IJITGEB

INTERNATIONAL JOURNAL IN INFORMATION TECHNOLOGY IN GOVERNANCE, EDUCATION AND BUSINESS
Vol. 6, No. 2, 2024
ISSN 2686-0694 (Print)
e-ISSN 2721-0030 (Online)

Learning Algebra via self-study using Mixed Reality and the Microsoft HoloLens 2 Headset

Fabiano Bruno Barros de Almeida
barros.f.aa@m.titech.ac.jp
Tokyo Institute of Technology, Japan

Jeffrey S. Cross
cross.j.aa@m.titech.ac.jp
Tokyo Institute of Technology, Japan

Abstract: This study explores the use of mixed reality (MR) glasses, specifically the HoloLens 2, in a pilot study for Algebra learning by university students. Students initially took a paper-based algebra test, followed by group-specific interventions for addressing their incorrect answers. The experimental group used the HoloLens 2 with an original step-by-step equation-solving software, while the control groups utilized paper-based or conventional study tools. Post-intervention, all groups retook the test. Despite no clear distinctions in improvement between groups through statistical analysis, qualitative feedback and comparisons showed promising trends. While ANOVA and t-tests did not reveal significant differences, students with higher initial scores overall showed lower improvement. Despite starting with the higher mean scores, the experimental group demonstrated better improvement than the paper-based control group and simi lar improvement to the group with freely chosen tools. Further research with a larger number of learners, focusing on students with similar initial test scores, is expected to show significant advantages of using the Hololens 2 as studies conducted in the field of medical education have shown.

Keywords: Algebra, Mixed Reality, Hololens 2, Self-study

INTRODUCTION

Mixed Reality (MR) is used to describe technologies that mix virtual elements with the real environment of the user in interactive ways. In these experiences, the user can usually use their hands, gaze, or voice commands to control elements that are layered on top of their environment using a MR device. Furthermore, it is also possible to see that the MR device is constantly generating a 3D mesh of the surrounding environment to simulate interactions between the virtual elements and elements in the physical world. Because of the ability to manipulate elements in 3D space, MR is often preferred over Virtual Reality (VR) in contexts where 3D interaction can benefit the user, such as in the viewing and manipulation of architectural elements and geometrical shapes.

WHY USE MR FOR TEACHING ALGEBRA?

The first developments in extended reality (XR) aimed at using it for solving problems that would otherwise be close to unsolvable without it. This sort of problem found its answer in XR with a level of fit that could not be equally achieved by other solutions, such as representing the columns in a 3D viewing software on a computer screen. The application of XR would provide the user with a level of ease and interactivity that can hardly be matched by means other than XR. Understanding this, researchers in the area reasonably aimed first at exploring this aspect of XR in which it stands unmatched, in which it provides an experience that cannot be rivaled. By doing so, even if other aspects of the technology were not quite favorable yet, such as cost and ease of use, its unmatched application would suffice as a reason to invest in it.

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

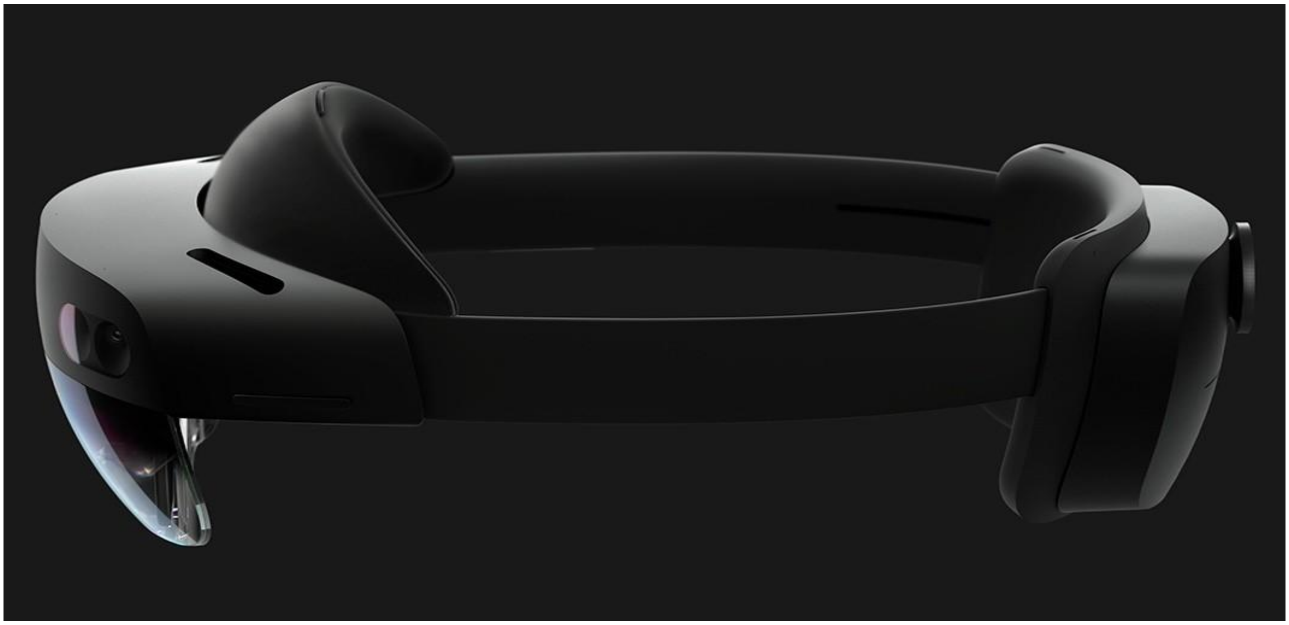
As the aforementioned "other aspects" of the technology slowly become more favorable, researchers bear the burden of developing an application and investigating its utility. MR can be utilized for the visualization of 3D objects (Agarwal 2024). MR provides online learners the chance to learn topics in an immersive experience which cannot be achieved using a web browser. For studying mathematics, the learning experience can significantly be enhanced for algebra, especially integrals.

HOLOLENS 2

The device used for this research was the Microsoft HoloLens 2, which is a mixed reality device and has been shown to applicable for teaching in the medical field (Zaccardi et al., 2023) as well as used for training (Microsoft, 2024). However, there are few studies that have used the Hololens 2 for linear algebra instruction. A prior study with students using augmented reality on the topic of spatial transformation matrices did show improvement compared to a control group (Shaghaghian et al., 2024). Figure 1 below shows a side view of the HoloLens 2, where the glass component is on the lower left side and head strap adjustment in on the right side of the image.

Figure 1

HoloLens 2 side view Note. Image from Microsoft (2023a)



Rather than fundamentally changing how math is studied by most students using a pen or a pencil and paper, this pilot study aimed at utilizing a MR tool that could enhance learning without disturbing it. Therefore, one of the conditions for that was to have a device that could be controlled without the use of the learner's hands. The fact that students could interact with the device from beginning to end using only gaze and voice commands allowed for a more natural experience in which the glasses provided significant assistance without causing distraction.

As this technology becomes more available, MR glasses with the necessary eye-tracking, head-tracking, and voice-command tools for this sort of interaction will undoubtedly become more common, but at the current moment, few devices match the capability of the HoloLens 2.

PEYETHAGOREAN

The software used by the members of the experimental group was custom-made for this research project by the first author. We aptly decided to name this software *Peyethagorean*. Peyethagorean was deployed to the HoloLens 2 as an app built in Unity, cross-platform game development software.

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

Many packages and frameworks were used in Unity, so to mention only some of the most relevant: MRTK3 (Microsoft, 2024) was used for the design language and the necessary basic tools to create an environment and player that could properly interact with the specifics of the HoloLens 2, TEXDraw (Wello Soft, 2024) was used to display equations using LaTeX format, and OpenXR (Khronos Group, 2024) was the framework used for accessing XR capabilities in Unity, as the HoloLens 2 and many other XR devices use it.

Peyethagorean worked almost exclusively with voice commands. It was designed that way so students could interact with the software without ever needing to drop their pencils/pens while writing their responses on paper. A list of the voice commands used for the experiment is given in Table 1.

Table 1

Command list for Peyethagorean

Command	Action triggered by the command
“Partial”	Begins the partial decomposition of a fraction
“Integral”	Integrates an expression
“Next” “Forward”	Moves to the next step of the solution
“Back” “Previous”	Moves to the previous step of the solution
“Solution”	Moves to the last step of the solution
“End” “Finish”	Finishes displaying solution
“Reset”	Returns the app to the starting condition

For some of the commands that were used more often, two options were given as triggers. For instance, to move forward a step, subjects could either say “next” or “forward”. This was useful especially as this project had subjects from many countries with many different accents, all non-native speakers of English. We discuss this in further detail in the subsection that presents feedback on the software.

The two most important commands were “partial” and “integral”. When these commands were issued, the front-facing camera of the HoloLens 2 was activated, and a picture was taken of what was in front of the subjects. The user is notified of this through the flashing LED of the front-facing camera and a balloon notification.

For some of the commands that were used more often, two options were given as triggers. For instance, to move forward a step, subjects could either say “next” or “forward”. This was useful especially as this project had subjects from many countries with many different accents, all non-native speakers of English. We discuss this in further detail in the subsection that presents feedback on the software.

The two most important commands were “partial” and “integral”. When these commands were issued, the front-facing camera of the HoloLens 2 was activated, and a picture was taken of what was in front of the subjects. The user is notified of this through the flashing LED of the front-facing camera and a balloon notification.

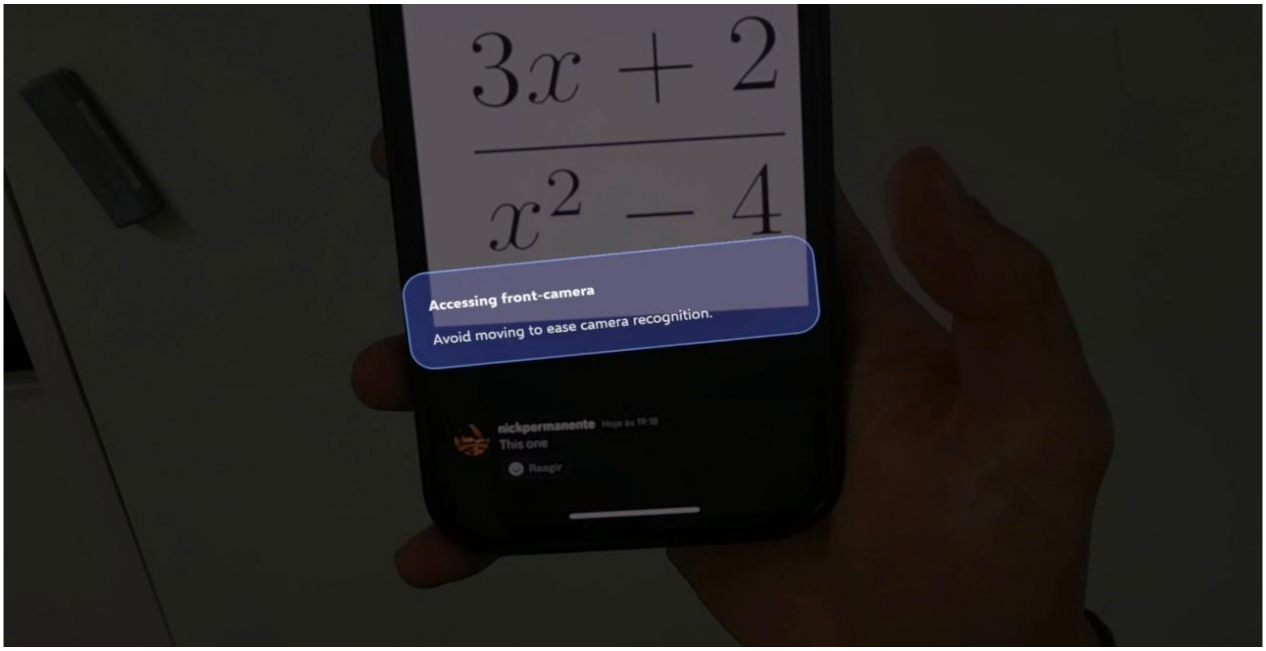
This moment can be seen in Figure 2. In the example below, the user is looking at an expression presented on his cell phone and asking for its partial fraction decomposition.

The picture taken was then sent to the Mathpix, (2024) API. This API recognized equations in the image and if the equation was printed or not, allowing for differentiation between the equations written by students and those printed in the quiz. If the image had no equation in it, or the API could not recognize an equation with a confidence

level above 75%, an error message was returned. It would show the equation in LaTeX format, which was then displayed as a text balloon.

Figure 2

App Displaying the Notification Balloon on the Glasses



For some of the commands that were used more often, two options were given as triggers. For instance, to move forward a step, subjects could either say “next” or “forward”. This was useful especially as this project had subjects from many countries with many different accents, all non-native speakers of English. We discuss this in further detail in the subsection that presents feedback on the software.

The two most important commands were “partial” and “integral”. When these commands were issued, the front-facing camera of the HoloLens 2 was activated, and a picture was taken of what was in front of the subjects. The user is notified of this through the flashing LED of the front-facing camera and a balloon notification.

This moment can be seen in Figure 2. In the example below, the user is looking at an expression presented on his cell phone and asking for its partial fraction decomposition.

The picture taken was then sent to the Mathpix, (2024) API. This API recognized equations in the image and if the equation was printed or not, allowing for differentiation between the equations written by students and those printed in the quiz. If the image had no equation in it, or the API could not recognize an equation with a confidence level above 75%, an error message was returned. It would show the equation in LaTeX format, which was then displayed as a text balloon.

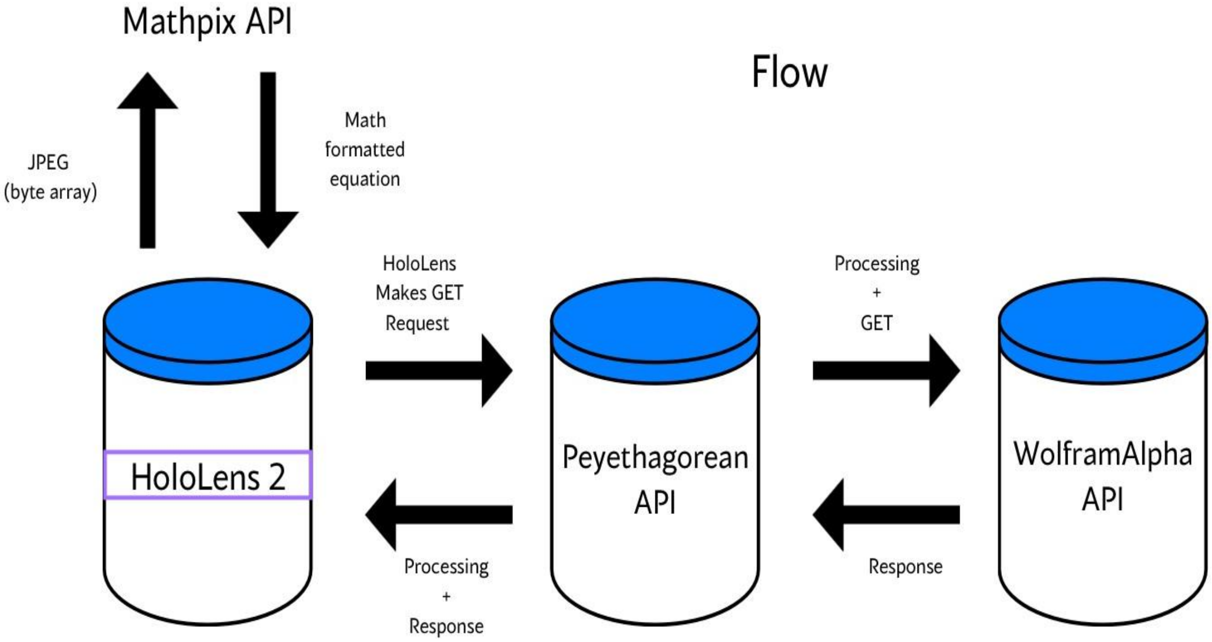
Once the app received this, it made a call to a server built specifically for this project. This server did some parsing and then made a request to the Wolfram Alpha Full Results API (2024). This API returned to the server an object containing several elements, among them a step-by-step solution of the problem sent to the API in MathML format.

This solution was then processed in the server and converted into several strings in LaTeX format. These strings and information about them were returned as a response to the app, which started the display of the solution. A flowchart describing the process just explained is shown in Figure 3.

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

Figure 3

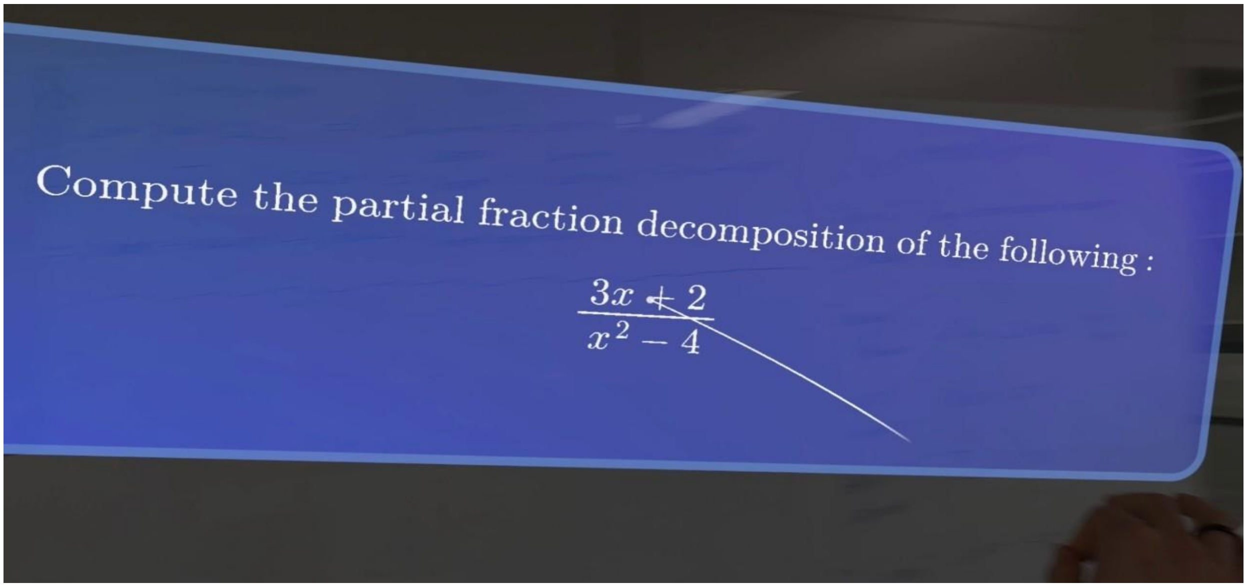
Flowchart Describing the Transfer of Data between Devices and API's



A new, interactable panel containing the first step of the solution to that problem would then appear in front of the user. The user could move this panel to position it wherever they desired. Figure 4 shows the user moving the panel.

Figure 4

Hololens Panel Displaying the First Step of the Solution



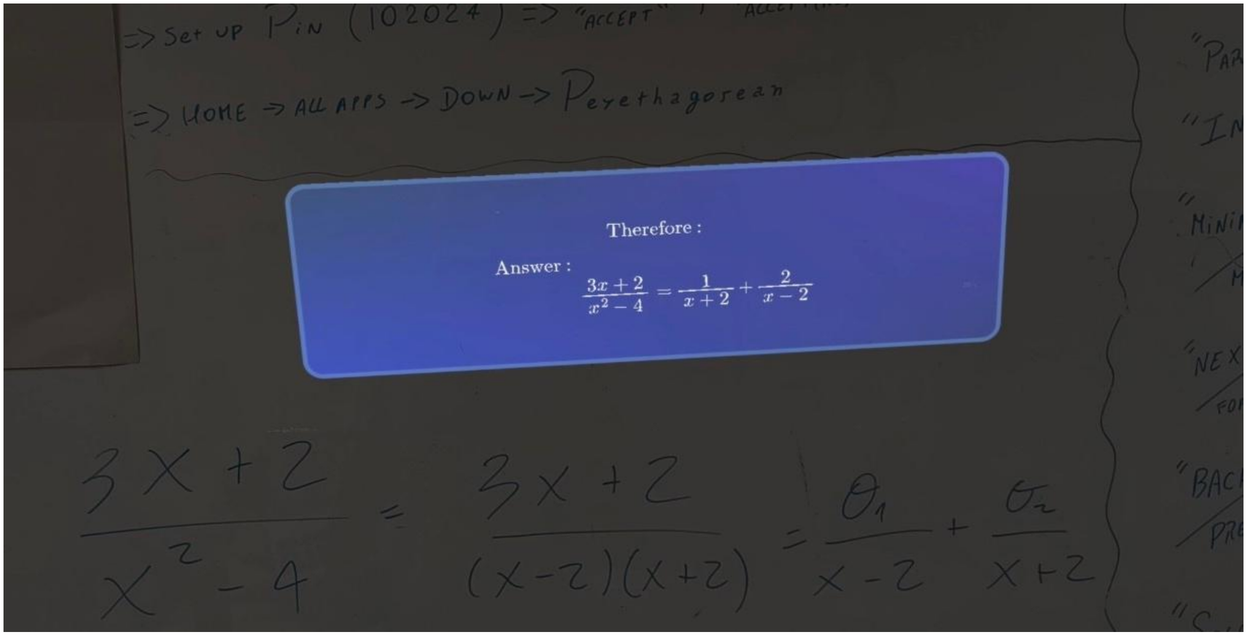
Compute the partial fraction decomposition of the following :

$$\frac{3x+2}{x^2-4}$$

Students could then navigate through the steps of the solution, one at a time, until they felt that they had understood the steps to solving it. They could also skip to the solution in case they just wanted to check the answer. Figure 5 below shows the solution made by the user and the panel displaying the solution. A demonstration of the Peyethagorean (2024) software usage with the Hololens 2 is available as a Youtube video.

Figure 5

Panel Displaying the Solution



Therefore :

Answer :

$$\frac{3x+2}{x^2-4} = \frac{1}{x+2} + \frac{2}{x-2}$$

=> Set up Pin (102024) => "ACCEPT"
=> HOME -> ALL APPS -> DOWN -> Peyethagorean
"PA
"IN
"MINI
"NEX
"BAC
$\frac{3x+2}{x^2-4} = \frac{3x+2}{(x-2)(x+2)} = \frac{\theta_1}{x-2} + \frac{\theta_2}{x+2}$

Once they were done with the solution, they could issue either the "end" or "finish" commands to stop displaying the solution and set the app ready to solve new problems. Since the solutions were generated procedurally at runtime, students could also write down an integral they wished to check and use the app to solve those as well.

LITERATURE REVIEW

Most of the research available at this point in the education of mathematics using MR focuses on using the 3D capabilities of MR to some extent. Hence, a lot of the literature on the area concerns either the teaching of 3D models, structural design, or 3D geometry. The current research aims only at understanding whether the change brought by MR tools in learning mathematics, specifically algebra, leads to positive learning outcomes, not on exploring some feature of learning using MR that is unique to MR, such as the interactivity with 3D objects in a mixed-reality environment. In that sense, learning algebra using the MR device is in some ways comparable to learning using traditional methods, such as looking at solutions in a textbook or even typing equations on software that can yield solutions. This implementation of MR, however, is novel and has no significant body of research regarding it yet published. Therefore, the closest research papers we can reference concern either the use of MR in general in higher education or, more broadly, the education of mathematics in VR.

EXTENDED REALITY IN HIGHER EDUCATION

Although the usage of XR in higher education has gained traction in the last decade progress has been slow. As pointed out by Radianti et al., 2020, although interest in the area is promising, a lot of research has not progressed

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

past the experimental state. The burden associated with the implementation of these technologies and the limited observed improvements in learning makes it so that moving from an experimental state to implementation is hard. Furthermore, as mentioned in the subsection regarding market overview, metadata analysis of research conducted on the topic shows that it is often employed in the medical field or in developing tools for teaching concepts in spatial geometry and design (Park et al., 2021). Research specifically on the latter and more broadly in mathematics often shows promising results in terms of student engagement and learning experience (Coimbra et al., 2015). (Montoya et al., 2021) provides an extensive metadata analysis of research conducted in recent years involving XR and Mathematics. Their analysis too shows the aforementioned preference for using XR in the context of learning topics that require some level of spatial visualization. Specifically, in the area of higher-education Algebra, one work worthy of mention is the work of (Kang et al., 2020). Their work explored, with similar methodologies to the current research project, the use of head-mounted displays (HMD) and VR for evaluating the impact of VR on the visualization of partial derivatives. Their research found that students performed worse after treatment in some questions and elaborated as to why that may have happened. This discussion helped us predict in the current research which questions could benefit less or even not benefit from the use of our software. Based on their results, they proposed useful tips for VR developers, some of which were taken into consideration for the development of the tool used in the current research.

Even in cases where XR showed itself similar to traditional methods in terms of content acquisition, other measures such as student motivation, learner experience (Stepan et al., 2017), and learning competency (Sattar et al., 2019) often showed improvements. Furthermore, immersive VR often shows positive results not only in learning gain but also in concentration and enjoyment if compared to non-immersive VR (Mahmoud et al., 2020), results that may extend to the use of MR in similar applications. Moreover, even in contexts of limited improvement, these tools may play an important part in preserving time for lecturers and teachers with consistently increasing numbers of students while offering a way to learn concepts on their own that is still motivational and engaging (Martín-Gutiérrez et al., 2015).

RESEARCH OBJECTIVE

This research investigates how the use of MR can be employed to teach concepts of algebra, especially those in the realm of Integrals, to university students. Comparatively, once calculators and online tools that could solve equations became widespread, learning algebra was severely affected by the fact that students could now have easier access to solutions and support material. Believing that the next step in the field of learning algebra will come by allowing students to have almost immediate feedback and solutions to problems by MR-powered tools, this research aims to assess how that will impact students' learning and to propose ways to better implement this technology.

RESEARCH DESIGN & METHODS

Designing experiments to assess new educational tools, especially in the field of educational technology (edtech), is often an arduous task. One of the main reasons for this difficulty stems from the problem of establishing what is a good comparison in itself. In the case of this study, comparing the MR technology to having no assistance in solving mathematical problems or having access to limited solutions would probably contribute to showing the effectiveness of the technology.

However, the same comparison can be made against more modern assistance tools, such as software that provides computational solutions, as is the case with popular web apps such as WolframAlpha (Wolfram Alpha LLC, 2023) or GeoGebra (GeoGebra, 2023). Furthermore, dividing control groups into several types reduces the number of samples that can be obtained for each group. This has the potential to affect the validity of the results in small sample sets and generate a Type II error, that is, having a false negative result on the alternative hypothesis (Serdar et al., 2021).

With these aspects in mind, for this experiment, we will follow the two-group pretest-posttest design procedures outlined in Dimitrov et al. (2003).

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

TWO-GROUP PRETEST-POSTTEST DESIGN

The experiment's participants were all students currently enrolled at the institute. All subjects were rewarded for participation. Their reward was made up of two parts: a base hourly pay and a pay proportional to their success in the exam. Participants were assigned them to a control group or experimental group based on the composition of both groups at the moment the subject finishes the pre-test. Other than pre-test scores, their seniority in college was also considered when assigning students to a group, as students in the first years of the undergraduate course were expected to perform better than those in later years.

To calculate an effective number of subjects for this experiment, we followed the procedures outlined in Serdar et al., 2021 for a two means study type. Furthermore, the ratio of the sample sizes r is 1 since the control group and the experimental group will have the same size. Hence, equation 1 was used to calculate the ideal number of participants as follows:

$$\begin{aligned} N &= \frac{(r+1)(Z_{\alpha/2}+Z_{1-\beta})^2\sigma^2}{rd^2} \\ \Rightarrow N &= \frac{(1+1)(1.96+0.84)^2\sigma^2}{d^2} \\ \Rightarrow N &= 15.68\left(\frac{\sigma}{d}\right)^2 \end{aligned} \tag{1}$$

where:

N = Sample size

r = Ratio of the sample sizes

$Z_{\alpha/2}$ = 1.96 for $\alpha = 0.05$

$Z_{1\ \beta}$ = 0.84 for power of 0.80

σ = Pooled standard deviation

d = Difference of means of 2 groups

For this sort of experiment, values anywhere between 1.5 and 3 can be expected for $(\sigma/d)^2$. Hence, numbers in between that range could be fair sample sizes for this experiment. Given experimental conditions and timeline, we chose 34 as an ideal sample size. However, as explained ahead in the section regarding quantitative results, the control group will be divided into two groups, and those two groups will be individually compared to the test group. Therefore, we must count the control group, which counts for half the sample size, twice. As a result, 51 participants would be an ideal sample size for this experiment, being 17 subjects assigned to the experimental group, 17 assigned to the paper-based group, and 17 assigned to the free group.

Since this project was done as a pilot study, we gathered 8 subjects for the experimental group, 7 for control group A, and 6 for control group B. The participants were assessed quantitatively and qualitatively, and the procedure for each assessment is explained in the next section.

METHODOLOGY

Prior to the beginning of experiments, an application for the approval of the Human subject research ethics committee at the institute outlining the research processes involved, data storage and measures to ensure the anonymity of participants was submitted. The participants were informed their participation was voluntary and they could withdraw from the study at any time. The research study participants received compensation for their time to participate in this study based upon the Tokyo, Japan hourly wage at the time the study was undertaken.

1. Participants take the math exam for the first time. This is to evaluate initial knowledge of the topics being assessed.

Although from now on, the word “groups” will be used, due to only having one device (HoloLens 2) available, all the experimental procedures happened with one participant at a time.

2. After the initial testing, the participants (*N*) were divided into three groups of equal size, one for the experimental group, one for the control group henceforth referred to as *control group A*, and another for the control group henceforth referred to as *control group B*

3. Control Groups:
   We wish to compare MR-assisted learning to two other methods of learning: a traditional paper- based approach and a *free* approach. The paper-based approach was implemented with control group A, and the free approach was implemented with control group B.

The paper-based approach consisted of giving students a resource containing the full solutions to the problems of the exam. This resource will be henceforth referred to as “cheat sheets”. These cheat sheets containing full step-by-step solutions, the same as the ones used by the experimental group, were procedure-generated by the Wolfram Alpha Full Results API, the same API generating answers for the experimental group. Hence, control group A and the experimental group will have access to the exact same answers only through different mediums.

The *free* approach used by control group B consisted of allowing students to use whichever methods they had at their disposal. This scenario aims to more accurately simulate a comparison between the experimental group and the tools students have in their daily study sessions. We expected students to use the aforementioned apps, look up solutions on web search engines, or even use generative software to generate the answers. In the feedback sheet, we ask students what tools they used during the study phase. The results of this inquiry will be discussed in detail in the results section, but students indeed used mainly the tools mentioned previously.

Hence, the control group was further divided into two groups, and the participants of each group used one of the aforementioned tools.

1. Experimental Group:
   All members of the experimental group used the HoloLens 2 and Peyethagorean as the method through which they learned how to solve the questions.
2. Participants in both groups underwent a 1-hour study phase in which they used the tools given to each group respectively to learn how to solve the questions presented during the exam and also check whether their initial solutions were correct or not.
3. Students then took the exam a second time. Given that the exam applied was the same, improvements were expected in all groups. Hence, rather than simply observing whether there were improvements or not, we aimed to compare the level of improvement across groups.
4. The results of each group were then compared using mean differences, ANOVA, and t-test scores.

THE TEST

The test used for the quantitative analysis had questions on varying degrees of difficulty and varying topics. Varying degrees of difficulty are important as item difficulty and length of solution can play a key role in producing clear results in pretest-posttest design experiments (Dimitrov et al., 2003). The topics were used to measure the student’s understanding of specific techniques within the field of Algebra, more specifically, Integrals. That is, groups of questions on the same topics, but with questions of varying difficulty, were given to participants to assess the extent to which they were able to acquire knowledge of the construct that underlines the question. For instance, if the construct being analyzed is the technique of *trigonometric substitution*, participants were first given a question that illustrates the simplest use-case of this technique, and then the subsequent questions required using the technique in increasingly creative ways. This approach to the design of the test allowed us to compare not only overall results but also improvements in understanding specific constructs and techniques within the field.

The tests used during the experiment were in English and Japanese, respectively. Subjects were able to freely choose the language they felt most comfortable with, as the content of the questions in both tests was the same.
After finishing the post-test, subjects were given a paper-based questionnaire about some of their impressions of the experience. Subjects in all groups were asked questions, but the questionnaire given to the experimental group included questions about the hardware used, the software used, and the tool overall. That is, they will evaluate separately their experience with the HoloLens 2, their experience with Peyethagorean, and their experience with the tool, which is the combination of hardware and software.

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

By asking questions to members of the control group, we aim to check whether some characteristic of the experiment has stood out in any way that could affect results, asses their perception of the test, their previous knowledge of the constructs being analyzed, and what tools they usually use to study Algebra.

QUESTIONNAIRES

The members of the control groups answered the questionnaires in English and Japanese. The experimental group answered the questionnaire outlined in English and Japanese.

RESULTS

As shown in the qualitative assessment sheets, the questions were divided into two sections. One section concerned the subject and the test, and the other section concerned the tool used. Since all participants answered questions about the test, we will first look into those.

EVALUATION OF TEST AND INDIVIDUAL KNOWLEDGE

The specific formulation of each question in English or in Japanese. Table 2 shows results for the questions that used a Likert scale in the "About the Test" section. The values are rounded to one decimal place.

Table 2

Descriptive Statistics of the 5-Point Likert Scale Results (1 – Very Hard and 5 – Very Easy)

Question	Mean	Median	S.D.
Test Difficulty	2.0	2	0.7
Knowledge of Polynomial Long Division	4.1	4	0.9
Knowledge of Partial Fraction Decomposition	4.1	4	0.7
Knowledge of Substitution Method	3.8	4	0.9
Knowledge of Trigonometric Substitution	2.7	3	1.0

Students, on average, considered the test difficult. The mean for prior knowledge of each construct matches the mean higher in their scale of knowledge. The coefficient of variation (standard deviation divided by mean) was below one for all items, suggesting that the values for each item were not widely spread.

Wolfram Alpha tops the list as the most commonly used tool. It is also notable that other apps that provide a similar service to Wolfram Alpha (2024) were mentioned i.e. Photomath, Symbolab, Integral-calculator, and Geogebra. Photomath, 2024 is especially worthy of mention, as it has one significant similarity to Peyethagorean. Photomath uses a cellphone camera to capture images and identify equations in them. Although it does not implement any level of immersive technology, it has the added benefit of being often faster than writing equations manually. It should also be noted that the total number of mentions in the table above is 60. That count represents the fact that most students mentioned more than one tool.

Similarly, the subsequent question asked subjects for any comments on the exam. To summarize those, we listed, same as above, overarching themes in comments and how often they were mentioned in Table 3 below. Not all

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

students provided answers to this question, and some answered “no comments”. In the case of multiple themes in a single answer, they were all individually counted.

The most common comments referred to the test being either too long for the given time or too challenging. The test was made to be entirely completed in one hour and a half. Subjects were only given, however, 1 hour in both attempts. Limiting students’ time was a decision made with the intent of removing time from consideration in the posterior analysis of improvements. That is, if students were given time in excess, observing overall improvement could be harder as students with extra time and higher initial scores would not benefit from the extra time as would students with lower initial scores. A few students with higher initial scores commented that they would be able to solve more questions were they given more time, but students overall had the chance to properly implement the knowledge from the study session during the second attempt at the exam.

Table 3

Topics Mentioned by Subjects in Comments and Their Respective Occurrences

Comment	Mentions
Not enough time	4
Challenging	3
Some problems were too difficult	3
Did not remember trigonometric relation	2
Fun	2
Good question progression	2
Seeing solutions made solving easier	2
Solutions were overly complicated	2
AI had a hard time solving some problems	1
Appropriate difficulty	1
Appropriate time	1
Cover-up method helped	1
Solutions were too long	1
Space for solutions was too small	1

Note. Sample size $n = 21$

Furthermore, some students commented that the solutions were either too long or too complex. However, we designed the test to use questions of different solution lengths and hoped to observe how questions of different lengths could be a better or worse fit for the tool used in this experiment. The questions in the first part of the exam, the partial fraction decomposition questions, did not progress significantly in difficulty from one to another. They did, however,

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

progress in terms of length. Whereas the first question required solving a linear system comprised of only 2 equations and 2 variables, the last one required solving a linear system of 4 equations and 4 variables. We expected that these questions would translate poorly to the tool being used in the experiment since most of the steps in the solution involved simply linear combinations[8] of rows and columns in the matrix that represent the linear system.

The choice itself of working with matrices is not one favored by students for linear systems of smaller dimension, as simply substituting one equation into another until the value for the first variable is discovered is often a faster solution.

Another comment by members of the experimental group and members of control group A, those that were given solutions, was that they often did not remember steps "in between steps". That is, when solving integrals, Wolfram assumed that the person solving had full knowledge of trigonometric relations, and would use those as a basis for some of the substitutions it made. This, however, was not true for some subjects, especially those who were not in the first years of the undergraduate course. On the other hand, students who were freely allowed to use whichever tools they wished to use often spent time using tools that were not helpful. For instance, one subject who relied on ChatGPT (OpenAI, 2024) for the entire study session mentioned that some of the questions were not properly solved by it, or that ChatGPT "hallucinated" some steps.

EVALUATION OF MR TOOL, HARDWARE, AND SOFTWARE

A total of eight subjects participated in the experimental group. Besides the items discussed in the previous section, the members of the experimental group also answered questions about the MR experience. The first seven of those were Likert-scale items. The results of those items are summarized in Table 4.

Table 4

Descriptive Statistics of Responses to the MR Experience Questionnaire (1 – Difficult and 5 – Easy/Good)

Questions	Mean	Median	S.D.
Overall experience with the *tool*	3.9	4.0	0.8
Ease of use of the *tool*	3.9	4.0	1.0
How the was learning affected by the *tool*	3.3	3.0	0.7
How likely are you to recommend the *tool*	4.1	4.5	1.1
How likely are you to recommend the *tool* to an educational institution like XXXXX	3.6	4.0	0.9
Rate your comfort using HoloLens 2, the hardware	3.1	3.0	1.0
Rate your experience using Peyethagorean	3.9	4.0	0.6

Note. Sample size $n = 8$

Of all the items above, the one with the lowest mean was the item asking about comfort while using the HoloLens 2. Their specific complaints will be looked at in further detail during the analysis of the comments in Table

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

5. As for the software, despite being a prototype, it was still positively received by subjects, suggesting that improving the hardware could improve the overall experience significantly.

The average for the overall experience was 3.9, where 4 stood for "good". With a standard deviation of only 0.8, we can affirm that over 70% of subjects had an experience between neutral and good. A very similar result was observed when it comes to "ease of use" of the tool. In the item concerning whether the tool made learning easier or harder, the median was 3, suggesting that for a majority of subjects, the tool affected neither negatively nor positively the learning experience. The mean, however, was 3.3, slightly above neutral.

The last question of this section was open-ended. It prompted subjects to write down any comments on the tool, the hardware, and the software. Same as done for the comments regarding the exam in the previous section, we have summarized the results and grouped them by topics approached in the comments. These grouped topics and the number of times each was mentioned are presented in the tables below. Table 5 has topics concerning the hardware. Table 6 has topics concerning the software.

Table 5

Comments Made by Subjects about the Hardware and Their Respective Occurrences

Comment	Mentions
Eye strain	3
Hardware was easy to use	3
Display was too bright	1
Felt dizzy after some time	1

Note. Sample size *n* = 8

Table 6

Comments Made by Subjects about the Software and Their Respective Occurrences

Comment	Mentions
Software could improve	3
Wish there were touch commands	2
Wish software could recall previous questions	2
Software was easy to use	1

Note. Sample size *n* = 8

In terms of comments regarding the hardware, the two most common topics were *eye strain* and *ease of use*. The comment on eye strain is especially concerning as subjects were not wearing the hardware for longer than 1 hour and a half. Furthermore, subjects were interacting with the panel displayed by the glass and with the pen and paper simultaneously, so they were not looking at VR objects throughout the entire time. There are guidelines on how to properly adjust the hardware to users, but despite following both the guidelines provided by Microsoft and the initial setup processes, half of the subjects made complaints about the display.

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

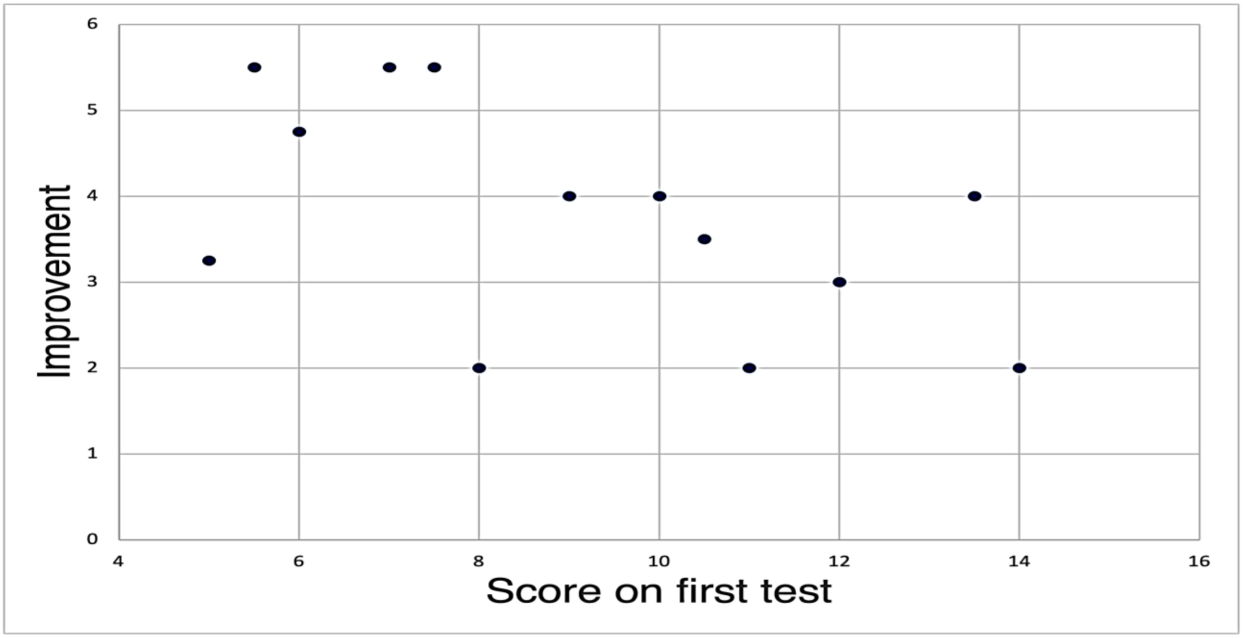
The most common comment regarding the software was that it could be improved. The subjects gave verbal details upon conversation after finishing the experiment. Two of the 3 students who mentioned that the software could improve mentioned that the software would not always recognize commands and that the software was somewhat slow. Among the subjects of the experimental group, there were subjects native to Japan, France, Malaysia, Thailand, Indonesia, and India. Since none of the subjects were native speakers of English, this often led to the software not picking up their pronunciation of certain words. Foreseeing that this might happen, two different voice commands were attached as activators to the actions that were used more often. We believe this strategy worked successfully because in observing study sessions and recordings of the study sessions, we noticed that students often picked a "favorite" command out of the two when two options were given. For most students that command was the quickest and smallest one, that is, students favored "next" and "back" over "forward" and "previous". However, students from two countries could not properly get the word "next" to be recognized every time, and, as a result, ended up settling for the word "forward", despite this word being longer. There were also comments on the fact that users wished to have touch commands on top of voice commands.

Lastly, regarding comments about the tool, the majority of comments mentioned that it was easy to use. Given that the tool was straightforward and its capabilities quite limited, these comments make sense from the perspective that users could quickly understand the tool's capabilities and implement them.

One important aspect to keep in mind when looking at score improvements is that participants who had a lower initial score on the first test, showed greater improvement. That is, if subjects of one of the groups had a higher number of subjects with lower scores on the first exam, they are expected to have some advantage in improving since there are more questions and easier questions for students to learn how to properly solve. This is evidenced in the graph in Figure 6. The graph below shows that the mean improvement of subjects, represented in the y-axis, decreases as the initial grade, represented by the x-axis, increases. We did not attempt to draw a curve that could represent this trend as the questions had varying levels of difficulty, and, therefore, the fewer questions there were left to correctly solve after the first exam, the harder were the questions left unsolved, making it so that such a trend would not be linear, and would depend on a proper evaluation of the difficulty level of each question.

Figure 6

Improvement over Study Phase in Relation to Initial Score with a Sample Size of n=21.



Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

The result on the graph above, however, reinforces the value of the data shown in Table 7 below. Despite having the highest mean initial score, the experimental group still had improvement, on average, better than the control group *A*, which had the same solutions, but in written format.

Table 7

Mean Scores for Each Group for the First Exam and Improvement over the Study Session.

Mean scores	Exp	Ctrl *A*	Ctrl *B*
First exam	9.7	8.2	7.8
Improvement	4.1	3.1	4.4

Note. Exp (*n* = 8), Ctrl *A* (*n* = 7), Ctrl *B* (*n* = 6)

As the table above shows, control group *B* had a higher improvement if compared to both other groups. This group, however, also had the lowest mean initial scores, hence, wider room for improvement if compared to other groups.

To perform an analysis of variance (ANOVA), our null hypothesis (H0) is that there was no significant difference in learning outcomes from the different methods employed in the study session. Hence, our alternative hypothesis (H1) is that there was a significant difference between the methods employed. We set our α value to 0.05. The results are displayed in Table 8.

Table 8

One-way ANOVA Test between All Three Groups.

Sample variation	SS	df	MS	F	P value	F-crit
Between groups	6.012	2	3.006	0.779	0.474	3.555
Within groups	69.440	18	3.858			
Total	75.452	20				

Note. $\alpha = .05$

Furthermore, since in this research, we wanted to observe possible changes from altering only the medium, comparing the groups individually, especially the experimental group and control group A, could be of relevance. Hence, we performed f-tests and t-tests between the groups, setting our α value once again to 0.05. The results of these tests are presented in Table 9.

Table 9

F-test and T-test Results Comparing Groups Two at a Time

Groups compared	F-test	t-test
Exp and Ctrl *A*	0.951	0.372
Exp and Ctrl *B*	0.691	0.785
Ctrl *A* and Ctrl *B*	0.741	0.250

Note. $\alpha = .05$

These results indicate that a significant difference was not observed between the groups. We attribute the lack of significant difference between all groups and between paired groups to the initial scores. As shown in Table 7,

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

members of the control group *B* had considerably lower initial scores, which enabled them to have higher improvements over the study session. A fair comparison would indeed require students to start from equal initial scores. Given the current sample, however, that is not possible, as students' initial scores were quite scattered. It stands, however, that despite considerably higher initial scores, the experimental group had improvement better than that of control group *A* and comparable to that of control group *B*. Hence, we still believe that there is merit in experimenting again, either by creating a model for correcting improvement by adding weight in proportion to the initial score or simply by expanding the pool of subjects.

DISCUSSION

The results of this study highlight the nuanced impacts of utilizing a mixed reality (MR) device specifically the HoloLens 2, for teaching students' Algebra. While quantitative analyses did not demonstrate statistically significant differences in learning outcomes among the groups, qualitative feedback revealed the use of MR tools fostered an engaging and novel learning environment. Despite higher initial test scores, the experimental group showed comparable improvements to those in control groups, suggesting that MR-based interventions might bridge gaps in educational engagement rather than outperform traditional methods purely on score improvements.

The qualitative findings indicate that the immersive nature of MR technology enhanced student motivation and offered a unique, hands-free way to engage with algebraic problem-solving. However, challenges such as eye strain and discomfort with prolonged device usage were noted, underscoring the need for improved ergonomic design in MR hardware. Software usability and response time were also highlighted as areas for potential enhancement to make MR tools more effective and seamless in learning applications.

Overall, while the use of MR tools such as the HoloLens 2 did not surpass traditional methods in quantitative learning improvements, the wide range of initial scores limited the effectiveness of this analysis. Nonetheless, both the qualitative and qualitative data support the potential of MR as an engaging supple- mental tool that could be optimized for better academic support in higher education settings.

There are few reports in the peer-reviewed published literature on using the Hololens 2 to teach Algebra to compare these results to. Prior research on using the Hololens 2 for medical education has been shown to be effective (Zaccardi et al., (2023)). There is one reported article on teaching math in social media (McNiell, 2018) using the Hololens 2. The reason for the lack of usage in education is the that the Hololens 2 is relative expensive compared to other head mounted MR displays. There are reports of using MR to teach mathematics with various degrees of success (Agrawal, 2024), (Fernández-Enríquez & Delgado-Martín, 2020).

CONCLUSION

The qualitative evaluation was quite insightful in better understanding how subjects interacted with the test and with the MR tool. Overall, the students who used the tool were pleased with their experience and willing to implement it in their self-study sessions. Furthermore, students' answers when asked what tools they were using confirmed our expectation that step-by-step procedure-generated solution software ranks as the most common tool used by students, although the exact tool varied between students. Another important aspect revealed by the questionnaire is that comfort still seems to be an issue for a large number of subjects. Despite only using the hardware for slightly over one hour, many subjects made complaints related to comfort and usage. Moreover, the comments about the software can be useful guidance for improving UX for apps developed for this purpose in the future.

Although research data did not show significant differences between the different intervention methods, that is, the different methods used by the groups for the study session, the increase in difficulty to increase improvement shown by Figure 6 and the comparison between mean first scores and mean improvements shown in Table 7 are a strong indicator that experiencing the solutions in an immersive and seamless environment may have a positive effect for students learning. However, observing that effect without comparing students with the same initial test score or without normalizing their improvement in comparison to their initial score has its limitations. Overall, this research was a pilot study consisting of a limited number of participants, and a further MR study is needed on a larger group of participants to validate the observations.

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

ACKNOWLEDGEMENT

Special thanks are given to the 21 participants in this undergraduate research project and for their feedback. The participants were compensated for their time using a Japan Society for the Promotion of Science (JSPS) *Kakenhi* basic research grant entitled, Personalized Online Adaptive Learning System, grant # 20H01719.

The first author posthumously expresses his deepest gratitude to fellow Japan MEXT scholarship student Julio César Tapara Tejada for his unwavering support, guidance, and inspiration. Though no longer alive, his impact and inspiration lives on with the author and his friends.

The Microsoft Hololens 2 was provided for research use by the Center for Innovative Teaching and learning (CITL) Online Content Research and Development section's professor and the Super Smart Society Doctoral Degree Program.

This paper was presented at the 2024 World Immersive Learning Labs Symposium (WILLS) in Tokyo, Japan, in March 2024. The authors express their gratitude for the questions and feedback received at the symposium.

REFERENCES

Agrawal, A. (2024). A mixed reality environment for Mathematics. *International Journal of Scientific Research in Engineering and Management*, *8*(03), 1–5. https://doi.org/10.55041/IJSREM29739

Coimbra, M. T., Cardoso, T., & Mateus, A. (2015). Augmented reality: An enhancer for higher education students in math's learning? In *Proceedings of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* (pp. 332–339). Procedia Computer Science, 67. https://doi.org/10.1016/j.procs.2015.09.277

Dimitrov, D. M., & Rumrill, D. P., Jr. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159–165.

Fernández-Enríquez, R., & Delgado-Martín, L. (2020). Augmented Reality as a Didactic Resource for Teaching Mathematics. *Applied Sciences*, *10*(7), 2560. https://doi.org/10.3390/app10072560

GeoGebra. (2023). *What is geogebra?* https://www.geogebra.org/about

Kang, K., Kushnarev, S., Wei Pin, W., Ortiz, O., & Chen Shihang, J. (2020). Impact of virtual reality on the visualization of partial derivatives in a multivariable calculus class. *IEEE Access*, *8*, 58940–58947. https://doi.org/10.1109/ACCESS.2020.2982972

Khronos Group. (2024). *Openxr*. https://www.khronos.org/openxr/

Mahmoud, K., Harris, I., Yassin, H., Hurkxkens, T. J., Matar, O. K., Bhatia, N., & Kalkanis, I. (2020). Does immersive VR increase learning gain when compared to a non-immersive VR learning experience? In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies. human and technology ecosystems* (pp. 480–498). Springer International Publishing. https://doi.org/10.1007/978-3-030-50506-6_33

Martín-Gutiérrez, J., Fabiani, P., Benesova, W., Meneses, M. D., & Mora, C. E. (2015). Augmented reality to promote collaborative and autonomous learning in higher education [Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Networks Era]. *Computers in Human Behavior*, *51*, 752–761. https://doi.org/10.1016/j.chb.2014.11.093

Mathpix. (2024). https://mathpix.com

McNeill, S. (2018, August 13). *Teaching Math with Microsoft Hololens*. https://samuelmcneill.com/2018/08/13/teaching-maths-with-microsoft-hololens/

Microsoft. (2024). *HoloLens 2*. https://www.microsoft.com/pt-br/hololens/hardware

Microsoft. (2023b). *HoloLens 2*. https://www.microsoft.com/en-us/hololens

Microsoft. (2024). *Mixed reality toolkit*. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/

Montoya, D. B., Plascencia, M. L., & Herrera, L. M. (2021). The role of reality enhancing technologies in teaching and learning of mathematics. *Computers & Electrical Engineering*, *94*, 107287. https://doi.org/10.1016/j.compeleceng.2021.107287

OpenAI. (2024). *ChatGPT*. https://openai.com/blog/chatgpt

Park, S., Bokijonov, S., & Choi, Y. (2021). Review of microsoft hololens applications over the past five years. *Applied Sciences*, *11*(16). https://doi.org/10.3390/app11167259

Photomath. (2024). https://photomath.com

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024

Peyethagorean. (2024). *Hololens 2 demo Youtube video*. https://youtu.be/YjL-PanAufE?si=XixvclhW08aIEijw

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Sattar, M. U., Palaniappan, S., Lokman, A., Hassan, A., Shah, N., & Riaz, Z. (2019). Effects of virtual reality training on medical students learning motivation and competency. *Pakistan Journal of Medical Sciences*, *35*(3), 852–857. https://doi.org/10.12669/pjms.35.3.44

Shaghaghian, Z., Burte, H., Song, D., et al. (2024). An augmented reality application and experiment for understanding and learning spatial transformation matrices. *Virtual Reality*, *28*, 12. https://doi.org/10.1007/s10055-023-00904-x

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical, and laboratory studies. *Biochemia Medica (Zagreb)*, *31*(1), 010502. https://doi.org/10.11613/bm.2021.010502

Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., & Iloreta, A. (2017). Immersive virtual reality as a teaching tool for neuroanatomy. *International Forum of Allergy & Rhinology*, *7*(10), 1006–1013. https://doi.org/10.1002/alr.21986

Unity Technologies. (2024). *Unity Software*. https://unity.com

Wello Soft. (2024). *Texdraw*. https://assetstore.unity.com/packages/tools/gui/texdraw-51426

Wolfram. (2024). *Wolfram|Alpha full results api reference*. https://products.wolframalpha.com/api/documentation

Wolfram Alpha LLC. (2023). *About wolfram|alpha*. https://www.wolframalpha.com/about

Zaccardi, S., Frantz, T., Beckwée, D., Swinnen, E., & Jansen, B. (2023). On-Device Execution of Deep Learning Models on HoloLens2 for Real-Time Augmented Reality Medical Applications. *Sensors*, *23*(21), 8698. https://doi.org/10.3390/s23218698

Received Date: December 11, 2024 / **Revised Date:** December 27, 2024 / **Accepted Date:** December 30, 2024