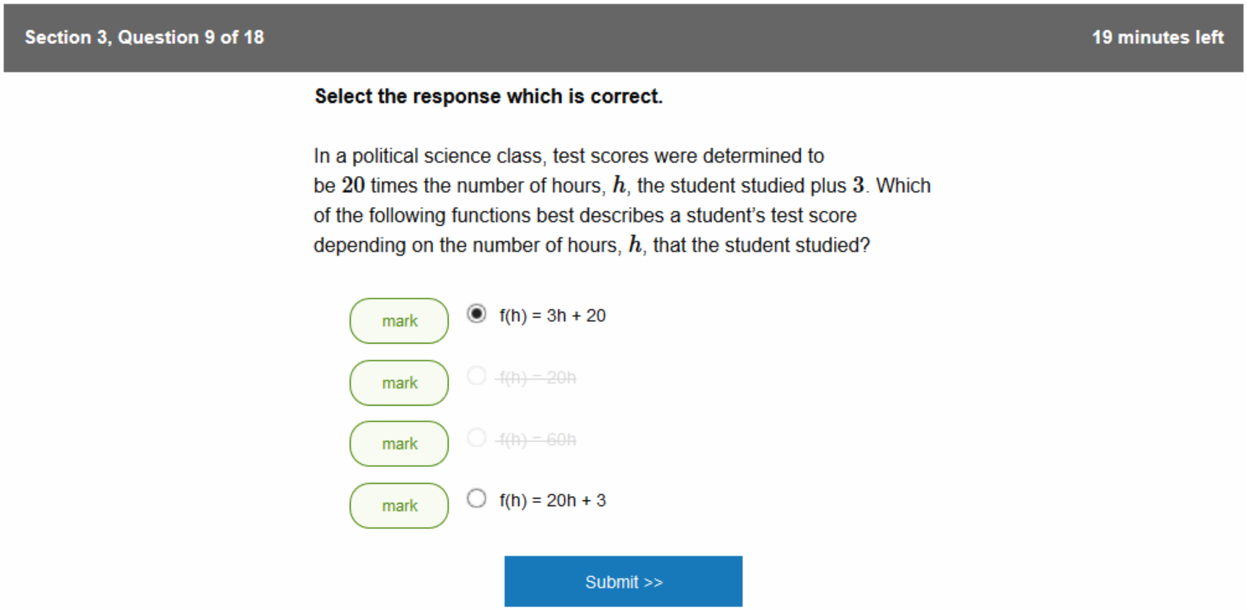


Figure 4. Example of an OMCT and IRM on the CTS being used in the required-mark condition.

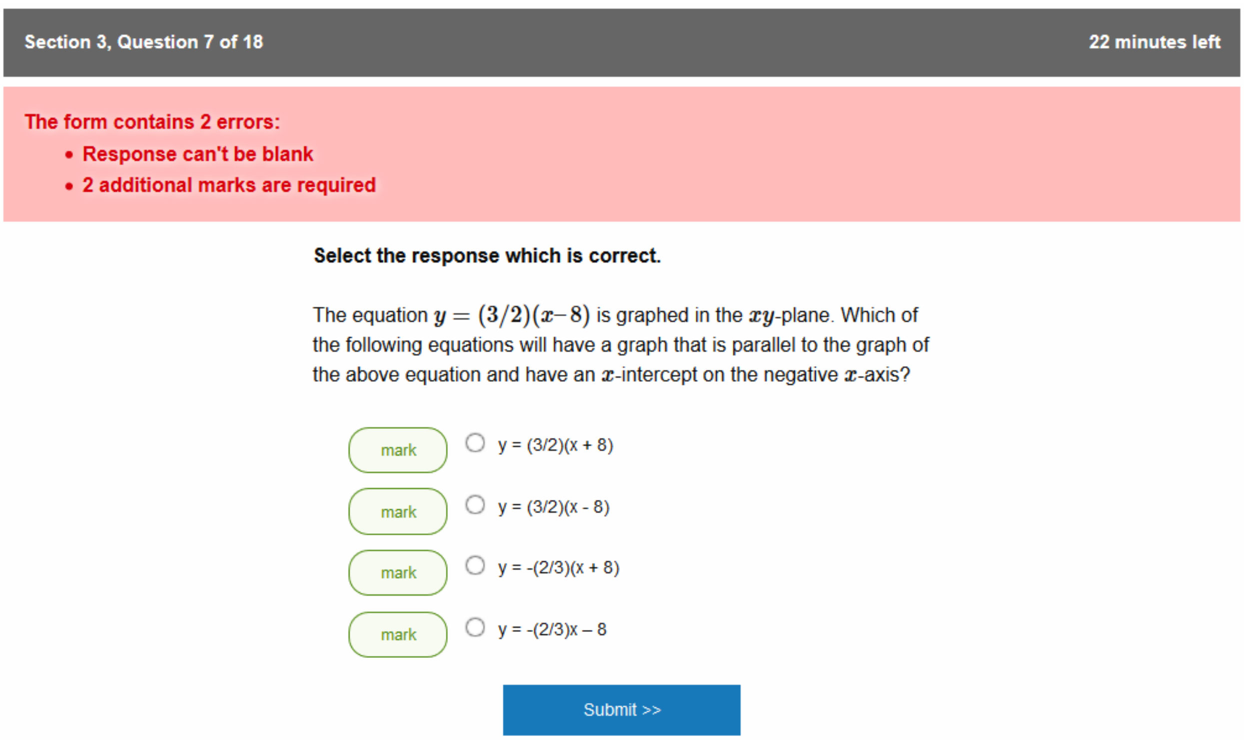


Figure 5. Example of an OMCT and force IRM in the required-mark condition producing an error.

Test Anxiety Measure. The Test Anxiety Inventory Short-form (TAI-5) was a template used to measure test anxiety in participants. The TAI-5 is a paper based test which was converted into an item format compatible with the CTS. This newly formatted TAI-5 for the CTS is designated as the TAI-5C and has a few key differences, primarily in how TAI-5C questions are displayed. On the TAI-5C, each question was individually displayed (see Figure 6), unlike the TAI-5 which displayed all test items simultaneously on a single side of paper. However, because of this change, there are no prior validity or reliability data for the TAI-5C itself.

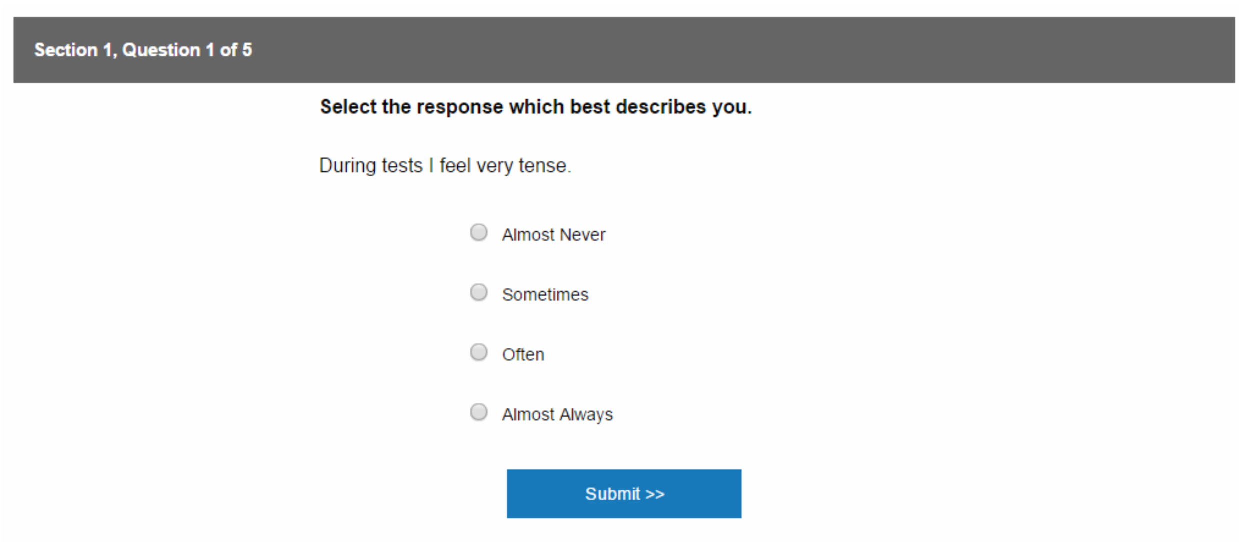


Figure 6. Example of a TAI-5C question as displayed by the CTS.

However, psychometric information is available for the TAI-5. According to Taylor and Deane (2002), the TAI-5 has good internal consistency (α=.87), generates similar distributions to the full Test Anxiety Inventory (TAI), has strong correlation with the TAI (r=.94), has moderate correlation with the State-Trait Anxiety Inventory (STAI; r=.39), and has strong correlations with the STAI Short-form prior to an actual (r=.50) and a hypothetical testing situation (r=.66). The TAI-5 and the TAI-5C are composed of 5 questions intervaled from 1 (*almost never*) to 4 (*almost always*) on a Likert-type scale (see Appendix A; Szafranski et al., 2012). Of these five questions, two measure emotionality, two measure worry, and one is used in the total calculation of the TAI-5 score (Taylor & Deane, 2002).

To create scores comparable to the full TAI, which is composed of 20 questions, the TAI-5 score is multiplied by 4. This manipulation generates a score between 20 (lower test anxiety) to 80 (higher test anxiety) with mean scores of 35.81 ($SD = 10.34$; Taylor & Deane, 2002). Using this same score manipulation on the TAI-5C within this study, the TAI-5C was found to have a good internal consistency ($\alpha = .88$) with mean scores of 52.39 ($SD = 16.74$). Also, the TAI-5C was found to have strong internal consistency on the emotionality subscale ($\alpha = .80$) and acceptable internal consistency on the worry subscale ($\alpha = .71$).

Objective Multiple-choice Test. To measure test performance in participants, an objective multiple-choice test (OMCT) was used. To reflect materials college students should be knowledgeable of, the OMCT in this study used 18 SAT heart of algebra practice questions (see Appendix C) hosted by Khan Academy (2017) and approved by the College Board (2017). Altogether there are no published validity or reliability measures for this specific sample of questions, in general, the math sections of the SAT has demonstrated an internal consistency of .68 to .81 and an alternative-form reliability of .91 (Ewing, Huff, Andrews, & King, 2005). Additionally, participants were given a 23 minute timer to complete the OMCT in the study to mimic the time constraints of a traditional SAT no-calculator permitted math section (College Board, 2017; Heimbach, 2015; Ivy Global, 2015).

OMCT Training Question. Before starting the OMCT, a training question (OMCT-TQ) was used to familiarize participants with forced IRM on the CTS. During this OMCT-TQ, participants in the required-mark condition were instructed and required to complete the question similarly to how they would during the OMTC within the study. For example, a participant in the required-mark condition was required to make at least two marks and select a final response before continuing on, while a participant in the control condition was only required to select a final response. Also only in the required-mark condition, the OMCT-TQ contained a small description of how to use the IRM functionality on the CTS and the minimum marking requirements established by forced IRM (see Figure 7; see Appendix B).

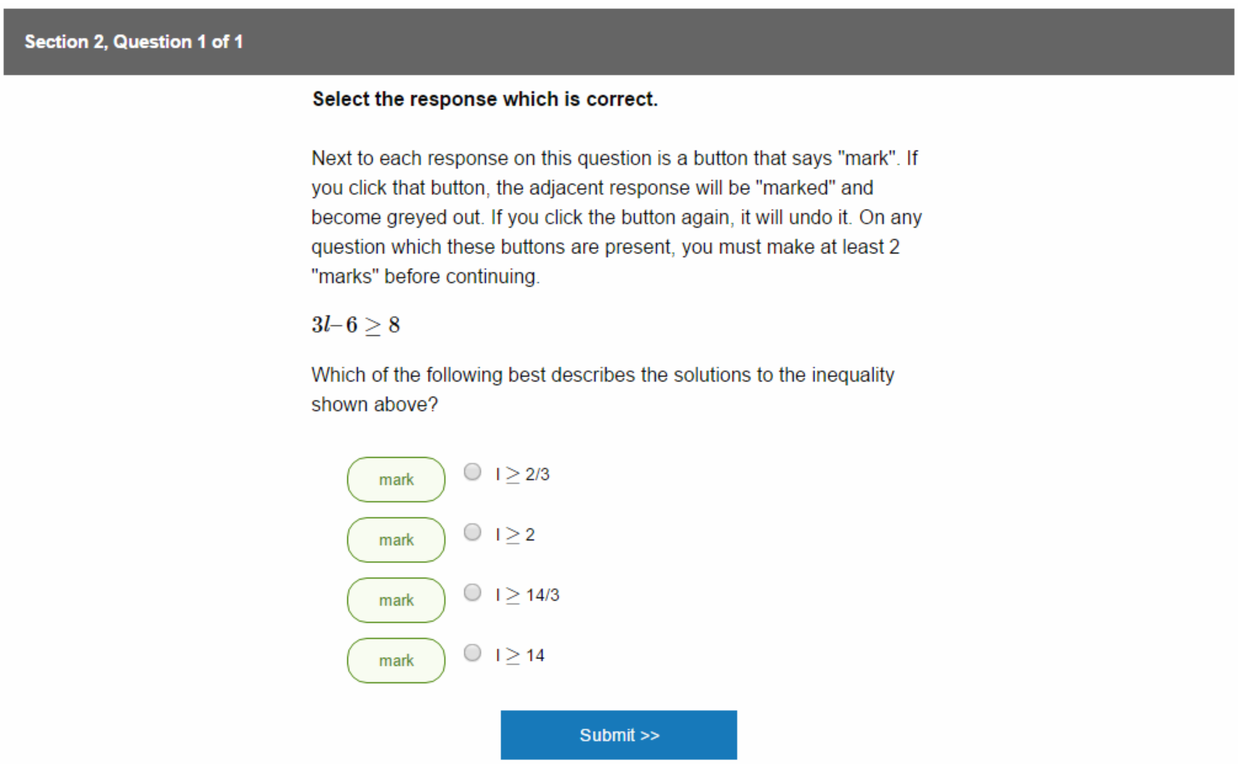


Figure 7. OMCT-TQ as it appeared in the required-mark condition.

Demographic Questionnaire. A demographic questionnaire (DQ) was used to collect participant information. The DQ asked participants their age, gender, ethnicity, and formal education in mathematics. Also, the DQ asked if they used any external resources to help complete the study, what their experiences were with the OMCT, and their feelings about the features provided by the CTS (see Appendix D). Lastly, the DQ had no required responses, meaning that participants could have chosen to skip the DQ entirely or only chosen to answer certain questions as they saw fit.

Information and Thank-you Pages. Since this study was conducted online, an information page was presented at the beginning of the study for all participants and contained a "Start Study" button on the bottom of the page (see Appendix E). Additionally, a thank-you page was presented at the end of the study which contained the purpose of the study, its hypotheses, and contact information for the principle investigator and faculty sponsor (see Appendix F).

Procedure

After approval from the Human Subjects Review Council at CWU, a description and link to the CTS was posted on the Department of Psychology SONA system which allowed CWU students to participate in this study. After a participant registered for the study through SONA, they were given a link to click which redirected them to the information page of the CTS. Also after clicking the link, the participant was assigned by the CTS to whichever condition (i.e., control or required-mark) had the fewest number of completed participants. If each condition had an equal number of completed participants then the participant was randomly assigned to either condition by the CTS. Nevertheless, after clicking the “Start study” button on the information page, the participant began the study with the first question of the TAI-5C.

The study itself was partitioned into four sections: the TAI-5C, the OMCT-TQ, the OMCT, and the DQ in that order. Each participant experienced each section in the same order and forced IRM and the IRM functionality were only enabled on the OMCT-TQ and OMCT sections. All other sections of the study had the same item format as the control condition. During the OMCT-TQ and OMCT sections only, a participant in the required-mark condition was presented the question, all potential responses for that question, and the IRM functionality; with forced IRM requiring these participants to make at least two marks and select one final answer. But in the control condition during these sections, a participant was only presented the question and all potential responses, not the IRM functionality, and they were only required to select one final answer.

After completing the requirements for their appropriate condition on an item, participants clicked “Submit” to save their answer to the CTS database and the next study question was automatically displayed. However, if a participant did not meet the given requirements, then an error was produced on their screen informing the participant of what still needed to be done before they could continue (e.g., “2 additional marks are required”). Furthermore, on the TAI-5C, OMCT-TQ, and DQ, participants were given unlimited time to complete each question. However, if a participant did not complete the OMCT section within the 23 minute time limit, then the CTS would automatically redirect the participant to the first question of the DQ. After completing the DQ, the CTS would automatically redirect the participant to the thank-you page of the study. On the thank-you page, the participant was presented with relevant information about the study in addition to their test session being ended by the CTS.

Statistical Analysis Overview

A multiple regression analysis was appropriate for this study. The first independent variable was the condition in which a participant was assigned (i.e., required-mark or control) and was categorical in its foundation. The second independent variable was the level of test anxiety that a participant was experiencing and was measured by the TAI-5C using an interval scale. Lastly, the dependent variable was test performance defined as the total number of correct responses given by the participant during the OMCT. We expected to find a significant regression and interaction on test performance. Specifically, we expected (a) as test anxiety increases in the control condition test performance would decrease and (b) as test anxiety increases in the required-mark condition test performance would increase.

CHAPTER IV

RESULTS

Descriptive Statistics

Most participants were unfamiliar with the OMCT questions (M = 2.17, SD = 0.54), did not use external resources to help them answer questions on the OMCT (M = 1.90, SD = 0.34), agreed that the OMCT was stressful (M = 2.19, SD = 1.14), and expected to perform poorly on the OMCT (M = 3.92, SD = 0.93). Additionally, participants appeared to be experienced with standardized tests (M = 1.39, SD = 0.53), computer-administered tests (M = 1.11, SD = 0.36), and used traditional marking strategies on paper tests (M = 2.15, SD = 0.93). Also, participants did not find the features of the test helpful in either the required-mark (M = 3.41, SD = 0.98) or control (M = 3.33, SD = 0.88) conditions (see Figure 8). Lastly, performing independent t test between participants who dropped out (n = 23) vs. did not drop out (n = 174) of the study during the OMCT, we found no significant differences between condition assignment (i.e., control vs. required-mark; $p > .05$) or test anxiety levels ($p > .05$).

Analysis

Within this study we hypothesized that (a) in the control condition, as the level of test anxiety increased participants performance would decrease, and (b) in the required-mark condition, as the level of test anxiety increased participants performance would increase.

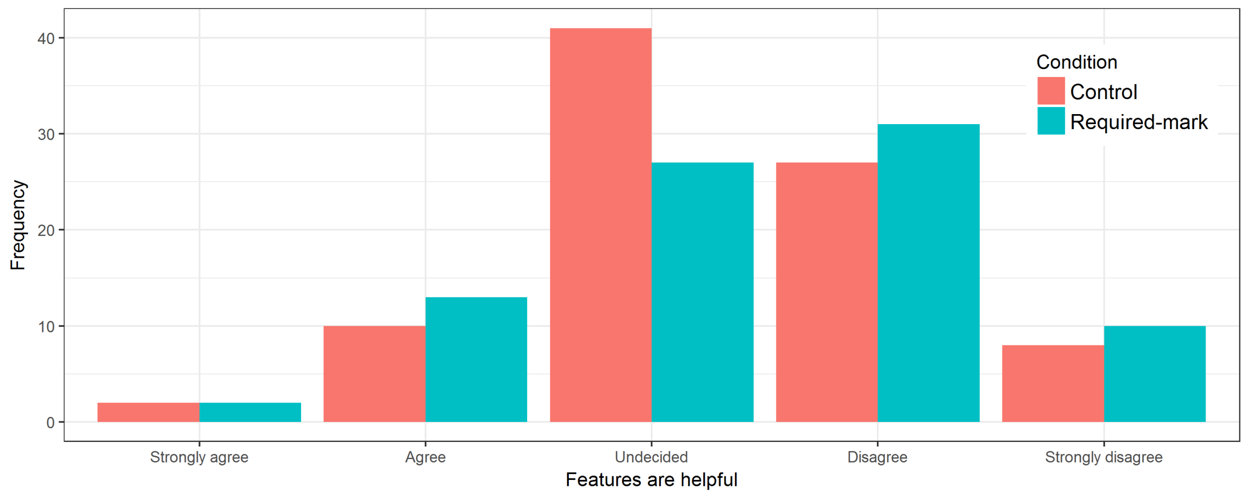


Figure 8. Frequency graph of responses regarding how much participants agreed that the features of the test were beneficial, grouped by condition.

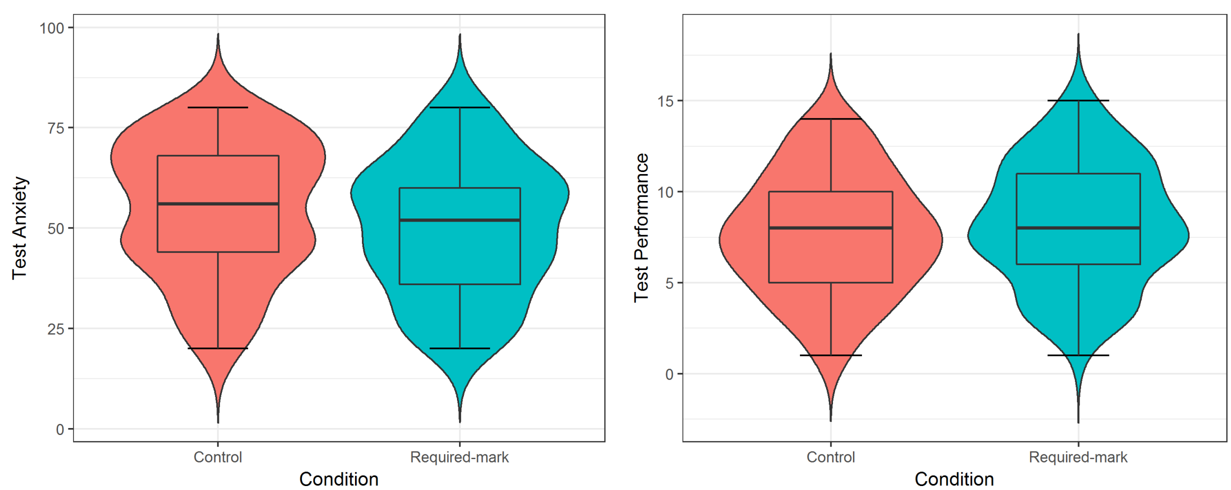


Figure 9. Test anxiety (left) and test performance (right) grouped by condition.

Using the TAI-5C to measure test anxiety (M = 52.39, SD = 16.74; see Figure 9) and the OMCT to measure test performance (M = 7.94, SD = 3.30; see Figure 9) in participants, we performed a multiple regression using test anxiety, condition (i.e., required-mark vs. control), and their interaction to predict performance. Looking at Figure 10, we see that (a) there is no curvilinear relationship within our regression, (b) the data is normally distributed, (c) homoscedastic, and (d) there exist no notable outliers that significantly alter our regression line. Next, Table 1 shows that there is no collinearity between test anxiety and condition within our study. Examining the regression itself, we found the regression to be significant, R^2 = .08, $F(3, 169) = 5.17$, $p < .01$. However, only test anxiety was found to significantly impact performance, β = -.06, $p < .01$, no significant difference in test performance was found between the required-mark (M = 8.24, SD = 3.33) and control (M = 7.66, SD = 3.27) conditions and neither of our hypotheses were supported. Forced IRM did not interact with test anxiety to influence test performance, as presented in Table 2.

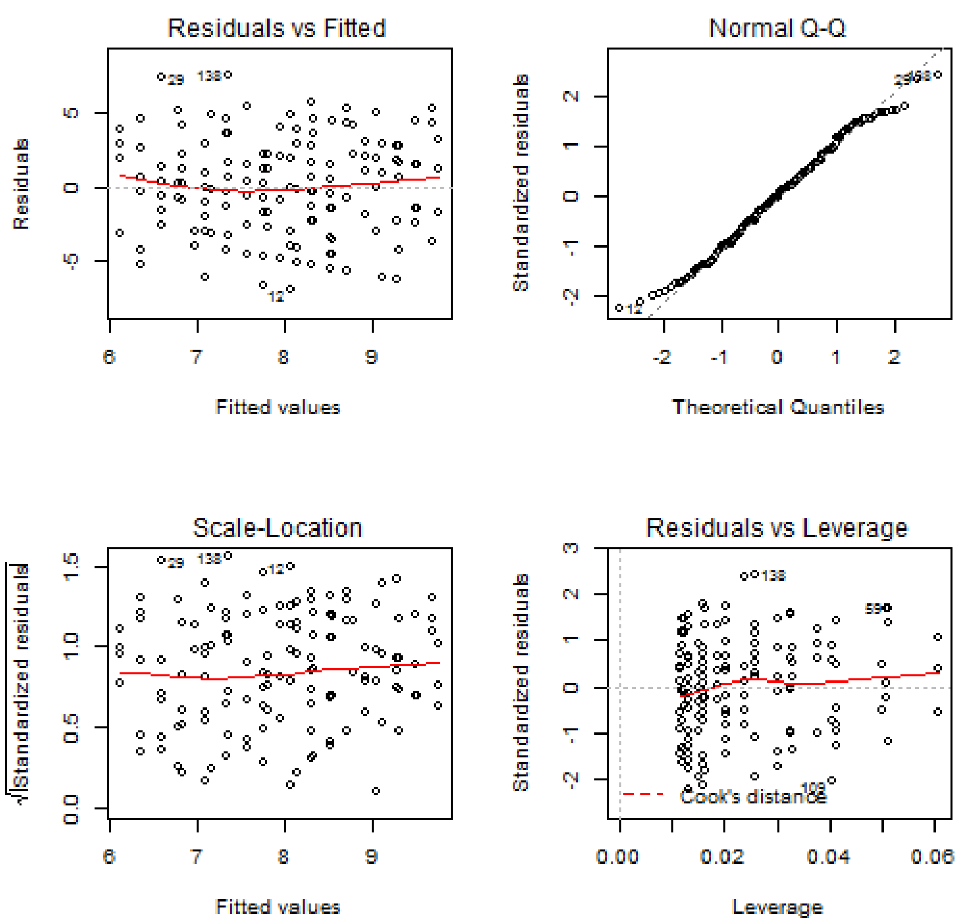


Figure 10. Model fit of multiple regression.

Table 1

Correlation Table.

	Pearson	Spearman	Kendall
	r	r_s	r_τ
Condition – Test Anxiety	-.14	-.14	-.12

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 2

Multiple Regression Table for the OMCT.

Predictor	β	$t(169)$	p
Intercept	10.98	9.40	$< .001$
Test Anxiety	-0.06	-2.97	.003
Condition	-0.32	-0.20	.841
Test Anxiety × Condition	0.01	0.42	.672

Notes: $R^2 = .08$, $F(3, 169) = 5.17$, $p < .01$

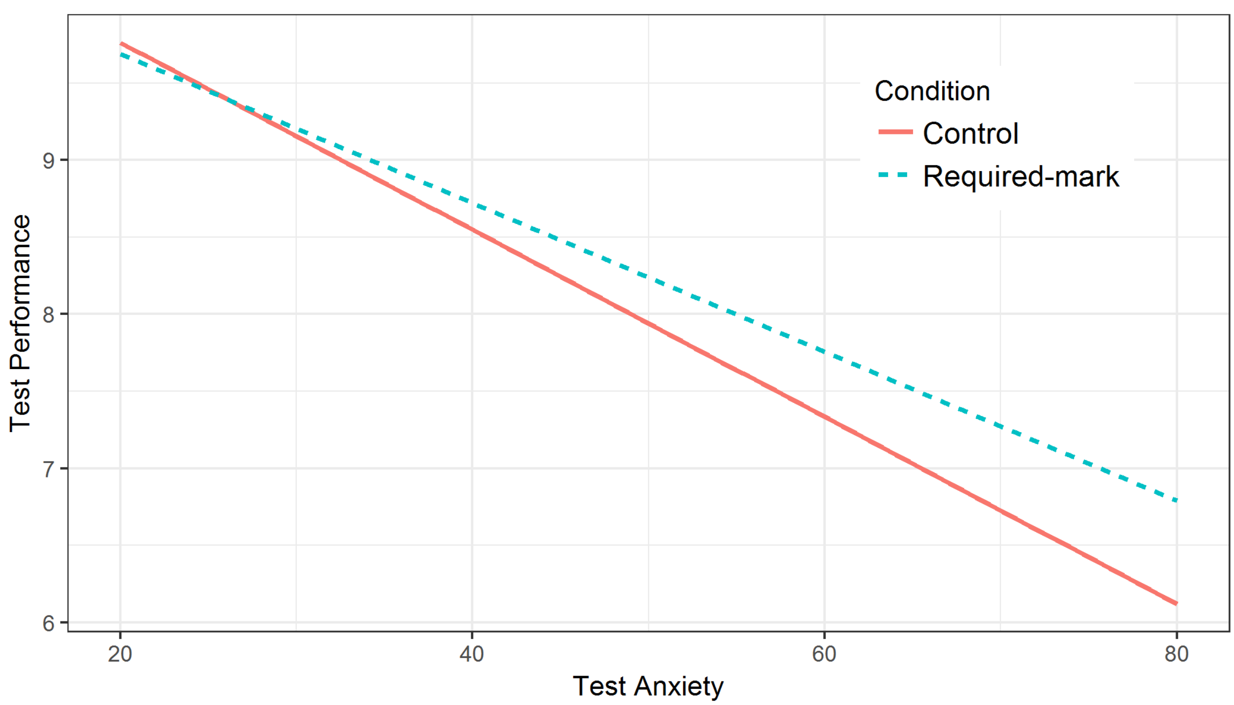


Figure 11. Graph of test anxiety on performance by condition on the OMCT.

Supplemental Analyses

First-Last Half Regression. Calvin et al. (1957) found that marking behavior benefited higher test anxiety students the most during the second half of tests. Performing multiple regressions using test anxiety, condition, and their interaction on test performance on the first and last halves of the OMCT, we only found a significant regression for the first half of the test, $R^2 = .07$, $F(3, 169) = 4.60$, $p < .01$, as seen in Table 3. Thus, we did not find a significant regression for the last of the OMCT as seen in Table 4. On the first half of the OMCT, however, we found a negative effect for test anxiety only, β = -.04, $p < .01$, as seen in Figure 12.

Table 3

Multiple Regression Table for the First Half of OMCT.

Predictor	β	$t(169)$	p
Intercept	6.77	8.54	< .001
Test Anxiety	-0.04	-2.79	.006
Condition	-0.19	-0.17	.864
Test Anxiety × Condition	0.01	0.28	.778

Notes: $R^2 = .07$, $F(3, 169) = 4.60$, $p < .01$

Table 4

Multiple Regression Table for the Last Half of OMCT.

Predictor	β	$t(169)$	p
Intercept	4.20	6.88	< .001
Test Anxiety	-0.02	-2.06	.041
Condition	-0.01	-0.16	.872
Test Anxiety × Condition	0.01	0.44	.658

Notes: $R^2 = .04$, $F(3, 169) = 2.53$, $p > .05$

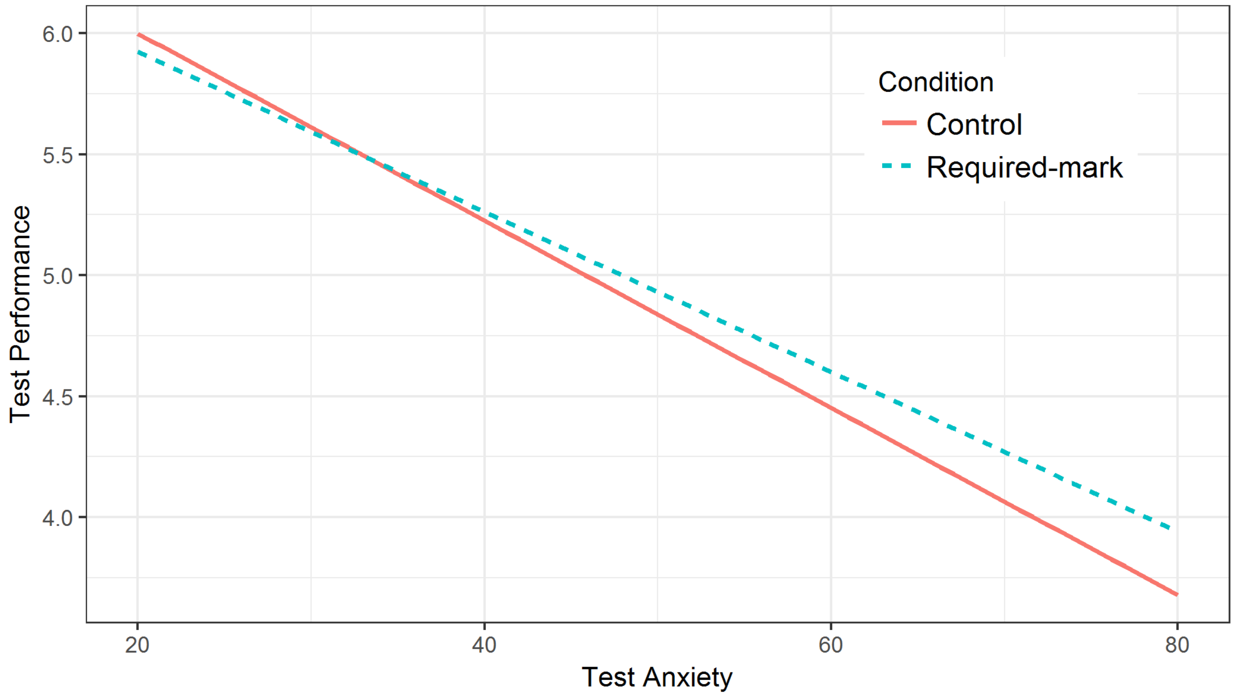


Figure 12. Graph of test anxiety on performance by condition on the first half of the OMCT.

Sex as a Predictor. Using sex (i.e., female vs. male) as a predictor variable within our regression, we found a significant regression, $R^2 = .11$, $F(4, 166) = 5.31$, $p <$ .001 with main effects of test anxiety, β = -0.06, $p <$ .01, and sex, β = 1.28, $p <$ .05, on test performance (see Table 5). Looking at Figures 13 and 14, we note that females (M = 8.25, SD = 3.19) are performing significantly better on the OMCT than males (M = 7.17, SD = 3.50).

Table 5

Multiple Regression Using Test Anxiety, Condition, Test Anxiety × Condition, and Sex on Performance.

Predictor	β	$t(166)$	p
Intercept	8.85	6.05	$< .001$
Sex	1.28	2.36	.019
Test Anxiety	-0.06	-3.06	.003
Test Anxiety × Condition	0.01	0.20	.841
Condition	-0.04	-0.02	.983

Notes: $R^2 = .11$, $F(4, 166) = 5.31$, $p < .001$

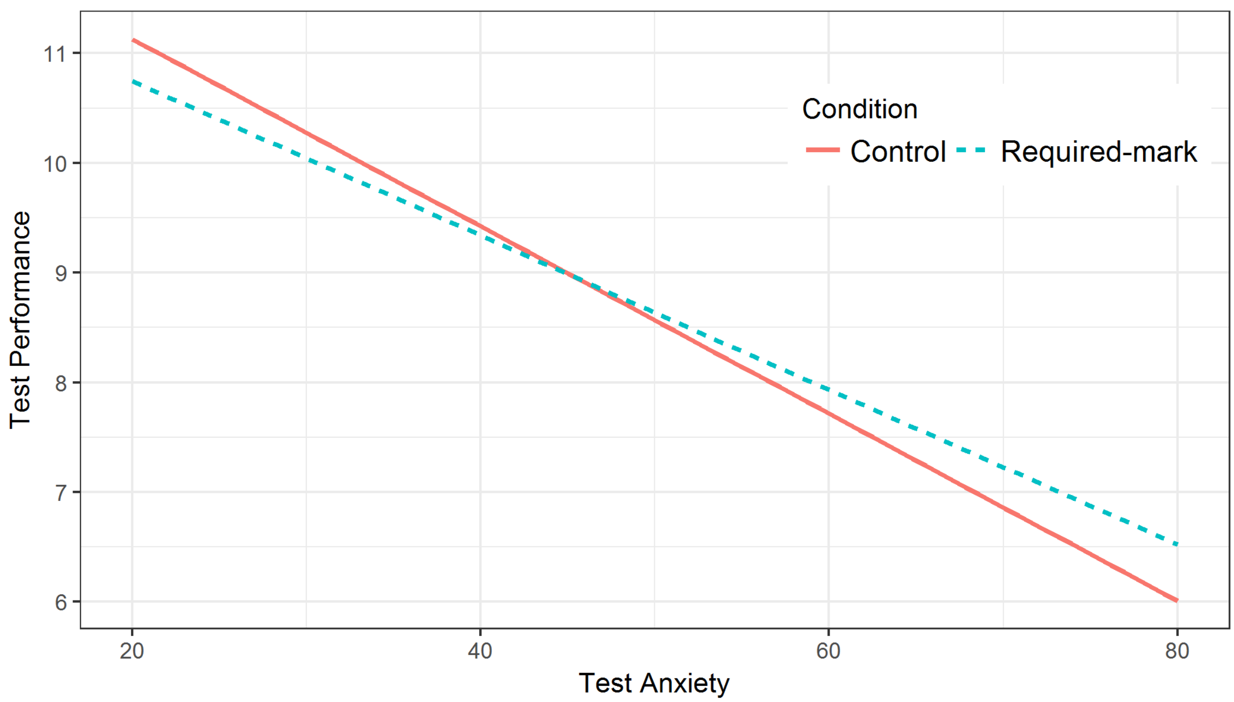


Figure 14. Graph of test anxiety on test performance by condition for female participants.

Experienced Stress and Test Anxiety. Smith and Rockett (1958) reported that how anxiety provoking a test was mediated the effects of test anxiety on performance. Performing a multiple regression using test anxiety on self-reported stress invoked from the OMCT, we found a significant regression, $R^2 = .15$, $F(1, 170) = 30.78$, $p < .001$,

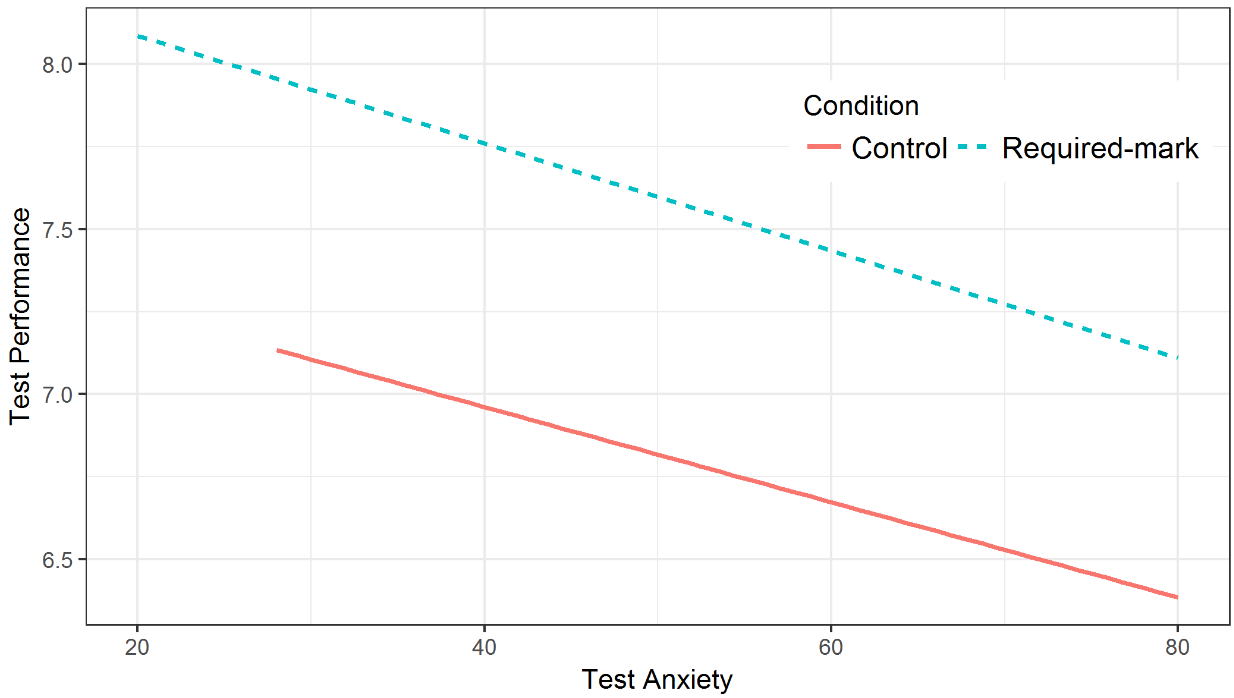


Figure 13. Graph of test anxiety on test performance by condition for male participants.

demonstrating that test anxiety is a significant predictor of self-reported stress during the OMCT, β = -.03, $p < .001$ (see Table 6). Furthermore as test anxiety increased, so did self-reported stress provoked from the test (see Figure 15).

Table 6

Multiple Regression Table for Stress Felt During the OMCT Using TAI-5C.

Predictor	β	$t(169)$	p
Intercept	3.56	13.55	$< .001$
Test Anxiety	-0.03	-5.55	$< .001$

Notes: $R^2 = .15$, $F(1,170) = 30.78$, $p < .001$

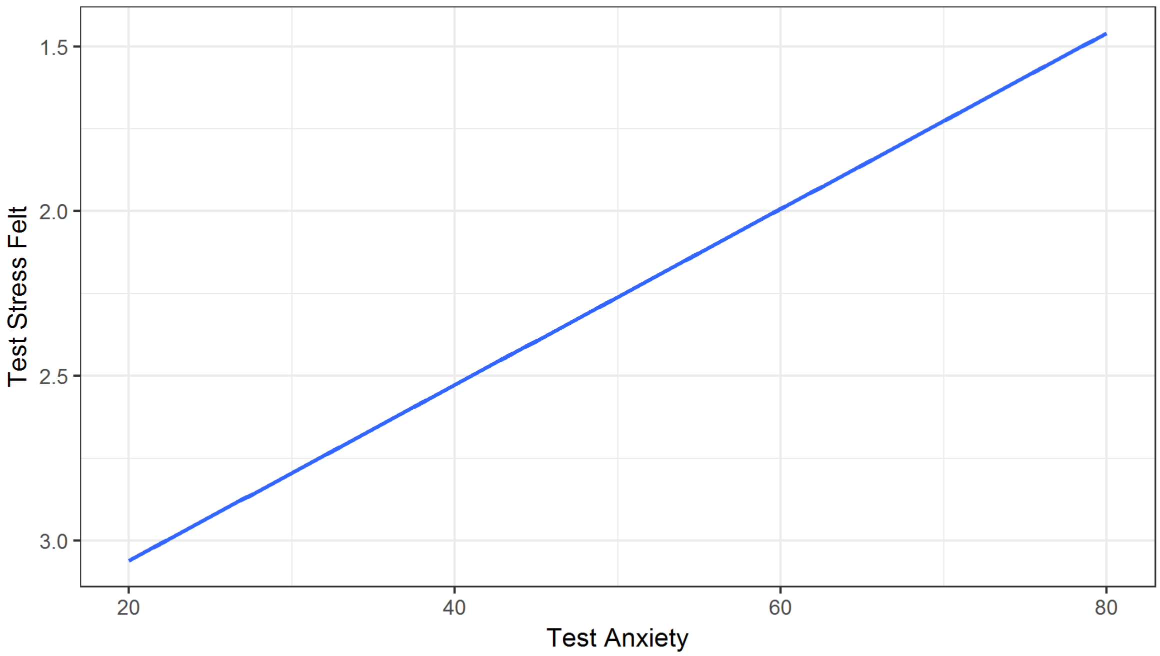


Figure 15. Graph of stress felt during the OMCT from 1 (*Strongly Agree*) to 5 (*Strongly Disagree*) with inverted y-axis.

CHAPTER V

DISCUSSION

This study predicted that test anxiety would interact with forced marking behavior to increase performance when present and to decrease performance when it was not afforded. However, this interaction was not supported. These findings indicate that the absence or presence of required marking behavior does not influence student performance on computer-administered tests, and therefore does not support either of our hypotheses. Instead, we found that test anxiety decreased performance regardless of the affordance of marking behavior, suggesting that test anxiety negatively impacts students even during low stakes tests, which is a contrast from the literature.

Smith and Rockett (1958) found a significant interaction between test anxiety levels and performance on low vs. high anxiety provoking tests; a distinction which requires us to reevaluate the student experience within our study. Despite our thinking that the OMCT was going to provoke low levels of anxiety because (a) it used content college students should be familiar with and (b) it was a voluntary testing situation, this may have not been the case. From the demographic information we collected, we noticed that the students self-reported feeling that the OMCT was anxiety provoking. A possible explanation for why we found no interaction between anxiety and forced marking behavior on test performance similar to Smith and Rockett (1958) was because students did not find the OMCT a low anxiety provoking test. This study also found that test anxiety predicted the amount of stress felt during the OMCT (see Figure 15). This finding highlights a relationship between test anxiety and the anxiety provoking attributes of tests found by Smith and Rockett (1958) that need to be further explored.

Another effect not replicated in our study was the difference between the first-last half of required-marking behavior on test performance. Calvin et al. (1957) found that when marking behavior was afforded on paper tests, higher test anxiety students benefited the most from this affordance during the second half of classroom exams. Performing an exploratory analysis on the first and last halves of the OMCT, however, we found no significant effects of test anxiety, condition, or their interaction on performance during the second half of the OMCT (see Table 4). This failure to replicate the findings of Calvin et al. (1957) may point to test anxiety changing with respect to time, an example being a student becoming acclimated to the testing environment. However, this may also be attributed to the ecological concerns associated with the study.

The number of questions used on the OMCT is a concern. Traditional no-calculator permitted SAT math sections are 20 questions with a 25 minute timer. However, not wanting to burden participants with too many questions, we reduced this number to 18 questions with a 23 minute timer to make the study more manageable; a reduction which may ultimately limit the generalizability of this study to more standardized computer-administered tests. Additionally, the information page told participants that we would be evaluating their performance on math questions (see Appendix E). Though other studies used math questions to observe the effects of test anxiety and marking on performance (e.g., Hong et al., 2006), divulging the content of the upcoming task may have contributed to the higher test anxiety levels reported in our study than by Taylor and Deane (2002) in the past.

Building from this idea further, researchers like McKeachie et al. (1955), Calvin et al. (1957), and Herman (1996) used undergraduate psychology exams instead of math questions which may be content undergraduate psychology students find less stressful to answer. This difference in content may partially explain why we did not find results similar to Smith and Rockett (1958), and perhaps suggests a relationship between test content and students' preferred knowledge areas. Furthermore, Herman (1996) noted that marking helps students the most during difficult questions, specifically finding that test performance was positively associated with marking as test items became more difficult on classroom exams. In this study, we used practice SAT questions from Khan Academy (2017); however, no difficulty information was associated with each individual item. Factors such as task content and difficulty should be further explored moving forward.

Sarason et al. (1952) demonstrated that the effects of test anxiety on performance are mediated by the importance of the test itself. Sarason et al. (1952) defined high stakes tests by students' feeling of need to complete the entire test (e.g., a classroom exam). While low stakes tests are defined by students' feelings of not needing to complete the entirety of the test. Taking this idea further, future studies may want to measure the students' feeling of need to complete the test. This factor of test importance exaggerates the effects of test anxiety and may provide more insight into how marking and test anxiety interact with each other (Sarason et al., 1952). As such, future research should explore if forced marking is beneficial on higher stakes tests, but not lower stakes tests as found in this study.

Moreover, sex is another factor that needs to be further examined. As seen in Figures 13, 14, and Table 5, there are notable differences in test performance with relation to test anxiety between males and females. Granted, this may be due to the small number of male participants in this study ($n = 46$). But given the results of this study, sex is another factor that needs to be evaluated more with respect to test anxiety, performance, and marking behaviors.

Furthermore, future studies also need to evaluate the different types of marking strategies afforded by computer-administered tests. Forced marking behavior was not found to be beneficial to students, but McKeachie et al. (1955) found that affording optional elaboration, even if a student did not used it, improved their performance. So in a similar fashion, future studies need to evaluate the different implementations of marking, such as optional elaboration or optional IRM. These alternative marking functionalities may be more representative of traditional paper strategies that students use than forced IRM.

Although no interaction between forced marking behavior and performance was found, this study did find that test anxiety negatively affected the performance of students on low stakes computer-administered tests. Additionally, previous research has shown that the effect of test anxiety on performance is proportional to the importance of the test itself (Freedman, 1982; Sarason et al., 1952) and that traditional testing strategies which benefit these higher test anxiety students the most (Smith & Rockett, 1958) are disappearing (Macedo-Rouet et al., 2009). If this past research on paper tests is translative to computer-administered tests similarly to test anxiety, then higher test anxiety test takers, such as students in the classroom or professionals seeking certification, are being disadvantaged within the modern testing environment. Therefore, it is the responsibility of test developers to implement and evaluate alternative strategies to forced IRM or to discover new strategies which benefit these test takers during computer-administered tests. Novel interactive features of computer-administered tests that should be developed, tested, and incorporated into testing software to improve the test taking experience on standardized tests.

REFERENCES

Ashcraft, M. H. (1988). Math anxiety: Personal, education, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5), 181-185. doi: 10.1111/1467-8721.00196

Calvin, A. D., McGuigan, F. J., & Sullivan, M. W. (1957). A further investigation of the relationship between anxiety and classroom examination performance. *Journal of Educational Psychology*, *48*(4), 240-244. doi: 10.1037/h0040661

College Board. (2017). *Sample questions.* Retrieved from https://collegereadiness.collegeboard.org/sample-questions.

Ewing, M., Huff, K., Andrews, M., & King, K. (2005). *Assessing the reliability of skills measured by the SAT.* Retrieved from ERIC Institute of Education Sciences. (ERIC ED562595).

Freedman, E. S. (1982, July). *On improving the format of examination papers.* Paper presented at the Annual Meeting of the United Kingdom Reading Association, Newcastle upon Tyne, England.

Heimbach, A. (2015). *Can you take the SAT online? A guide to computerized testing.* Retrieved from https://blog.prepscholar.com/can-you-take-the-sat-online-a-guide-to-computerized-testing.

Herman, W. E. (1996, August). *An analysis of multiple-choice test items booklets.* Poster presented at the 104th Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Hong, E., Sas, M., & Sas, J. C. (2006). Test taking stratagies of high and low mathematics achievers. *Journal of Educational Research*, *99*(3), 144-155.

Ivy Global. (2015). *The SAT from 2016 onward.* Retrieved from http://sat.ivyglobal.com/new-vs-old/.

Khan Academy. (2017). *Official SAT practice.* Retrieved from https://www.khanacademy.org/mission/sat/practice/math.

Kim, Y. H., & Goetz, E. T. (1993). Strategic processing of test questions: The test marking responses of college students [Abstract]. *Learning and Individual Differences*, *5*(3), 211-218. Abstract retrieved from Science Direct database. doi: 10.1016/1041-6080(93)90003-B

Macedo-Rouet, M., Ney, M., Charles, S., & Lallich-Boidin, G. (2009). Student's performance and satisfaction with web vs. paper-based practice quizzes and lecture notes. *Computers & Education*, *53*, 375-384. doi: 10.1016/j.compedu.2009.02.013

Mavilidi, M., Hoogerheide, V., & Pass, F. (2014). A quick and easy strategy to reduce test anxiety and enhance test performance. *Applied Cognitive Psychology*, *28*, 720-726. doi: 10.1002/acp.3058

McKeachie, W. J., Pollie, D., & Speisman, J. (1955). Relieving anxiety in classroom examinations. *The Journal of Abnormal and Social Psychology*, *50*(1), 93-98. doi: 10.1037/h0046560

Nield, A. F., & Wintre, M. G. (1986). Multiple-choice questions with an option to comment: Student attitudes and use. *Teaching of Psychology*, *13*(4), 196-199.

Osterhouse, R. A. (1975). Classroom anxiety and the examination performance of test-anxious students. *The Journal of Educational Research*, *68*(7), 247-250. doi: 10.1080/00220671.1975.10884763

Sarason, S. B., Mandler, G., & Craighill, P. G. (1952). The effect of differential instructions on anxiety and learning. *Journal of Abnormal Psychology*, *4*(*Suppl.* 2), 561-565.

Smith, W. F., & Rockett, F. C. (1958). Test performance as a function of anxiety, instructor and instructions [Abstract]. *The Journal of Educational Research*, *52*(4), 138-141. Abstract retrieved from APA PsychNet database. http://dx.doi.org/10.1080/00220671.1958.10882554.

Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. *Anxiety, Stress & Coping*, *25*(6), 667-677. doi: 10.1080/10615806.2012.663490

Taylor, J., & Deane, F. P. (2002). Development of a short form of the test anxiety inventory (TAI). *The Journal of General Psychology*, *129*(2), 127-136.

Zatz, S., & Chassin, L. (1985). Cognitions of text-anxious children under naturalistic test-taking conditions. *Journal of Consulting and Clinical Psychology*, *53*(3), 393-401. doi: 10.1037/0022-006X.53.3.393

APPENDIXES

Appendix A

TAI-5C

Figure 16 contains the test item questions and responses from the TAI-5C transferred from the TAI-5 on a scale from 1 (*Almost Never*) to 4 (*Almost Always*; Taylor & Deane, 2002).

section_number	question	response_1	response_2	response_3	response_4	response_5
1	During tests I feel very tense.	Almost Never	Sometimes	Often	Almost Always	
1	I wish examinations did not bother me to much.	Almost Never	Sometimes	Often	Almost Always	
1	I seem to defeat myself while working on important tests.	Almost Never	Sometimes	Often	Almost Always	
1	I feel very panicky when I take an important test.	Almost Never	Sometimes	Often	Almost Always	
1	During examinations I get so nervous that I forget facts I really know.	Almost Never	Sometimes	Often	Almost Always	

Figure 16. TAI-5C questions and possible responses.

Appendix B

Objective Multiple Choice Test Training Question (OMCT-TQ)

Figure 17 contains the questions used for the OMCT-TQ within the control and required-mark conditions.

section_number	question	response_1	response_2	response_3	response_4	response_5	response_correct
2	$$ 3l – 6 \geq 8 $$ Which of the following best describes the solutions to the inequality shown above?	l \(\geq \) 2/3	l \(\geq \) 2	***l \(\geq \) 14/3***	l \(\geq \) 14		3
2	Next to each response on this question is a button that says "mark". If you click that button, the adjacent response will be "marked" and become greyed out. If you click the button again, it will undo it. On any question which these buttons are present, you must make at least 2 "marks" before continuing. $$ 3l – 6 \geq 8 $$ Which of the following best describes the solutions to the inequality shown above?	l \(\geq \) 2/3	l \(\geq \) 2	***l \(\geq \) 14/3***	l \(\geq \) 14		3

Figure 17. OMCT-TQ questions with MathJax formatting.

Appendix C

Objective Multiple Choice Test (OMCT)

Figures 18, 19, 20 contain all of the questions which will be used in the OMCT section of the study, in addition to the possible responses for each question.

section_number	question	response_1	response_2	response_3	response_4	response_5	response_correct
3	$$ 5.5B + 4R = 28 $$ The above equation is true if Amit buys \(B \) pounds of blueberries and \(R \) pounds of raspberries at a farm where blueberries cost \($5.50 \) per pound and raspberries cost \($4.00 \) per pound. According to the equation, how much does Amit spend in total on both types of berries?	9.5	22	***28***	56		3
3	$$ 6 = -s + 77 $$ Given the above equation, what is the value of \(1 + 5(77 – s) \)?	-739	−29	***31***	741		3
3	Dimitri is helping to plan the school talent show. Each performer for the talent show has \(6 \) minutes for his or her performance, which includes transition time between performances. If the introduction for the talent show is \(24 \) minutes long and the entire show, including the introduction, will last \(150 \) minutes, how many different performances can the talent show accommodate?	***21***	24	25	29		1
3	$$ S = 40,000 + 500c $$ Caden started a new job selling dental chairs. He earns a base salary plus a commission for every chair he sells. The equation above gives Caden's annual salary, \(S \), in dollars, after selling \(c \) dental chairs. Based on the equation above, what is Caden's base salary?	$39500	***$40000***	$40500	$50000		2
3	More than \(450 \) students traveled to a state park for a field trip. The school allowed \(6 \) students to travel by car, and the rest traveled on \(11 \) buses, each of which held the same number of students. If there were \(s \) students in each bus, which inequality best represents this situation?	***11s + 6 > 450***	11s + 6 < 450	6s + 11 > 450	6s + 11 < 450		1
3	The property taxes in a town decrease as the distance from the local elementary school increases. The greatest property taxes are \(4.5% \), and for every \(10 \) miles from the school, property taxes decrease by \(0.5 \) percentage points. If a house's property taxes are \(3% \), what is the distance of that house from the school?	10 miles	20 miles	***30 miles***	40 miles		3
3	The equation \(y = (3/2)(x – 8) \) is graphed in the \(xy \)-plane. Which of the following equations will have a graph that is parallel to the graph of the above equation and have an \(x \)-intercept on the negative \(x \)-axis?	***y = (3/2)(x + 8)***	y = (3/2)(x - 8)	y = -(2/3)(x + 8)	y = -(2/3)x – 8		1

Figure 18. OMCT questions 1 through 7 with MathJax formatting.

section_number	question	response_1	response_2	response_3	response_4	response_5	response_correct
3	A distributor ships DVDs to several stores. The shipping boxes contain several DVDs (in their cases) plus a layer of padding at each end of the box. The DVD cases and layers of padding can be arranged neatly inside each box. Each DVD case is \(14\text{mm} \) (millimeters) thick, each layer of padding is \(10\text{mm} \) thick, and the length of the interior of the box is \(132\text{mm} \). If \(d \) represents the number of DVDs that the distributor can fit into one box with two layers of padding, which of the following inequalities best models the situation?	10d + 14 ≤ 132	***14d + 20 ≤ 132***	20d + 14 ≤ 132	14d + 10 ≤ 132		2
3	In a political science class, test scores were determined to be \(20 \) times the number of hours, \(h \), the student studied plus \(3 \). Which of the following functions best describes a student's test score depending on the number of hours, \(h \), that the student studied?	f(h) = 3h + 20	f(h) = 20h	f(h) = 60h	***f(h) = 20h + 3***		4
3	Which of the following is an equation of the line \(A \) graphed in the \(xy \)-plane that passes through the point \((-1, 3.5) \) and is perpendicular to the line \(B \) whose equation is \(x + 4.5 = 0 \)?	x = −1	x = 3.5	***y = 3.5***	y = 4.5		3
3	$$ 4c + 5 < 4c + 3 $$ Which of the following best describes the solutions to the inequality shown above?	All real numbers	c < 1/2	c > 1/4	***No solution***		4
3	A \(300 \)-room hotel collects \($75 \) per occupied room and does not collect any money for vacant rooms. Which of the following functions best represents how many dollars, \(d \), the hotel generates if there are \(v \) vacant rooms in the hotel?	***d = 75(300 − v)***	d = 75(300 + v)	d = 300(75 − v)	d = 300(75 + v)		1
3	Anouk is an engineer planning sound and lighting for a free concert in the park. The concert was advertised with a promise to use no more than \(108 \text{kilowatts} \) (kW) of power. It was determined that the main contributors to power usage, speakers and floodlights, use \(1.8 \text{kW} \) and \(2.2 \text{kW} \), respectively. Anouk also must keep within her budget of \($3,300 \). The rental company is charging \($75 \) for each speaker and \($42 \) for each floodlight. Which of the following combinations meets Anouk's requirements?	40 speakers and 30 floodlights	12 speakers and 54 floodlights	***26 speakers and 13 floodlights***	38 speakers and 22 floodlights		3

Figure 19. OMCT questions 8 through 13 with MathJax formatting.

section_number	question	response_1	response_2	response_3	response_4	response_5	response_correct
3	Vijay needs to take a taxi, which costs a flat fee of \(3 \) dollars, plus an additional \(4 \) dollars per mile. If Vijay has a dollars with him, which inequality shows the number of miles, \(m \), he can afford to travel in the taxi?	0 ≤ m ≤ 4a − 3	***0 ≤ m ≤ (a/4) − (3/4)***	4a − 3 ≤ m	(a/4) − (3/4) ≤ m		2
3	$$ z + w - 3 = k $$ $$ 6z - 10w = 8 $$ Consider the system of equations above, where \(k \) is a constant. For which value of \(k \) are there infinitely many \((w, z) \) solutions?	-19 / 5	5	8	***None of the above***		4
3	$$ -5x - 4y = 2a $$ $$ 4x - 5y = 2 $$ Which of the following choices of \(a \) will result in a system of linear equations with exactly one solution?	***\(a \) can be any number***	\(a \) can be any number except 0.8	\(a \) can be any number except -0.8	\(a \) = 0.8		1

3	The owner of a landscaping company is developing a proposal to maintain the grounds of a building. It is estimated that \(75 \) gardening hours and \(25 \) foreman hours will be required. The total budget for these hours is \($1,600 \). The hourly wage for a foreman is \(30% \) more than a gardener plus an additional \($1.65 \) per hour. Which of the following systems of equations can be used to determine the hourly wages of a gardener, \(g \), and a foreman, \(f \), so the total wages are \($1,600 \)?	25g + 75f = 1,600
		       f = 1.3g + 1.65
			***25f + 75g = 1,600
			       f = 1.3g + 1.65***
				25g + 75f = 1,600
				       f = 1.3f + 1.65
					25f + 75g = 1,600
					       f = 1.3f + 1.65
							2

3	A caterer is determining how many forks she will need to buy for her upcoming event. Each adult needs \(5 \) forks, and each child needs \(2 \) forks. If the event will host \(764 \) adults and children in all, and the caterer ordered \(2, 992 \) forks, how many adults and how many children are expected to attend?	207 adults and 557 children	276 adults and 488 children	***488 adults and 276 children***	557 adults and 207 children		3

Figure 20. OMCT questions 14 through 18 with MathJax formatting.

Appendix D

Demographic Questionnaire

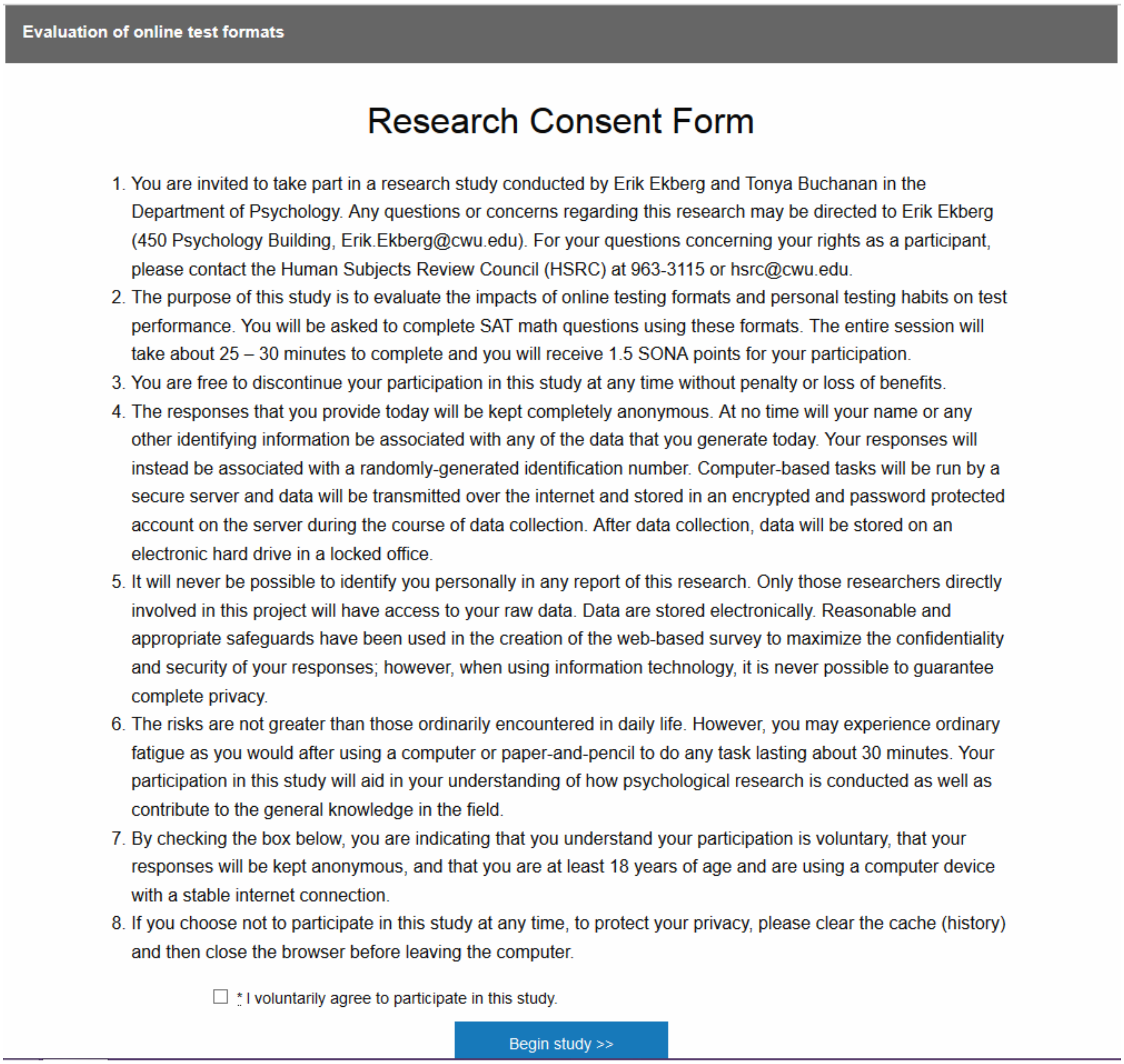
Figure 21 contains the list of demographic questions used within the study, with questions seven to ten being ranked on a Likert-type scale from 1 (*Strongly agree*) to 5 (*Strongly disagree*).

section_number	question	response_1	response_2	response_3	response_4	response_5
4	Have you seen any of the questions on this study before?	Yes	No	Not sure		
4	Did you use any external resources (e.g. websites, forums, other people) to help complete questions on this study?	Yes	No	Not sure		
4	What kind of formal mathematics education do you have?	Precalculus or higher	Intermediate algebra	Introductory algebra	Pre-algebra or below	
4	In the box below, enter your age.					
4	What is your biological sex?	Male	Female	Other		
4	Please specify your ethnicity.	Asian or Pacific Islander	Black or African American	Hispanic or Latino	Caucasian or White	Other
4	I found this test stressful	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree
4	I performed well on this test	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree
4	I found the features of this test helpful	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree
4	During paper tests, I frequently use marking stratagies	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree
4	Have you taken an online standarized test, such as the ACT?	Yes	No	Not sure		
4	Have you taken an computer administered test before?	Yes	No	Not sure		

Figure 21. The demographic questions and responses asked participants in the study.

Appendix E

Information Page



Evaluation of online test formats

Research Consent Form

1. You are invited to take part in a research study conducted by Erik Ekberg and Tonya Buchanan in the Department of Psychology. Any questions or concerns regarding this research may be directed to Erik Ekberg (450 Psychology Building, Erik.Ekberg@cwu.edu). For your questions concerning your rights as a participant, please contact the Human Subjects Review Council (HSRC) at 963-3115 or hsrc@cwu.edu.
2. The purpose of this study is to evaluate the impacts of online testing formats and personal testing habits on test performance. You will be asked to complete SAT math questions using these formats. The entire session will take about 25 – 30 minutes to complete and you will receive 1.5 SONA points for your participation.
3. You are free to discontinue your participation in this study at any time without penalty or loss of benefits.
4. The responses that you provide today will be kept completely anonymous. At no time will your name or any other identifying information be associated with any of the data that you generate today. Your responses will instead be associated with a randomly-generated identification number. Computer-based tasks will be run by a secure server and data will be transmitted over the internet and stored in an encrypted and password protected account on the server during the course of data collection. After data collection, data will be stored on an electronic hard drive in a locked office.
5. It will never be possible to identify you personally in any report of this research. Only those researchers directly involved in this project will have access to your raw data. Data are stored electronically. Reasonable and appropriate safeguards have been used in the creation of the web-based survey to maximize the confidentiality and security of your responses; however, when using information technology, it is never possible to guarantee complete privacy.
6. The risks are not greater than those ordinarily encountered in daily life. However, you may experience ordinary fatigue as you would after using a computer or paper-and-pencil to do any task lasting about 30 minutes. Your participation in this study will aid in your understanding of how psychological research is conducted as well as contribute to the general knowledge in the field.
7. By checking the box below, you are indicating that you understand your participation is voluntary, that your responses will be kept anonymous, and that you are at least 18 years of age and are using a computer device with a stable internet connection.
8. If you choose not to participate in this study at any time, to protect your privacy, please clear the cache (history) and then close the browser before leaving the computer.

☐ * I voluntarily agree to participate in this study.

Begin study >>

Figure 22. The information page which appeared at the beginning of the study.

Appendix F

Thank-you Page

Thank you for participating

TO PROTECT YOUR PRIVACY, PLEASE CLEAR THE CACHE (HISTORY) THEN CLOSE THE BROWSER BEFORE LEAVING THE COMPUTER

Debriefing Statement

Thank you for participating in this experiment. We hope you found it to be an interesting experience and that you learned something about how psychological research is conducted. As a result of participating in this study, we hope that it has aided in your understanding of how psychological research is conducted as well as contributed to the general knowledge in the field. You will receive partial course credit for your participation.

In this study, you were asked to complete 3 different types of questsion: (1) questions about test anxiety, (2) math questions, and (3) demographic and evaluation question. We are examining how test anxiety and question formatting influence each other. We are also interested in how well you thought the formatting of the questions influenced your testing experience.

Due to the on-going nature of this research, we would like to ask for your cooperation in not revealing any details of this study to others (e.g. friends, classmates) who might eventually participate in this study. These details could affect the way they perform in this experiment, which would adversely affect the nature of our study. If someone does ask, you can just tell them that you were asked to participate in a study on exam item formats, rather than providing specific details about the study.

If you have further questions, please contact the experimenter listed on your consent form or the Human Subjects Review Council (963-3115; hsrc@cwu.edu). Given that test anxiety and math questions were discussed within this study and can possibly be a sensitive subject, we encourage you to talk with a professional if you have strong negative feelings. Counseling services are available from the Student Medical & Counseling Clinic (SMaCC, 963-1391). Should you be interested in reading research related to this work, you can get more information from:

- Herman, W. E. (1996, August).An analysis of multiple-choice test items booklets. Postersession presented at the 104th Annual Convention of the American PsychologicalAssociation, Toronto, Ontario, Canada.
- McKeachie, W. J., Pollie, D., & Speisman, J. (1955). Relieving anxiety in classroomexaminations.The Journal of Abnormal and Social Psychology,50(1), 93-98. doi:10.1037/h0046560

Figure 23. Thank-you page presented at the end of the study.