

Can Delivery Modality Influence Test Performance?  
Comparing Traditional and Digital Formats

By

Gage H. Moyer

B. A., University of West Florida, 2018

A thesis submitted to the School of Psychology  
Usha Kundu, MD College of Health  
The University of West Florida  
In partial fulfillment of the requirements for the degree of  
Master of Arts

2020

## **THESIS CERTIFICATION**

Candidate Gage Moyer defended this thesis on 25 March 2020. The members of the thesis committee were:

Professor	Jane Halonen	Ph.D, Committee Chair
Professor	David B. Strohmets	Ph.D, Committee Member

The University of West Florida Graduate School verifies the names of the committee members and certifies that the thesis has been approved in accordance with University requirements.

## TABLE OF CONTENTS

Abstract	iv
Introduction. Can Modality Influence Test Performance? Comparing Traditional and Digital Formats	1
A. Perceptions of Student Preparedness and Test Difficult	3
B. Types of Questions	5
C. Actual Differences in Testing Performance	7
D. Test Performance on Digital Modalities	8
E. Student Modality Preference	10
F. Testing Anxiety	11
G. Current Study Hypotheses	13
Method	14
A. Participants	14
B. Materials	15
a. Demographics	15
b. Exam 1 and 2	15
c. Preference	16
d. Attitudes Towards Computer Testing Questionnaire	16
e. Anxiety Scale	17
f. Devices used on Canvas	17
C. Procedure	18
Result	19
A. Performance	19
B. Time	20
C. Anxiety	21
D. Attitudes Towards Computer Based Testing	21
E. Preference	22
Discussion	22
References	31
Appendix	36

## **Abstract**

This paper presents a mixed group experimental design to examine whether how students take a test within the classroom affects exam performance and test delivery preferences. In this study, I randomly assigned 35 students in a social psychology course to take their first exam in either a paper-based format or an online mode in the same context. Students switched to the alternative format for their second exam. I measured average exam scores, speed of completion, attitudes towards computer testing, testing anxiety, and modality preference. As predicted, the testing format made no difference in mean test performance on each exam. There also was no majority preference for one modality over the other when given a choice of how students wanted to take the final exam in the class. Format preference was unrelated to testing anxiety or exam performance. However, attitudes towards computer-based testing appear to correlate with modality preference. Therefore, online-based testing modalities do not appear to have any significant disadvantages when compared to paper-based formats and can possibly serve as a convenient, resource saving alternative.

*Keywords:* Assessment, Testing Strategies, Computer-based Testing, Testing Anxiety

## **Introduction: Can Delivery Modality Influence Test Performance?**

### **Comparing Traditional and Digital Formats**

Assessment of academic knowledge has been a part of academia in the United States of America since the mid 1800s (National Educational Association, 2019). Over the past century knowledge assessments have developed into paper-based, and more recently computer-assisted, select option style assessments consisting of multiple-choice and true-false style questions. The current testing trend in academia involves transitioning to fully online mediums at both the collegiate and K-12 levels (Backes & Cowan, 2018). This trend has arisen to meet educational concerns such as speed in getting feedback to students. One of the major drawbacks of paper-based testing is the turnaround time necessary for instructors to hand grade or run test documents through a grading machine, which can lead to student exam feedback being delayed. The move to computer-based testing has several advantages such as faster, if not immediate, turnaround of exam scores, a reduction in student opportunities to cheat due to randomized exam item order, and additional flexibility being granted to instructors in how to design test items (Backes & Cowen, 2018).

Optical character recognition technology (OCR) revolutionized testing upon its widespread modernization by the company Scantron in 1972 (Cortez, 2016). OCR testing involves combinations of multiple-choice and true-false style questions in which the test taker must bubble in the correct answer choice on a provided Scantron answer sheet. The answer sheets are later run through a machine that compares each selected answer against a programmed answer key. OCR testing allows instructors in K-12 and university settings alike to administer, grade, and give feedback to students much more quickly than previously used hand-grading techniques (National Educational Association, 2019). According to Cortez (2016), in 2012 over

fifteen million students were estimated to be graded by OCR technology for testing purposes in higher education alone. With so many students being administered multiple exams per class per term, the implementation of efficient grading procedures is vital to ensure students receive exam feedback in a timely manner so they may better prepare for later material. OCR testing via Scantron represented the first step in the evolution of computer-assisted testing since machines completed the hard work of grading but the test taking medium remained an interaction that was paper-based. The more streamlined a grading method is, the less turnaround time between exam delivery and result delivery, regardless of class size (Paul et al., 1994). While OCR methods do decrease the length of time it takes instructors to grade examinations, the procedure itself does not ensure students receive timely feedback.

Switching to computer-based strategies facilitates a possible solution to the shortcomings of feedback speed imposed by traditional paper-based testing (Madeja, 2004). Not only are there clear advantages for test administration, but computer-based testing has been reported to be more popular with test takers than paper-based testing. Early research on computer-based testing found that test-takers have positive attitudes towards computer-based testing and perceived testing on the computer to be easier than testing on paper (Wang et al., 2008).

Many instructors have concerns that moving examinations away from traditional paper-based testing modes (ie, Scantron) will have adverse implications on student learning outcomes, including negatively affecting student performance in terms of exam scores. In response to these concerns, researchers have begun to analyze potential benefits and drawbacks of switching to online testing modes, but these concerns appear to be unfounded; for example, online testing modalities do not vary significantly from paper-based modes in terms of student performance or average exam scores in a university setting (VanPatten et al., 2015). VanPatten et al. compared

paper-based and online testing in a university level language course and found no differences in scores. The effect seems to be similar in K-12 populations (Poggio et al., 2005). A target population of 644 middle school students were randomly assigned to either a computer-based testing form or paper-based testing form, with results reflecting no differences in student exam performance. Backes and Cowen (2018) described research from the Partnership for Assessment of Readiness for College and Careers (PARCC) that analyzed the differences between testing modes on a sample of 538,375 K-12 students while controlling for differences in student academic levels. The PARCC found an insignificant decrease in scores when the exams were delivered using the online modality which disappeared in a majority of participant scores by the end of the second year of the study, suggesting that online modalities do not significantly affect exam score averages (Backes & Cowan, 2018).

This study seeks to determine whether or not the delivery modality of an academic examination can influence student performance and modality preference. How well students do on any given exam is the result of a combination of factors (Oz & Ozturan, 2018). Variables affecting student exam performance and exam format preference include perceptions of student preparedness for examination and of test difficulty, types of questions used on exams, student preparation (sleep, intensity of study), context issues (timing and location of testing), actual test difficulty, and testing anxiety.

### **Perceptions of Student Preparedness and Test Difficulty**

Research on student perceptions of preparedness for examination and of test difficulty as influences on testing performance focuses on measuring attitudes towards exam modalities, comparing students' perceived level of preparedness to their actual performance, and measuring student perceptions of question and exam difficulty. When comparing student perceptions of

their performance between computer-based and paper-based modalities, 77% of students in one study (Flowers et al., 2011) reported perceiving their performance as better on the computer-based modality than on the paper-based modality. Despite student perceptions, students' actual performance was not correlated with their perceived performance. The lack of correlation between perceived performance and actual performance suggests that student perceptions of preparedness are not related to actual student performance on academic examinations. Furthermore, student reports of perceived mental effort required to answer exam questions and perceived level of item difficulty do not vary significantly between online and paper testing modalities (Prisacari & Danielson, 2017). This finding further shows that student perceptions of exam item difficulty and overall exam difficulty also do not relate to actual testing performance. Nor do positive attitudes toward CBT correlate with increased performance on the computer-based modality over the paper-based modality (Ebrahimi, 2019). Positive perceptions towards computer-based testing do not significantly influence students' actual performance.

Carnegie (2017) found that 62.2% of students perceived answer trends when several questions in a sequence appear to have the same letter answer on an OCR answer form and may take a second look at the questions associated with the perceived answer trend. These trends do not influence students to alter the manner in which they select their answers; however, they do prompt some students to reevaluate their answer selections. Although the findings of this study do not support the idea that perceived patterns significantly influence student test performance, it does pose the question of what is actually responsible for testing difficulty. Perceptions of preparedness for examination nor perceptions of test difficulty appear to influence actual test performance; however, other factors, such as which styles of questions an examination consists of, may be an influence on student testing performance.



## Types of Questions

Current academic knowledge assessments used at the collegiate level tend to rely on multiple-choice style answers consisting of a stem statement with four or five possible answers. Another heavily used question type is the true-false question, which typically involves the test taker reading a statement and determining whether or not that statement is correctly or incorrectly written (Mafinejad et al., 2017). The downside of both multiple-choice and true-false questions is that they provide cues and context that allow test takers to narrow down their answer and ultimately guess when they do not know the answer, which fails to properly measure knowledge acquisition.

Brassil and Couch (2019) argued that a newer question type, multiple-true-false style (MTF) questions, allow for better measurement of student understanding of presented material. MTF questions are true false style questions consisting of multiple part mini-statements, each of which must be selected as either “true” or “false”. MTF questions are stylistically hybrid questions involving aspects of both traditional multiple-choice and traditional true-false style questions. The requirement of MTF questions is to break down and rate individual parts of a multi-part statement as true or false. The advantage of MTF questions is that they allow student understanding to be measured more accurately than traditional TF in which simply guessing can result in the correct answer. For example, students may earn partial credit by accurately selecting true for the first part of the question but miss out on the second part of the statement. Another major advantage of multiple-true-false style questions over traditional MC or TF questions is the lack of embedded cues that allow test-takers to narrow down possible answers almost immediately and then take a best guess at the correct answer. According to Brassil and Couch,

the student testing strategy of eliminating obvious distractor answers does not exemplify student understanding of the material.

However, the push for examinations procedures to move to online, digital formats dictates that multiple-choice style and true-false style questions continue to be used due to the ease of access in digital modalities. Couch et al. (2018) compared traditional multiple-choice questions to multiple-true-false questions. The results suggest that traditional multiple-choice style questions overestimate student mastery of material, with MTF questions offering a more comprehensive measure of student understanding of tested material.

In a study conducted by Brunfaut et al. (2018), an English language comprehension exam involving essay-style long response questions was compared across both paper-based and online testing modalities to find any differences in student performance when essay-style questions were delivered using an online testing modality. Their findings suggest that there are no significant performance differences between online and paper essay-style testing modalities in any of the participant groups. Brunfaut's study addresses possible concerns with how digital testing modalities may affect student performance on essay style questions, which tend to require much more critical thinking and articulation to answer fully.

Mafinejad et al. (2017) gave 159 students a critical thinking skills test consisting of multiple-choice, true-false, and short-response written essay questions. Written essay questions and traditional multiple-choice questions appear to have the strongest relationship with student critical thinking skills, which is an important factor involved in accurately assessing student acquisition of learned material.

The purpose of academic knowledge assessments is to measure student learning as effectively as possible using the tools available. The research done by both Brunfaut and

Mafinejad suggests that switching to online testing modalities does not negatively affect the measurement of student learning or student critical thinking abilities. These studies illustrate the many potential moderating variables involved in determining whether or not a particular style of question appropriately measures student learning. Using a variety of question styles may be the most appropriate method of measuring student learning; however, the types of questions used to measure student learning do not account for actual differences in student testing performance.

### **Actual Differences in Testing Performance**

Actual testing differences appear to be related to a handful of various factors including the time the exam takes place, the primary language of the test taker, and study habits. According to Groen and Pabilonia (2019), another factor that relates to exam performance is at what time an exam occurs during a school day. In their study, high school students who attend schools with later start times tended to get more sleep and achieve higher scores on reading tests. Later school start times may share more of a direct relationship with students receiving an adequate amount of sleep, which is a moderating variable associated with achieving higher test scores. Sievertsen et al. (2016) similarly found that students who take exams later in the day tend to perform marginally worse on their exams for every hour they have been awake. This finding suggests that student cognitive function in relation to the number of hours an individual has been awake is among the factors that actually affect exam performance.

Students who access online materials more frequently and for longer periods of time perform better on both online and paper-based exams than students who spend less time studying material, suggesting that student preparedness plays a role in student performance outcomes (Spivey & McMillan, 2014). As more academic testing transitions onto online mediums, it is important to understand what delivery factors actually affect student test performance.

## **Test Performance on Digital Modalities**

Current research shows that there are generally no statistically significant differences between student exam performance when comparing traditional paper-based testing methods with online-based testing methods (Candrlic et al., 2014). This outcome is generally found to be true at both the university and secondary educational levels; however, high school students tend to perform relatively better when taking examinations using online formats when compared to traditional paper-based examinations (Butters & Walstad, 2011).

Oz and Ozturan (2018) demonstrated that online-based examinations can be delivered to various testing groups and still achieve similar results regardless of the modality of exam delivery. Their study entailed randomly assigning 97 undergraduate students to take an academic knowledge examination using either a paper-based mode or a computer-assisted mode. Then the students were administered a survey on their attitudes towards computer-assisted testing. The difference in student performance was not statistically significant between paper-based and online-based testing formats, meaning that administering exams using online testing modalities is comparable to paper-based administration.

Candrlic et al. (2014) compared paper-based testing to online testing using digitally generated test forms created from test banks on an online learning management system. The study compared the results of 753 paper exams and 678 online exams from three separate university classes. The mean test scores did not vary significantly between groups. Students scored similarly across both paper-based and online testing modalities when comparing their scores on either format.

Khoshsima et al. (2017) compared the performance of Iranian English language learners using paper-based and computer-based assessments. The researchers were interested to see if

participant gender played a role in student performance on computer-based examinations. Statistical differences were absent when exams were delivered via online modalities, regardless of participant gender.

In terms of language comprehension, students taking examinations written in English tend to perform significantly better when their primary spoken language is English (Retnawati, 2015). Similar results have been found in other European and Asian cultures (Seo & Jong, 2015; Retnawati, 2015); regardless of the test-taker's culture, as long as the test being taken is in a familiar language, there should be no difference in performance due to format. According to VanPatten et al. (2015), the lack of differences between formats extends to the learning language as well; the culture and language of an examinee overall fails to produce differences in student performance when digital testing formats are used.

Priscari and Danielson (2017) showed that there are no statistical differences between paper and computer-based examinations even in subjects such as chemistry and other hard sciences. Steinberg et al.'s (2014) research compared examination scores between paper-based and computer-assisted modalities of the Pre-Professional Skills Test (PPST) taken to earn a teaching certification in America. The research by Steinberg was aimed to uncover any achievement gaps when using online-testing methods and contrary to Priscari and Danielson, found that students who were administered the computer-assisted exam performed significantly better than students who were administered the paper-based mode. According to Steinberg et al. (2014) the benefits of online testing also extends to standardized licensure examinations such as the Graduate Readiness Exam (GRE), Medical College Admission Test (MCAT), National Council Licensure Examination (NCLEX), and similar examinations. This finding demonstrates

that even high stakes testing can benefit from the advantages of online examinations since online-testing modalities do not disadvantage student performance.

The research findings on mode of delivery effects generally demonstrate that there are no significant differences between paper-based and online testing modalities with only a handful of studies finding either a marginally positive or negative difference in performance. With exam performance seeming to be a non-issue when switching to digital testing modalities, the next factor to consider is which modality students seem to prefer and what that preference is based on.

### **Student Modality Preference**

Current research shows that students tend to prefer online-based testing over paper-based testing due to its accessibility and ease of use (Amalia & Herlambang, 2017). Results from a study done by Ebrahimi (2019) suggest that test takers develop a more positive attitude towards the features of computer-based testing (CBT) after initial exposure to the modality, with 65% of participants stating they enjoyed taking the test in that format and 47.5% of participants stating that they believed the CBT to be a more accurate measure of the tested knowledge.

Cisar et al. (2012) created a 5-point likert-type questionnaire to measure participant exam modality preference. Over 82.23% of participants responded either “Agree” or “Strongly agree” to the statement “I prefer the computer-based test” (Cisar et al., 2012). These findings suggest that presentation on a computer screen has an effect on student satisfaction with computer-based testing. Further, when online format exam results are immediately shared with students, there is a significant correlation between both positive and negative emotional release (Daniels & Gierl, 2017). Immediate exam score feedback may be a determining factor in student preference for online-based testing formats, largely because most traditional paper-based examinations require turnaround time between exam delivery and result delivery.

Online-based examinations also benefit students who may have difficulty taking paper-based examinations. Online exams are easily navigable by students with disabilities documented by their reported preference for this format and perceived higher performance when using online examinations (Flowers et al., 2011). This advantage translates to students without disabilities as well, with exam navigation being as simple as scrolling with a mouse and clicking the desired answers (Fluck, 2019).

Due to increased ease of navigation on online-formats, students who take exams using online-based formats tend to complete exams six minutes faster on average when compared to paper-based formats (Bayazit & Askar, 2011). In accordance with Bayazit and Askar's research, decreased amount of time necessary to complete online examinations may be another factor involved in student preference towards online testing formats. Contrary to the research study conducted by Cisar et al. (2012), Boeve et al.(2015) found that fifty percent of students preferred paper-based examinations over computer-based examinations, with only about twenty-five percent of participants preferring computer-based examinations after taking both formats. Boeve et al. (2015) described the explanation of the majority for preferring the paper-based mode was the familiar structure of paper-based exams, which suggests that preference for online exam formats may be related to student comfort level with technology as a whole. Unfortunately, very few studies have contained a measure of student comfort level with technology, which may in fact be related to testing anxiety.

### **Testing Anxiety**

Testing anxiety generates a physiological stress response prior to or during an examination procedure and may be related to the duration, frequency, and effectiveness of student study strategies (Cipra & Muller-Hilke, 2019). In particular, the more comprehensive and

strategic student study strategies are, the less anxiety students report during testing procedures, which also correlates with higher examination scores. Their study involved surveying a sample of undergraduate medical students on their experienced anxiety levels in testing scenarios. The students' cortisol levels were measured before and after each examination and their exam results were compared to their self-reported anxiety levels. Students who scored higher on effectiveness in a study habits inventory tended to have lower ratings of testing anxiety and performed better on examinations than students who did not use strategic study habits. Their findings suggest that students who experience less testing anxiety prepare for examinations more effectively than their peers who report higher testing anxiety, suggesting that testing anxiety is less of an influence on testing performance, but rather an indicator of student perceptions of preparedness for an exam. This research prompts the question as to whether or not there is a relationship between testing anxiety and student perceptions of online testing modalities that affect student exam format preference.

Research on testing anxiety and its actual influence on student performance shows negligible results related to testing format. Students who report preferring examinations administered on paper rather than online state that the reason for their preference was additional anxiety associated with taking an exam using an online format (Washburn et al., 2017). Their reported anxiety was not related to the actual exam, but to issues perceived to be outside of the students' control, such as computer battery issues, internet associated issues, and inability to mark directly on electronic exam documents. Students' perceptions of additional testing anxiety when taking the exam online did not significantly affect mean scores on either testing modality.

Self-reported testing anxiety seems to be related to students having a lack of control over aspects of the testing procedure. In fact, Students who report high levels of testing anxiety have



been found to perceive greater control over the examination procedure when examinations allow students to select the difficulty of their test items as they are taking the exam. Wise (2019) tested students using the student adaptive tests (SAT) format during the student's selection did not actually modify the difficulty of the exam content, but did allow for greater perceived control over the examination, which alleviated student testing anxiety. However, the perception of greater control over the examination procedure through student perception of selecting the exam difficulty of questions did not appear to influence actual exam item difficulty or overall exam difficulty, indicating that adaptive digital testing modalities such as SATs mitigate student anxiety while maintaining the content and challenge of the examination.

Although student perception of control or lack thereof does not appear to significantly affect student performance, the research on student perception of control over exam proceedings and the relationship to testing anxiety begs the question of whether student modality preference and levels of testing anxiety in terms of comfort level with technology are related.

### **Current Study Hypotheses**

The main purpose of the current study is to determine if there are any significant differences in student test performance and exam format preference when a test is given electronically using a learning management system versus being given traditionally on paper using evaluated scantron technology. Previous research has primarily focused on testing differences between online and paper-based testing modalities, but has largely failed to examine format preference, especially in relation to testing anxiety. The current study compares paper-based and online-based testing modes to examine differences between formats in terms of average test scores, speed of exam completion, test anxiety levels, and attitudes towards computer testing.

I hypothesize that there will be no significant difference in performance between the two testing modalities, with average exam scores not varying statistically across either format. This hypothesis is grounded in the research of Oz & Ozturan (2018) and Candrljic et al. (2014) showing no differences in performance between online and paper testing modalities. Additionally, I hypothesize that students reporting higher levels of testing anxiety will perform worse on both exam formats than students who report minimal testing anxiety. This hypothesis reflects the findings of Cipra & Muller-Hilke (2019) that student reports of anxiety are related to student preparedness for examination. I hypothesize that students reporting higher testing anxiety and lower attitudes towards computer testing scores will prefer the paper-based exam over the online modality, just as the students in Washburn et al. (2017) preferred paper over online testing due to the possibility of technology failure. I also hypothesize that the online-based format will allow students to complete the exam more quickly than on paper, similar to the results of Bayazit and Askar's (2011) study in which online examinations took 6 minutes less on average to complete. Finally, I hypothesize that our results will mirror those of Cisar et al. (2012) with a majority of participants preferring the exam online over the paper and that attitudes towards computer testing will influence student exam format preference.

## **Method**

### **Participants**

The original sample consisted of 40 individuals. One participant was taking the course as a part of a senior citizen learning opportunity that did not require taking any examinations. One participant dropped the course prior to the first examination and was also dropped from the study. The remaining three excluded participants were students who had documented testing accommodations and were administered their examinations separately from the main class; thus

their data were withheld from the study. Participants were offered credit in the university's psychology research pool for participating in the study. Therefore, the final sample consisted of 35 student volunteers from an undergraduate level social psychology course at a Southeastern regional comprehensive university. The sample contains 11 males and 24 females ranging in age from 18 to 52 ( $M = 22.03$ ,  $SD = 6.34$ ). 71.14% of participants indicated being psychology majors. The sample consisted of primarily Caucasian participants (63%) with the second predominant group being African Americans (17%). The remaining participants reported either mixed race (17%) or Asian (3%). The mean GPA reported by participants was 3.0857 with a standard deviation of 1.04. Most students stated that they were in their third year of school ( $M = 2.91$ ,  $SD = 0.82$ ).

## **Materials**

***Demographics Form.*** A demographic form administered at the beginning of the study captured information such as gender, age, race, major, year in school, and current GPA range. Gender, race, age, and major were fill-in-the-blank format, with year in school and GPA range requiring participants to check the most appropriate response from a selection of ranges (ie, Freshman, Sophomore, Junior, Senior).

***Exams 1 and 2.*** Both exams were traditional 50 item multiple-choice question tests. Each item allowed for four possible answers to each question with only one answer per question being correct. Both exams consisted of conceptual, definitional, and application questions. Exam 1 included topics such as research methods, social judgement, and behavior and attitudes. The second exam included topics such as conformity, persuasion, and group influence. The exam was delivered within the same classroom environment in which the participants attended lectures. Exams were given in either paper format with a scantron sheet provided on which participants

could mark and write on their test forms or on the university's eLearning infrastructure "Canvas." The Canvas format exams were identical to the paper format exams in terms of content and structure. The questions were delivered in the same order in both formats. Both exams consisted of instructor-made questions directly relating to the textbook and lectures. Both pen and paper and Canvas times were recorded by the test administrator using a designated digital clock to denote start time for the exam and by marking the end time on each paper test form as they are handed in.

**Preference.** Exam format preference was measured simply by allowing participants to choose their preferred format for the third and final exam. Students received a simple preference choice form that consisted of two questions: "What is your name?" and "Which format would you like to take the final exam on?" The preference form was administered along with the Attitudes Towards Computer Testing Questionnaire.

**Attitudes Towards Computer Testing Questionnaire.** As an additional component of measuring preference, the Attitudes Towards Computer Testing Questionnaire validated participant exam choice as a clear indicator of preference. The ATCT-Q is an adaptation of Cisart et al's (2012) questionnaire on attitudes towards testing. The adaptation is a ten-item questionnaire using a five-point Likert scale with responses ranging from 1 *strongly disagree* to 5 *strongly agree* with a neutral point of 3 *neither agree nor disagree*. The items on this questionnaire are adapted to measure whether participants' attitudes towards computer-based testing are positive or negative. Items entailed comparisons between paper-based and computer-based testing with participant responses indicating whether or not they agreed with the statement. Examples of statements are "Taking an online exam in class is more convenient than taking a paper exam in class"; "It was easier to read the questions on paper than on canvas"; and "It was easier to review my answers

on canvas compared to on paper and pencil.” Items varied in their possible framing of whether or not paper or computer testing was preferential. This scale had a Cronbach’s Alpha of .912.

**Anxiety Scale.** The Test Anxiety Questionnaire by Nist and Diehl (1990) is a ten-item Likert scale with answers ranging from 1 *never* to 5 *always*. The ten items on the questionnaire are statements that require the participant to indicate how often they relate to the particular statement. All items are directed to measure general testing anxiety with 50 being the highest possible score, indicating the highest level of anxiety felt towards testing. Nist and Diehl’s (1990) 10 item testing anxiety scale was subjected to reliability testing and was found to have a Cronbach’s alpha of .868, suggesting that all 10 items were consistently reliable with one another at an internal level.

**Devices used on Canvas.** I asked students to supply their own device to use for their computer-based exam format, including a laptop (Windows or Macintosh), a tablet of their choice (either Android or Apple), or a cellular smartphone (Android or Apple). Participants downloaded the university eLearning Canvas infrastructure phone application if taking the exam on a tablet or phone. Participants taking the exam on a laptop used a standard web browser. Participants had the option of checking out a laptop from the library if they didn’t have their own computer.

### **Experimental Design**

Participants were randomly assigned to either the control group (paper) or the experimental group (online) using a random number generator. This experiment is of mixed design inspired by a similarly designed study by Washburn et al. (2017). The current study counterbalances test performance as the within factor and testing method as the between factor. The independent variable for both within and between groups is testing format (either Canvas or Paper/Scantron) with student performance compared to themselves across formats acting as the

within dependent variable and student performance compared across the control and experimental group acting as the between dependent variable. The dependent variables are broken down into measures of speed of completion time, actual test scores, and format preference. I also sought to measure the possible covariates testing anxiety rating and attitude towards computer testing score.

## **Procedure**

Participants completed an informed consent form in compliance with university protocol. Please see the Appendix for further information on institutional review board approval. Participants also filled out a demographics form inquiring their year in school, major, GPA, age, gender, and race. Prior to taking any examinations the participants took an adaptation of the 10 item Test Anxiety Questionnaire by Nist and Diehl (1990). This scale was also compared to participant exam performance on both exam formats to determine if general testing anxiety affected performance on either or both formats. Measurement of format preference was conducted using an adaptation of the Attitudes Towards Computer Testing used in a similar study by Cisar et al. (2012).

The study involved the use of two 50-item multiple choice tests with exam 1 being administered during week 5 and exam 2 being administered during week 9 of the semester. Each exam was allotted an hour and fifteen-minute time block. For the first exam, half of the participants were randomly assigned to the control group (paper & scantron testing) and half were assigned to the experimental group (online testing) using a random number generator. The instructor told the students to be prepared to take the first exam using either format and assigned their testing method the day of the first examination. For the second 50-item exam participants received the opposite testing format from their first testing experience. The final exam format

was selected by the participants with their choice representing a measurement of student satisfaction; however, data from the third exam were not used in terms of performance or completion time. Students taking the exam on Canvas were asked which device (smart phone, tablet, laptop) they used to take the exam as an additional item by the exam proctor upon turning in the exam; this choice did not enter into the data analysis. Completion time in minutes was recorded by the experimenter as students handed in paper exam forms or showed the online exam completion screen.

Students also completed an adaptation of the 10 item Attitudes Towards Computer-Testing Questionnaire (ATCT-Q) as an additional measure of student satisfaction at the same time preference was polled after completion of both exam formats during week 10 of the semester. Data were collected in the form of speed of completion time (in minutes), actual test scores, and preference for testing format. I measured testing preference by allowing students to select their preferred format for their third examination one week prior to the third exam as well as with the adaptation of the ATCT-Q. Participants were debriefed following the third examination.

## **Results**

### **Performance**

My first hypothesis stated that there would be no significant difference in performance between the two testing modes in terms of average scores. In order for student performance to be compared between the two exams, students' actual test scores were converted into standardized scores, or  $z$ -scores, for all comparisons.  $z$ -scores from the first exam were compared between Paper-based and computer-based testing modes using an independent samples  $t$ -test. The results found that students in the exam 1 paper condition ( $M = .272$ ,  $SD = .934$ ) scoring relatively better

than the online exam condition ( $M = -.257$ ,  $SD = 1.017$ ),  $t(33) = 1.597$ ,  $p = .12$ ,  $d = 0.542$ , 95CI [-.14, 1.20]. The same comparison was made for the second exam using an independent sample t-test. There were no differences between the exam 2 paper condition ( $M = .108$ ,  $SD = 1.074$ ) and the exam 2 online condition ( $M = -.115$ ,  $SD = .935$ ) which supported my hypothesis,  $t(31) = .635$ ,  $p = .53$ ,  $d = 0.221$ , 95CI [-0.49, 0.94].

To ensure the results of my previous comparisons of the exams separately were accurate, I conducted a paired samples t-test using combined exam one z-scores and exam two z-scores from both modalities. Exam one and two combined paper scores ( $M = .142$ ,  $SD = .982$ ) were found to vary from the combined online scores ( $M = -.1736$ ,  $SD = .978$ ),  $t(32) = 2.45$ ,  $p = .020$ ,  $d = 0.322$ , 95CI [0.05, 0.58]. In terms of my performance hypothesis, the results of the paired samples t-test suggest that when comparing individual student performance to that of the entire sample, students tended to perform relatively better (roughly 1-3 points higher) when taking the exam on paper.

## **Time**

I hypothesized that the online-based format would allow students to complete the exam more quickly than on paper. Completion time analysis began with converting time measured in minutes into z-scores to standardize values. An independent t-test comparing paper time z-scores ( $M = .217$ ,  $SD = .955$ ) and online time z-scores ( $M = -.183$ ,  $SD = 1.079$ ) was conducted with times across formats showing significant differences,  $t(32) = -2.62$ ,  $p = .013$ ,  $d = 0.449$ , 95CI [-0.71, -0.9]. Students took approximately half a standard deviation longer (3-5 minutes on average) when completing an exam using the paper/scantron format than the online Canvas format, thus supporting my time related hypothesis that the online-based format will allow students to complete the exam more quickly than on paper.



## **Anxiety**

I hypothesized that students who reported higher levels of testing anxiety would perform worse on both exam formats than students who reported minimal testing anxiety. A Pearson's correlation test was conducted to determine the relationship between students' reported testing anxiety and student performance, as well as testing anxiety and format preference. Results were inconclusive in terms of my hypothesis, with correlations between mean anxiety score and paper score,  $r = -.125$ ,  $p = .473$  and mean anxiety score and mean canvas score showing no significance,  $r = -.193$ ,  $p = .282$ . Results were also inconclusive in terms of testing preference, with anxiety mean and preference mean showing no significance,  $r = -.278$ ,  $p = .106$ . My results did not allow us to draw a clear conclusion regarding self-reported testing anxiety and exam performance; there does not seem to be a significant relationship between the two variables.

To assess whether or not reported anxiety levels influenced student format preference, an independent t-test was run comparing mean scores on the anxiety scale to students' self-stated format preference. Those who preferred paper ( $M = 2.522$ ,  $SD = .786$ ) did not report significantly different anxiety levels than those who preferred canvas ( $M = 2.359$ ,  $SD = .607$ ),  $t(33) = .686$ ,  $p = .315$ ,  $d = 0.232$ , 95% CI [-0.32, 0.65]. The results of this t-test were in line with the Pearson's correlation, suggesting that there may not be a significant relationship between reported testing anxiety and exam format preference.

## **Attitudes Towards Computer-based Testing**

I hypothesized that student attitudes towards computer testing would influence student exam format preference. An independent sample t-test was run to determine the relationship

between attitudes towards computer testing and student format preference. In terms of format preference my results did not find a majority of participants to prefer one format over the other. Those who preferred the paper format ( $M = 2.094$ ,  $SD = .554$ ) and those who preferred the canvas format ( $M = 3.753$ ,  $SD = .476$ ) varied significantly in their attitudes towards computer based testing,  $t(33) = -9.479$ ,  $p = .000$ ,  $d = 3.21$ , 95% CI [-2.01, -1.30]. Students that preferred the online format indicated strong attitudes towards CBT with students preferring the paper modality indicating strong attitudes against CBT. Based on the effect size, I believe that students' attitudes towards computer testing affirmed my hypothesis that students' attitudes towards computer testing would influence student format preference.

### **Preference**

I hypothesized that student performance would influence their preferred testing modality. An independent sample t-test was conducted comparing overall paper z-scores to students' stated format preference. Those who preferred paper ( $M = .3643$ ,  $SD = .882$ ) did not score significantly differently on paper than on canvas ( $M = -.0487$ ,  $SD = 1.08$ ),  $t(33) = 1.24$ ,  $p = .72$ ,  $d = 0.418$ , 95% CI [-0.26, 1.09]. Another independent t-test was run comparing overall canvas z-scores to students' format preference. Those who preferred paper ( $M = -.409$ ,  $SD = .690$ ) did not score significantly differently on canvas ( $M = .077$ ,  $SD = 1.18$ ),  $t(31) = -1.45$ ,  $p = .131$ ,  $d = 0.502$ , 95% CI [-1.17, 0.20]. These results suggest that student performance is not related to format preference.

### **Discussion**

In summary, my research question encompasses whether there are any differences in student performance and format preference when examinations are delivered using online-based and paper-based methods. My hypotheses included that there will be no significant difference in

performance between modalities, with average scores remaining similar in either mode. Also, students who report higher levels of testing anxiety will perform worse on both formats in comparison to students reporting minimal testing anxiety. I also hypothesized that students with higher testing anxiety scores and lower attitudes towards computer testing scores would prefer the paper-based examination to the online. The online format was hypothesized to allow students to complete the exams more quickly than on paper. I also hypothesized that the majority of students would share preference for the online modality and that attitudes towards computer-testing would influence format preference.

Contrary to my hypothesis that there would be no statistically significant differences in performance between formats, students performed relatively better when taking the paper condition when individual performance was compared to that of the whole sample. My finding varies with those of research mentioned in Candrljic et al. (2014) and Oz and Ozturan (2018); both studies compared paper-based and online-based exam modalities and found that student exam scores did not vary significantly between modalities. My finding of a relative difference in performance differs from previously conducted research, which prompts the question as to what factors influence relative performance differences between formats.

My finding of a relative performance difference between formats is opposite to the findings of Butters and Walstad (2011), in which students tended to perform relatively better using online modalities when compared to paper. Unlike the current study, Butters and Walstad's consisted of a different age participant group (eighth and ninth graders), with a majority of the online exam modality participants being the older aged ninth graders. I speculate that students introduced to digital testing mediums at an earlier age will adapt to it more readily, particularly if introduced during a critical period of development. The overall implication of my results

regarding student performance is that using online-based modalities may have a relative impact on the assessment of student learning, although it is possible that students introduced to online-testing mediums at a younger age may adapt to it more easily than students introduced to it later in life.

Completion time between exam formats did vary relatively between conditions. This finding is nearly identical to the findings of Bayazit and Askar's research (2011); they found that students take around 6 minutes longer when using paper-based testing methods. They speculated that the increased time spent on the paper mode may be in relation to participant familiarity with the testing method. Similar to Poggio et al's (2005) study on format differences, the current study's online exam forms were exact reflections of their sister paper forms. Due to the simplicity of the online-medium and the similarity to the paper-based format, it is not likely that student familiarity with the online-format played a role in the relative time differences.

A possible explanation for the relative difference in format times may be the affordances of paper-based examination methods that online testing modalities lack. Paper-based exams allow the test-taker the ability to mark through incorrect answers, circle or highlight key words in the questions, and write-notes in the margins, which are common strategies for taking multiple-choice exams that aid in increased student performance outcomes (Boeve et al., 2015). The ability to mark on paper-based exams and the need to double-check correct marking of answers on optical character recognition sheets may contribute to the additional time spent with the paper-based exams over the computer-based exams. Affording students the ability to mark on digital test forms may aid in ensuring that student learning is appropriately measured by online testing methods.

My findings did not support my hypothesis that students reporting higher levels of testing anxiety would perform worse on both exam formats than students reporting minimal testing anxiety, nor was my hypothesis that students reporting higher testing anxiety and lower attitudes towards computer-based testing scores would prefer the paper-based exam modality supported. My results do not allow us to draw clear conclusions about the impact of testing modality as a moderator of testing anxiety, because there did not appear to be a relationship between performance on either format and reported levels of testing anxiety. Cipra and Muller-Hilke (2019) suggested that participants who report higher testing anxiety actually indicate a perception of lesser preparedness for examination rather than actual higher levels of cortical stress response in the body. Since my study contained only a global measure of testing anxiety at the beginning of the study rather than a measure delivered immediately after participants were assigned their exam modality, I can only assume that the results of the current study may be reflective of the findings of Cipra and Muller-Hilke that student self-reports of anxiety are merely indicative of initial student perceptions of their own preparedness for examination. With neither of my anxiety hypotheses being supported, it is likely that rather than being related to preference, that testing anxiety is a moderator of another variable such as the degree to which students have prepared for examination procedures.

The hypothesis that students reporting more positive attitudes towards computer-based testing would prefer the online modality was supported; it does appear that student attitudes towards computer-based testing influence student modality preference. Students who scored more favorable attitudes towards computer-based testing were more likely to indicate a preference for the computer-based modality. Students who had negative attitudes towards CBT scored lower on the attitudes scale and were more likely to indicate that they preferred the paper-

based modality. My finding mirrors the findings of Ebrahimi (2019) that test takers with previous exposure to computer-based testing modalities develop a more positive attitude towards the modality as a whole, although I cannot draw any conclusions without further research. In terms of best measuring student learning, student attitudes towards computer-based testing do not indicate that online-based testing methods are any less preferable than paper-based testing.

My findings resulted in a nearly 50-50 split between exam preference, which is similar to Boeve et al's (2015) finding that fifty percent of students prefer paper-based examinations. Cisar et al. (2012) suggested that participants tend to prefer one format over the other. Consequently, this finding did not support my hypothesis that a majority of students would prefer one modality over the other. My results regarding preference are contrary to previous research findings suggesting that students tend to have a preference for one modality over the other (Amalia & Herlambang, 2017).

With no significant relationship between preference and actual or perceived exam performance, it is possible that other variables act as a moderator of student comfortability with online testing such as student cognitive power or student preparedness. For example, it could be that students who are especially thoughtful and tend to perform well in academic work may be more inclined to interact physically with the testing materials through marking out incorrect answers or circling keywords in exam questions and as such, may be more comfortable with paper-based testing methods. Measuring the relationship between student thoughtfulness and physical interaction with exam forms may be done through comparison of consistently high performing students with lower performing students by constructing a variable that measures the number of words or markings a student makes on an actual paper exam form compared to the number of markings or words a student writes on a piece of scratch paper when taking a digital

exam. Students who consistently perform well may produce more marks or words on their exam forms than lower performing students.

These results also indicate that there are no significant differences in format preference, although student attitudes towards computer-based testing are indicative of format preference. The connection between student attitudes and format preference signifies a possible relationship between students who are not technologically inclined and a preference for paper-based testing modalities that allow them to more accurately show their competency with the material. In contrast, individuals who have grown up in more digitally involved environments may be more fluent at interacting with digital test modes.

Despite the relative differences in performance between paper-based and online testing modalities in this research, instructors may still want to consider transitioning examinations onto their institution's digital learning management system. Transitioning examinations onto learning management systems can drastically cut the cost of testing at the departmental level by eliminating the need for expensive OCR testing forms such as Scantron forms (which cost 15-20 cents per sheet in 2012 (Cortez, 2016)) as well as the substantial amount of paper needed to administer several examinations each semester. The funds saved by going digital could be reallocated to other worthy expenses.

Several design limitations also influence interpretation of the results. Although I did poll students as to which device they used to take their online exams (smart phone, tablet, or laptop), I did not consider whether or not device choice for the online modality would influence exam performance or preference. The inability to supply all participants with an identical device for online testing may have potentially influenced the results in terms of relative performance. Differences in devices such as screen size and input method (touch versus keyboard) may affect

student ability to successfully read and navigate online exams (Bayazit & Askar, 2012). Ideally, all students would be able to take the computer-based examinations on uniform devices. If possible, a replication of the current study in which all students use uniform testing devices capable of digital marking of test documents (ie, iPads with Apple pencils or Microsoft Surfaces with Stylus) may indicate a clearer picture as to if testing device influences either exam performance or preference. In future studies, an alternative to supplying identical devices may simply be providing scratch paper alongside the computer-based exams to supplement physical test form interaction.

Additionally, there was an unintended difference between exams one and two that may have influenced participant perception of exam difficulty between exams. Although the instructor intended for both exams one and two to be on a similar level of difficulty, there were several questions on the second exam that needed to be excluded from the final results based on the post-examination item analysis of the questions. The eliminated questions did not create a statistical difference between exam one and exam two scores due to the raw exam score conversion into z-scores; however, this variation may have created a perception within the participants that the second exam was much more difficult than the first. Although students did not formally report perceptions that the second exam was more difficult than the first in my rating scales, multiple students verbally commented to both the instructor of the course and the exam proctor that the second exam was more challenging than the first exam. The possible perception of examination difficulty may have influenced in part student attitudes towards examination format depending on which format had been used. However, I did not anticipate this problem or collect perceptions in a formal way to clarify whether the difference could produce a confounding influence.



Future research would benefit from an addition of a scale measuring student comfort level with technology, which would provide greater context to student attitudes towards computer-based testing. Student comfort level may be a better indicator of why students may prefer paper-based testing modalities versus online mediums and vice versa, which may possibly be moderated by individual differences mentioned earlier such as student cognitive power or student preparedness for exams.

The next research question should consider how the testing environment itself (classroom versus home environment) affects testing performance and retention of material. Administering examinations in the home environment may lead to a decrease in performance due to in home distractions such as pets, roommates, or alternative activities.

Another possible research question involves whether level of student cognitive power influences desire for physical manipulation and interaction with testing forms. Students who have higher cognitive power may prefer the ability to mark on test forms as a method of critical thinking.

A final question ponders to what degree student preparedness for examination relates to student comfortability with technology. Student preparation for examination in some cases involves student reception of learning materials through digital mediums. Perhaps students who spend more time preparing for examinations using digital means would report a higher comfort level when taking examinations online.

In conclusion, the increasing norm for learning management systems capable of constructing and moderating online exam delivery being used for fully online and hybrid courses dictates that further research in the field focus on improving the manner of which student performance is gauged using digital mediums. On a practical level, the finding that paper-based

and digital formats are relatively different in terms of performance suggests that instructors opposed to technological integration should not feel a need to convert to digital testing modalities as it does not appear to produce advantages in terms of student testing outcomes. Similarly, instructors that readily adopt technology may be able to embrace digital test administration for the convenience and resource saving that accompany the format.

## References

- Amalia, F., & Herlambang, A.D. (2017). Acceptance of online testing as an alternative evaluation tool in higher education. *Institute of Electrical and Electronics Engineers*, 127-136. <https://doi.org/10.1109/ICACISIS.2017.8355022>
- Backes, B., Cowan, J. (2018). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, 68, 89-103. <https://doi.org/10.1016/j.econedurev.2018.12.007>
- Bayazit, A., Askar, P. (2011). Performance and duration differences between online and paper-pencil tests. *Asia Pacific Educational Review*, 13, 219-226. <https://doi.org/10.1007/s12564-011-9190-9>
- Boeve, A.J., Meijer, R.M., Albers, C.J., Beestma, Y., & Bosker, R.J. (2015). Introducing computer-based testing in high stakes exams in higher education: Results of a field experiment. *PLOS ONE*, 10, 1371-1383. <https://doi.org/10.1371/journal.pone.0143616>.
- Brunfaut, T., Harding, L., & Batty, A.O. (2018). Going online: The effect of mode of delivery on performances and perceptions on an english L2 writing test suite. *Assessing Writing*, 36, 3-18. <https://doi.org/10.1016/j.asw.2018.02.003>
- Butters, R.B., & Walstad, W.B.(2011). Computer versus paper testing in precollege economics. *The Journal of Economic Education*, 42, 366-374. <https://doi.org/10.1080/00220485.2011.606087>
- Candrlic, S., Asenbrenner, K., & Dlab, M. H. (2014). Online vs. paper-based testing: A comparison of test results. *MIPRO 2014 Conference* (pp. 657-662). Opatija: Croatia. Retrieved on January 20, 2019 from <https://ieeexplore.ieee.org/document/6859649>
- Carnegie, J.A. (2017). Does correct answer distribution influence student choices when writing

- multiple choice examinations? *The Canadian Journal for the Scholarship of Teaching and Learning*, 8, 1-13. Retrieved on September 13, 2019 from [https://ir.lib.uwo.ca/cjsotl\\_rcacea/vol8/iss1/11/?utm\\_source=ir.lib.uwo.ca%2Fcjsotl\\_rcacea%2Fvol8%2Fiss1%2F11&utm\\_medium=PDF&utm\\_campaign=PDFCoverPages](https://ir.lib.uwo.ca/cjsotl_rcacea/vol8/iss1/11/?utm_source=ir.lib.uwo.ca%2Fcjsotl_rcacea%2Fvol8%2Fiss1%2F11&utm_medium=PDF&utm_campaign=PDFCoverPages)
- Cipra, C., & Muller-Hilke (2019). Testing anxiety in undergraduate medical students and its correlation with different learning approaches. *PLOS ONE*, 14, 1-11. <https://doi.org/10.1371/journal.pone.0210130>
- Cisar, S. M., Pinter, R., Radosav, D., & Cisar, P. (2012). Comparison of examination methods based on multiple-choice questions using personal computers and paper-based testing. *JITA*, 2, 22-31. <https://doi.org/10.7251/JIT1201022C>
- Cortez, M. B. (2016, December). *What is a scantron: How scantron technology paved the way for testing tools*. EdTech. Retrieved on October 24, 2019 from <https://edtechmagazine.com/higher/article/2016/12/scantron-technology-creates-more-efficient-testing>
- Cushen, P.J., Hackathorn, J., Vazquez Brown, M.D., Rife, S.C., Joyce, A.W., Smith, E.D., Bordieri, M.J., Anderson, P.W., Karlsson, M.E., & Daniels, J. (2019). “What’s on the test?”: The impact of giving students a concept-list study guide. *Teaching of Psychology*, 46, 109-114. <https://doi.org/10.1177/009862831984171>
- Daniels, L. M., & Gierl, M. J. (2017). The impact of immediate test score reporting on university students’ achievement emotions in the context of computer-based multiple choice exams. *Learning and Instruction*, 52, 27-35. <https://doi.org/10.1016/j.learninstruc.2017.001>
- Ebrahimi, M. R. (2019). Score equivalence, gender differences, and testing mode preference in a comparative study between computer-based testing and paper-based testing. *International*

- Journal of Emerging Technologies*, 14, 128-142. <https://doi.org/10.3991/ijet.v14i07.10175>
- Flowers, C., Kim, D., Lewis, P., & Davis, V.C. (2011). A comparison of computer-based testing and pencil-and-paper testing for students with a read-aloud accommodation. *Journal of Special Education Technology*, 26, 1-13. Retrieved on May 20, 2019 from [http://explore.bl.uk/primo\\_library/libweb/action/display.do?tabs=detailsTab&gathStatTab=true&ct=display&fn=search&doc=ETOCRN289281172&indx=1&recIds=ETOCRN289281172](http://explore.bl.uk/primo_library/libweb/action/display.do?tabs=detailsTab&gathStatTab=true&ct=display&fn=search&doc=ETOCRN289281172&indx=1&recIds=ETOCRN289281172)
- Fluck, A.E. (2019). An international review of eExam technologies and impact. *Computers & Education*, 132, 1-15. <https://doi.org/10.1016/j.compedu.2018.12.008>
- Khoshsima, H., Hosseini, M., & Toroujeni, S. M. H. (2017). Cross-mode comparability of computer-based testing (CBT) versus paper-pencil based testing (PPT): An investigation of testing administration mode among Iranian intermediate EFL learners. *English Language Teaching*, 10 (2), 23-31. <https://doi.org/10.5539/elt.v10n2p23>
- Madeja, S. S. (2004). Alternative assessment strategies for schools. *Arts Education Policy Review*, 105, 3-13. Retrieved on May 23, 2019 from <https://eric.ed.gov/?id=EJ704725>
- Mafinejad, M. K., Arabshahi, S. K., Monajemi, A., Jalili, M., Soltani, A., & Rasouli, J. (2017). Using multi-response format test in the assessment of medical students' critical thinking ability. *Journal of Clinical and Diagnostic Research*, 11, 10-13. <https://doi.org/10.7860/JCDR/2017/24884.10607>
- National Education Association. (2019). History of standardized testing in the United States. Retrieved on October 24, 2019 from <http://www.nea.org/home/66139.htm>
- Nist, P, Diehl, M. (1990). Test Anxiety Questionnaire. Retrieved on January 20, 2019 from [https://web.ccsu.edu/crp/precollegiate/eop/files/LESSON\\_5\\_-](https://web.ccsu.edu/crp/precollegiate/eop/files/LESSON_5_-)

[TEST ANXIETY QUESTIONNAIRE.pdf](#)

- Oz, H., & Ozturan, T. (2018). Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies, 14*, 67-85. Retrieved on May 23, 2019 from <http://www.jlls.org/index.php/jlls/article/view/878/345>
- Paul, J., Lewis, S. B., & Supon, P. A. (1994). Improving education through improved assessment. *IEE Frontiers in Education Conference 1994* (pp. 17-21). San Jose, CA, USA. Retrieved on May 20, 2019 from <https://ieeexplore.ieee.org/document/580459>
- Poggio, J., Glasnapp, D.R., Yang, X., Poggio, A.J. (2005). A comparative evaluation of score results from computerized and paper & pencil mathematics testing in a large scale state assessment program. *The Journal of Technology, Learning, and Assessment, 3*, 1-31. Retrieved on May 23, 2019 from <http://jtla.org>.
- Prisacari, A. A., & Danielson, J. (2017). Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use. *Computers in Human Behavior, 77*, 1-10. <http://doi.org/10.1016/j.chb.2017.07.044>
- Retnawati, H. (2015). The comparison of accuracy scores on the paper and pencil testing vs. computer based testing. *The Turkish Online Journal of Educational Technology, 14*, 135-142. Retrieved on May 23, 2019 from <http://www.tojet.net>.
- Seo, D.G., & Jong, G.D. (2015). Comparability of online- and paper-based tests in a statewide assessment program: Using propensity score matching. *Journal of Educational Computing Research, 52*, 88-113. <https://doi.org/10.1177/0735633114568856>
- Spivey, M.F., & McMillan, J.J. (2014). Classroom versus online assessment. *Journal of Education for Business, 89*, 450-456. <https://doi.org/10.1080/08832323.2014.937676>

- Steinberg, J., Brennehan, M., Castellano, P.L., & Miller, S. (2014). A comparison of achievement gaps and test-taker characteristics on computer-delivered and paper-delivered *Praxis I* Tests. *ETS Research Report Series, RR-14-35*.  
<https://doi.org/10.1002/ets2.12033>
- VanPatten, B., Trego, D., & Hopkins, W.P. (2015). In-class vs. online testing in university-level language courses: A research report. *Foreign Language Annals, 48*, 659-668.  
<https://doi.org/10.1111/flan.12160>
- Wang, S., Jiao, H., Young, M., Brooks, T., & Olson, J. (2008). Comparability of computer-based testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement, 68*, 5-24.  
<https://doi.org/10.1177/0013164407305592>
- Washburn, S., Herman, J., & Stewart, R. (2017). Evaluation of performance and perceptions of electronic vs. paper multiple-choice exams. *Advanced Physiological Education, 41*, 548-555. <https://doi.org/10.1152/advan.00138.2016>
- Wise, S. L. (2019). Controlling construct-irrelevant factors through computer-based testing: disengagement, anxiety, & cheating. *Educational Inquiry, 10*, 21-33.  
<https://doi.org/10.1080/20004508.2018.1490127>

## Appendix A

### Institutional Review Board Approval



Mr. Gage Moyer

February 04, 2019

Dear Mr. Moyer:

The Institutional Review Board (IRB) for Human Research Participants Protection has completed its review of your proposal number IRB 2019-122 titled, "A Comparison of Traditional versus Computer Based Testing in a Social Psychology Course," as it relates to the protection of human participants used in research, and granted approval for you to proceed with your study on 01-27-2019. As a research investigator, please be aware of the following:

- \* You will immediately report to the IRB any injuries or other unanticipated problems involving risks to human participants.
- \* You acknowledge and accept your responsibility for protecting the rights and welfare of human research participants and for complying with all parts of 45 CFR Part 46, the UWF IRB Policy and Procedures, and the decisions of the IRB. You may view these documents on the Research and Sponsored Programs web page at <http://research.uwf.edu>. You acknowledge completion of the IRB ethical training requirements for researchers as attested in the IRB application.
- \* You will ensure that legally effective informed consent is obtained and documented. If written consent is required, the consent form must be signed by the participant or the participant's legally authorized representative. A copy is to be given to the person signing the form and a copy kept for your file.
- \* You will promptly report any proposed changes in previously approved human participant research activities to Research and Sponsored Programs. The proposed changes will not be initiated without IRB review and approval, except where necessary to eliminate apparent immediate hazards to the participants.
- \* **You are responsible for reporting progress of approved research to Research and Sponsored Programs at the end of the project period 01-26-2020. If the data phase of your project continues beyond the approved end date, you must receive an extension approval from the IRB.**
- \* If using electronic communication for your study, you will first obtain approval from the authority listed on the following web page:  
<https://uwf.edu/offices/institutional-communications/resources/broadcast-distribution-standards/>.

Good luck in your research endeavors. If you have any questions or need assistance, please contact Research and Sponsored Programs at 850-857-6203 or [irb@uwf.edu](mailto:irb@uwf.edu).

Sincerely,



Dr. Matthew Schwartz, Interim Assistant Vice President  
Research Administration



Dr. Carla Thompson, Chair, IRB for  
Human Research Participant Protection