

Spring 2019

Choosing a Challenge: Building a Choice-Adaptive Points System for a Digital Game

Annie Chen
Yale College, New Haven Connecticut

Abstract:

As technology gains popularity and prominence in education, software has made it possible for more schools to implement personalized learning. But, by putting the choice of what to learn and how fast to learn it on the student, personalized learning may disadvantage students who might be inclined to choose less challenging work that slows their academic progress. Educators and technologists must research how to help students learn to choose challenging problems. This work documents the development and iteration of an adaptive choice-based computer game for learning multiplication facts. Through a novel point system that places a higher value on difficult problems they have previously avoided, the game attempts to encourage students to practice more challenging problems. The game structure and user interface was developed through the creation of paper and digital prototypes, and through feedback from students and teachers on those prototypes. The final game was tested on a pilot study of 18 children. The data show a trend towards the novel point system increasing the number of difficult questions attempted. This adaptive incentive structure shows potential to shape student choices, and should be further studied in different educational environments.

Suggested Citation: Chen, A. (2019). *Choosing a Challenge: Building a Choice-Adaptive Points System for a Digital Game* (Unpublished Education Studies capstone). Yale University, New Haven, CT.

This capstone is a work of Yale student research. The arguments and research in the project are those of the individual student. They are not endorsed by Yale, nor are they official university positions or statements.

Choosing a Challenge:
Building a Choice-Adaptive Points System for a Digital Game
Annie Chen

Senior Capstone in Education Studies
Advisors: Professor Mira Debs and Professor Julian Jara-Ettinger
April 18, 2019

Abstract

As technology gains popularity and prominence in education, software has made it possible for more schools to implement personalized learning. But, by putting the choice of what to learn and how fast to learn it on the student, personalized learning may disadvantage students who might be inclined to choose less challenging work that slows their academic progress. Educators and technologists must research how to help students learn to choose challenging problems.

This work documents the development and iteration of an adaptive choice-based computer game for learning multiplication facts. Through a novel point system that places a higher value on difficult problems they have previously avoided, the game attempts to encourage students to practice more challenging problems. The game structure and user interface was developed through the creation of paper and digital prototypes, and through feedback from students and teachers on those prototypes. The final game was tested on a pilot study of 18 children. The data show a trend towards the novel point system increasing the number of difficult questions attempted. This adaptive incentive structure shows potential to shape student choices, and should be further studied in different educational environments.

Table of Contents

Abstract	1
Table of Contents	2
Introduction	3
Literature Review	6
Factors in Students Choosing Challenges	6
Interventions Towards Growth Mindset Behaviors	9
Choice Based Assessment	10
Putting it All Together	14
Game Development	16
Exploration of how students choose math problems to work on when given a choice	16
Paper Prototype	16
Digital Prototype	18
Teacher Observations	23
Observing at Elm City	23
Discussion at St. Thomas	23
Final Game Design	24
Efforts to shape student choices, while maintaining a sense of agency.	24
Points Algorithm	28
Experimental Hypothesis	29
Method	29
Findings	31
Qualitative	31
Quantitative	32
Limitations	35
Conclusion	36
Acknowledgements	38
References	38
Appendix: Statistical Model	41

Introduction

The internet has democratized knowledge. Before, students had to learn from teachers, who would generally decide what content to cover and how students would learn that content. Computer systems have made it possible to scale the spread of knowledge. Massive Open Online Courses (MOOCs) now enroll tens of millions of learners (“By The Numbers: MOOCs in 2017”, 2018). In developing countries, digital video is a way to bring students into the class, allowing them to teach themselves and work independently (“In Latin American Classrooms, a Television Triples Enrollment”, 2014). The potential outcomes are exciting: students who were previously denied a full education for a variety of reasons may now have access.

For some, individualized learning is the ideal form of education. Early proponents of personalized learning using technology often cite a 1984 study in which students who received one on one tutoring scored higher than 98% of students taught in a lecture format (Bloom, 1984). The logistical nightmare of getting each student the equivalent of a one on one tutor kept Bloom’s findings from being implemented widely in schools. Without digital systems, most differentiated instruction for each student requires a small student to teacher ratio, and a large amount of teacher effort and attention.¹

But advances in software capability and accessibility make it increasingly possible to scale differentiated instruction, and many organizations are implementing computer-based personalized learning systems (Kamenetz, Feinberg, & Mason, 2018). One example, Summit Learning, a charter network with its own personalized learning system where students direct their

¹Differentiating instruction has been successfully accomplished in Montessori education through a sequence of hands-on materials that allow students to learn individually in a sequence of progressively more difficult exercises? (Rinke, Gimbel, & Haskell, 2013)

own learning through an online platform, prides themselves on allowing students to choose what to work on next (“Summit Learning: Homepage, n.d). Students are given ‘self directed learning time,’ where they choose which projects or content areas to work on. Another example, Khan Academy, began as supplementary material for students to practice outside of the classroom, but recently began building full courses to be used in the classroom (“School district reports”, 2018). A two year study by the RAND corporation found ‘suggestive evidence’ that personalized learning practices may be related to more positive effects on achievement, but the gains are not large enough to be conclusive, and full effects may take more time to emerge (Pane 2017).

These systems’ dependence on technology is controversial. As education technology is hailed by some as a powerful tool for improving learning, skeptics are raising various issues with technologies in classrooms. Some parents in Silicon Valley, worried that the risks of addiction and stunting development are too high, have banned screens completely for their children (Bowles, 2018). Educators have raised concerns that unequal access to technology has widened the achievement gap (Phelps, 2018). At the same time, various educational technologies, from iPads to Chromebooks to educational software, are already in many schools (Manning, 2017). For instance, the Summit Learning program will reach more than 72,000 students in the 2018-2019 school year. Khan Academy reached 12 million learners every month in 2017 (“Khan Academy 2017 Annual Report”, 2018).

Many of these new personalized learning programs offer students the power to control the pace, direction, and content over their own learning process (Richmond, 2014). Perhaps with an ideal curriculum, every student would find every subject and skill appealing to learn. But in the real world, students enjoy some subjects less than others. They may be less personally

interested, have less prior experience, find a topic particularly difficult, or enjoy something else more. Given the ability to choose, students would decide to avoid the less enjoyable subjects.

The disadvantage of this freedom is that students no longer have an external force encouraging them to work on a specific skill or content area. A challenge of increased student autonomy arises: now that students have the freedom to make more educational choices, how can we ensure that students are making good choices and choosing appropriately challenging content material? Educators have expressed worry that allowing children to choose their own learning paces will lead to lower outcomes for students who don't know how to push themselves (Meyer, 2014).

This capstone seeks to develop an intervention for a specific choice students might make that inhibits learning: the choice to work on something that they already know, or to challenge themselves to work on an area of improvement.

How can we encourage learners who tend to choose to spend time on topics and skills that are easier for them to push themselves to spend time on more difficult topics, in the way an external influence might? Digital technologies allow us to consider a game-based approach, where the choice options are adapted to each student. My capstone proposes a personalized incentive structure within a math practice game. Instead of using a static definition of difficulty, which assumes every student finds the same problems difficult in assigning points/rewards, the incentive structure uses students' own choices to move them towards practicing the problems they are avoiding. In the same way a great teacher might notice a student's hesitation towards a particular subject and nudge them to focus on it, the points system prompts individual students to practice where they might most need it.

Literature Review

My capstone is situated at the intersection of technology-enabled choice-based assessment, the psychology of student motivation, and instructional game design. By combining findings from each field, it explores how we can give students choices while also incentivizing students to choose problems they would otherwise avoid.

Factors in Students Choosing Challenges

Researchers have long tried to understand why some students choose challenges while others avoid it. One explanation is that teacher input matters. In an exercise completing a shape-matching game, Boggiano, Main, and Katz (1988) found that self-reported perceptions of academic competence and personal control positively correlated with intrinsic interest in schoolwork and preference for challenging activities, but only when a “highly evaluative controlling directive,” a verbal cue that told students they ought to do well, was given. When no controlling directive was given, there was no difference between the groups in terms of preference for challenge. In the no controlling feedback group, students were told, “These next two trials are for real.” Students in the high controlling feedback group were told, “These next two trials are for real. I’ll bet you want to do well this next time as you should, as you ought to.” These findings suggest that academically strong students may tend towards taking on challenge, but also reveal the importance of the feedback-giver, often the teacher. They claim that academic

competence and personal control do not predict preference for a challenge without teacher leadership.

Others point to student ability as being the most critical factor. In a math-specific context, Zydney, Diehl, Grincewicz, Jones, and Hasselbring, 2010 found that students with lower math skills would choose to start with a lower difficulty level while students with higher math skills would begin the program by choosing a question with a higher level of difficulty. This study shows that prior math achievement influences student willingness to choose more difficult problems. Unlike Boggiano et al., 1988, this study does not measure nor discuss teacher leadership.

Inoue (2007) complicates the claim that academic competence is the main driver for challenging school activities, finding that students with higher individual interest levels in solving puzzles, as self reported in a survey, chose more interesting puzzles. High levels of perceived competence did not predict whether students chose difficult puzzles. The difference in findings could be due to age group, although a conclusion cannot be drawn from just three studies with different experimental designs: Boggiano et al., 1988 studied children age nine to eleven, Zydney et al., 2010 studied high school students, and Inoue, 2007 evaluated college students. There seems to be no definitive conclusion that a certain set of factors cause students motivated to choose challenges.

Alongside questions of academic competence, others have suggested that psychological mindsets may explain willingness to tackle challenges. In recent years, psychology research around growth and fixed mindset, a binary classification where people either believe intelligence is static (fixed mindset) or malleable (growth mindset), has entered the public domain, pioneered

by Carol Dweck who summarized her research in *Mindset* (2008). According to Dweck, people with a growth mindset seek out challenges, setting learning goals. People with a fixed mindset set performance goals, aiming to do well on tasks rather than challenging themselves, thus avoiding tasks they are not confident in. This research has given us a framework for understanding why some students are willing to challenge themselves.

The results are significant for student achievement and have influenced education practice, despite accompanying criticisms. Blackwell, Trzesniewski, and Dweck, 2007 found that over two years of junior high school, growth mindset predicted an upwards trajectory in grades, while fixed mindset predicted a flat trajectory. Claro, Paunesku, and Dweck, 2016 found that students from low income families were less likely to have a growth mindset, but that those who did performed significantly above average for their income percentile. According to the research, the stakes are incredibly high. If we can help students develop a growth mindset and choose challenges, we are likely to increase achievement. Of course, there are deeper systemic issues in our education systems that cause inequality, and attempting to change the learning mindset of an individual student cannot solve the problems embedded in the systems.

Indeed, some prominent scholars have critiqued these growth interventions. For instance, Dr. Luke Wood argues that educators should affirm both student effort and achievement, because boys and men of color may never have received positive messages about their ability from educators (Hilton, 2017).

Further research indicates that students' mindset, and thus their subsequent behaviors and academic performance, can be taught. Mueller and Dweck, 1998 found that students who were given praise for their intelligence displayed less task persistence, enjoyment, and performance

than children praised for their effort. Aronson, Fried, and Good, 2002 attempted to reduce the effects of stereotype threat on African American college students by encouraging them to see intelligence as malleable rather than fixed capacity. They found that those students reported greater enjoyment of the academic process and higher grade point averages than those in the control group. Various curricula have been developed to teach these concepts to students, in hopes that understanding the psychology will help students develop a growth mindset.

Interventions Towards Growth Mindset Behaviors

Various researchers have tested many different interventions meant to cultivate a growth mindset, with generally successful results. For instance, Blackwell et al., 2007 tested an intervention where students took classes on growth mindset and found that it increased classroom motivation, while students in a control group displayed a downward trajectory in grades. Yeager et al., 2016 applied user centered design and A/B testing to design an intervention, which involved reading various articles about growth mindset. Initial studies found that the intervention improved students' grade point averages. O'Rourke, Haimovitz, Ballweber, Dweck, and Popovic, 2014 created Refraction, an educational game with an incentive structure designed to promote growth mindset by giving points based on effort, use of strategy, and incremental progress, and found children who had sessions with this incentive structure showed more persistence.

Many of the growth mindset interventions already explored explicitly teach students about what a growth mindset is and its benefits. However, students are not given a choice, and are instead presented with what they might reasonably perceive as the 'correct' way to approach

schooling. A growth mindset is presented as strictly better than a fixed mindset. Even Refraction, which uses game structures to encourage growth mindset behaviors, has an initial screen with an animation that explicitly teaches players about the growth mindset. In these interventions, students are not given a chance to figure out how the problems should be solved. Instead, they are presented with the best method. Is it possible to lead students to the best methods without plainly telling them what those methods are?

To my knowledge, there has not been significant work done on stealthy interventions towards growth mindset behaviors. By ‘stealthy,’ I mean an intervention that does not explicitly teach students about brain plasticity and the psychology of growth mindset behavior, but instead gives them the choice to perform growth mindset behaviors without necessarily revealing the mechanics of what they’re trying to change. In removing the explicit knowledge transfer, we create potential to use the performance and motivation enhancing power of choice. Thus, I see a gap in the research: Could a stealthy intervention, one that doesn’t seem explicitly educational to the student, also succeed in producing growth mindset behaviors? In this case, I’m examining whether the computer program that nudges students to try harder problems can help students take on challenging problems and also produce growth mindset behavior.

Choice Based Assessment

One method of creating this nudge is through the format of choice-based assessment. The concept of choice-based assessments, tests which measure student choices and how they affect learning, was popularized by Dan Schwartz at Stanford. In the book *Measuring What Matters Most*, Schwartz and his co-author, Dylan Arena, argue that choice is a central concern in

education (Schwartz & Arena, 2013), because parents, school leaders, and employers all want schools to teach students to make choices that will help them succeed. In order to know whether an education is accomplishing that goal, we must be able to assess the choices students make.

Until relatively recently, Schwartz claims, choice based assessment was interesting in theory but not practical to implement at scale, again due to the logistical challenges of documenting and adapting curriculum based on each student's actions. However, new technologies allow us to "provide for learner choice while still supplying ample opportunities for assessment and learning" (Schwartz & Arena, 2013). In the same way that digital technologies have created the opportunity for increased student choice, they give educators and researchers the ability to measure and assess student choice.

Ultimately, in order to measure how much the student has learned, it's necessary to assess their learning via an assessment. Digital choice based assessment allows us to further examine the impact of motivation in specific contextual conditions. Where traditional knowledge-based assessments focus only on cognition, and thus do not give any insight on motivation, choice-based assessments can generate behavioral data on motivational profiles (Schwartz and Arena, 2013). By tracking what students click on and choose to do in digital assessments, researchers have discovered interesting patterns that link small student choices to their success in their academic courses. Specific examples are explored below.

Cutumisu, Blair, Chin, and Schwartz, 2015 developed a game-based assessment to measure students' self-regulated learning choices. Specifically, the game asked students to design a poster. After the poster was complete, students could choose either positive or negative feedback. After they received the feedback, they could choose to go back and revise, or to submit

the prior poster. They found that the frequency of choosing negative feedback and to revise before submitting again correlated with learning graphic design principles. In addition, seeking negative feedback correlated with higher performance on standardized achievement tests of reading and mathematics.

This supports the results found in another game-based assessment, where students' choice to engage in critical thinking to learn about color mixing in light predicted 35% of the variance in their mathematics grades (Chase, Chin, Opezzo, Schwartz, 2009). Chase, Chin, Opezzo, and Schwartz measured learner choices in a game where students had to create and iterate on mental models for concepts in ecology. The game environment also allowed students to chat online with each other and to play a game which tested their model but did not help it improve. The students were split into two groups: one group was told that the mental model was meant to help them learn ecology, while the other group was told that the mental model was an agent they had to teach. Students in the second group spent more time studying and evaluating their model, whereas students in the first group spent more time chatting and in the game. The choice the student made determined whether they learned the concepts of the model.

Baker, Gowda, and Corbett, 2010 found that how students choose to ask for help within an intelligent tutoring system predicted subsequent learning. These new insights into how student behaviors predict achievement could not necessarily have been proven without digital choice based assessment. They also confirm what educators have long intuited, that the way students choose to learn and revise is important for academic success.

There is evidence that allowing students to choose enhances intrinsic motivation. Iyengar and Lepper (1999) find that American students show more motivation and learning when given

opportunity for personal choice. This finding is further corroborated by Lomas et al (2017) who find that when participants in a game with self-selected difficulty level, moderate difficulty levels were most motivating. When difficulty was assigned randomly, with students unable to select their level, less difficulty was more motivating for students. Students were willing to take on higher difficulty when they had chosen the difficulty themselves.

The effects of choice have also been studied in physical education. When ten-year old children were taught ballet positions, the group that was able to choose video demonstrations showed greater improvement and enhanced learning in comparison to the control group, which watched the videos the first group chose (Lemos, Wulf, Lewthwaite, & Chiviacowsky, 2017). When seventh- and eighth-grade girls participated in two fitness units, one with choice, and one with no choice, motivation was higher in the choice unit. In addition, students who participated in choice first and then no choice had lowest levels of motivation (Ward, Wilkinson, Graser, & Prusak, 2008). Though physical education class is a very different educational setting than math class, these findings suggest that choice can be a powerful tool for increasing motivation.

Beyond what can be predicted with choices children make without intentional influence, Schwartz and Arena, 2013 discuss the ethical issue of shaping choices of students. They describe a paradox: to help students learn to choose for themselves, it's necessary to shape their choices. By shaping their choices, we're actually removing choices. Despite having identified the paradox, the AAALab has not published any work on attempts to shape student choices.

Thus, there is a gap in the field: few people are studying attempts to shape student choices in the context of digital choice-based assessment. In fact, Schwartz and Arena, 2013 explicitly call out the gap: "What is missing, besides new types of assessments, is research on

how to effectively guide learning choices.” To my knowledge, the only study to look at choice as an outcome rather than an input is Jeong et al., 2008. They used Hidden Markov Models, a machine learning algorithm, to characterize student behaviors in a learning environment. Students who were given tips (hints or other nudges to encourage beneficial student behavior) showed transfer, the ability to make harder choices on their own; i.e., when students subsequently made choices in the environment without the tips, they still chose the same things they had in the environment with the tips, indicating that they had learned to make the beneficial choices. Furthermore, students who received tips on making learning choices learned more in both the initial learning period and the next session six weeks later. This study shows that choice shaping environments, with prompts to encourage self-reflection, can be effective.

Putting it All Together

In traditional educational psychology, difficulty and challenge are similar concepts. Lomas et al., 2017 define difficulty as “the probability of task failure,” while abstaining from defining challenge, because it is a “more nuanced and complex concept.” Digital technologies allow us to define and measure a new rating: how likely a student is to avoid it, or ‘avoidance likelihood.’ This is a more useful concept for self-driven personalized learning, because it identifies the areas that a student would not learn or practice. I believe there is no existing research exploring how different mindsets affect which problems would have what avoidance likelihood for a specific student.

Many students avoid problems they know will be difficult for them. These are precisely the areas where students should be spending their time in order to maximize learning gains. How

can we encourage students to spend their time on problems with high avoidance likelihood while giving them agency?

There is existing work on educational games that aim to give students agency. Chen, n.d. built an online digital assessment called ActiveQuiz, which demonstrates how principles of flow used to create engaging video games can be applied to a traditional quiz experience. In ActiveQuiz, once students get a question right on a level, they are able to select between two questions to answer. This student choice between which problem to solve next has potential to illuminate whether students tend to pick questions that are easier for them. Matuschak, n.d. builds upon ActiveQuiz, articulating that offering choice is a useful concept in digital education software: algorithms can guess what a student should do next, but they're not perfect, and it is useful to let students choose.

There is potential for a growth mindset intervention that maintains student agency. I propose developing an intervention that directly addresses the fixed mindset behavior of avoiding personally challenging tasks by modifying the incentive structure of the activity, without explicitly explaining the purpose of the intervention. My capstone aims to increase the the behavior of choosing to work on areas of improvement, rather than content/skills one has already mastered.

Game Development

To develop the game, I first needed to understand how students choose challenges in the very limited context of basic math facts.²

To measure student choices, I had to create a learning activity where students had a choice. The basic idea I started with was to always have two different questions and have students choose only one to complete. I worked with students and teachers to refine that idea into the final game.

Exploration of how students choose math problems to work on when given a choice

Paper Prototype

I had three questions for this stage in my exploration:

1. How do students decide what problems to avoid or confront in a choice based environment?
2. What math questions would be best to ask to create that environment?
3. Is it possible to include a measurement of attitudes towards math in my final design?

Process

To begin, I created two different levels of math worksheets: 1 to 12 addition and 1 to 12 multiplication. Each worksheet was formatted with twenty questions split into two columns.

² From the literature review, we also know that factors traditionally said to influence desire for a challenge, such as intrinsic motivation, self-perceived competence, and personal interest in the subject likely have an influence on the choices students would make in my constructed experiment. Those factors though clearly complex and influential, were unfortunately outside the scope of my study.

Questions were placed randomly within the columns and rows. The instructions on the worksheet said “Complete the worksheet. For each row, pick **one** question to do.”

I used these worksheets to do user testing with 6-10 year olds at the Peabody Museum. Participants approached the booth if they were interested in the study. Their parents or guardians signed a waiver. After completing the activity, participants chose a small prize to take.

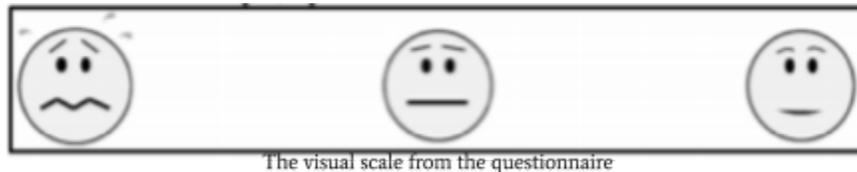
Complete the worksheet. For each row, pick **one** question to do:

$3 \times 4 =$	$7 \times 9 =$
$5 \times 9 =$	$9 \times 10 =$
$7 \times 8 =$	$11 \times 12 =$
$6 \times 11 =$	$2 \times 7 =$
$12 \times 0 =$	$4 \times 0 =$
$4 \times 7 =$	$3 \times 9 =$
$6 \times 8 =$	$7 \times 2 =$
$8 \times 2 =$	$9 \times 2 =$
$6 \times 6 =$	$1 \times 7 =$
$6 \times 6 =$	$8 \times 12 =$

I began by asking the participant if they had learned their times tables in school. Participants who said yes were given the multiplication worksheets, and others were given the addition worksheet. I then asked children to complete the worksheet by choosing one question from each row. Participants were not given feedback on whether their answer was correct, and if a participant struggled on both problems in a row, I told them it was alright to skip that row.

After they completed the worksheet, I gave them a shortened version of the Child Math Anxiety Questionnaire, which asks students to rate how they feel when (1) taking a big test in their math class, (2) you’re getting your math book and seeing all the numbers in it, and (3) you get called on by the teacher to explain a math problem on the board. The Questionnaire uses a

visual scale with a distressed face, neutral face, and smiley face. I used the exact images from the official document. I read the questions to them and asked them to point to where along the scale they felt they were. I showed these worksheets to 10 children. Afterwards, I asked them why they chose the questions they did.



The visual scale from the questionnaire

Takeaways

I found that when the material was easy for students, they went straight down the left column, without looking at questions in the right column. If there were some questions they didn't know, or they were less comfortable with the material in general, they looked at both options and selected either the easier one, or if they only knew one, that one. From this, I realized that the order in which the questions are presented matters. In the digital game, I made left-right ordering of the questions random.

The sliding scale for math anxiety did not work at all. Students were very confused by the faces, interpreting the middle face as bored rather than neutral. They also didn't seem to understand the questions. Thus, I removed this portion from the final study.

Digital Prototype

I knew that the final game would need to be done through a computer game rather than via paper worksheet. I needed the following features to implement my learning environment at a small scale (a single class of students):

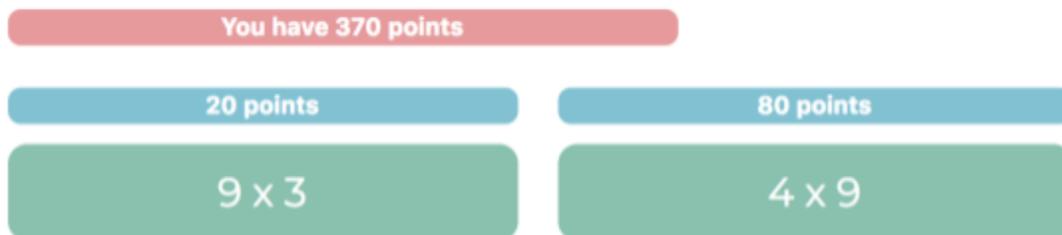
- Instant feedback to the students about whether their answer was correct
- Ability to randomize which problem a child sees first (unlike the static worksheets, which had a left and right row).
- Most importantly, instant and scale-able adaptation of points, customized to which questions each student is avoiding.

Process

For the second step in developing the game, I created a digital version of the worksheets. My goal was to understand:

1. How students interact differently with the computer vs. the paper
2. How to frame the activity as a learning experience rather than an evaluative one

The game was coded in React and Django. It had the same general concept as the worksheet, and two versions, one for addition and one for multiplication.



a screenshot from the game

For each round, participants could choose between two different questions each round to solve. Each question was worth a random number of points, and participants could see how many points they had at each time. They also got immediate feedback when they submitted the

question, and the game didn't let them move onto the next question until they had gotten the current one correct. I chose this design for simplicity, since I did not have a good sense of how I should assign the points, nor how difficult each question was likely to be. Thus, I was able to observe students choosing between different combinations of difficulty and point values. There were also instances where both questions were the same, with different point values, or each question was different, but had the same point value.

I again brought the game to the Peabody, where I tested it on 11 children aged 7 to 11 (and one five year old who just really wanted to play, so I walked her through the addition version). Again, participants were volunteers who walked up to the booth. Thus, there was a selection bias, as students who would go out of their way to participate in a math game probably enjoyed math.

I had each participant play the game on my computer for around 7 minutes (they were of different ages, so time per problem varied). I introduced the game to them by saying "This is a game to help you practice your [multiplication|addition]. Each round, you can pick one question to solve." I did not mention anything about trying to maximize points. There were no hints available in this version of the game. Only one participant had trouble with the questions, and I helped her count on her fingers. Afterwards, I asked them why they chose the questions they did.

Takeaways

I saw similar trends in choosing behavior in the digital version as I had in the paper version. Again, the biggest factor in how students chose seemed to be their comfort with the material. Students who knew their times tables well chose quickly, and mostly chose the question with the higher point value. When asked why they chose the question they did, they said they

picked the one with higher points. Students who were less comfortable with the material were more hesitant in their choosing. They seemed to weigh their likelihood of being able to solve the question more than the point values in the game, and when asked how they picked questions, said something along the lines of “If I knew the answer, I picked the one with the higher points. If I didn’t, I picked the other one.”

One main difference between the digital game and the paper mocks was that in the digital version, students had choose a problem to solve before they solved it, since they clicked a problem before they were taken to the submit screen, and then could not go back.

I realized that if a student didn’t know the answer to a question, no amount of point incentive would motivate them to choose that question, since they likely would not get it correct.

Thus, I added a hint option to the final game. I intended this to help frame the activity differently, making the game a formative assessment (an assessment for the purposes of fostering improvement), rather than a summative assessment (an assessment meant to determine their final understanding of the material). If students think the game is formative assessment or a chance to practice their skills, it makes sense for them to choose to work on areas they need improvement, rather than trying to prove what they know.

I also realized I needed to do the second part of my experiment somewhere other than the Peabody, because children at a museum with their parents on a Saturday are likely not very willing to practice their math skills. Therefore, I decided to conduct my formal experiment on students in a classroom.

When having students play the game individually at the Peabody, it was easy to tell how familiar the student was with the subject material. I knew that students, even in the same

classroom, would definitely be at different levels of comfort with whatever math problem I chose to test. Thus, I considered making the game adapt to their level of comfort by adjusting the difficulty of the questions asked overall. This might have looked like some mechanism that detected if a student got 10 single digit multiplication questions right in 2 minutes and changed all the questions to be double digits. I ultimately decided against this sort of adaptation because it would make the implementation of the game and the data analysis much more complicated, and my sample size of students would not be large enough to come to any conclusions. This issue is further discussed in the Limitations section.

Finally, students were generally much more motivated by the points than I had expected. The text showing the point values of each question, and the total points they had earned thus far were both visually much smaller than the question. When explaining the game, I did not put any emphasis on the points. Yet, all of them mentioned the points in their explanation for how they picked the problem. One even said that the questions with more points were harder, even though points were randomly assigned. This confirmed my intuition that points alone, even if not linked to an external reward like candy or toys, can be a powerful motivator.³ Thus, I used points as the incentive mechanism to shape student choices.

There were also some usability issues with the game itself that I had to fix. This project was my first time building a computer game from scratch, and I recognized after this round of testing that I needed to make the buttons bigger for kids to click on them, and that kids in general were slower typers than I had expected.

³ I had felt this myself when I realized I was completing activities on Khan Academy just to earn badges and streaks.

Teacher Observations

In addition to testing prototyping with students, I sought to learn from teachers. I observed Susan Clark's third grade classroom at Elm City Montessori School during math time for an hour and a half to help me understand how elementary math is taught in an environment where students can choose what to work on. I also spoke with Robin Reymond, a third grade teacher at St. Thomas Day School, showing her mocks of the game and explaining how the algorithm would work. Robin's students played the game for the final experiment.

Observing at Elm City

I watched Susan Clark teach a group of four students about converting fractions into mixed numbers. A student next to the group worked on a times table board, placing tiles into a grid to do multiplication facts up to 9×9 . My main takeaways were that students really enjoyed using the physical pie pieces to count out wholes of fractions, even if it took a long time. I talked with Susan afterwards, and she confirmed that they would always prefer using the physical pieces. Ultimately, I chose to build hints into the digital game rather than having students solve the problems with physical pieces, because putting hints directly in the game allowed me to log how long students spend completing the hint.

Discussion at St. Thomas

In February, I reached out to Robin Reymond of St. Thomas, hoping that I might be able to test the game in her classroom. My goal in this conversation was to evaluate if my game design was viable. Robin confirmed several intuitions I was hesitant about. First, I asked whether

some population of third graders would shy away from harder questions. She confirmed that much of her class were perfectionists, and would avoid things they might get wrong. Second, I was unsure if the point system would be enough to motivate students if points didn't correlate to a real world primary reward. Robin said that it would, and that students who would be most likely to shy away from tackling problems would also be the ones most motivated by points. She also suggested that levels might be even more motivating, because they otherwise wouldn't understand what was a "good score."

Finally, she gave me important logistical knowledge. I learned that at St. Thomas, students begin learning their times tables at the end of second grade, and that around 85% of her students are scoring 90% on the times tables, but that they are less sure on division. They find the 6, 7, and 8 times tables the hardest. They also need work on carry over adding and subtracting. She explained the various strategies her students are taught to solve the problems.

The final game deviated from Robin's advice because I chose to simplify it. She suggested multiple strategies her students used, but I implemented only one so that the hint type did not become an independent variable in my analysis. Likewise, though the prototypes were done with both multiplication and addition problems, the final game tested only multiplication problems, so that I could compare results directly across my sample of students.

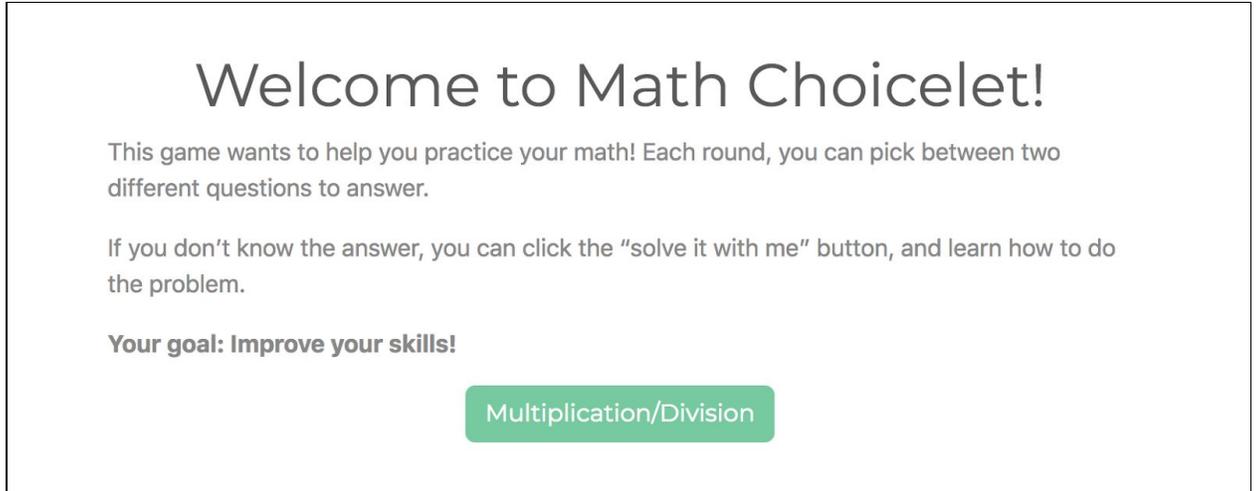
Final Game Design

Efforts to shape student choices, while maintaining a sense of agency.

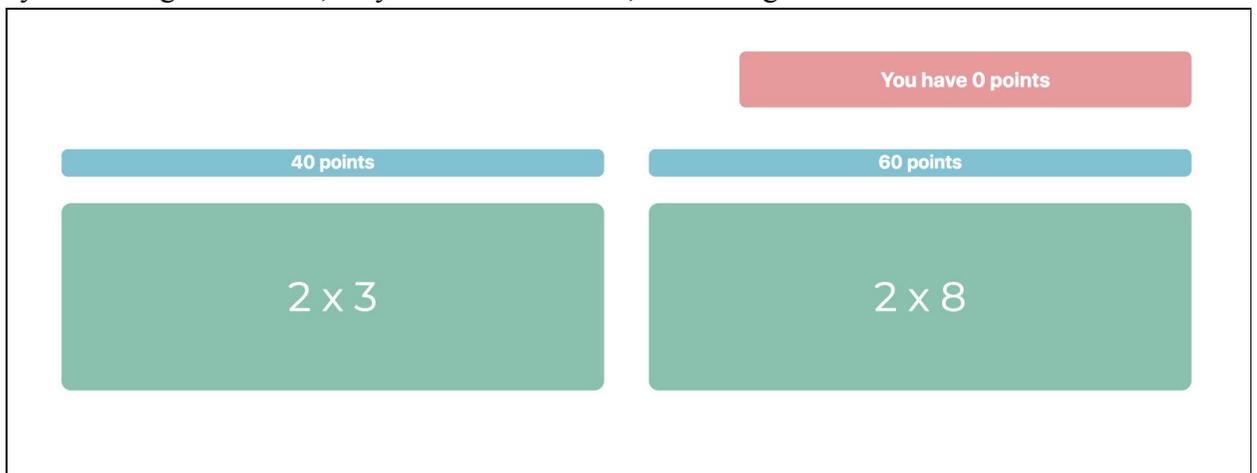
The second portion of my capstone is a quantitative study measuring the effect of an effort to shape student choices in a math practice game.

How the game works from the student point of view

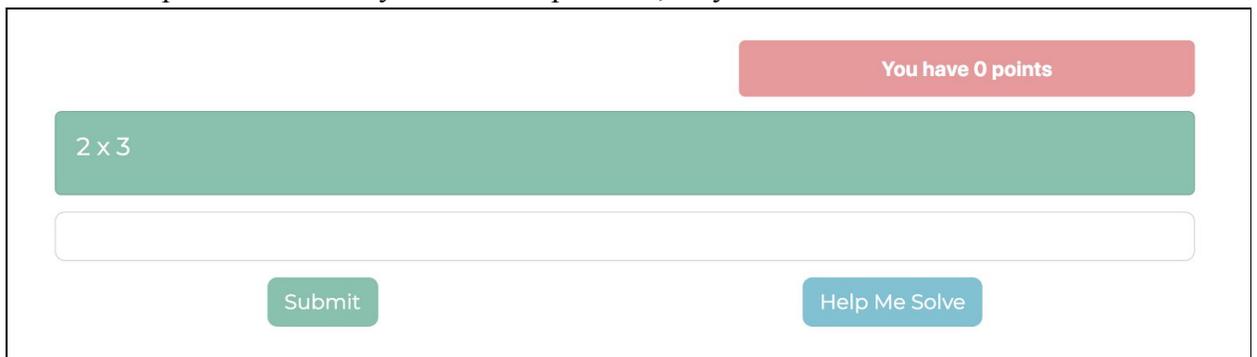
Once a user enters their access code, they see this screen.



After they click the green button, they see the first round, which might look like this:



Users will be able to choose between two questions to answer every time, each of which is worth a certain number of points. When they click into a problem, they see this:



Users can either type in their answer and click “Submit”, or click the “Help Me Solve” button.

If they get the answer incorrect, they’ll see this message:

The screenshot shows a math problem interface. At the top right, a red box says "You have 0 points". Below this, a green box contains the problem "2 x 3". Underneath is a white input field containing the number "5". Below the input field is an orange error message box that says "Not quite! Try again." with a close button (X). At the bottom, there are two buttons: a green "Submit" button and a blue "Help Me Solve" button.

At this point, the user can either continue inputting answers, or click for a hint.

The hint mechanism breaks down the multiplication fact into all the times tables leading up to it. In the explanation below, refer to the first number in the problem as i , the second as j , and the current row number as k . In the example below, $i=2, j=3$.

The screenshot shows a math problem interface with a hint mechanism. At the top right, a red box says "You have 0 points". Below this, a green box contains the problem "2 x 3". Underneath is a table with three columns. The first column contains multiplication facts, the second column contains the previous row's answer plus the first number, and the third column contains the user's input. The table has three rows: the first row shows "2 x 1", "0 + 2", and "2"; the second row shows "2 x 2", "2 + 2", and "4"; the third row shows "2 x 3", "4 + 2", and an empty input field. Below the input field is a green "Submit" button.

2 x 1	0 + 2	2
2 x 2	2 + 2	4
2 x 3	4 + 2	<input type="text"/>

There are three columns: The first is the multiplication fact for that row, $i \times k$. The second column is the previous row's answer, $(i \times (k - 1)) + i$. The final column is where the user inputs the answer for that row. The user cannot move onto the next row until they get the answer for the previous row correct. The game continues generating rows until $k = j$.

Once the user completes that row, which was the original question, they see a button that moves them onto the next question

The screenshot shows a quiz interface. At the top right, a red box displays "You have 0 points". Below this, a green box contains the question "2 x 3". Underneath the question is a table with two rows and three columns. The first row contains "2 x 1", "0 + 2", and "2". The second row contains "2 x 2", "2 + 2", and "4". At the bottom of the interface, a green box displays the feedback message "Correct! Go To Next Question".

2 x 1	0 + 2	2
2 x 2	2 + 2	4

After they successfully complete a question, they get that number of points added to their total, and they see another round of two questions.

The screenshot shows a quiz interface with two questions. At the top right, a red box displays "You have 40 points". Below this, there are two columns of questions. The left column has a blue box labeled "50 points" above a green box containing the question "4 x 2". The right column has a blue box labeled "60 points" above a green box containing the question "7 x 3".

After a set amount of time, the user is shown the below screen, indicating the end of the session.

The screenshot shows a final screen with a large, bold, black text message: "Congratulations! You're done!". Below this message, there is a green text link that says "Logout".

Points Algorithm

On the backend, the game has two modes, experimental and control, to determine what questions are presented and how many points they're worth. The mechanism for determining what mode a user is in at what time is described in the Methods section.

In both modes, the system keeps track of what question a student avoids. For each user, we keep a To-Push list, a list containing pairs of 'hard'⁴ questions that the user has been presented with and not selected, and their next point value increase. Every time a user avoids a hard question, the next point increase is incremented by 20.

In the control mode, both questions are selected randomly. Each question is randomly assigned a point value from 40 to 60. Difficulty of the question and user choice has no influence on the point value, and the To-Push list is not used.

In the experimental mode, if there are questions in the To-Push list, one of the questions in the round will be the question with the highest point value increase on the To-Push list. The other question will be a random question with a random point value. We set the pushed questions point value by taking the random point value of the other question and increasing it by the Next Point Value increase number from the To-Push list. For instance, on the first time a question is avoided, if the other problem has a randomly set point value of 50, it would set the avoided problem to 70. If the student avoided again, and next one has randomly set value of 40, it would set the avoided problem to 80. An avoided question with an added point value is referred to as a **pushed question** further in this paper.

⁴ Multiplication facts that have 6, 7, 8, or 12, but not 0, 1, 2, and 10 are labeled hard. 9 was not included because the students learned a finger counting trick for the 9s fact.

Experimental Hypothesis

Children in the experimental mode will pick more challenging problems than children in the control mode.

This might be shown in different indicators: clicking the ‘Help Me Solve’ button more frequently, picking the pushed problem, or picking more difficult problems in general.

This would suggest that assigning more points to the choice causes students to pick that choice, and more broadly, that it is possible to encourage students to work on problems they find challenging, rather than practicing what they already know. (There is a small possibility that students click the ‘Help Me Learn’ button even though they know they could calculate the answer on their own, because they consider it less work. I believe that it is not likely, because going through the problem step by step would be slower and require more work than just answering the question if the student knows the answer.)

Method

The experimental hypothesis was tested on 18 third grade students at St. Thomas Day School, a private school in New Haven, CT, during class time on a Friday afternoon. The game was introduced as “a computer game to help you practice your times tables.” Due to limited computers, 9 students played the game for the first 15 minutes, while the other 9 students did a three digit addition and subtraction worksheet. Then, the groups switched.

Students were given the above sheet of instructions, along with a random access code. 10 of the codes were to accounts that started in the experimental version of the game, and then switched to the control version 8 minutes into the game. 8 of the codes were for accounts that started in the control version, and switched to the experimental version 8 minutes in. Participants played for a total of 15 minutes.

After the student saw the “Done” screen, they raised their hands. I then showed them a piece of paper with questions on it, asked them to read it silently, and then answer the questions on the back of their sheet. These questions were not meant for formal analysis, but rather to give me a sense of how students thought they were making choices in the game, and whether they found it enjoyable. The first group saw the question “How did you choose which question to pick?”. The second group saw the questions “How did you choose which question to pick?” and “On a scale from 1 to 10, how enjoyable was the game?”

After both groups completed the activity, I held a brief discussion with the entire class, asking them what they thought the game was aiming to test. This was again for informal qualitative data.

Rather than having each participant play entirely under either the control or experimental modes, each participant played in both modes. For my main experimental hypothesis, I used data only from the first task each participant did. Thus, potential order effects were avoided, because behavior on the first task cannot be retroactively influenced by the second task.

The second task was added to gather data to potentially measure whether the behavior of choosing a challenging task transfers, i.e., participants who have the experimental version first

and then the control version choose more challenging problems during the control version than participants who had the control version first.

Findings

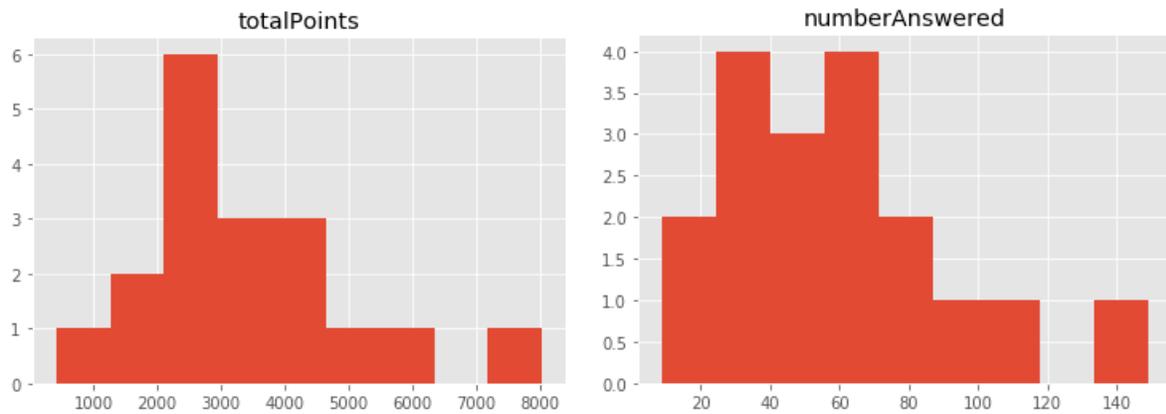
Qualitative

Overall, the students enjoyed the game much more than I had expected. Out of the 9 students who were asked to rate the game from 1 to 10 on enjoyment, 7 of them gave it a 10/10, and the average rating was a 9.72. Comments included “It’s a really good game. If I could, I’ll do it again” and “Loved it!!!” Some children even asked to keep the sheet of paper because they wanted the link to play the game at home.

Some students were confused by the hint mechanism. They would click “Help Me Solve” and then not read the new broken down question, likely because it was in smaller text. Then, they would try to input an answer to the original question, and be confused when the game told them it was incorrect. In further iterations, the size of the main question should decrease and the size of the current question should increase in the hint mechanism.

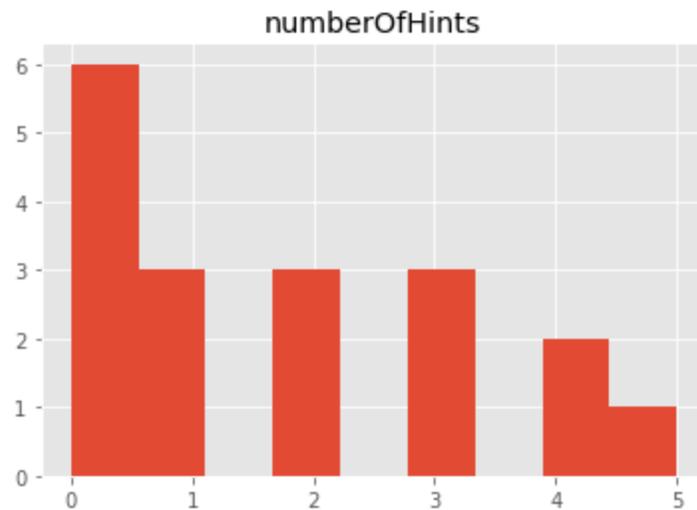
One student learned to “game the system” by refreshing the page. The game was designed to give a new set of questions every time it was loaded, so if he didn’t like either of the questions, he would just refresh the page until he got questions he did like. In further iterations, the software should be altered to prevent this from happening.

Quantitative



Histograms of the number points earned and number of questions answered by the 18 students.

In total, the 18 students answered 1065 questions in 15 minutes, for an average of 59 questions per student, and around 4 questions per minute.



Histogram of number of hints used by each student

31 hints were used in total, with 6 students never using a hint, and 3 students using just 1 throughout the session. Thus, instead of analyzing whether students used hints to test my

experimental hypothesis, I decided to examine whether students chose difficult problems. In the below analysis, a **frame** is a set of two problems that students were choosing between. A **pushed question**, a question *pushed* by the algorithm, is a difficult question that a student previously avoided, and now has a higher score value in this frame.⁵

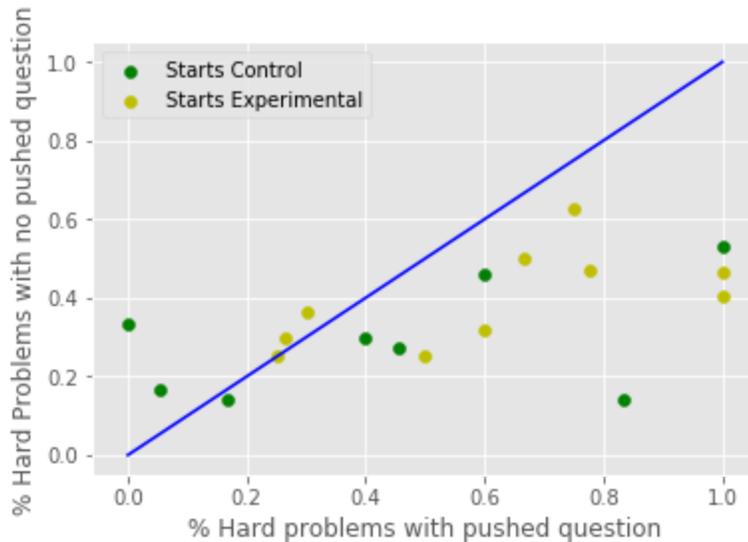
Out of the 1065 frames in the dataset, there were three possible conditions. (1) neither question was difficult, (2) only one question was difficult, and (3) both questions were difficult. To test my hypothesis, I narrowed the set down to the 535 frames where only one question was difficult, because those were the conditions where students were choosing explicitly between easy and difficult. Out of those 535 frames, 266 frames were in the control mode, with no pushed questions, and 269 were in the experimental mode, with the possibility of a pushed question.

In the control mode, students choose difficult questions 99 times out of 266, or 37.2% of the time. In the experimental mode, students chose difficult questions 120 times out of 269, or 44.6% of the time. A one tailed T-test yields a p-value of 0.04, indicating statistical significance.

Further analysis grouped by each student indicates promise as well. For the analysis below, rather than splitting between control and experimental modes, I split between frames with a pushed question and frames without pushed questions.⁶ The below chart shows the 18 students plotted by the percentage of the time they chose hard problems in a frame with a pushed question and the percentage of the time they chose hard problems in a frame without a pushed question. Points below the blue line indicate that they chose harder problems more often with pushed questions than without. 13 points are below the line, and 5 are above or on the line.

⁵ Every time a pushed question is avoided again, its score goes up.

⁶ The only difference between this divide and the control/experimental mode divide is that if there is no pushed question to give on a frame in the experimental mode, it behaves like a frame in the control mode. In other words, the category *frames without pushed questions* contains every frame from the control mode, and some frames from the experimental mode.



Thus, we see that the points incentive was strongly effective for a portion of the 18 students, and not effective for a smaller portion. Framed another way, this graph shows that only 1 of the 18 students chose hard questions 60% or more of the time without the pushed question point incentive. With pushed questions, 9 of 18 students chose hard questions 60% or more of the time. There is no clear difference in effect between the students who started in the experimental mode and the students who started in the control mode.

Statistical analysis shows that pushing a question has a significant effect on participants likelihood to choose it. A generalized linear mixed model was fit by maximum likelihood predicting whether students chose the left question, q1, or right question, q2, using (1) the difference in difficulty between q1 and q2, (2) whether the frame had a pushed question, and (3) the interaction between factors 1 and 2.⁷ Each student was assumed to have a fixed baseline likelihood of picking a difficult question, which was accounted for using a random effects term. The model finds that the first two factors are significant, with a p value of 0.001 and 0.003

⁷ The model used the full set of 1065 frames, because it used the difference in difficulty between two questions and so could take into account the frames with two easy questions or two hard questions.

respectively. On average, a pushed question makes a student more likely to choose the left question (which is always the pushed question). Without a pushed question, students are more likely to choose the easier question. The interaction term is not significant.

Ultimately, the data set in the pilot study was too small to draw any firm conclusions, as I discuss in the Limitations section below. However, the analysis by frame indicates trends towards the experimental mode, and specifically frames with pushed questions causing students to choose more difficult questions, and thus the points algorithm should be studied further.

Limitations

The largest limitation is the narrow scope to which this experiment was bound, in order to complete building the game, testing it, and analyzing the data within one semester. This has several implications for the broader applicability for the results.

The main limitation that could not be addressed in a short term study is that much of the research on children choosing challenges finds that personal relationships with teachers and parents are very influential. This was too complex to measure or imitate in a game.

Furthermore, I worked with a very limited age group and subject area: eight to nine year old students and elementary math. Times tables were chosen because they have one right answer, so I could avoid complicated natural language processing solutions for deciding whether a student is correct. They're also basic facts which improve with practice. I chose eight to nine year old students because they are studying basic multiplication facts in school. Though there are multiple strategies children are taught to learn their multiplication facts, I limited the hint style to a single strategy. Thus, some students may not have found the hint strategy helpful. This was a necessary limitation to reduce independent variables in the experiment.

The experiment was limited to one session, and I did not connect their behavior in my study to their performance in school. I also did not measure longer term changes in student behavior after the study. St. Thomas Day School is an independent school that charges tuition, and has small class sizes. Further testing with different groups of students would be necessary to extrapolate results to a larger population. Finally, because the main portion of my project ultimately evolved into documenting the design process, the actual data shows the viability of the design more than it draws strong conclusions about the effects of the intervention.

Conclusion

In this capstone, I show a promising new incentive structure for positively impacting children's behavior in an educational game. The structure, implemented in a learning environment where students choose what to learn, keeps track of the difficult problems that a student has avoided and presents them later with an increased point incentive. The results from the preliminary study, in which the incentive system was implemented in a multiplication game for third graders, are promising. 18 students played the game for 15 minutes, with each student spending half the time in experimental mode, and half the time in control mode. The incentive structure in the experimental mode increased the likelihood that a student would choose a hard question. Students also self-reported high enjoyment of the game.

While the initial results are promising, this study has limitations that would need to be addressed to fully understand the effects and potential implementations of the incentive structure. The students played for a very limited amount of time, in a very narrow educational domain. It

was not possible to measure how the intervention impacted student learning outside of the gameplay.

Despite these limitations, the preliminary results suggest potential that should be explored further. This work could be expanded by applying the incentive structure to different subjects and grade levels, and implementing it in games already used for longer term learning in the classroom. With more data, it would be possible to address whether the incentive structure's effect on likelihood of picking hard problems transfers to behavior once the incentive structure is not in place.

Finally, the design process of creating the game demonstrates the importance of working closely with students and teachers when developing tools for learners. At some Silicon Valley education technology companies, product managers and engineers work less directly with students and teachers than one might expect (Advani, 2018). At every stage of development for this project, I sought input from teachers or watched students play the game to inform the changes I made. Without working closely with teachers and students, I could not have developed a learning environment to test the incentive effectively.

As education technology moves towards allowing students to choose their own learning pathways, educators and software developers must consider how to help students learn to learn. The incentive system and design process of this capstone provide a starting point towards reaching that goal.

Acknowledgements

I wouldn't have been brave enough to pursue this project without the advice of Professor Julian Jara-Ettinger, who encouraged me to design a capstone with quantitative results and helped me analyze my results. I am thankful to Professor Mira Debs, who supported my more unconventional capstone and helped me improve both my writing and my writing skills. Robin Reymond's wisdom shaped the development of the game immensely, and I am especially grateful that she gave me precious classroom time to have her students play it. Her students' enthusiasm and feedback gave me the energy to complete the project.

I could not have done this project without Shivam Sarodia, who provided advice on software design, but also pivotal support and encouragement at the most difficult points. Finally, I'm immensely appreciative of the fellow scholars of the 2019 cohort.

References

Academic papers and Books

- Advani, Veena. *User Research in Online Education: The Challenges of Accurately Understanding Student Needs in a Global Classroom*. Yale Education Studies. April 2018. [https://educationstudies.yale.edu/sites/default/files/files/Veena Advani Capstone.pdf](https://educationstudies.yale.edu/sites/default/files/files/Veena%20Advani%20Capstone.pdf).
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on african american college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2), 113–125.
- Baker, R. S., Gowda, S., & Corbett, A. (2010). Automatically detecting a student's preparation for future learning: help use is key. In *Educational data mining 2011*.
- Balint, T. A. (2016). *An in-depth analysis of problem-solving profiles of students in open online environments* (Doctoral dissertation).
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an

- intervention. *Child development*, 78(1), 246–263.
- Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6), 4–16.
- Boggiano, A. K., Main, D. S., & Katz, P. A. (1988). Children's preference for challenge: the role of perceived competence and control. *Journal of Personality and Social Psychology*, 54(1), 134.
- Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protege effect: increasing the effort towards learning. *Journal of Science Education and Technology*, 18(4), 334–352.
- Chen, J. (n.d.). About Activequiz. Retrieved September 30, 2018, from www.jenovachen.com/flowingames/implementations/math/mathabout.htm.
- Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences*, 113(31), 8664–8668.
- Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology*, 88(4), 715.
- Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2015). Posterlet: a game-based assessment of children's choices to seek feedback and to revise. *Journal of Learning Analytics*, 2(1), 49–71.
- Dewey, J. (1928). *Progressive education and the science of education*. Progressive Education Association.
- Dweck, C. S. (2008). *Mindset: the new psychology of success*. Random House Digital, Inc.
- Inoue, N. (2007). Why face a challenge?: the reason behind intrinsically motivated students' spontaneous choice of challenging tasks. *Learning and individual differences*, 17(3), 251–259.
- Iyengar, S. S. & Lepper, M. R. (1999). Rethinking the value of choice: a cultural perspective on intrinsic motivation. *Journal of personality and social psychology*, 76(3), 349.
- Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden markov models to characterize student behaviors in learning-by-teaching environments. In *International conference on intelligent tutoring systems* (pp. 614–625). Springer.
- Lemos, A., Wulf, G., Lewthwaite, R., & Chiviawowsky, S. (2017). Autonomy support enhances performance expectancies, positive affect, and motor learning. *Psychology of Sport and Exercise*, 31, 28–34.
- Lomas, J. D., Koedinger, K., Patel, N., Shodhan, S., Poonwala, N., & Forlizzi, J. L. (2017). Is difficulty overrated?: the effects of choice, novelty and suspense on intrinsic motivation in educational games. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 1028–1039). ACM.
- Matuschak, A. (n.d.). Student agency and flow. Retrieved September 30, 2018, from klr.tumblr.com/post/146190595653/student-%20agency-and-flow
- Mueller, C. M. & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of personality and social psychology*, 75(1), 33.
- Pane, John F., Elizabeth D. Steiner, Matthew D. Baird, Laura S. Hamilton, and Joseph D. Pane,

- How Does Personalized Learning Affect Student Achievement?. Santa Monica, CA: RAND Corporation, 2017. https://www.rand.org/pubs/research_briefs/RB9994.html.
- O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., & Popovic, Z. (2014). Brain points: a growth mindset incentive structure boosts persistence in an educational game. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 3339–3348). ACM.
- Pane, J. F. (2017). *Informing progress: insights on personalized learning implementation and effects*. RAND.
- Rainey, R. G. (1965). The effects of directed versus non-directed laboratory work on high school chemistry achievement. *Journal of Research in Science Teaching*, 3(4), 286-292.
- Rinke, C. R., Gimbel, S. J., & Haskell, S. (2013). Opportunities for inquiry science in montessori classrooms: learning from a culture of interest, communication, and explanation. *Research in Science Education*, 43(4), 1517–1533.
- Roberts, J. D., Chung, G. K., & Parks, C. B. (2016). Supporting children's progress through the pbs kids learning analytics platform. *Journal of Children and Media*, 10(2), 257–266.
- Schwartz, D. L. & Arena, D. (2013). *Measuring what matters most: choice-based assessments for the digital age*. MIT Press.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759.
- Ward, J., Wilkinson, C., Graser, S. V., & Prusak, K. A. (2008). Effects of choice on student motivation and physical activity behavior in physical education. *Journal of Teaching in Physical Education*, 27(3), 385–398.
- Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., . . . Roberts, A., et al. (2016). Using design thinking to improve psychological interventions: the case of the growth mindset during the transition to high school. *Journal of educational psychology*, 108(3), 374.
- Zydney, J. M., Diehl, L., Grincewicz, A., Jones, P., & Hasselbring, T. S. (2010). Empowering learners to choose the difficulty level of problems based on their learning needs. *Journal of Systemics, Cybernetics, and Informatics*, 8(4), 8–13.

Online Articles

- Bowles, N. (2018, October 26). A Dark Consensus About Screens and Kids Begins to Emerge in Silicon Valley. Retrieved from <https://www.nytimes.com/2018/10/26/style/phones-children-silicon-valley.html>
- By The Numbers: MOOCS in 2017 - Class Central. (2018, November 21). Retrieved from <https://www.class-central.com/report/mooc-stats-2017/>
- Hilton, A. A. (2017, November 17). Prominent Scholar Calls Growth Mindset A "Cancerous" Idea. Retrieved from https://www.huffingtonpost.com/entry/prominent-scholar-calls-growth-mindset-a-cancerous_us_5a07f046e4b0f1dc729a6bc3
- In Latin American Classrooms, a Television Triples Enrollment. (2014, October 22). Retrieved from <http://www.worldbank.org/en/news/feature/2014/10/22/en-aulas-de-latinoamerica-un-televisor-logra-triplicar-la-matricula/>
- Kamenetz, A., Feinberg, R., & Mason, K. C. (2018, November 16). The Future Of Learning? Well, It's Personal. Retrieved from

- <https://www.npr.org/2018/11/16/657895964/the-future-of-learning-well-it-s-personal>
Khan Academy 2017 Annual Report. (2018). Retrieved from
<https://khanacademyannualreport.org/>
- Manning, E. (2017, August 01). Out with the old school? The rise of ed tech in the classroom. Retrieved from
<https://www.theguardian.com/small-business-network/2017/aug/01/schools-slowly-edtech-sector-cubetto-kahoot-firefly>
- Meyer, D. (2014, June 25). Don't Personalize Learning. Retrieved from
<https://blog.mrmeyer.com/2014/dont-personalize-learning/>
- Phelps, S. (2018, June 8). Ending the Digital Divide is Key to Educational Equity. Retrieved from
<https://www.nextgenlearning.org/articles/ending-the-digital-divide-is-key-to-educational-equity>
- Richmond, E. (2014, October 24). When Students Take Over the Classroom. Retrieved from
<https://www.theatlantic.com/education/archive/2014/10/what-happens-when-students-control-their-own-education/381828/>
- School district reports test scores rise with mastery learning on Khan Academy. (2018, September 06). Retrieved March 26, 2019, from
<https://www.khanacademy.org/about/blog/post/177799922435/school-district-reports-test-scores-rise-with>
- Summit Learning: Homepage. (n.d.). Retrieved from <https://summitlearning.org/>

Appendix: Statistical Model

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

['glmerMod'] Family: binomial (logit)

Formula: chosenQ ~ dif * hasPushedQuestion + (1 + q1Dif + q2Dif | user_id)

Data: data

AIC	BIC	logLik	deviance	df.resid
1433.0	1482.7	-706.5	1413.0	1055

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.5541	-0.9072	-0.6126	1.0059	2.2820

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
user_id	(Intercept)	0.16025	0.4003	
	q1Dif	0.02962	0.1721	-0.25
	q2Dif	0.02318	0.1522	-0.21 -0.89

Number of obs: 1065, groups: user_id, 18

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.08817	0.09312	-0.947	0.34369
dif	0.15892	0.05009	3.173	0.00151 **
hasPushedQuestionTrue	-0.78993	0.27278	-2.896	0.00378 **
dif:hasPushedQuestionTrue	0.05354	0.08656	0.619	0.53624

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	dif	hsPsQT
dif		0.097	
hsPshdQstnT	-0.351	0.064	
df:hsPshdQT	0.069	-0.282	-0.708