

THE EFFECT OF DIGITAL ASSESSMENT IN A
HIGH SCHOOL CHEMISTRY CLASSROOM

by

Benjamin Jasper Swen

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2020

©COPYRIGHT

by

Benjamin Jasper Swen

2020

All Rights Reserved

DEDICATION

I would like to dedicate this paper to my loving wife Nicole, son William, and daughter Lily, who have sacrificed much time to help fill my bucket. To my parents who have always stood behind me and supported me with love and guidance.

TABLE OF CONTENTS

1. INTRODUCTION AND BACKGROUND	1
Context of the Study	1
Focus Question.....	2
2. CONCEPTUAL FRAMEWORK	3
Theoretical Support.....	3
Advantages and Disadvantages of Online Assessment	3
Online Versus Paper-Format Assessment.....	4
Question Format.....	5
Comparison of Methodologies.....	5
3. METHODOLOGY	9
Demographics	9
Treatment	10
Data Collection and Analysis Strategies.....	11
4. DATA ANALYSIS.....	14
Results.....	14
Student Perception	14
Student Performance	18
5. CLAIM, EVIDENCE, AND REASONING	25
Claims from the Study	25
Student Performance	25
Student Perception	26
Impact of Action Research on the Author	27
Value of the Study and Consideration for Future Research	28
REFERENCES CITED.....	30
APPENDICES	33
APPENDIX A: Institutional Review Board Exemption.....	34
APPENDIX B: Pre-Treatment Likert Survey.....	36
APPENDIX C: Post-Treatment Likert Survey	38
APPENDIX D: Post-Treatment Interview Questions.....	40

LIST OF TABLES

Table	Page
1. Data Triangulation Matrix	10

LIST OF FIGURES

Figure	Page
1. Student Perceived Amount of Assessments Per Format Across Core Classes	15
2. Student Perceived Performance Based on Format of Assessment Before and After Treatment	16
3. Student Preference in Format of Assessment Before and After Treatment.....	17
4. Student Formative Assessment Scores.....	19
5. Student Summative Assessment Scores.....	20
6. Student Cumulative Assessment Scores.....	21
7. Grouped Student Scores Based for Cumulative Paper Grades Compared to Cumulative Digital Assessment Grades	21
8. Student Cumulative Percentage on Paper Assessment Compare to Cumulative Grade	22
9. Student Cumulative Percentage on Digital Assessment Compare to Cumulative Grade	23

ABSTRACT

Today's students are being assessed in a multitude of formats compared to the traditional paper-format quizzes and tests. As digital formatting of mathematical and chemical symbols is difficult for students to complete in the chemistry setting, the use of calculation-based open response questions in digital assessments has diminished. The value of such written response questions remains a valuable tool for instructors as it provides assessment information on student understanding and their communication processes. The purpose of this research project was to determine if there was any significant difference in performance between paper and digital assessments. As well, a look at the student perception of performance and preference in format were evaluated. Students were given instruction on how to perform equivalent digital formatting for chemical and mathematical notation where possible. They were then given instruction in various units of study, and assessed using either paper or digital means. Grades on formative and summative assessments were compared to determine if a significant difference in performance existed for either format type. Student surveys were also administered to determine preference for either format of assessment. Student preference shifted slightly towards paper format due to the amount of effort required for input using digital means. During the learning phase, performance on formative assessments showed a significant difference in student results favoring paper assessments. Whereas assessment data from the summative learning assessments showed that the students learned the chemistry objectives with no significant difference using either format. Qualitative data collected suggested methods to change the online delivery method in the future to bridge the gap between performance and preference.

INTRODUCTION AND BACKGROUND

Context of Study

Since my career began slightly over a decade ago, the manner of student assessment has gradually been changing to include more assessments completed in a digital format. Most courses benefit from the ease of word processing and feedback tools that are readily available. There are also many assessment platforms that allow for data gathering and assessment of various question types (e.g., multiple choice, true/false, numerical-response) across all subjects. However, the numerical notation of math and science-based courses have made longer responses on assessments much more difficult to complete using digital tools.

In the past, I have frequently used many digital tools to help my students increase their understanding of the learning outcomes and to further their passion in the sciences. Using Learning Management Systems (LMS) as an online collaborative environment, I have used online practice assessments as a way to get my students to review course material. These materials contained a variety of question types, but from casual observation, the students struggled to complete those questions that required them to show their work for all of their calculations.

Many studies have looked at digital assessment in the classroom, but few have looked at longer term studies and comparing individual student results as opposed to one treatment group compared with another. My research was focused on my eleventh-grade chemistry students' use of formative and summative digital assessment to determine if their academic performance was significantly affected when compared to traditional

paper-format assessments. I also wanted to gain insight into student preference in assessment format and ways to improve assessments and my teaching practices.

Focus Question

My focus question was, What is the effect of digital format assessment on student performance?

My sub-questions include the following:

1. How does digital assessment affect performance on a high-stake assessment?
2. How do students view and perceive their performance on paper compared with digital assessments?
3. What are the implications for me as a teacher and how do they impact my professional practices?

CONCEPTUAL FRAMEWORK

Theoretical Support

Prior studies provide a foundation of support for this research project. Some provide support and others indicate deficits of paper assessments and those of online assessments. A comparison of methodologies helped to implement a successful treatment in this research project.

Advantages and Disadvantages of Online Assessment

The use of digital assessment has been increasing over the past decades. Research has identified advantages and disadvantages of digital assessment when compared to the traditional off-print (paper) format assessment. Some advantages were identified in a study in which students, teachers, and administrators were interviewed after the implementation of digital assessment (Bergen, Fili, & Nordberg, 2015). The students that provided feedback identified faster grading and feedback, less stress due to handwriting, and fairer grading because of anonymity as reasons why they preferred digital assessment over paper-format counterparts. The time to grade the assessment decreased by approximately 50% compared to grading a paper exam. This was one benefit put forth by instructors along with less chances of losing exams. As well, teachers also stated that grading became fairer as they did not have to overcome poor student handwriting. School administrators pointed to the reduction of photocopying costs as an advantage of digital assessments (Bergen, Fili, & Nordberg, 2015).

Certain negative aspects of online assessment include time to implement for teachers, costly software or subscription for assessment software for school leaders (Bergen, Fili, & Nordberg, 2015). Other disadvantages that have been identified in a different study include a change in perceived reliability by students and teachers when hardware or software is unreliable or crashes (Noyes & Garland, 2008).

Online Versus Paper-Format Assessments

Previous studies have also tried to identify the effects of online assessment in relation to student performance. In one study, students took part in one of three different methods of assessment while writing either paper or online. The students were given either a training-homework task, an evaluation-unproctored exam, or an evaluation-proctored exam. After the treatments were completed and results compared, the researchers found no significant difference in assessment performance between paper-format or online-format assessments (Ardid, Gomez-Tejedor, Meseguer-Dueñas, Riera, & Vidaurre, 2015).

Similar results were found in another study in which digital assessment results were compared to off-print assessment while also comparing question types. The researchers narrowed the focus by examining three question types: algorithmic, conceptual, or definition questions. They found that there was no significant difference in student performance between the question types when completed online or on paper (Prisacari & Danielson, 2017).

In a second journal article by the same authors, the researchers examined the mental load on students and their use of scratch paper when taking assessments either on

paper or online. They once again found no significant difference in performance by their students online or on paper, but also no difference for each of the question types. One helpful insight they found as a result of looking at the use of scratch paper, was that students who wrote their exam on paper were more likely to use scratch paper while answering the questions (Prisacari & Danielson, 2017).

In a meta-analysis study of 26 journal articles spanning a decade, the use of digital assessment tools was found to significantly improve the quality and quantity of written work as compared to paper-based assessments (Goldberg, Russel, & Cook, 2003). This research further supports the use of online assessment as the implementation of my action research should not negatively affect student performance.

All three of these papers had relevance to my study, as their results were in alignment with the data collected from my action research project in which there was no significant difference in student performance for online or off-print assessments. The increased use of scratch paper by those writing paper-format exams, as identified by Prisacari and Danielson (2017), was something that was also brought up during my initial student interviews. One of my students indicated that it would be helpful for fair implementation of my action research project to include paper for students to complete algorithmic or conceptual type questions, as completing calculations in the online environment are challenging.

Question Format

Many standardized exams and STEM courses have employed multiple choice questions as they can be quickly machine graded and can be administered easily on paper

or digitally (DeAngelo, Hurtado, Pryor, Santos, 2009; Stanger-Hall, 2012). Another study suggests that the extra time required to grade, the need for student writing skills, and grading inconsistencies are pitfalls of open-ended question types (Martinez, 1999). Other researchers argue that open-ended questions are better at authentically capturing student knowledge as students come up with their own responses rather than choosing one from a list of options (Haudek, Prevost, Moscarella, Merrill, Urban-Lurain, 2012; Milton, 1979). I support the argument that open-ended response question types are a more genuine method for students to demonstrate their understanding. This is the reason for my action research project focused on using both closed-ended and open-ended response question types.

Comparison of Methodologies

In an article comparing various online assessment strategies, the researchers made use of Likert-scale data and a staff and student survey. The Likert scale data was tabulated and results from the strongly agree and agree were presented. The staff and students in the study indicated many quality elements of instruction, however, several were deemed more essential than others. Continual feedback, comparable rigor, use of a variety of instruction strategies, and thought-provoking discussion threads were the most effective tools identified (Gaytan & McEwen, 2007). For my action research project these results helped to implement my online assessments successfully. I used a similar Likert scale to collect quantitative data to further the interpretation of the results.

The methods used by Aakre (2017) in his capstone project were also of great influence on the design I had chosen to follow. He used a pre- and post-Likert survey as

well as two formative assessments followed by a summative assessment. I followed suit with this format, but over a longer period of time using multiple units of study as opposed to one unit of study. My treatment also differed as I used a single class where each student received both treatments in alternating fashion in comparison to Aakre's three classes that were randomly divided into treatment groups.

Both of the studies by Prisacari and Danielson (2017) have similar elements of design that I utilized as well. Both of the studies were implemented in undergraduate chemistry classes and used two in-class quizzes, an optional practice test, and then a third unit test. The students alternated their mode of assessment for the first two quizzes and then were randomly selected for online or print assessment for the exam (Prisacari & Danielson, 2017). For my study, I used a similar format with three quizzes, instead of two, and one test per unit to gather data to compare student performance throughout the term of study. The greatest difference between Prisacari and Danielson and my action research project, is that it took a longitudinal approach. The longer timeframe of my study in which I alternated the online and off-print assessment for the four units of study as opposed to only one unit.

Prisacari and Danielson (2017) also provided insight into maintaining validity and reliability of my future action research project. In their methodology, they indicated that not only were the questions alike for the online and paper assessments, they also chose to give it the same look. They maintained one question per page on both modes of assessment so as to eliminate an advantage for either type.

Besides the data collection process, the literature has provided different ways to analyze and present the data. The student grade on the assessment was compared to their grade point average (GPA) to see the correlation between the mode of assessment and their achievement (Aakre, 2017). This correlation was replicated for my action research project and was used to determine if there was any difference in performance of a student using paper or digital assessments (see Data Analysis).

Likert scale data was the primary mode of data collection for another study in which the results were tabulated and the mean response and broad agreement percentage was provided (McCarthy, 2015). I also collected Likert scale data and followed suit with this study by providing a similar reporting method. Besides the Likert scale survey, the students in the study by McCarthy (2015) were given open-ended questions to determine their preference of audio, video, or written feedback on online assessments. The study was an excellent example of incorporating thirty student feedback quotes from qualitative data collection (McCarthy, 2015). The use of qualitative data helps to tell the story of the treatment, student preferences, and further the quantitative data's interpretation. The method of student qualitative feedback and quotations was one that I also employed to convey my action research project's results.

METHODOLOGY

Demographics

Coalhurst High School, CHS, the school where I currently teach is in Coalhurst, Alberta, Canada. CHS is a public 7-12 secondary school with a population of approximately 193 students. Coalhurst is a bedroom community close to Lethbridge, the nearest center. Most of the parents commute to work as there are few local businesses.

The community consists mainly of dual income families with low to medium household incomes. Median income of Coalhurst dual-income families is 13.4% lower than the Alberta median. For single-income households the difference increases to 19.8% lower than the Alberta median. The CHS student body consists of approximately 94.7% Caucasians, 3.8% Indigenous, 2.6% Asian, and 0.5% African American.

Many of the students have been in school together since starting school with few new to the community, few to leave, and most staying within the surrounding community for post-secondary studies. The students in my sophomore and senior classes tend to be mid to high achievers. As most of the students work towards entering either college or university, they need to be prepared for the types of assessments that are being used at those levels. Increasingly, the post-secondary programs utilize digital assessments in order to save time for instructors when grading, save money from printing, and increases the rate which students receive feedback. Having my student's complete online assessments in secondary school will prepare them for coming post-secondary assessment practices, while at the same time allowing me to develop the following questions that I wanted to investigate throughout my action research project.

Treatment

Twenty-two students started with me in the selected Chemistry 20 class in September and all chose to participate in the action research project. Two of those students moved away during the first month of instruction, this caused a difference in number when comparing the pre- and post-treatment qualitative data. The treatment protocol is shown in Table 1 and is described below.

Table 1. Data Triangulation Matrix

Focus Question	Primary Question: 1. What is the effect of digital format assessment on student performance?	Sub-Question 1: 2. How does digital assessment affect performance on a high-stake assessment?	Sub-Question 2: 3. How do students view and perceive their performance on paper compared with digital assessments?	Sub-Question 3: 4. What are the implications for me as a teacher and how do they impact my professional practices?
Treatment	Two Formative Quizzes One Summative Quiz	Online Provincial Field Test	Pre-Likert Survey Post-Likert Survey	Pre-Likert Survey Post-Likert Survey
Post-Treatment	Summative Exam	Comparison of Data to School-based Results	Student Interview	Student Interview

The treatment protocol implemented involved teaching the curriculum with four assessments in each unit of study. In the first and third units of study paper-format assessments were used, whereas digital-format assessments were used in the second and fourth units. Chemistry 20 has four units of study in its curriculum. The first two units are of approximate equal difficulty as the recommended weighting for each set out by provincial guidelines is approximately the same (Unit A 20%, Unit B 16%). In addition, the last two units carry the exact same recommended weighting of 32% making them of

equal difficulty. The students received instruction in each of the four units of study and for each unit there were two formative quizzes, a summative quiz, and a summative unit exam as part of the treatment. Other assessments took place, but were not included in the action research project. This portion of the treatment protocol was completed over a 4-month period of the fall semester.

Data Collection and Analysis Strategies

During the data collection process, I collected the assessments as I would in any other course and graded them. In order for the students to still use them as learning opportunities, for the paper-based assessments I scanned them and kept a digital copy to review, but handed back the original to the student. In this fashion, the students were able to learn from the assessment for the summative quiz and exam in each unit, but the data remained intact for use in data analysis. The digital assessments were performed using the Schoology Learning Management System (LMS) as the students were already familiar with this platform since several teachers at CHS have used it as their online classroom to access materials. The digital assessments performed online were reviewable by students once completed at any time. These assessments were also reviewable by the teacher and the data was stored electronically in order to analyze the data.

Pre-treatment and post-treatment Likert surveys were performed digitally using a Microsoft Form in order to easily gather the data. The questions were almost identical in the pre- and post-treatment surveys so as to reflect any possible change in attitude or perception towards the treatment. The Likert-style surveys also included open-ended

questions that were used along-side student interviews to provide qualitative data to help improve future implementation and action research studies for the researcher.

Due to the relatively small class size, I interviewed each of the students after the completion of the treatment. This provided me with feedback from students with a range of abilities. One of the greatest aspects of the interview process that was of particular interest to me was to gain insight on how the process can be improved for future classes.

I also wanted to expand on my longitudinal approach one step further and collected data from the students that continued into the twelfth grade Chemistry 30 course for the first unit of study in the first month of the spring semester. The provincial assessment branch offers field testing for our high stakes diploma exam for each unit of the Chemistry 30 course; offered only in digital format. As a result, the students were given the chance to complete a digital field test in the unit of Thermochemistry to give me further insight into whether the use of digital assessment within the classroom had matching performance results in a high stakes exam situation. The biggest difference between the field test and the assessments used in the regular treatment protocol was that there were no long answer questions on the field test.

To ensure validity and reliability of the instrumentation throughout the treatment protocol the sources used were vetted. As well, feedback was provided from colleagues, my project supervisor and science reader. Lastly, the triangulation of several instruments to collect various sets of supporting data for this project's research questions.

The research methodology for this project received an exemption by Montana State University's Institutional Review Board (Appendix D) and compliance for working with human subjects was maintained throughout the course of the study.

DATA ANALYSIS

Results

The results of the action research project are described below and have been broken into two categories relevant to the research questions: student perception and student performance.

Student Perception

At CHS we have transitioned from a traditional computer lab to having Chromebooks on carts at a 2-to-1 ratio. Along with this, the Wi-Fi capabilities have increased significantly and access is available to any student on any device, school-owned or personal. With this increase in availability and access, I first wanted to gauge to what level the teachers at CHS are using this added functionality in their classrooms to assess their students. The results of the Likert-style survey given early in the term compared to the end of the first semester, as shown in Figure 1, demonstrates that our students feel that there is about an equivalent amount of assessments done on paper as opposed to digitally.

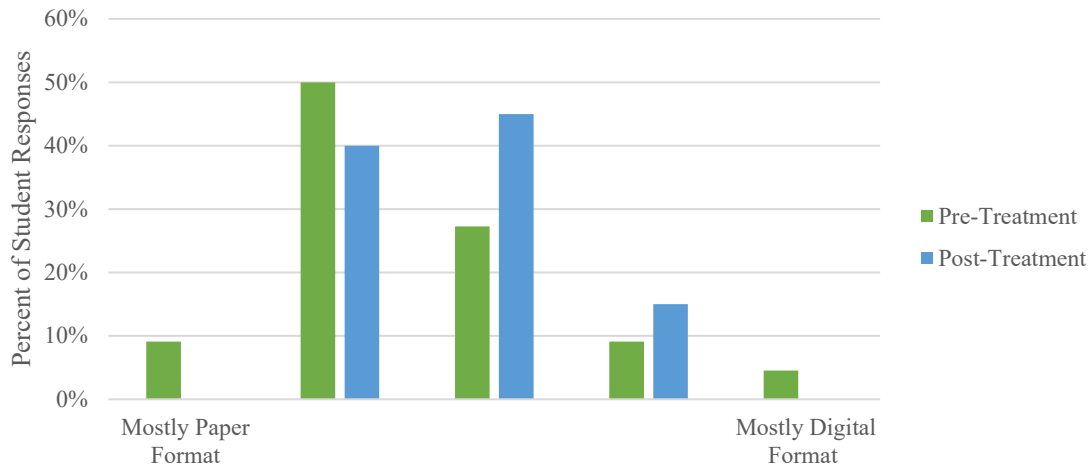


Figure 1. Student perceived amount of assessments per format across core classes, ($N=22$ pre-treatment, $N=20$ post-treatment).

When compared to how our four senior high teachers responded, the results were very similar. The English teacher responded that she does approximately 75% of her assessments (formative or summative) in digital format, our Social Studies teacher and myself (sciences) use digital format assessments for approximately half of our assessments. Finally, the mathematics teacher uses digital format approximately 30% of the time. The results in Figure 1 indicate that the students recognize a similar amount as pointed out by my colleagues. This is in line with the purpose for this action research project as it highlights the need for digital fluency in all subject matters including subjects like chemistry. It is important to note that the data collection for both pre- and post-treatment data was conducted prior to the COVID-19 school closures. Since that time, all classes transitioned fully online for the remainder of the school year and thus 100% of the assessments were then completed online.

Item 4 of the pre- and post-Likert surveys was used to determine the students perceived performance when comparing paper-format and digital-format assessments, the results are below in Figure 2.

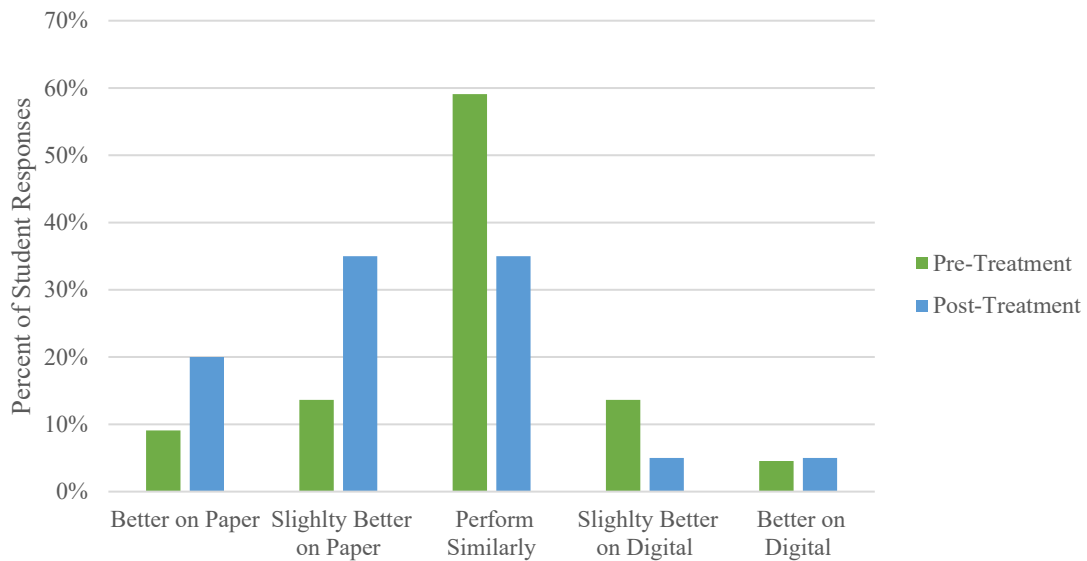


Figure 2. Student perceived performance based on format of assessment before and after treatment, ($N=22$ pre-treatment, $N=20$ post-treatment).

Using a scale of 1-5, with a one indicating they believe they perform better on paper, and a five meaning they feel they perform better on digital assessments, the results from the fourth question pre-treatment had a mean of 2.91 ($N=22$), whereas the same question post-treatment had a mean of 2.40 ($N=20$). Students initially felt that they perform almost identically on either type of assessment. However, after the treatment over the course of the semester, the students felt they performed slightly better on paper than using digital assessments. During post-treatment interviews students shed light for some of the discrepancy as one student put it, “early on it [digital assessment] took some getting used to, but at the end I would say that it was fair and I performed about the same

on both.” I shall return to the analysis of Figure 2 to compare student perception along with the actual assessment results.

A much larger change in results occurred when looking at student preference in assessment format. Figure 3 shows prior to the start of the study, students had a very slight preference for paper-format assessments with a mean of 2.73. A mean of 3, would represent no difference in preference for either format. This changed significantly to a mean of 1.8 with many shifting their preference in format to paper post-treatment.

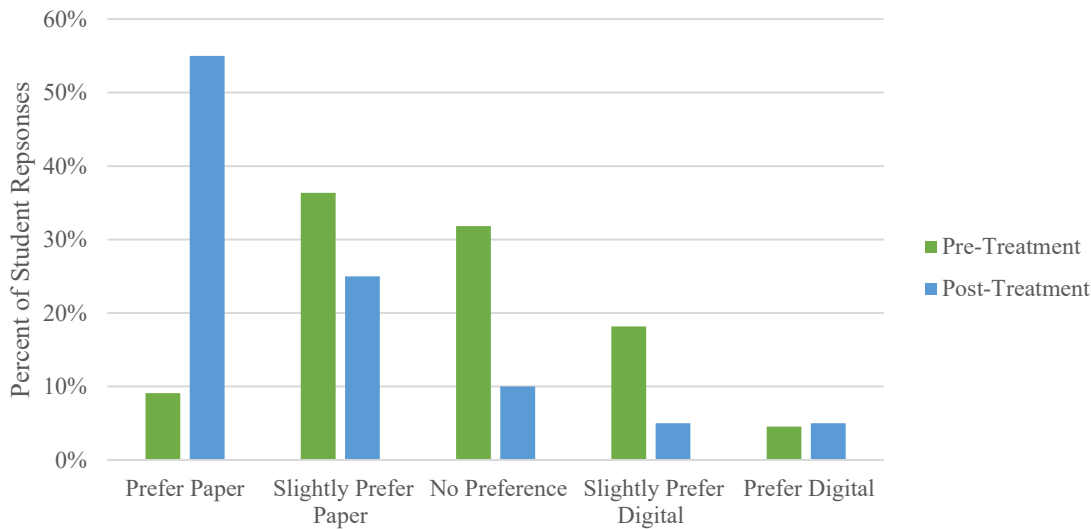


Figure 3. Student preference in format of assessment before and after treatment, ($N=22$ pre-treatment, $N=20$ post-treatment).

The Likert survey and interviews also provided ample reasons for the shift as seen above in Figure 3. Question 7 of the Likert survey had a 50% response rate in which students indicated the difficulty in formatting chemical equations in long answer-style questions as the reason for their preference for paper-format assessments in chemistry. When further probed during the interviews, students responded that their indicated preference was mostly based on the type of question; those that are calculation-based and

long answer. Other qualitative responses from the Likert surveys provided more evidence for their preference. Several students provided similar statements such as the following,

I can't show my work, so the possibility of extra marks are out the window. On paper, everything is together and I can show my work beside questions, whereas online, I use scrap paper and if I need to go back and see if I made an error on a previous question, it can get messy doing that.

Another student stated, “It can be difficult to show equations on a computer program and can be slower.” This was very evident after giving the assessments in the second unit of study as the students took what seemed a significantly longer time to complete the assessments digitally. The time to complete the digital assessments did decrease over the course of the study, but was not tracked by the researcher.

Student Performance Results

As there were four units of study and four assessments within each unit of study, there are many aspects of comparison that can be made from the data collected. Figure 4 below shows the comparison of all of the formative assessments for each of the four units of studies. Due to my small sample size, a Spearman’s rank correlation was done for the formative data to see if there was a connection between the students’ performance on paper and that done online. The students performed significantly better on paper format formative quizzes as indicated by a mean of 85% when compared to a much lower mean of 80% on the digital format quizzes ($\rho=0.810$).

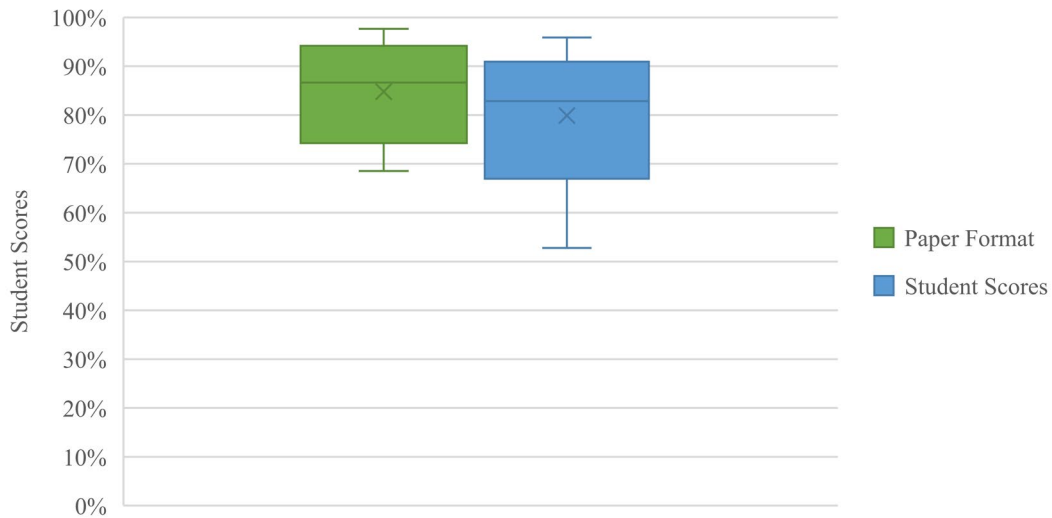


Figure 4. Student formative assessment scores, ($N=20$).

When comparing to the Likert data as shown in Figure 2, this is one of the key aspects of the results which the students' perception of performance matched the results of the study. They felt they performed better on paper than using digital assessments and for the formative assessments this is evident. One of the students pointed out that "there was some learning that had to take place each unit, but by the time we got to the summative assessments, it didn't matter what format we used."

The discrepancy in performance was not seen when looking at the results for the summative assessments, as shown in in Figure 5 below. The students performed almost identically on paper, mean of 86%, when compared to digital format which had a mean of 85%. The Spearman's rank correlation, $\rho=0.873$, also demonstrates a much stronger correlation between the summative assessments than what was present in the formative assessments. The response from the student above demonstrates that the students experienced a learning curve with respect to applying digital skills that correspond to

each unit of study, but once the skills were practiced in an online format, they were able to perform with similar results on the summative assessments.

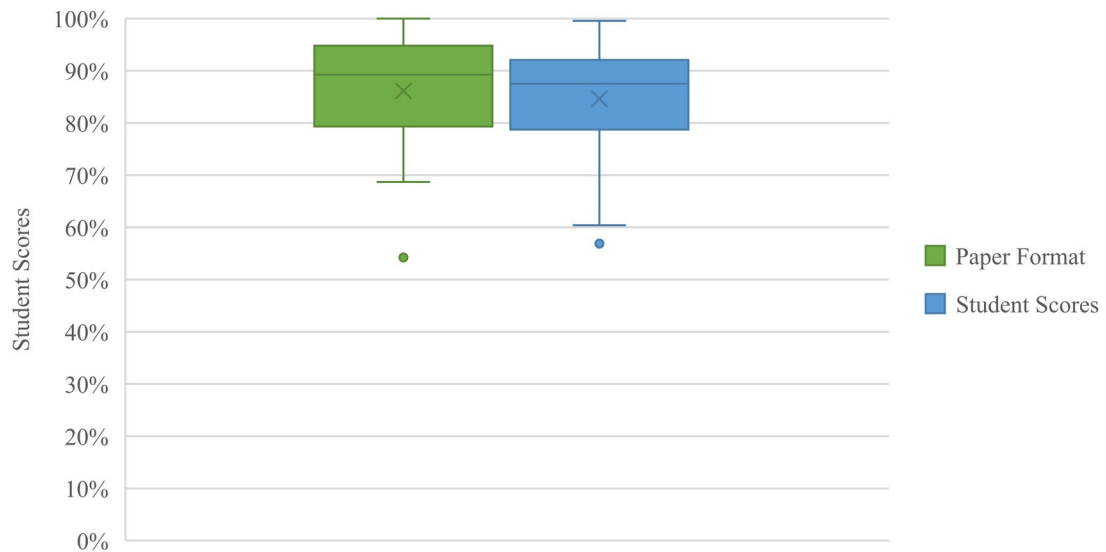


Figure 5. Student summative assessment scores, ($N=20$).

The next comparison is between the cumulative grades, both formative and summative assessments combined, for the units performed on paper ($\bar{x} = 85\%$) and the one done digitally ($\bar{x} = 82\%$). The means, as shown in Figure 6 below, have a difference of approximately 3 percent between formats in which students performed better on paper. This difference is likely due to the significant difference in formative assessment performance as opposed to the results from the summative assessments.

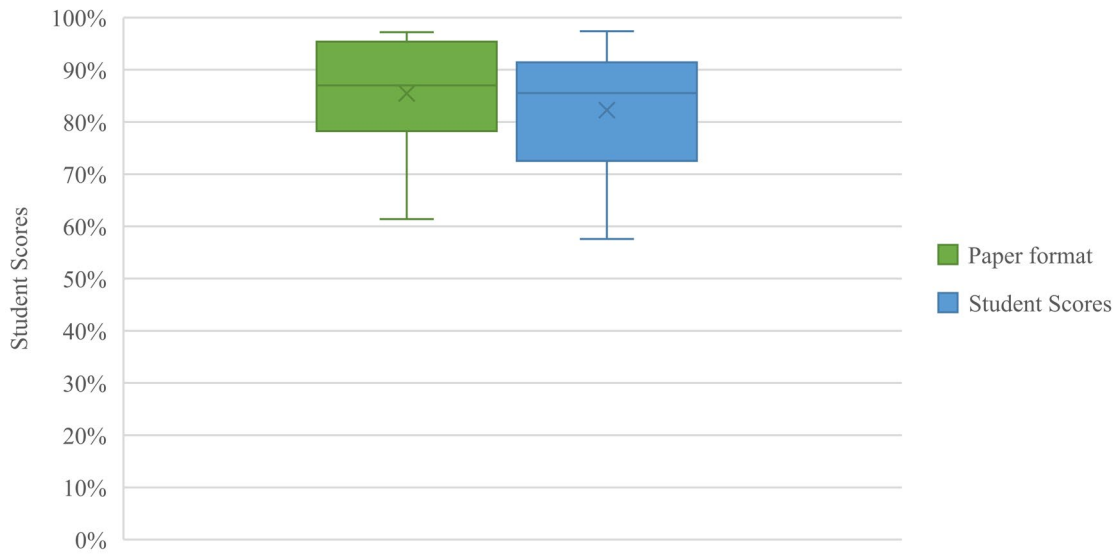


Figure 6. Student cumulative assessment scores, ($N=20$).

When divided into grade groups, the data result seems similar, but there seems to be some differences for individual performance between paper-based and digital-based assessment, as shown in Figure 7 below. Two students performed better on paper as opposed to the digital format throughout the treatment period and as a result moved into a lower grade rank grouping.

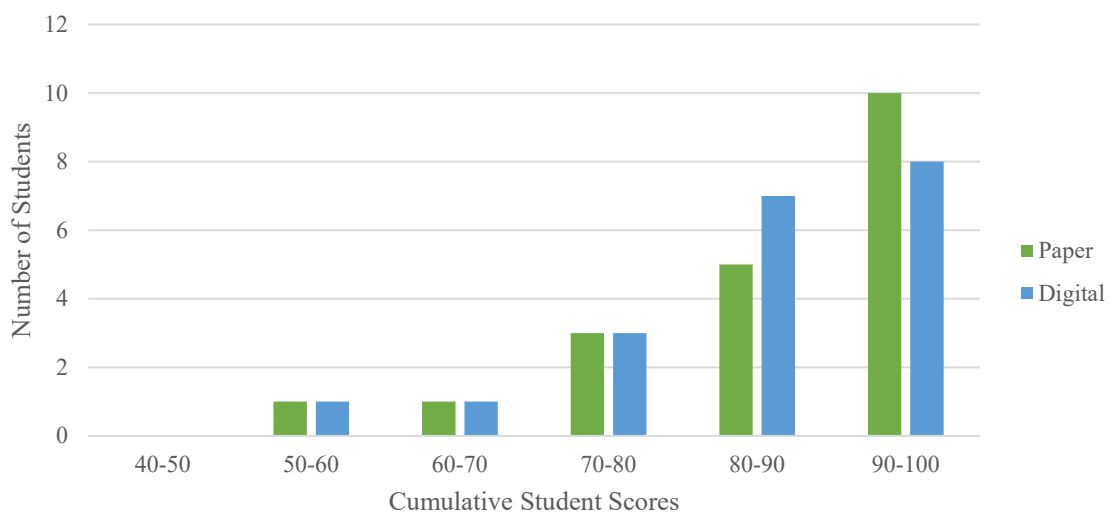


Figure 7. Grouped student scores based for cumulative paper grades compared to cumulative digital assessment grades, ($N=20$).

Another Spearman's rank correlation was done for the cumulative data to see if there was a connection between the students' performance on paper and that done digitally ($p=0.931$). With a high p -value, a strong positive monotonic relationship seems to exist between the performance on paper-based and digital-based assessments for this small sample group. This is also evident in Figure 8 and Figure 9, as there is a direct relationship between the performance on either type of assessment and the student's cumulative percent grade.

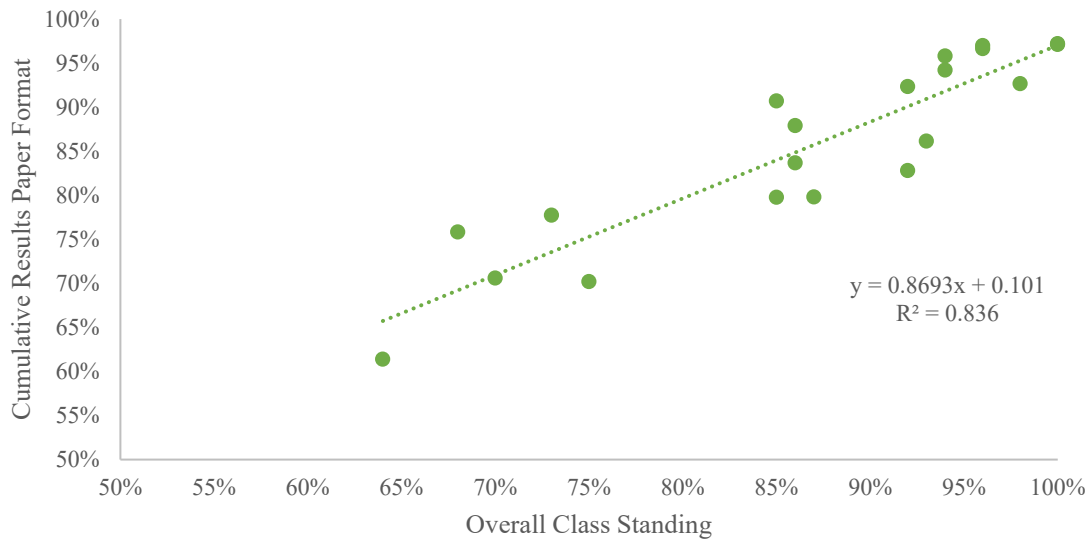


Figure 8. Student cumulative percentage on paper assessment compare to cumulative grade, ($N=20$).

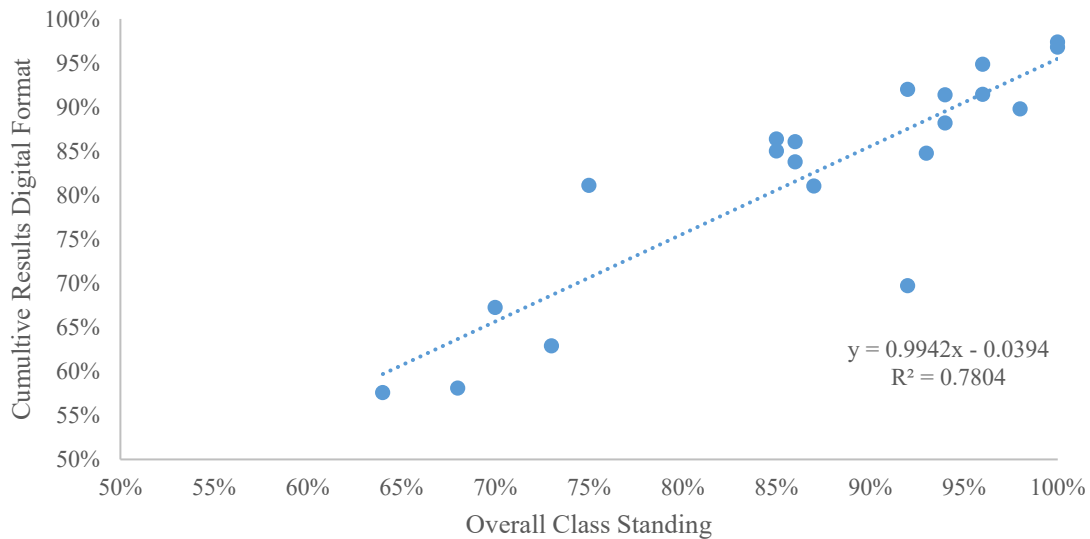


Figure 9. Student cumulative percentage on digital assessment compare to cumulative grade, ($N=20$).

The strong correlation of data ranking, as shown by the Spearman correlation, the high R-value and tight clustering of data points, in both figures above, is indicative of the strong relationship between performance regardless of the format of assessment. These provide good evidence within this small sampling that there is not a significant difference in performance between assessments completed on paper or digitally.

In the results, there were two outliers, both of these students when interviewed indicated a distinct preference for the paper format. The one indicated that in order to perform using digital means, she had to use scrap paper to first complete the question on paper and then transferred her answer to the digital assessment platform. She believed this resulted in some minor errors during transference. The other student is a student who needs paper when her mind wanders and is frequently creating doodles to help her focus. As a result, she found that the lack of paper meant that she found focusing more difficult during the online assessment process.

After semester change, thirteen of the twenty students continued on to take the twelfth-grade chemistry course. Each of these students were given instruction in a fifth unit of study, thermochemistry, given the same number of assessments as in the treatment and then asked to perform a simulated high stakes field test exam which was conducted online. The students ranking was compared to their final eleventh-grade standing as a comparison and the results of the field test were comparable to their rankings at the end of the sophomore class. A Spearman's rank correlation was performed and there was a high correlation between the ranking from the field test results and their final rankings as suggested by a high ρ -value ($\rho = 0.97$). These results followed the results of the summative assessment data during the treatment suggesting no significant difference between student performance on paper when compared to digital format assessment.

CLAIM, EVIDENCE, AND REASONING

Claims from the Study

In the following section, claims arising from the results are described and are broken into the two areas of interest: student performance and student perception.

Student Performance

The first question addressed by the action research project was to determine if there was a significant difference in student performance for assessments done on paper versus those online. The results from the study seem to indicate that two separate results exist. In the first situation, students performed significantly better on paper while performing formative assessments when compared to ones completed online as can be seen in Figure 4. This seems to indicate that during the learning phase of the chemistry objectives students are better able to demonstrate their understanding directly on paper. However, since the students were learning both chemistry objectives and how to display this in a digital format, there seemed to be a second state of learning that occurred. The extra time to not only learn the content, but also how to show it in the online environment led to the second situation.

When comparing only the summative assessment results, as shown in Figure 5, the students seemed to perform with no significant difference. These results were similar to those of Aakre (2017), as well as Prisacari and Danielson (2017). In the case of my action research project, this was likely the result of having time to practice the chemistry learning objectives and the time to learn how to properly input this in a digital format. The two formative quizzes provided several situations for the students to learn both sets

of skills, chemistry and formatting, and the feedback provided from these assessments allowed the students to improve enough so that results for the summative assessments were comparable to the students' results on paper.

The students were able to demonstrate similar results in standing when compared to the high-stake provincial field test. The results of the action research project were therefore helpful in preparing the students and providing key skills for them to complete the assessment. The more difficult topics of study in the senior course reflected the greatest difference in performance as the students ranking in performance was similar to the eleventh-grade chemistry course throughout the treatment.

Student Perception

Student perception of performance, in Figure 2, seemed to correlate well with the cumulative assessment results as shown in Figure 6. Prior to the treatment, the students gave an almost impartial view of how they would perform in either format, but with a slight favoring towards paper. The results of the cumulative data due to the student's poorer performance online for formative assessments meant that their post-treatment perception was correct, their overall performance was significantly less using digital assessments. I believe that as they were learning two sets of skills, the difficulty they experienced to not only learn the chemistry, but also the skill to demonstrate their learning online, suggests that the students felt the learning phase was negatively impacted by having it done online. To remediate this in future classes, a series of activities should be built in at the start of the course as well as a few throughout to ensure that the skills

needed to accurately report online are practiced separate from the learning of the outcomes of the course.

Suggestions from student interviews also provided helpful ways that the treatment could be improved. Some indicated that using tablet style computers with digital pens would give an almost similar feel/mode of input to that of paper. Another suggestion was for an application that can read pen input and turn formulas into the correct digital representation to save time on the formatting required by the students.

Impact of Action Research for the Author

This action research project was an invaluable stepping stone for both myself and my students. Almost directly following the treatment, the outbreak of the COVID-19 pandemic occurred. As a result of preparing my students for what I believed to be for post-secondary, the skills learned were immediately implemented for online learning at home. As a parent of one of the participants noted, “you are ahead of everybody with the online program.” The students involved were able to use the learned skills and transition almost immediately with great success for all of their core classes.

The results of this study were also very rewarding in many different ways. With little difference in performance for summative assessments, future classes will use similar aspects of the online assessments. In the coming year I will take a small step back by moving to a hybrid model of assessment. I will use online assessments for both formative and summative learning tools. However, any open-ended question that requires the use of equations or chemical notations will be done separately on paper. I am going to also approach our IT department to request either a trial or a class set of touch-capable devices

that are able to use a digital stylus. If I am successful in procuring these devices, I will provide training in-class for my students to use these devices. I will also include training with Equatio, which is an application that is able to convert hand-written inputs into digitally formatted mathematical or scientific equations. This application came to light near the end of my action research protocol and when combined with the needed hardware may prove to be what is needed to increase the student preference towards completing all assessments digitally. I will continue to collect similar data used in this action research project to evaluate the effectiveness of those devices and software. I hope to ultimately move to using mostly digital assessments in a few years. By using the approach of action research methodology and implementing a defined treatment and procedure to collect data within the classroom the process has been instrumental in providing beneficial and rewarding professional development for the researcher and will be used in coming years.

Value of the Study and Consideration for Future Research

Similarities in summative assessment results suggest that online assessment continues to be an effective technique to be used in the chemistry classroom that will not negatively impact student assessment results. Suggestions from the post-Likert survey and interviews suggest various modifications could be tested in the future to close the gap for student preference towards digital format. Several students brought forth the idea of a hybrid model in which the question type would determine the assessment format. They proposed to have written response questions that are used to demonstrate process, but also require inclusion of formulas to be done on paper. Whereas paragraph style answers,

multiple choice, numerical response and other quick entrance format questions could be done online.

Other avenues of pursuit will be to use different hardware inputs to bridge the gap between the speed of using pen or pencil and the slower input of having to type symbols necessary for chemical equations and formulas. As mentioned previously, this may include the use of a touch-based computer or laptop that has a digital pen. Software applications such as Equatio are another area of future study. These programs are able to convert finger or digital pen drawn inputs into a typed formula or equation. Each of these suggestions, which came from students in the study or colleagues, would be great additions or modifications to the treatment used for future action research projects.

REFERENCES CITED

- Aakre, H. (2017). The effect of transitioning to paperless assessment in a high school biology course. (Professional paper submitted for Master of Science in Science Education in Bozeman, MT). Retrieved from <https://scholarworks.montana.edu/xmlui/handle/1/13625>
- Ardid, M., Gomez-Tejedor, J., Meseguer-Duenas, J., Riera, J., & Vidaurre, A. (2015). Online exams for blended assessment. Study of different application methodologies. *Computers and Education*, 81, 296-303.
- Berggren, B., Fili, A., & Nordberg, O. (2015). Digital examination in higher education - Experiences from three different perspectives. *International Journal of Education and Development using Information and Communication Technology*, 11 (3), 100-108.
- Coalhurst Census Profile. (2016). Demographics. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=POPC&Code1=1249&Geo2=PR&Code2=48&Data=Count&SearchText=Coalhurst&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=1249&TABID=1>
- DeAngelo, L., Hurtado, S., Pryor, J.H., Kelly, K.R., & Santos, J.L. (2009). The American college teacher: national norms for 2007-2008. Los Angeles: Higher Education Research Institute, University of California-Los Angeles.
- Gayton, J., & McEwan, B.C. (2007). Effective online instructional and assessment strategies. *The American Journal of Distance Education*, 21 (3), 117-132.
- Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *The Journal of Technology, Learning, and Assessment*, 2 (1), 1-30.
- Haudek, K.C., Prevost, L.B., Moscarella, R.A., Merrill, J., & Urban-Lurain, M. (2012). What are they thinking? Automated analysis of student writing about acid-base chemistry in introductory biology. *CBE-Life Sciences Education*, 11, 283-293.
- Martinez, M.E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207-218.
- McCarthy, J. (2015). Evaluating written, audio and video feedback in higher education summative assessment tasks. *Issues in Educational Research*, 25 (2), 153-169.
- Milton, O. (1979). Improving achievement via essay exams. *Journal of Veterinary Medical Education*, 6, 108-112.

- Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: are they equivalent? *Ergonomics*, 51 (9), 1352-1375.
- 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 Behaviour*, 77, 1-10.
- Prisacari, A.A., & Danielson, J. (2017). Rethinking testing mode: should I offer my next chemistry test on paper or computer? *Computers & Education*, 106, 1-12.
- Stanger-Hall, K.F. (2012). Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE-Life Sciences Education*, 11, 294-306.

APPENDICES

APPENDIX A

INSTITUTIONAL REVIEW BOARD EXEMPTION



**INSTITUTIONAL REVIEW BOARD
For the Protection of Human Subjects
FWA 00000165**

2155 Analysis Drive
c/o Microbiology & Immunology
Montana State University
Bozeman, MT 59718
Telephone: 406-994-4706
FAX: 406-994-4303
E-mail: cherylj@montana.edu

Chair: Mark Quinn
406-994-4707
mquinn@montana.edu
Administrator:
Cheryl Johnson
406-994-4706
cherylj@montana.edu

MEMORANDUM

TO: Benjamin Swen and Walter Woolbaugh

FROM: Mark Quinn *Mark Quinn CH*

DATE: October 4, 2019

RE: "The Effect of Digital Assessment in a High School Chemistry Classroom" [BS100419-EX]

The above research, described in your submission of October 3, 2019, is exempt from the requirement of review by the Institutional Review Board in accordance with the Code of Federal regulations, Part 46, section 101. The specific paragraph which applies to your research is:

- ☒ (b) (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.
- ☒ (b) (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation; and (iii) the information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects, and an IRB conducts a limited IRB review to make the determination required by section 16.111(a)(7).
- ☐ (b) (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if: (i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.
- ☐ (b) (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available, or if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects.
- ☐ (b) (5) Research and demonstration projects, which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.
- ☐ (b) (6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed, or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the FDA, or approved by the EPA, or the Food Safety and Inspection Service of the USDA.

Although review by the Institutional Review Board is not required for the above research, the Committee will be glad to review it. If you wish a review and committee approval, please submit 3 copies of the usual application form and it will be processed by expedited review.

APPENDIX B

PRE-TREATMENT LIKERT SURVEY

1. Think of all assessments in all of your classes for this question. To what extent are your assessments in digital format?

Mostly paper format 1 2 3 4 5 Mostly digital format

2. In your opinion, what are some advantages of taking assessments digitally when compared to completing them on paper?
3. In your opinion, what are some disadvantages of taking assessments digitally when compared to completing them on paper?
4. When taking assessments, I perform better on digital assessments than on paper assessments.

Strongly disagree 1 2 3 4 5 Strongly agree

5. Why did you answer the way you did in the above question?
6. When taking assessments, I prefer to take digital assessments more than paper assessments.

Strongly disagree 1 2 3 4 5 Strongly agree

7. If you answered less than five on the previous question, what are some of your reservations about using digital assessments?

APPENDIX C

POST-TREATMENT LIKERT SURVEY

1. Think of all assessments in all of your classes for this question. To what extent are your assessments in digital format?

Mostly paper format 1 2 3 4 5 Mostly digital format

2. In your opinion, what are some advantages of taking assessments digitally when compared to completing them on paper?
3. In your opinion, what are some disadvantages of taking assessments digitally when compared to completing them on paper?
4. When taking assessments, I perform better on digital assessments than on paper assessments.

Strongly disagree 1 2 3 4 5 Strongly agree

5. Why did you answer the way you did in the above question?
6. Now that we have completed a course where we alternated the use of paper-format assessments and digital assessments based on the unit of study, what is your preference for completing assessments?

Mostly paper format 1 2 3 4 5 Mostly digital format

7. Why did you answer the way you did in the above question?
8. Regardless of your preference, in what ways could the delivery of the digital assessment be improved?
9. Are the issues with digital assessment you have identified specific to the chemistry setting or do they exist in your other classes as well i.e., English Language Arts, Social Studies, etc.?

APPENDIX D

POST-TREATMENT INTERVIEW QUESTIONS

1. After having gone through the paper versus digital treatment this semester, did you find you have a preference for completing assessments? Elaborate.
2. Do you feel that you were assessed fairly when completing assessments digitally based on previous assessments in the course and/or previous classes?
3. Were you more likely to complete formative assessments when they were in digital format or paper? Why?
4. In the past, we used digital format assessments primarily for formative and practice quizzes. In your opinion, should the course continue to be designed this way, or should it shift to include more summative assessments, like the ones we did this semester, in the digital format?
5. Are there specific question types that should be done on paper only as opposed to completing them digitally?
6. Can you think of some improvements to how the assessments were delivered?