The Impact of User Interface on Games for Teaching Children

by

Maximilian Thompson

A Major Qualifying Project

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

in Computer Science

by

May 2021

APPROVED:

_________________________

Smith, Therese
Abstract

We attempt to experimentally identify a difference in how different user interface design elements effect different age groups' ability to learn new material, but we fall short due to a lack of responses and the limitations of our web-based method of experimentation. We hope to inspire further and more reliable research into the line of questioning we present.
# Contents

1 Problem Description .......................... 1
  1.1 Measuring Learning Outcomes ................. 2

2 Methodology ................................ 3
  2.1 Data Collection ............................ 4
  2.2 Anonymity and Participant Protection ......... 5
  2.3 Participant Selection Criteria ................. 5
  2.4 Data Analysis ................................ 6

3 Application Design .......................... 7
  3.1 Application for the Experiment ............... 7
    3.1.1 Structure of Application ................. 7
  3.2 Application for Data Analysis ................. 9

4 Results and Conclusions .................... 10
  4.1 Expected Results .......................... 10

5 Limitations and Future Work .................. 11

Appendices .................................... 13

A Application Source Code .................... 14
Chapter 1

Problem Description

There is a lack of work attempting to find whether there is a difference between children and adults with regard to how the quality of digital learning is effected by the design of a user interface and presentation of the learning material. Differences between what elements of user interfaces positively or negatively impact learning outcomes of separate age groups would indicate that, for the efficacy of games for teaching children, further research would need to be done into how common design principles change when working with children.

We intend to contribute toward resolving this by designing and running an experiment to measure a learning outcome that all ages capable of operating a computer may participate in. Our goal is to attempt to identify any differences in the impact of variations of the user interface present in the experiment on the learning outcomes of people of different age groups.

Unfortunately the ongoing COVID-19 pandemic necessitates a remote approach to running an experiment, which makes obtaining a controlled environment for running the experiment borderline impossible for us. We chose to adapt to the situation by developing a web application rather than a desktop application for the purpose of facilitating voluntary participants to self-administer the experiment from their own homes. This has the added benefit of bolstering the anonymity of participation.
1.1 Measuring Learning Outcomes

Kraiger, Ford, and Salas [KFS93] define learning outcomes with three conceptual categories: cognitive outcomes, skill-based outcomes, and affective outcomes. Cognitive outcomes include acquisition of verbal knowledge, knowledge organization, and cognitive strategies. Of the three, verbal knowledge is the most familiar; it represents what we traditionally view as the product of learning (increased knowledge of a subject) and it is by far the easiest to measure, as it may be measured using methods such as a standard multiple choice quiz or free recall quiz. These measurement methods lend themselves well to short form computerized applications, however teaching verbal knowledge within our constraints (a short web application) is prohibitively difficult; thus we choose to focus on generating and measuring declarative knowledge-based learning outcomes as it is something Kraiger, Ford, and Salas acknowledge to be similar to verbal knowledge-based learning outcomes. Thankfully, according to Dacin and Mitchell [DM86], declarative knowledge may be measured in much the same way.
Chapter 2

Methodology

Before the experiment begins, the participant is asked to identify which age group they belong to. If the participant is a minor, a parent or guardian is requested so that informed consent may be made. Following informed consent, the participant will be presented with instructions for completing the experiment.

The experiment consists of two phases: an initial learning phase which attempts to teach the participant the information they will need to complete the testing phase that follows. The goal is to attempt to measure how variations in the presentation of the learning phase effect recall and application of knowledge in the testing phase. Participants are anonymous and are not monitored by an observer while taking part in the trial. Participants are given the option to cease participation in their trial at any time for any reason with no consequence, and metrics collected by the system are only retained if the participant gave consent to do so at the end of the trial.

In the learning phase of the experiment, participants are randomly given one variation of the learning material. The variation between the learning material is defined by the random selection of specific options at each of the number of variable points. All options modify the presentation of the learning material.

The learning material consists of a set of number to color associations. How this information is presented varies in the following ways: The continue button will either resemble the back button in stylization or it will be larger and green. The color of the number will either match the associated color or it won’t. A static
cartoon character graphic “presenting” the learning material will either be present or it won’t. Each association is either present on its own page with a back and continue button, or all associations share a long page that must be scrolled down to reach a continue button that proceeds to the next phase.

After completing the learning phase of the experiment, a one minute break was intended to be enforced during which the participant is asked to count backward from one hundred. The intention is to limit the impact of short term memory on the results of the testing phase by interfering with rehearsal and delaying application of information. However, this was not implemented due to a technical oversight.

In the testing phase of the experiment, the participant is presented with a series of colored squares and asked to select the associated number. The participant is allowed as much time as they want. The participant is not informed of whether or not they selected the correct answer after they select their answer. There are thirty trials before the end of the testing phase of the experiment. The average time taken to respond is recorded.

2.1 Data Collection

Following the conclusion of the experiment, the participant will be presented with all of the data generated by their participation and will be asked, given that they still consent, to submit the data to a linked SurveyMonkey survey. This has been chosen over automatically recording the data due to the limitations of GitHub Pages, and the inability to use a database for data collection.

The data collected will be the participant’s age, the learning material variation options selected by the web application, and the average amount of time it took the participant to respond to each trial.
2.2 Anonymity and Participant Protection

While an argument may be made that the variation of the learning phase may result in an unique identifying combination of options, the fact that the experiment is administered by an automated system that does not monitor the participant in any way other than the collected metrics results in the dataset being completely anonymous. The dataset will not be published as, if it were, a participant could view the dataset and, given that they had a unique variation of the learning phase of the experiment and their age, identify their entry in the dataset (as the participant is the only observer of their own trial).

Additionally, before beginning the experiment the participant is informed of the structure of the experiment and their right to cease participation at any time with no penalty. They are also informed that their results will not be recorded until they have given consent and submitted the results through the provided survey following the conclusion of the experiment.

2.3 Participant Selection Criteria

Participants consist of WPI students and relatives of WPI students of ages 13 and up. Excluded from the research population are people who are blind, color blind, people who can not read English, and people who can not operate a computer unaided. Full color vision is required to complete the testing phase of the experiment. Additionally, because the experiment is administered by an automated system, the participant must be able read English to read the instructions and informed consent form. The participant must be able to operate the computer they use to participate unaided in order to properly participate in the experiment, but
children are not expected to understand the structure of the experiment without
the required assistance of a parent or guardian.

2.4 Data Analysis

The data will be analyzed by performing a polynomial linear regression \([Pec12]\) with the data. All options for the display of learning material will be coded as
dummy variables. Features will be selected using lasso regression utilizing cross
validation to identify the parameters for the lasso regression that provide the best
model. If there are fewer than 100 participants, the bootstrap method \([K+95]\) will
be applied to enhance the dataset.

Additionally, the dataset will be broken into age ranges with the above method
repeated to produce additional models for specific age ranges. The intent of doing
this is to see if the coefficients of the selected model for each age group differs from
the coefficients of the overall model and from each other.
Chapter 3

Application Design

There were intended to be two applications in this project: an web application for administering the experiment, and a local application used to analyze the generated data with machine learning. Due to an insufficient number of responses to the experiment, the application for data analysis was never produced. The source code for the experiment administration application may be found in appendix A.

3.1 Application for the Experiment

The application for conducting the experiment detailed in chapter 2. Methodology is a web application that is hosted using GitHub Pages that links the user to a SurveyMonkey survey at the end of the experiment for the submission of results. This method of data collection was chosen because it simplifies the application and eliminates hosting costs. The primary alternative was to pay for hosting and run a database that the web application would automatically submit results to.

3.1.1 Structure of Application

The application is a simple web application that is broken up into an HTML component, a vanilla javascript component, and a CSS component. The HTML component lays out a number of elements that are used to display each phase of the experiment. A majority of these elements are initially hidden, and over the course
of application execution are both shown and hidden, as well as have their direct styling modified. Both elements with classes and elements with identifiers are used where applicable.

The CSS component of the web application simply provides a baseline styling to the application which is further modified with direct styling in the javascript to produce the effects required for the different variants of the learning material. Of note, the portion devoted to the learning phase and testing phase assume and enforce a specific absolute pixel size of the content area, and center the content area with the intention of creating a consistent experimental environment across multiple platforms.

Finally, the javascript component of the web application is the driving force behind the administration of the experiment. It handles the dynamic components of each experimental phase as well as progression between experimental phases. Due to the small scale of the application, it was coded in vanilla javascript with no web-frameworks. At the beginning of execution, the pseudorandom number generator is used to generate trials for the testing phase as well as select options for the learning material variation. Then a number of listeners using both named functions and lambda functions where applicable manage the progression and functionality of each phase of the application. Changes to the page CSS are made at the beginning of the learning phase to implement the options selected at the beginning of execution, and the response time for each trial in the testing phase is tracked in milliseconds using the system clock. At the end of the experiment the results and the codified options selected for the learning material are displayed to the user and they are asked to submit it to the linked survey.

Of note, following the end of the study on May 13th, 2021, the application will be slightly modified to note that the study is no longer being run, and the link to
the survey will be removed. Otherwise the application will remain unchanged.

3.2 Application for Data Analysis

The application for data analysis was never produced because the number of responses required to actually utilize it was never met. If it were produced, it would have been a collection of python scripts utilizing scipy or some other python library that implements simple machine learning algorithms to generate a collection of linear regressions on subsets of the data defined by age ranges. It is likely that the bootstrap method would have be used to make sure that each subset had a similar amount of data points and improve comparability between the subsets.
Chapter 4

Results and Conclusions

Unfortunately we did not receive enough responses to our study to draw any reasonably founded conclusions. For more information on how we would have analyzed the resulting data, see section 2.4 for information on our data analysis methods and 3.2 for information on how we would have implemented these methods.

4.1 Expected Results

We obviously can not provide much more than speculation here, but it is safe to say that we expected there to be some difference between how different age ranges responded to the presence of different elements in the learning material. For instance, children of a younger age may have been more strongly effected by the presence of a cartoon character on the screen during the learning material. We hoped to see this in a difference between normalized response times of different groups given the presence of a specific feature or collection of features.

Finding evidence of this would have provided a springboard for further exploring how the design of a user interface in an application for teaching children may need to differ from one for teaching adults. So although this research would not have directly benefited the field of teaching children through games, it would have furthered research in a direction that may have one day caused great benefit. We hope that this line of questioning may be further explored in the future by other groups regardless of our lack of time/responses to produce reliable evidence.
Chapter 5

Limitations and Future Work

The primary limitations of the conducted study are the low resulting sample size, the short duration of the study, and the lack of a standardized laboratory environment. The design of the experiment unfortunately required a decently large sample size due to the intention to utilize machine learning to identify relations between multiple different factors and response times. Unfortunately, we only received a handful of responses which is so far below the number required that we were unable to draw any useful conclusions. The short duration of the study is the most likely culprit in the small resulting dataset, and if this study were run again with significantly more time and more rounds of advertisement to potential participants, it may have much better results. Finally, the lack of a standardized laboratory environment does give any resulting data collected less reliability and probably more variance overall because there are a number of factors (noise level around computer used to participate, size of screen, etc.) that we are unable to control for. Repeating the study with laboratory environment for participants would likely have significantly more reliable results.
Bibliography


Appendices
Appendix A

Application Source Code

The source files for the application in this project as well as the source files for generating this report with \LaTeX can be found in GitHub in the following repository:

https://github.com/ameliasthompson/MQP

Additionally, the live version of the web app can be found on the GitHub pages for the above repo:

https://ameliasthompson.github.io/MQP/