Fall 2021

Mathematics Online Assessment Tools

Eric Schultz

Mathematics Online Assessment Tools

Eric Schultz

Northwestern College

An Action Research Project Presented

in Partial Fulfillment of the Requirements

For the Degree of Master of Education

Abstract

This paper explores which online assessment tool is preferred within a classroom setting. Four Geometry classes were given weekly formative assessments using DeltaMath, Classkick, and Desmos. The students' scores from these assessments and opinions were collected and analyzed to find which program would be most beneficial for the classroom. The literature review covers a variety of resources that determine effective classroom teaching in which technology is used within the classroom. The articles vary in what practices most benefit the students when utilizing online assessments, and allows a better view into what processes should be considered before implementing online assessments into a classroom. This research was done in a specific classroom, and the results may not be the same if utilized in any other classroom, but the data gives some valuable insight to allow others to make a choice in what program may be most effective for their classroom as well.

Keywords: K-12 Online, Online Assessments, Distance Learning, Assessments, Geometry

Table of Contents

8-12 Mathematics Online Assessment Tools

In this ever-changing world of technology and evolving methods of teaching it is important to dive into the different possible methods to assessing Geometry students within a distance learning setting. With the recent events of a global pandemic there are many schools searching for the best methods for schools to assess students' knowledge in an efficient and integrity filled way. Assessments are referred to as the process done to obtain information that is used to make conclusions and decisions regarding students, curriculum achievement, implementation results of programs, and educational policies for any institution that organizes learning activities (Adri et al., 2021). Teachers have to adapt to the change in our world and find effective ways to assess and get a realistic view of what our students are comprehending.

The problem right now is that there is not enough data and research to identify which assessment program for distance learning will most effectively obtain the information needed to make conclusions about students. There are many different apps and online programs that use multiple choice or even short answer options, but mathematics is unique, it needs symbols and to write out the steps and process to get an answer. No matter what topics are studied in mathematics; students will be required to argue and defend their arguments through reasoning and proof (Novianda et al., 2021). This is the problem. There is not enough research or information on which programs have the mathematic capability to assess students and get a full picture of what they know.

The purpose of this research is to identify online programs that would make assessments for distance learning in math more efficient and equitable for all students. With this research other math teachers would be able to identify which programs would best assist them in getting

an accurate measure of their students' comprehension and understanding of their content. Given the importance of the role of mathematics, mathematics is studied starting from kindergarten, elementary, middle and high school (Ramadhan & Suhendra, 2021). This work is important, due to its wide-ranging use throughout different levels of mathematics. The educational world is changing, and educational tools need to keep up with those changes. Getting both quantitative and qualitative feedback from assessments and student reflections is the way to help teachers make the best decisions moving forward in this post COVID-19 world.

Review of the Literature

In a new realm of education where distance learning for the 8-12 student is becoming more and more popular, it is essential to identify the different methods of assessing students. Prior to the pandemic the distance learning model was not something the 8-12 or even K-12 educational world was prepared for. "During the 2020–2021 school year, districts in the United States each used a mind-boggling 1,449 different digital tools on average *per month*. The overall figure is up a substantial 9% over the spring 2020 ed tech adoption surge during the frantic exodus to remote learning as schools shut down for in-person instruction nationwide. And it's up 52% over 2019-2020 pre-pandemic levels" (2021). Students were thrust into a new education model that was not only new for them, but for a huge percent of the teachers as well. Technology of some kind has been handed to almost an entire nation of students and staff and the weight now falls on the schools to identify the most beneficial way to utilize these devices and best serve students with little to no training.

Mathematics has seen a gap in learning with this swift change of methodology, and something that was previously difficult for students has become an even greater challenge. Students have difficulty in using their mathematical reasoning skills to solve problems, because, in the learning process, they are given procedural learning which results different than developing their thinking and reasoning skills (Novianda et al., 2021). These reasoning skills and processing are relatively unique to mathematics, which is why the instruction and assessment methods of mathematics tend to be separate from its academic counterparts. Because its importance, mathematics is one of the core subjects to be offered to all students in every level of their education. Through mathematic learning students were practiced to develop logical thinking, analytical, systematic, critical, creative and the ability to work together

(Barut & Retnawati, 2020). The new change to 1-1 devices and online learning is a new opportunity to reevaluate our learning process and ways to assess our student population. Not only does it force us to reevaluate these processes, but it creates a new avenue to push students toward reasoning and thought process, instead of procedural learning we have grown accustomed to.

This literature review will describe some of the types of online instruction, and their impact. The review will look into the comparisons between online assessments compared to paper and pencil, as well as the students' opinions on which method is preferred or seen as more impactful. The review will also look into the online testing practices and compare the effectiveness of those differing methods. The possible drawbacks and differing views will also be discussed.

Online Instruction

Teaching online is a different experience from that of teaching in a face-to-face setting. Knowledge and skills developed for teaching face-to-face classes are not adequate preparation for teaching online (Yang, 2017). The process of good teaching can change not only from subject-to-subject, but from class to class. Online teaching is a new teaching mode that should be treated as such. The needed training for this new endeavor is something not only wanted, but necessary.

A key element of online instruction that teachers are learning about are classroom sizes in an online setting. The initial thought for certified staff was that they may lose their jobs due to the fact that one teacher could create and distribute lessons that could eliminate the need for multiple teachers in one subject matter. Miron and Gulosino (2016) found that 356 students were

enrolled in one virtual-school class, and that the average virtual-school class size was 35—far above the U.S. averages for face-to-face class sizes. Within the research from Lin et al. (2018) they found that the relationship between class sizes in math and students' performance was also parabolic: final grades increased as class size increased up to a maximum of 38 students, but decreased if class size rose beyond that point. This gives educators and school administration a good starting point to help students have the best opportunity to learn. It is also essential to note that each course within Lin et al. (2018) research had a different number of students to maximize learning. This research helps us understand that the world of "mega-teachers" is not going to be coming from a transition to online education, and can put teachers' minds at ease as job security and futures seem to be unsure in the new model of online teaching.

Online course design refers to the features that shape the overall structure of the course, including learning activities, sequence of content and communication, and structure of assignments. In most cases, course design drives the instructional strategy adopted in online courses (Yang, 2017). Explicit directions, clear expectations, learning objectives relevant to the assessment, and a format that guides the students towards next steps and understanding are essential, just like any face-to-face class. According to Yang's study (2017), instructional strategies such as case studies, video demonstration, instructor's notes, mini projects, and discussion forums were among the most effective instructional strategies. All of these methods described above sound like something that most teachers would agree upon whether it was online or in person.

Assessments for online learners are key part of developing an effective course. Assessment is a learning tool used to determine whether students are progressing in achieving their learning goals or not (Adri et al., 2021). The process in creating these assessments is often

overlooked, but there is a process necessary for online assessments to ensure assessments are valid. The test is carried out in four phases, where the first test is unit testing, then integration testing, system testing, and acceptance testing (Adri et al., 2021). Many times, teachers overlook the modifications necessary to move an assessment from its paper form into an online setting. Teachers need to look at the questions themselves and identify which type of answer they would like the students to produce, multiple choice, short answer, draw a diagram, move a manipulative, etc. These considerations affect not only what may be expected from the student but also the time it will take for the teacher to actually assess the student and give them timely feedback.

Student and Staff Input

Online assessments will not be as effective if those using and implementing them do not trust their validity or believe them to be necessary. Students need to be convinced of the usefulness of the transition to online assessment before they embrace it willingly. It is evident through this study and in literature that one of the most essential elements to the acceptance of online assessments is how the transition from traditional assessments is made (Khan & Khan, 2018). Instructors feedback from research by Bakhmat et al. (2021) states the main benefits of online education experience: time efficiency, mobility, clear control of student work, and individualization. This is a starting point for staff buy-in in implementing online learning, without these considerations a good program could fall apart due to lack of belief in the program.

Although it is sometimes assumed that today's students crave technology and digitalization, there is evidence that states otherwise. In research from Ramlo (2016) the results from this study indicate that the engineering technology students enrolled in his first semester Technical Physics course agree that they do not want to take classes like physics online and that

simulated science laboratories are not the equivalent to those that are hands-on and face-to-face. There are also difficulties faced when moving to an online format. This makes sense due to the large number of interactive activities and labs that require a student to physically manipulate an object in order to grasp how it moves and operates. Among the main problems, the following are indicated: low quality or no Internet connection, software and hardware, a drop of assignment and quality of student work, and computer and digital competency (Bakhmat et al., 2021) As stated above, student and staff buy-in is key to the success of any educational implementation, and some staff also struggle to embrace this new fully online movement. There is certain resistance among instructors of online education as a majority of respondents did not plan to integrate online segments in courses and did not recognize its importance in the long run (Bakhmat et al., 2021). This lack of time and preparation to move online causes a resilience to fully embrace the move to online education.

Online vs. Paper-Pencil

It is important to look at the correlation between paper-pencil testing and online testing for equitability. After an abrupt shift to online learning due to a pandemic, it is essential to look at the equity of online and paper-pencil testing. In research from Bayazit and Aşkar (2011) At the end of the study, no significant difference has been found between scores obtained by students at the online test and the paper–pencil test. This is something that may shock most, but has very important implications. It means that students' scores remain the same within different testing environments. This gives teachers a reason to possibly invest their time in looking into online assessments. This is also a key finding that would be beneficial in sharing with families who may be skeptical or wary of a change to technology since their personal beliefs tend to play a part in how their child may view this new shift.

Another factor that has influenced the shift to online assessments has been the idea that students who prefer paper-pencil testing will do poorer on their assessment due to the mental aspect of feeling anxious or upset about the change. The findings of Hewson (2012) have contributed evidence to support the validity of online assessment methods by showing that the performance of students taking an online or offline assessment does not differ depending on whether they are required to use their preferred or non-preferred mode. These findings help nudge students towards fully embracing a shift to online assessing, especially if the data shows that it does not affect their overall performance. It is also noteworthy to mention that the average age of the participants within Hewson's (2012) study was 26 years old. The data from this specific study was there was essentially no difference in the assessment scores of the groups expressing a preference for either online (m = 13.7) or offline (m = 13.2) modes or no strong preference (m = 13.6) (Hewson, 2012) This study helps support the move to online learning, but does not negate the affects and real anxiety that some students have when assessing both online and in person.

Although this data is promising there is another side to the debate between online and paper-pencil testing. A study about proctored and un-proctored assessments by Howard (2020) showed there was likely no significant difference in exam scores between those who were proctored and those who were not, even though the time values were significantly different, is somewhat connected to the first. The reason is because the nature of the exams meant looking answers up could slightly help a person who had not prepared well, but the concepts could still be readily mastered by someone who took the time to prepare. This is alarming due to the effects this may have on the inflation of students' grades, and something that would require more research.

In a separate study, performance and duration differences online test was completed at approximately 40.53 min, whereas the paper–pencil test was completed at about 34.26 min (Bayazit and Askar, 2011). Bayazit and Askar's study was a fully proctored assessment which limits the possibility of cheating, but with it only being a 30-question test (6 multiple-choice, 6 matching, 6 multiple-answer, 6 short-answer, and 6 long-answer), and only 6 needing more than a sentence for a response the 6-minute differential between tests is significant. Time management might play a factor in educators tweaking paper-pencil tests to fit the new online format. This may not seem like a large task initially, but with the past years of education systems trying to create classrooms that are only teaching the essential standards, this may further minimize the content that is being taught to students.

Types of Questioning

Teachers look to understand how to best utilize these online assessments and some of the most common testing methods are short-answer and multiple choice. Teachers have their own personal opinions about these types of testing and ÇIftçi (2019) did a study to find out those exact views. His findings on metaphors used to represent short-answer questions were that most metaphors were produced under "freedom in answering" and "requiring knowledge," while the least number of metaphors were under "number of questions" and "chance success" categories. This shows that short-answer questions involve more thought provoking answers, and the negative association section was due to anxious students and longer time for the teacher to grade. Within the same study, multiple-choice questioning showed most metaphors were produced under "offering choices" and "chance success," while the smallest number of metaphors were under "length of questions" and "level of difficulty" categories (ÇIftçi, 2019).

This shows that the teachers found multiple choice questions to not be very strenuous and represent full knowledge of the content.

Comparing the two results shows that the general feelings from teachers are that short-answer questions invoke more thought, while multiple-choice questions are more of a game of chance than fully comprehending the content. This is further backed by Burfitt's (2019) qualitative research on student thought processing during multiple choice tests. His findings showed many students indicated partial knowledge of concepts when asked to explain their thinking and when selecting incorrect options which contained some, but not all aspects of the correct response. Students admit to only having partial knowledge and using process of elimination instead of understanding to select an answer. This does not give the teacher accurate feedback into the student's true mastery level of the content. This is also troublesome when thinking about students moving into high-risk fields of study where partial knowledge may cause serious harm. Some examples of these fields could be doctors, engineers, chemists, dermatologists, etc. Partial knowledge in any one of these fields could cause a heart to stop, a building to fall, or an explosion in a lab.

The next driving factor that assists teachers in their selection of testing methods are the actual performance comparisons between testing options. In research from Attali et al. (2016) there is a comparison between multiple choice questioning and short answer testing. Their research was looking into immediate feedback options of each style test and their findings concluded, the results of this study suggest that this interactivity was more beneficial to the measurement of test-takers' ability with OE than MC items. These advantages could benefit different types of assessments, including formative and other low-stakes assessments. Other teachers share this viewpoint, it is believed that MCQ items do not tap higher orders of thinking

and allow for a higher probability of guessing correctly, which reduces reliabilities of the test for lower ability students (Melovitz Vasan et al., 2017). The data is showing that although the grading may take a little longer, the assessment itself is more valid if it incorporates some type of open-ended questioning. Looking at Melovitz's (2017) research, he finds that the OEQ (open-ended question) format created more challenging expectations, it appeared to motivate the students to develop and adopt an effective strategy for in depth learning (i.e., conceptual understanding rather than simple memorization). This is again, in comparison to multiple choice questions. This study looked into how the types of questions affect the ways students prepare for tests, and it shows the more in-depth preparation for open-ended questions compared to multiple-choice questions.

Teachers also look to improve the student learning, which should be a follow up of each of our formative assessments. Research from Puthiaparampil and Rahman (2020) shows that VSAQ (very short answer questions) employ directly asked questions, which require students to answer briefly and directly with no scope of guessing or reaching the answer by elimination. The information obtained by the teachers while scoring will help to modify the teaching and to give feedback to the students. This use of VSAQ was found to be a more viable alternative to either multiple choice questions or best answer questions. Another study found that students are less likely to generate a correct answer when asked to articulate a response to a clinical vignette than when they have to pick an answer from a list of options. This therefore raises the possibility that the VSA (very short answer) is a valid test of a student's knowledge (Sam et al., 2016). These studies point educators to put the extra time in to grade assessments utilizing some form of short answer questions to ensure that the scores being collected are valid.

The downfall of the research-based need for more short answer assessments are the difficulty in grading such assessments. A recent study showed it turned out that grading short-answer questions still requires that can understand the context and order of keywords, otherwise, only keywords check can be achieved in current study (Lu et al., 2021). This may turn some teachers off to this idea, but as educators the goal is to best serve students, and this short answer assessments may be a viable option to do so.

Although online instruction has been around for a while within the collegiate ranks, it is new to most high school teachers, which means there is a large learning curve. This literature review shows that online instruction needs to be approached in a different manner than in person. The views held by students and staff alike need to be taken seriously, and both need to be shown a purpose for this transition to online education. There are multiple different types of assessing, but short answer questions may be a good start for teachers looking to truly understand what their students know. This research is looking to fill a gap within mathematics in online assessments. Students and staff are looking for an option that benefits both and can still truly show what the students know. This research hopes to fill that gap in comparing different online assessment programs to help struggling teachers transitioning to online courses have more data to determine which program may best serve their students as well as efficiently assist them in assessing their students.

Methods

Research Questions

The purpose of the study was to identify which of three online assessment programs would be preferred by students and staff in a math classroom. This action research project focused on the following questions: Which program (DeltaMath, ClassKick, or Desmos) would produce higher average test scores? Which program would students prefer to use? Which program would a teacher prefer to use? Through this action research, the researcher will be able to identify which program will be most effective for their classroom.

Research Design

The research that was implemented over a six-week period was quasi-experimental. Students were given two assessments each week over the current topics, Angle Relationships and Coordinate Grid Transformations, in Geometry. Each assessment was 6-11 questions and students were given 20 minutes to complete the assessment. Each program was used for 3 assessments, and at the end of the research the students were asked three questions about the programs. The questions were: “On a scale from 1-5, one being extremely easy and 5 being very difficult, how difficult was it to use the online assessment? Which program did you prefer to use?”

Quantitative data was collected from the two assessments each week over the students’ performance. The assessments were short-answer responses that required the student to show their understanding without the help of possible choices, multiple-choice. The data was used to see if the students’ average scores showed a discrepancy between assessment programs.

Qualitative data was collected in the form of a Google Form survey completed after the completion of each program, three assessments. The same questions were asked after the completion of each program, and an additional survey asking the students which they preferred overall will be administered at the completion of the final program assessment.

Variables

The study will be analyzing the correlation between the variables related to each research question. The variables for the research question, "Which program (DeltaMath, ClassKick, or Desmos) would produce higher average test scores?", are the three different assessment tools being utilized and the student's test scores. Each program will be used 3 separate times to gain a larger data set to make a more accurate assessment of which program created higher results. Looking at the overall averages of the four classes, as well as the individual class testing scores will help provide insight into which program showed a larger growth. The assessment tools will then be used for the other research question, "Which program would students prefer to use?", the variable will be the students' demographics and individual feelings towards the assessment. The last research question, "Which program would a teacher prefer to use?", has the variable of the teacher's personal feelings.

Setting and Participants

The study was conducted at a Junior High School in Waukee, Iowa. The population consists of a mixture of rural and suburban families. The research was conducted in four classrooms. The classes consisted of 98 students: 51 males and 47 females. Nine of these students received extended learning program services, two students had 504's, and four students received Special Education services. Eight of these students identified as Asian, five identify as

African American/Black, one identifies as American Indian/Alaska Native, three identify as Mixed, and the remaining 81 identify as White.

Data Collection Plan

Students were given an online formative assessment twice each week. Each program collected data from 3 formative assessments, and the data was stored in an Excel Spreadsheet. The students' average scores were collected from each program. After all programs were administered, the students were given an overall survey to show which program they prefer.

Data Analysis Plan

The data analysis process included conducting a paired samples t-test comparing the students' average scores to see if there was a significant difference in scores between the different online assessments. I will use a comparative survey for the quantitative data in order to answer the question, "Which program would students prefer to use?" The data for each of these comparisons will be done in Microsoft Excel. Lastly, I will reflect and share my findings on how personally easy or difficult each program was to use as a teacher.

IRB

Northwestern College's IRB board granted an IRB exemption prior to data collection. The approval was granted due to the researcher using normal educational practices, students posed minimal risk, and the research was conducted within the classroom in our school.

Findings

Data Analysis

Average Scores

A one-way ANOVA, analysis of variance, test was conducted to determine if there was a significant difference in students' scores between DeltaMath, Classkick, and Desmos' formative assessments. This analysis was conducted through utilizing the data analysis tool in Microsoft Excel.

The students' scores throughout the five-week research period where the students were assessed on the current topic using one of the three online assessment programs. While using DeltaMath students scored a 73.66% (M = 73.66, SD = 20) on the three assessments. Students scored an average of 71.86% (M = 71.86, SD = 22) when using Classkick, and an average score of 79.11% (M = 79.11, SD = 18) was attained when assessing using Desmos.

Table 1

Anova Test - Average Test Score

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7955.88	2	3977.94	9.02	0.0001	3.01
Within Groups	362199.45	821	441.17			
Total	370155.33	823				

When looking at the data, the results showed that there is a significant difference between the groups. This is shown by the *p-value* = .0001 < .05. The initial results showed that Desmos resulted in the highest average scores. With these initial results, and three data sets being compared it was essential to run a post-hoc Tukey HSD test.

Table 2

Tukey HSD Test - Students' Average Test Scores

A = DeltaMath **B = Classkick** **C = Desmos**

Treatment Pairs	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	1.41	0.57	insignificant
A vs C	4.31	0.007	** p < 0.01
B vs C	5.76	0.001	** p < 0.01

The Tukey HSD test's p-value shows the comparison of more than two sample sizes. The p-value between DeltaMath and Classkick showed that they are closely related and that their scores show that they are very comparable and using one compared to the other would not create a significant difference in scores. The low p-value between DelatMath and Desmos, and Classkick and Desmos, show that Desmos' scores are much higher than the other two online programs.

The results of this study answer the first research question, "Which program would create higher scores?" While the p-value between DeltaMath and Classkick shows that there is no significant difference in testing with these programs, the research showed that using Desmos resulted in higher test scores.

Student Preference

After the assessments were completed, the students were asked a few questions to conclude which online assessment was preferred by the students. The first question asked the students to rank how easy the program was to use, one being extremely easy to five being very difficult. Although the average scores showed that the students answer ranked Classkick, *M* =

2.3, the most difficult, DeltaMath, *M* = 2.06, the next easiest, and Desmos, *M* = 1.96, the easiest to use, the p-value, as seen below, showed that there was not a significant amount of difference in their responses to say one assessment was easier than the other. The p-value was .o6, which being over .05 shows that the data was not different enough to warrant a significant result from these findings.

Table 3

Anova Test – Scores for User Friendly Program

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.14	2	3.57	2.83	0.06	3.03
Within Groups	310.75	246	1.26			
Total	317.89	248				

Although the students' views on how easy the programs were to use did not warrant any significant results, the students' preference was clear. The students were also asked which online assessment they preferred. The results showed that 29 students, 35.4%, chose DeltaMath, 19 students chose Classkick, 23.2%, and 34 chose Desmos, 41.5%. (The total number of students who took the survey on this date was lower than the data collected for the average scores due to multiple school sporting events and student illness.)

This data answers the second research question, "Which program would students prefer to use?" Desmos received the highest votes, with DeltaMath only 5 votes behind. This may not be a significant margin, but it was clear that the students did not prefer Classkick.

Teacher Preference

When looking at Classkick through the teacher's lens, it was very easy to set up and administer assessments. You can upload documents quickly into the program, and it allows you to see what students are writing as they write their work on their screen. The downfall of Classkick is its grading. Unless you are doing multiple choice questions, you have to go into each of the student's slides and grade each individually, which is essentially the same as grading things by hand. It was also difficult to read some of the students' answers and work due to the writing feature on Classkick. Students write using their touchpads, which causes students to take more time on assessments and creates a more challenging task for the teacher to decipher what the students has written.

Desmos has a wide variety of formative assessment created from other teachers that a teacher could pull from to assess students' learning. This is the only non-time consuming way to use Desmos as a teacher. The creation of a single assessment takes over triple the time it takes to create an assessment from scratch in Classkick, and even more than triple the time to create something compared to DeltaMath. This causes me to rule Desmos out. It did have the ability to see the students' progress as they are working and allowed you to set up short answer questions that the program would grade for you, if the student typed the short answer in the exact same format as you.

DeltaMath was the easiest and quickest program to utilize for a teacher. It auto-generates questions that are linked to the Common Core standards. You click on a common core standard and then can choose from an infinite number of auto-generated questions on that subject. Once the problems are selected, the tests are automatically graded and students get feedback as soon as they complete the assessment. The only flaw I saw in this program was that if a student accidentally clicks to submit the assessment prior to starting the first question, there was not a

way to go back and allow them another opportunity to take the assessment. This happened three times during my classes.

The last research question was, “Which program would a teacher prefer to use?” After concluding using each program there were a few aspects that made DeltaMath the preferred online assessment program for my classroom.

Discussion

Summary of Major Findings

The average test scores showed that students scored a significant amount higher using Desmos than Classkick and DeltaMath. The p-value from the TukeyHSD test resulting in a significant difference substantiates the scores. The data points towards Desmos yielding higher scores than either Classkick or DeltaMath.

The students were very clear in which program they did not prefer to use, Classkick. The data showed that there were only 5 votes separating DeltaMath and Desmos, which shows that they would be okay with using either of those programs in the future.

Lastly, it was very clear which program was easier to use as a teacher. DeltaMath created an easier process to create standard specific problems and get and give timely feedback on how well students were grasping concepts. It was the easiest program for a teacher to use, even without any real technology training.

Limitations of the Study

During the study there were a few things that were unavoidable. Student attendance caused an issue with collecting consistent data during this process. Students were absent due to illness, sporting events, vacations, medical appointments, quarantine, etc. These untimely absences caused some variance in the data.

The study was also limited due to the time the students had to get comfortable with the programs. Some students had used Classkick and Desmos, but none had used DeltaMath before this research. Students may have done poorer than expected due to using a new program.

The time to collect the data was another limitation that could have caused some changes in the data. Classes were taking formative assessments using the online programs twice a week, while also continuing to complete summative assessments and exit tickets. This caused some burnout in students and by then end of the study there were a good number of students who were verbally against taking another online assessment.

Further Study

This study could benefit from a more controlled group of study. Students who either volunteer or are specifically selected for the study would create a more consistent and clearer picture of the data being collected. Combining the scores of students from four Geometry classes may affect the data, and further study may need to be done with less variables.

In the future it would be beneficial to further assess the difference in feedback between programs and what students prefer. Classkick has a sticker system of feedback or teacher notes that are able to be written or stuck directly onto the students' screen while they are working. DeltaMath has the ability to give students feedback after each question or once the assessment is complete, and Desmos has features that allow students to get feedback after each question.

Now that this research has shown which program the students prefer it would be a good idea to see how Math scores would compare with paper-and-pencil tests versus Desmos. The literature review shows some research on this, but I had not seen any research comparing Desmos and paper-and-pencil testing.

Conclusion

When looking back at the research questions, the conclusion from this research is that students are shown to score higher using Desmos compared to DeltaMath and Classkick. Students prefer Desmos and DeltaMath over Classkick, Desmos slightly more than DeltaMath. Teachers may prefer DeltaMath over Desmos or Classkick.

It makes sense that students would prefer a testing program that results in higher test scores. Although that may be their selection, a teacher would most likely prefer the more standards connected DeltaMath due to its auto-generated questions, self-grading, quick feedback, and ease of use. Looking back on Attali et al.'s (2016) research, timely feedback is key to a student maximizing their potential and learning within a classroom. This will be something that I will be using in the future to assist in assessing and further assisting students of all levels within my classroom.

When looking towards the future of education it is getting more and more evident that technology is going to be key. This research allows teachers to make an informed decision on what online assessment program may help their Mathematics classroom.

References

Adri, H. T., Suwarjono, Sesrita, A., & Sudjani, D. H. (2021). The online assessment in education course. *Journal of Physics: Conference Series*, *1918*(5), 052086. https://doi.org/10.1088/1742-6596/1918/5/052086

Attali, Y., Laitusis, C., & Stone, E. (2016). Differences in Reaction to Immediate Feedback and Opportunity to Revise Answers for Multiple-Choice and Open-Ended Questions. *Educational and Psychological Measurement*, *76*(5), 787–802. https://doi.org/10.1177/0013164415612548

Bakhmat, L., Babakina, O., & Belmaz, Y. (2021). Assessing online education during the COVID-19 pandemic: a survey of lecturers in Ukraine. *Journal of Physics: Conference Series*, *1840*(1), 012050. https://doi.org/10.1088/1742-6596/1840/1/012050

Barut, M. E. O., & Retnawati, H. (2020). Geometry learning in vocational high school: Investigating the students' difficulties and levels of thinking. *Journal of Physics: Conference Series*, *1613*, 012058. https://doi.org/10.1088/1742-6596/1613/1/012058

Bayazit, A., & Aşkar, P. (2011). Performance and duration differences between online and paper–pencil tests. *Asia Pacific Education Review*, *13*(2), 219–226. https://doi.org/10.1007/s12564-011-9190-9

ÇIftçi, S. (2019). Metaphors on Open-Ended Question and Multiple-Choice Tests Produced by Pre-Service Classroom Teachers. *Lnternational Electronic Journal of Elementary Education*, *11*(4), 361–369. https://doi.org/10.26822/iejee.2019450794

Ed tech use continues growth beyond the peak of the pandemic. THE Journal. (2021, September 9). Retrieved December 4, 2021, from https://thejournal.com/articles/2021/09/09/ed-tech-use-accelerates-beyond-the-peak-of-the-pandemic.aspx.

Gulosino, C., & Miron, G. (2017). Growth and performance of fully online and blended K-12 public schools. *Education Policy Analysis Archives*, *25*, 124. https://doi.org/10.14507/epaa.25.2859

Hewson, C. (2012). Can online course-based assessment methods be fair and equitable? Relationships between students' preferences and performance within online and offline assessments. *Journal of Computer Assisted Learning*, *28*(5), 488–498. https://doi.org/10.1111/j.1365-2729.2011.00473.x

Hewson, C., & Charlton, J. P. (2018). An investigation of the validity of course-based online assessment methods: The role of computer-related attitudes and assessment mode preferences. *Journal of Computer Assisted Learning*, *35*(1), 51–60. https://doi.org/10.1111/jcal.12310

Howard, D. W. (2020). A Synopsis of Online Testing Integrity in a General Education Math Course: A Correlational Study. *Online Learning*, *24*(4), 225. https://doi.org/10.24059/olj.v24i4.2148

Khan, S., & Khan, R. A. (2018). Online assessments: Exploring perspectives of university students. *Education and Information Technologies*, *24*(1), 661–677. https://doi.org/10.1007/s10639-018-9797-0

Lin, C. H., Kwon, J. B., & Zhang, Y. (2018). Online self-paced high-school class size and student achievement. *Educational Technology Research and Development*, *67*(2), 317–336. https://doi.org/10.1007/s11423-018-9614-x

Lu, O. H. T., Huang, A. Y. Q., Tsai, D. C. L., & Yang, S. J. H. (2021). Expert-Authored and Machine-Generated Short-Answer Questions for Assessing Students' Learning Performance. *Educational Technology & Society*, *24*(3), 159–173.

Melovitz Vasan, C. A., DeFouw, D. O., Holland, B. K., & Vasan, N. S. (2017). Analysis of testing with multiple choice versus open-ended questions: Outcome-based observations in an anatomy course. *Anatomical Sciences Education*, *11*(3), 254–261. https://doi.org/10.1002/ase.1739

Novianda, D., Darhim, & Prabawanto, S. (2021). Analysis of students' mathematical reasoning ability in geometry through distance learning. *Journal of Physics: Conference Series*, *1882*(1), 012085. https://doi.org/10.1088/1742-6596/1882/1/012085

Puthiaparampil, T., & Rahman, M. M. (2020). Very short answer questions: a viable alternative to multiple choice questions. *BMC Medical Education*, *20*(1). https://doi.org/10.1186/s12909-020-02057-w

Ramadhan, R., & Suhendra. (2021). The effect of distance learning during Covid-19 pandemics on the mathematical learning results. *Journal of Physics: Conference Series*, *1806*(1), 012096. https://doi.org/10.1088/1742-6596/1806/1/012096

Ramlo, S. E. (2016). Students' Views About Potentially Offering Physics Courses Online. *Journal of Science Education and Technology*, *25*(3), 489–496. https://doi.org/10.1007/s10956-016-9608-6

Sam, A. H., Hameed, S., Harris, J., & Meeran, K. (2016). Validity of very short answer versus single best answer questions for undergraduate assessment. *BMC Medical Education*, *16*(1). https://doi.org/10.1186/s12909-016-0793-z

Weaver, C. C., Rivera, P. R., Xiang, L., Taylor, J. C., & Osborn, J. L. (2020). Open Problem-Based Instruction Improves Student Outcome on Short Answer Critical Thinking Physiology Questions. *The FASEB Journal*, *34*(S1), 1. https://doi.org/10.1096/fasebj.2020.34.s1.06674

Yang, D. (2017). Instructional strategies and course design for teaching statistics online: perspectives from online students. *International Journal of STEM Education*, *4*(1). https://doi.org/10.1186/s40594-017-0096-x