Wuzzit Trouble Feasibility Report

FOR INCLUSION IN IES SBIR PHASE I FINAL REPORT

Submitted October 23rd, 2015

Bryan Matlen
Sara Atienza
Kimkinyona Fox Cully

WestEd.org

Introduction

Wuzzit Trouble (WT) is a game-based app that is designed to provide a learner-friendly interface to mathematical thinking. It was created by BrainQuake in part under the theoretical assumption that a primary hurdle in learning mathematics is learning to read the symbolic notation (i.e., the '*symbol barrier*'). If this language aspect of mathematics is removed, students may be capable of solving mathematical problems well beyond what is traditionally considered typical of their grade-level. To this end, the game objects within the app are designed to provide direct representations of mathematical concepts (see Figure 1) so that it will support students' mathematics learning without the need to recruit symbolic representations. In addition, WT is hypothesized to support students' attitudes towards mathematics.

As a part of the Phase I SBIR grant, BrainQuake developed an adaptive engine that adaptively selects puzzles at the appropriate level of difficulty for students. In the summer of 2015, WestEd conducted usability research for BrainQuake on this version of WT. The goal of this research was to support the development of WT and its adaptive engine. In the fall of 2015, WestEd conducted a feasibility study that consisted of a mixed-methods (both qualitative and quantitative) approach that aimed to assess the feasibility of using WT in classroom contexts. The study was specifically designed to explore the following questions:

1. Does playing WT increase student learning and attitudes towards mathematics?
2. What are teachers' impressions of WT and how do they implement WT in their classroom?
3. What modifications to WT would maximize its classroom feasibility?

In this report we describe the methods and procedure used to explore the research questions above. We present both quantitative and qualitative results from the surveys and observations conducted during the study, and conclude with suggestions for future development of WT.

Figure1. Screenshot of Wuzzit Trouble.

Method

Participants

Participants in the feasibility study consisted of ninety-eight 5th grade students and four teachers from a public elementary school in Gilroy, CA – referred to as School A – and one hundred and seven 6th grade students and two teachers from a charter middle school in Oakland, CA – referred to as School B. Demographic information on these schools is included in Table 1. Both schools had a high percentage of students who qualified for Free and Reduced Price Lunch, as well as a high percentage of Hispanic students.

Table 1. Percentage of students by demographic category for each participating school.

	Free or Reduced Lunch	English as Second Language	Hispanic	Black	Asian	White	School Type
School A	75%	98%	89%	1%	2%	4%	Public Elementary School
School B	93%	39%	78%	12%	2%	1%	Charter Middle School

Table 2. Background information on participating teachers.

Teacher	School	Grade and Subject Taught	Yrs Teaching Exp.	Highest Level Education
Mrs. Addition	School A	4th - 5th split[1], General Ed	12 - 15 years	Master's
Mrs. Subtraction	School A	5th General Ed	12 - 15 years	Bachelor's
Mr. Multiply	School A	5th General Ed	9 – 11 years	Bachelor's
Mr. Division	School A	5th General Ed	12 - 15 years	Master's
Mrs. Squared	School B	6th Engineering	6 – 8 years	Master's
Mrs. Cubed	School B	6th Science	< 1 year	Bachelor's

[1]Only 5th grade students participated in the feasibility study

Though WT is a mathematics learning app, mathematics teachers at School B were not allowed to participate in the study because they were undertaking a new curriculum that prohibited use of other activities during their mathematics lesson. However, because mathematics is a subject that pervades many subject areas, Science and Engineering teachers were allowed to participate in the study. Inclusion of these teachers were meant to provide a more diverse picture of how WT is implemented in classroom settings. All participating teachers were provided a $200 stipend for participation. Background information on teachers are presented in Table 2.

It is important to note that the accessibility of technology within each school varied greatly. Specifically, students at School B had access to iPads and the school also offered free Wi-Fi to students. School B is a science, technology, engineering, arts, and math (STEAM) school focused on design thinking and leveraging technology to solve real world problems. In contrast, students and teachers at School A did not have access to Wi-Fi and the school did not offer access to mobile devices. Thus, students and teachers at School B were likely more familiar with technology, on average.

Measures

There were four teacher surveys and three student surveys administered throughout the course of the study. WestEd researchers also conducted an observation of four classrooms. These measures are described below.

Student Measures

Post-WT Student Survey: The post-WT survey consisted of 20 questions that asked students about their opinions related to WT's usability, feasibility, engagement, and ability to support student learning and motivation for mathematics. For each question, students were asked to rate the extent to which they agreed or disagreed with item statements on a 5-option Likert scale, ranging from Strongly Disagree to Strongly Agree and containing a No Opinion response. The survey took approximately 10 minutes to complete. This survey is provided in Table 3.

Pre and Post-Math Survey: The pre- and post-math surveys were identical and consisted of six multiple-choice items that were aimed at assessing students' mathematical ability. Four problems involved solving for an unknown quantity. Of these four items, two were word problems and two were equations. The two remaining items consisted of completing a pattern based on an initial sequence of numbers (e.g., 2, 4, 6, __). Students were allowed to use scratch paper to work out the math problems. The survey took approximately 10 – 15 minutes to complete. The math survey is included in the Appendix.

Table 3. Item statements in the post-WT survey.

Subscale	Question Number	Item Statement	Item Abbreviation
Engagement	1	I enjoyed using Wuzzit Trouble.	enjoyed
	2	I had fun playing Wuzzit Trouble even after using it multiple times.	had fun
	3	I would use Wuzzit Trouble even without someone telling me to.	would use
	4	Wuzzit Trouble was engaging.	engaging
Usability	5	I thought Wuzzit Trouble was easy to use.	easy to use
	6	It would be easy for other kids to learn how to play Wuzzit Trouble.	easy to learn
	7*	Playing Wuzzit Trouble was too confusing.	confusing
	8	I think most people would learn how to play Wuzzit Trouble very quickly.	learn quickly
Learning	9	I thought the puzzles were at the right level of difficulty.	right difficulty
	10	I feel like I learned about math from using Wuzzit Trouble.	learned math
	11	I feel like I learned more from Wuzzit Trouble than I would have learned from my teacher.	WT > teacher
	12*	The Wuzzit Trouble puzzles were too difficult for me.	too difficult
Feasibility	13	I would use Wuzzit Trouble again to learn more about math.	use again
	14	I would like to use Wuzzit Trouble frequently.	use frequently
	15	Using Wuzzit Trouble in the classroom is helpful.	helpful
	16	Playing Wuzzit Trouble in the classroom was easy to do.	classroom easy
Motivation	17	After playing Wuzzit Trouble, I am more interested in learning about math.	math interest
	18	After playing Wuzzit Trouble, I find learning about math more exciting.	math excite
	19	After playing Wuzzit Trouble, I am more motivated to learn about math.	math motivate
	20	After playing Wuzzit Trouble, I am more interested in attending my math class.	take math class

* Reverse coded

Teacher Measures

Teacher Background Survey: The teacher background survey consisted of questions about teachers' teaching and educational history, as well as questions that asked them to self-report their familiarity and ability to use technology for teaching (e.g., The extent to which teachers agree or disagree with statements like "I can choose technologies that enhance students' learning for a lesson”.). The survey was administered as a paper and pencil assessment and took approximately 5 – 10 minutes to complete. Teacher completed the survey at the Teacher Orientation.

Teacher Logs: The teacher log was administered once at the end of the first week and again at the end of the second week. The log asked teachers to describe how they implemented WT in their lesson that week as well as the amount of time and frequency with which they assigned WT. The logs also consisted of open-ended questions about whether teachers believed their students learned from WT, and how engaged students’ were in playing the game, as well as any tech issues experienced. Teachers completed the survey online, and each log took approximately 5 minutes to complete.

Teacher Post-WT Survey: This survey consisted of identical questions to the student post-WT survey, but reworded for a teacher respondent (e.g., “My students enjoyed playing WT”, etc.) (Motivation questions 16 – 20 were excluded from the teacher survey). In addition, teachers were asked to provide open-ended responses to questions that asked about 1) how they would describe WT to a friend, 2) whether the puzzles were at the right level of difficulty for their students, 3) whether there were any obstacles associated with using WT in the classroom, 4) how they would use WT if they were to do the study over again. This survey was administered online and took approximately 10 minutes to complete.

Classroom Observations

WestEd researchers observed one 45-minute class period for four of the six participating teachers (i.e., a total of four classroom observations were conducted). During the observations, researchers documented the engagement of students while they played WT, how teachers implemented WT in their class (including whether/how they interacted with students during game-play), and whether concepts from WT were integrated into lesson content.

Procedure

The study took place at the beginning of the Fall 2015 semester. The study spanned two consecutive weeks. During this two-week period, teachers were asked to assign WT to students in class for at least 10 minutes, on at least 3 days for each of the two weeks.

Table 4. Overview of study tasks and approximate timeline of administration.

	First Week		*Second Week*	
	Beginning of Week	*Mid - End of Week*	*Beginning of Week*	*Mid - End of Week*
Students	Math Pre Survey	Play WT at least 10 minutes on at least 3 days	Play WT at least 10 minutes on at least 3 days	Math Post Survey WT Post Survey
Teachers	Teacher Background Survey	Teacher Log 1		Teacher Log 2 WT Post Survey

Before the study, WestEd researchers conducted an in-person study orientation with all participating teachers. During the orientation, teachers were briefed about the interface and basic features of WT, including how their students could log in and out of the game. Teachers were also briefed about the study tasks. As a part of the study, teachers were instructed to assign WT to their students in class for at least 10 minutes, 3 times a week, for each of the two weeks of the study. Teachers were told that they could implement WT in whatever way was most feasible for their class (e.g., as a warm up activity, as part of mathematics instruction, etc.).

Teachers were then provided a timeline of study tasks (see Table 4). At the conclusion of the study orientation, teachers filled out a brief background survey, which asked them about demographic and teaching background, as well as their familiarity with technology.

On the first day of the study, teachers were instructed to administer the pre-math survey to their students, which consisted of six questions aimed at assessing their baseline mathematical ability. In addition, teachers assigned each student a Wuzzit login name and password, provided by BrainQuake, so that students could log into the game and save their progress each time they played. Because School A did not offer Wi-Fi, BrainQuake developed an analogous version of WT that could save students' progress without the need for Wi-Fi. All game features in the Wi-Fi and non-Wi-Fi versions were identical.

For the remainder of the first week, teachers were expected to assign WT for at least 10 minutes during at least three class periods. At the end of the week, teachers filled out the first weekly log. During the second week of the study, teachers again were instructed to assign WT for at least 10 minutes during at least three class periods. At the end of the second week, teachers administered two surveys to students: 1) the post-math survey – which was identical to the pre-math survey – and, 2) the post-WT survey, that consisted of likert questions asking students about their opinions of WT in terms of its usability, feasibility, and ability to support their learning and interest in mathematics. Teachers also completed the second weekly log, and and the Teacher Post-WT survey online.

In addition to completing these surveys and assessments, four of the six participating teachers (two School A and two School B teachers) were observed by WestEd researchers during one of the class periods in which they implemented WT.

Results

Results are presented by first examining the different kinds of classroom implementation models used by teachers. We then explore student data, including responses to the post-WT survey, student usage data, and pre- and post-math surveys. Teacher data from the post-WT teacher survey are then examined.

Classroom Implementations

Teacher logs and classroom observations were used to determine the types of implementations that teachers used. The majority of teachers decided to implement WT for more than the minimum time and frequency requirements. For instance, five of the six teachers implemented WT everyday for the two-week period. On average, teachers assigned WT almost every day each week (M = 4.11 days out of 5) and for almost double the minimum requirement each class period (M = 19.67 minutes). School A shared one classroom set of iPads across four classrooms. Due to accessibility, some teachers were able to use the app during math class where as the other teachers used it during another class period.

We observed a variety of different ways that teachers used WT, and this information was used to create different categories of implementation models. These models varied greatly in relation to how actively the teachers assisted students while playing the game and the extent to which WT was incorporated into the lesson. We describe these models below.

Implementation Models

WT as Warm-up Activity. Mrs. Cubed decided to use WT primarily as a warm-up activity at the beginning of class. Each day of the study, Mrs. Cubed began her class by having students play WT individually for 10 minutes. As students played WT, the teacher took a 'hands-off' approach towards assisting students: The teacher primarily let the students play by themselves without help and used the time to organize her lesson and paperwork for the day. Mr. Division also used this model.

WT as Individual Work Station. Mr. Multiply assigned his students to play WT individually everyday in his class during his math lesson. The students in the one half of the room played the game, while the other half completed worksheet practice on the current math topic. This teacher actively assisted students who were *not* playing the game, giving them individualized attention on their worksheets. After about 15 minutes of game-play, the teacher asked the half of the class who was using the iPads to begin the worksheets and those previously completing the worksheets, to play on the iPads. The process of individualized attention for worksheets began again, this time with the students in the other half of the room.

WT as Supporting Students' Mathematical Strategies. Mrs. Addition assigned her students to play WT individually everyday before their mathematics lesson. This teacher took an active role in assisting her students during game play. As her students played the game, the teacher walked around the room, actively encouraging students to use mathematical strategies to solve the puzzles. Instead of telling students how to solve the puzzles, the teacher provided assistance in the form of self-explanation prompts. For example, the teacher would say, "*Can you tell me more about your strategy? What's your plan?*" or "*What operation are you using, is it addition, subtraction, multiplication?*" The teacher would also provide encouraging comments such as, "*I knew you could do it*." After about 30 minutes of game-play, the teacher transitioned to her math lesson for the day.

WT as Link to Algebraic Notation. Mrs. Squared tried to directly connect WT puzzles to algebraic notation. Under this implementation model, the teacher asked students to "*Write an equation*" for each puzzle they played. The students were expected to complete this task individually, but the teacher took an active role in supporting students in their generation of equations. For instance, below is a dialogue between the teacher and a student:

T: *So this gear has 4 teeth. And when I press on the gear, I can see where my keys are, right?*

S: *It tells you the numbers where the keys are*

T: *So now, what's the equation I need to use to get to 12?*

S: *12 divided by 4*

T: *So can you write down the equation you need to write*.

S: [Writes out the equation on the board. "12 ÷ 4 = 3"].

T: *So what do I have to do to unlock the key?*

S: *Spin it 3 times*

Across all classroom observations, we observed high levels of engagement. Students would frequently make exclamations such as "*Yes!*" when completing levels, and students overall remained on task throughout the time they played WT. Students primarily played WT individually, but we often observed discussions amongst the students about what puzzle level they were on and some students would provide help when other students were stuck.

In one elementary classroom (Mrs. Addition) 15 students participated in a focus group, conducted by the researcher at the end of the observation. In the focus group, the researcher asked students their impressions of WT (i.e., whether they liked WT, whether it helps them learn math, whether they would play it at home outside the study, and what could improve WT).

There was unanimous agreement to the questions of whether the students liked WT, and whether they thought it helped them learn math. Some notable quotes were:

> "*Two thumbs up. Really like it*"
>
> "*It's kind of hard but fun at the same time*"
>
> "*It makes your brain work and it's fun*".

About 80% of the students said that they would play WT at home. One student mentioned that he had downloaded WT onto his mom's iPad. When asked how to improve the game, students said things like more levels and adding a bonus game in between levels.

Student Data

Post-WT Survey

To further examine student reactions to WT, we analyzed student responses to the post-WT survey, which aimed to measure students' opinions about WT's usability, feasibility, engagement, and ability to support learning and motivation towards mathematics. Respondents for each item of the Post-WT survey ranged from 150 – 159 students, due to some missed items and absences during the post-test administration. The percent of responses by each agreement category are presented in Figure 2. Inspection of the responses reveal that responses were mostly in the positive direction (i.e., there was higher agreement M = 53% than disagreement M = 23% across the question types), suggesting that students overall took away positive impressions of WT.

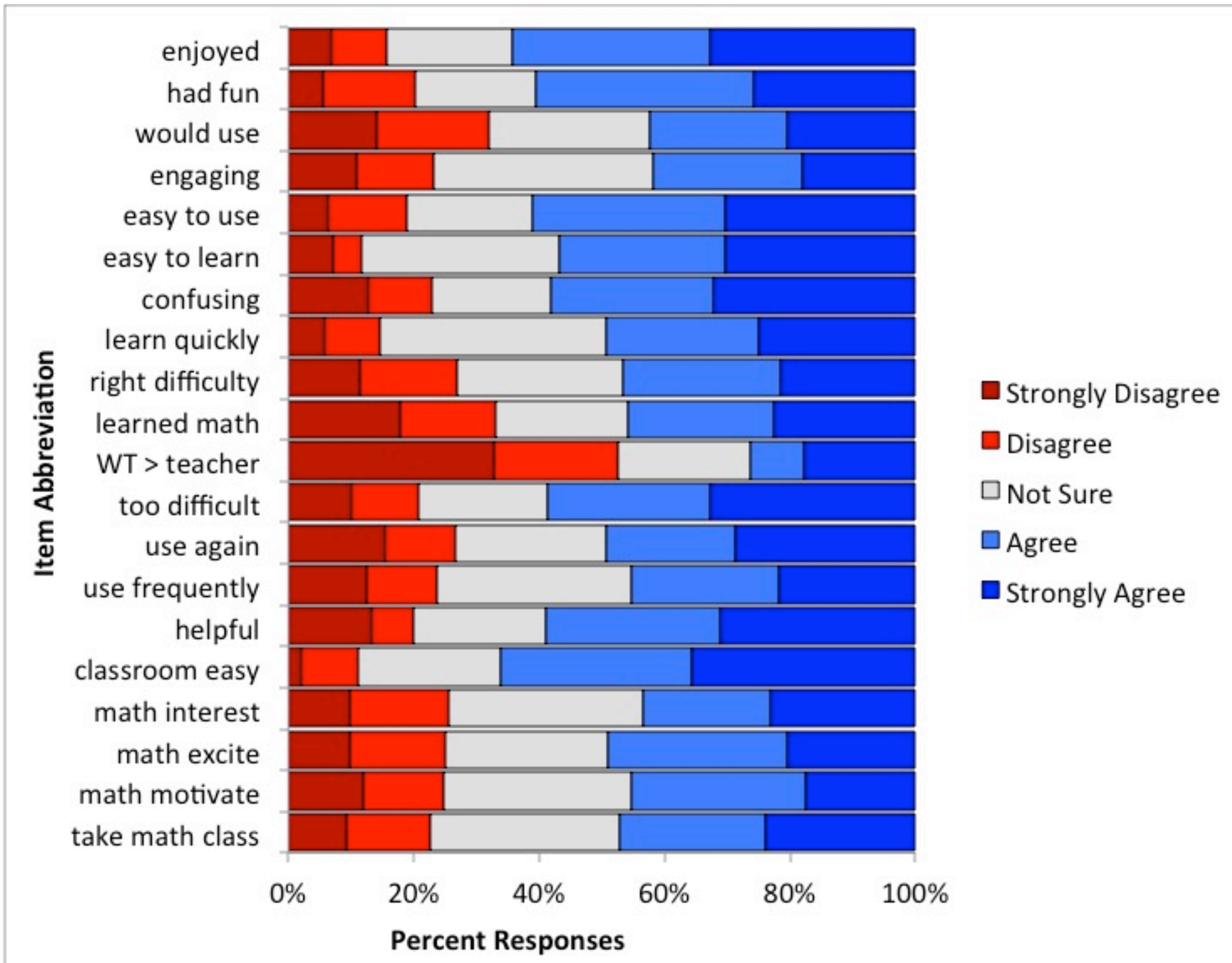
To determine whether responses were statistically different from what would be expected if response rates were equal across the five categories (i.e., neutral responding), chi-square analyses were conducted on each of the questions. This analysis revealed that all questions were statistically different from a neutral response pattern at the alpha $< .05$ level, with the exception of questions 3, 9, and 10 (Question 18 (math excite) "*After playing Wuzzit Trouble, I find learning about math more exciting*" was also was marginally significant in the positive direction at the alpha $\leq .05$).

All responses trended in the positive direction, except for Question 11 (WT > teacher): "*I feel like I learned more from Wuzzit Trouble than I would have learned from my teacher*" which trended in the negative direction. This question was not aligned with the goals of WT (i.e., it is not the goal of WT to teach students mathematics better than their teacher). In addition, this question showed poor item fit with the learning subscale in reliability analyses (described, below). Therefore, this question is not considered in subsequent analyses.

The overall positive response rates to all remaining questions indicate that students found WT to be well suited for classroom learning. For instance, responses were positive to questions like Q2 (had fun): "*I had fun playing Wuzzit Trouble even after using it multiple times*" and Q4 (engaging): "*Wuzzit Trouble was engaging*", suggesting that WT was perceived as engaging by students, even after playing the game for a two week study period. Responses were also positive to questions related to usability like Q5 (easy to use): "*I thought Wuzzit Trouble was easy to use*" and Q8 (learn quickly): "*I think most people would learn how to play Wuzzit Trouble very quickly*", suggesting that WT is user-friendly. Responses also indicated that students perceived WT as feasible for classroom learning, as indicated by positive response patterns to questions like Q13 (use again): "*I would use Wuzzit Trouble again to learn more about math*" and Q14 (use frequently): "*I would like to use Wuzzit Trouble frequently*". Finally, WT appeared to exhibit motivational benefits for students, as indicated by positive responses to questions like

Q17 (math interest): “*After playing Wuzzit Trouble, I am more interested in learning about math*” and Q19 (math motivate): “*After playing Wuzzit Trouble, I am more motivated to learn about math*”.

Figure 2. Students’ percent of agreement by question across the five response categories. Negatively worded items are reverse coded.



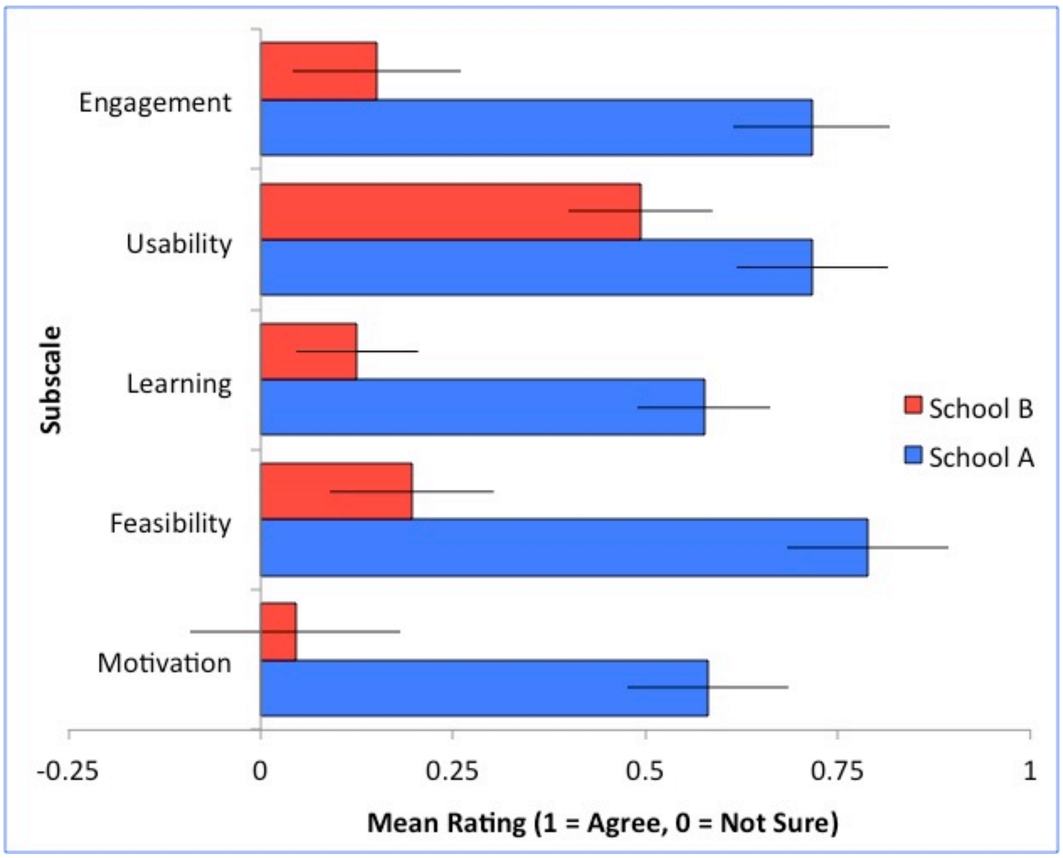
One question that was not statistically different from neutral overall was Q9 (right difficulty): “*I thought the puzzles were at the right level of difficulty*”. This finding may suggest that the adaptive engine could be improved. However, responses trended in the positive direction for this question: 47% of students agreed or strongly agreed with this statement, whereas only 27% of students disagreed or strongly disagreed with the statement. 26% of students were not sure. Students did exhibit positive responses to the Q12 (too difficult): “*The Wuzzit Trouble puzzles were too difficult for me*” (i.e., positive responses in this case refers to higher disagreement with the question), suggesting that they did not perceive the puzzles to be difficult. Taken together, these findings may imply that students perceived the puzzles as too easy; however, this conclusion is tentative since rate of agreement was overall high for question 9 (right difficulty). We explore further how difficult puzzle play was for students when analyzing the student usage data, below.

Analyses by Subscale

Responses were also analyzed by grouping questions into their respective subscales (i.e., Motivation, Feasibility, Learning, Usability, and Engagement). For this analysis, all survey responses were recoded, with responses of "strongly agree" coded as 2, "agree" coded as 1, "no opinion" coded as 0, "disagree" coded as -1, and "strongly disagree" coded as -2. Thus, a positive score is representative of on average more positive responses, and a negative score is representative of on average more negative responses. The data are presented in Figure 3.

Reliability analyses were then conducted for each of the 5 subscales in the survey. These analyses revealed moderate to good reliability across the subscales (average Cronbach's alpha ranged from .67 - .89; Motivation = .81, Feasibility = .67, Learning = .71[1], Usability = .80, Engagement = .89), suggesting that the items within each subscale were related overall as a group.

Figure 3. Average ratings by subscale and school. A rating of 1 indicates an average response of "Agree", whereas a response of 0 indicates an average response of "Not Sure". Error bars represent 95% confidence intervals.



Though responses were overall positive at both schools, the data showed a school effect: Students at School A provided overall more positive responses than students at School B. There are a number of possible reasons for the existence of a school effect in the present study (e.g., grade level of study participants, the demographic makeup of the school, experience level

[1] Question 11 was not included in this analysis as it showed poor item fit as indicated by low corrected item total correlation (.31) and higher Cronbach's alpha if the item is deleted (.71 vs. .67).

of teachers across the schools, etc.). However, one difference that may have influenced perceptions of WT is the familiarity and/or the availability of technology within the school. Students at School B are accustomed to using technology such as iPads and laptops to assist in their learning. In fact, using technology to assist classroom learning is consistent with the school's mission statement. In contrast, students (as well as teachers) at School A are less experienced with using technology in the classroom, and may have been more excited to integrate the app and iPads. This explanation remains only a hypothesis.

Despite the observed school differences, the results were overall positive for both School A and School B across the range of the survey's subscales, suggesting that students perceived WT as a usable, feasible, and engaging way to support their mathematics learning.

Student Usage Data

We next analyzed student usage data. Because the build was modified to account for the lack of Wi-Fi at School A, School A's data were stored in each individual iPad and needed to be hand coded. It was not feasible to hand code the usage data from School A in a short period. School B's data were captured remotely. Therefore, we analyze School B's usage data, below.

Of particular interest was whether the adaptive engine assigned problems at appropriate levels of difficulty. One measure of difficulty embedded in the game is the "Star count". On any given puzzle in WT, players receive Star(s) for their performance. The number of Stars awarded is based on the number of moves that students' took to complete a level, relative to the optimal number of moves that it takes to complete that level. Stars range from 1 to 3, with 3 representing the optimal number of moves performed.

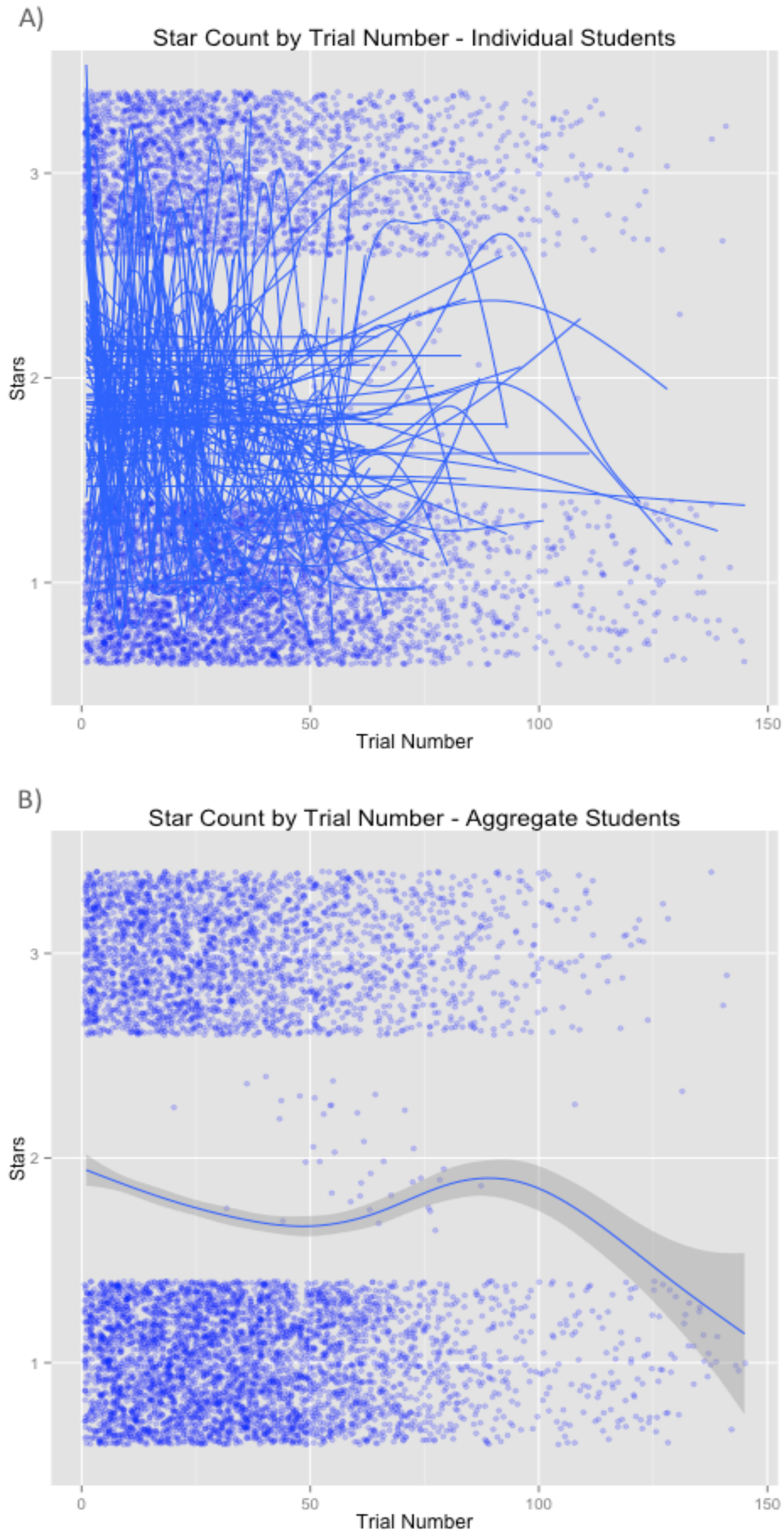
As students progress through each successive puzzle, they receive Stars that reflect their performance. If the adaptive engine assigns puzzles at the right level of difficulty, we expected to observe Star counts that wavered in between 3 and 1 Stars. In other words, Stars counts should not remain high across successive puzzles (suggesting that the engine is assigning puzzles that are too easy) nor should remain low across successive puzzles (suggesting that the engine is assigning puzzles that are too difficult).

To examine this possibility, the Star counts for each student were plotted for each successive puzzle that students played in Figure 4A and 4B. In Figure 4A and 4B, the density of Stars are represented by the shading, with darker shading indicating higher density of Stars (i.e., there is lower density at higher Trial Numbers because not all students played the same number of puzzles). These data were also fit with a natural spline model, which follows the density of points in the data and estimates curvature in the data. The only difference between Figures 4A and 4B are the individual student or aggregate spline models that are fit to the data (i.e., the points are identical).

As can be observed from Figure 4B, the spline model is centered in between 1 and 3 stars across the Trials. The model dips towards the 1 Star count for higher Trial numbers, suggesting that the engine might be too difficult for students playing many Trials. With the exception of these higher Trials, these findings are consistent with an adaptive engine that assigns puzzles at an appropriate level of difficulty. That is, it does not appear that the majority of students experienced successive periods of 1 or 3 Star counts.

One surprising observation was that there is a strikingly low number of 2 Stars counts awarded relative to 1 and 3 Stars counts awarded. This observation is not necessarily cause for concern. However, future research might explore whether the algorithm for assigning stars is overly liberal in assigning students to either high or low Star counts, as well as whether there are more intermediate puzzles that can be assigned that are of medium difficulty for students and that would therefore generate more 2 Star performances.

Figure 4. Stars are plotted by increasing Trial Number for each student at School B. Points are jittered around each of the Star levels (1, 2, or 3). The density of Stars at each Trial Number is represented by shading: Darker blue indicates higher Star density. The data are fit with a natural spline model **A)** for each student, and **B)** in aggregate.



Pre- and Post-Math Survey

To further explore whether WT exhibited learning benefits, the Pre- and Post-Math Surveys were analyzed by assigning a 0 or 1 code based on whether responses were incorrect or correct, respectively, and summing responses to create a total score (out of 6). A linear regression analysis was then conducted, using the test phase (pre- or post-survey, with pre-survey as the baseline), gender (male or female, with female as the baseline), and school (A or B, with School B as the baseline) as predictors of each students' total score. This analysis did not reveal any significant main effects for test phase (B = .03, p = .87), gender (B = .10, p = .59), or school (B = -.09, p = .62). It is important to note that, due to missing data resulting from absences and missing student IDs, sample attrition was high for these analyses (i.e., only 82 students of the total original sample of 205 were included in the regression analysis). Therefore, due to high sample attrition the analyses are likely underpowered for small effects (we would expect only small effects after a short, two week intervention).

A linear regression was also conducted using post-scores as a dependent variable and question type (Word, Equation, or Pattern Completion) as a predictor, controlling for pre-scores. Pattern Completion questions were used as the baseline. This analysis revealed that students performed better on algebraic questions (Word and Equation) relative to Pattern Completion questions (Word B = .28, p < .005; Equation B = .40, p < .001; i.e., going from Pattern Completion to Equation questions resulted in a .40 point increase in students' post-test score, and this increase was significant at the alpha < .001 level). In addition, effect sizes for pre- to post-gains were largest for Equation questions, followed by Word questions, and then Pattern Completion questions (Cohen's ds = .10, .04, and -.09, respectively). This trend is at least consistent with the hypothesis that WT helps to '*break down the symbol barrier'*.

Though we did not observe significant learning differences on the pre- and post-tests, sample attrition was high in the student pre- and post-data, therefore we do not draw conclusions from null findings in these analyses. Given the student responses about their beliefs regarding WT's potential to support their learning, as well as other research showing learning improvements from WT (Pope, Boaler, and Mangram, in prep), it is possible that WT results in measureable increases in mathematics learning after longer exposure to WT.

Teacher Data

Post-WT Survey

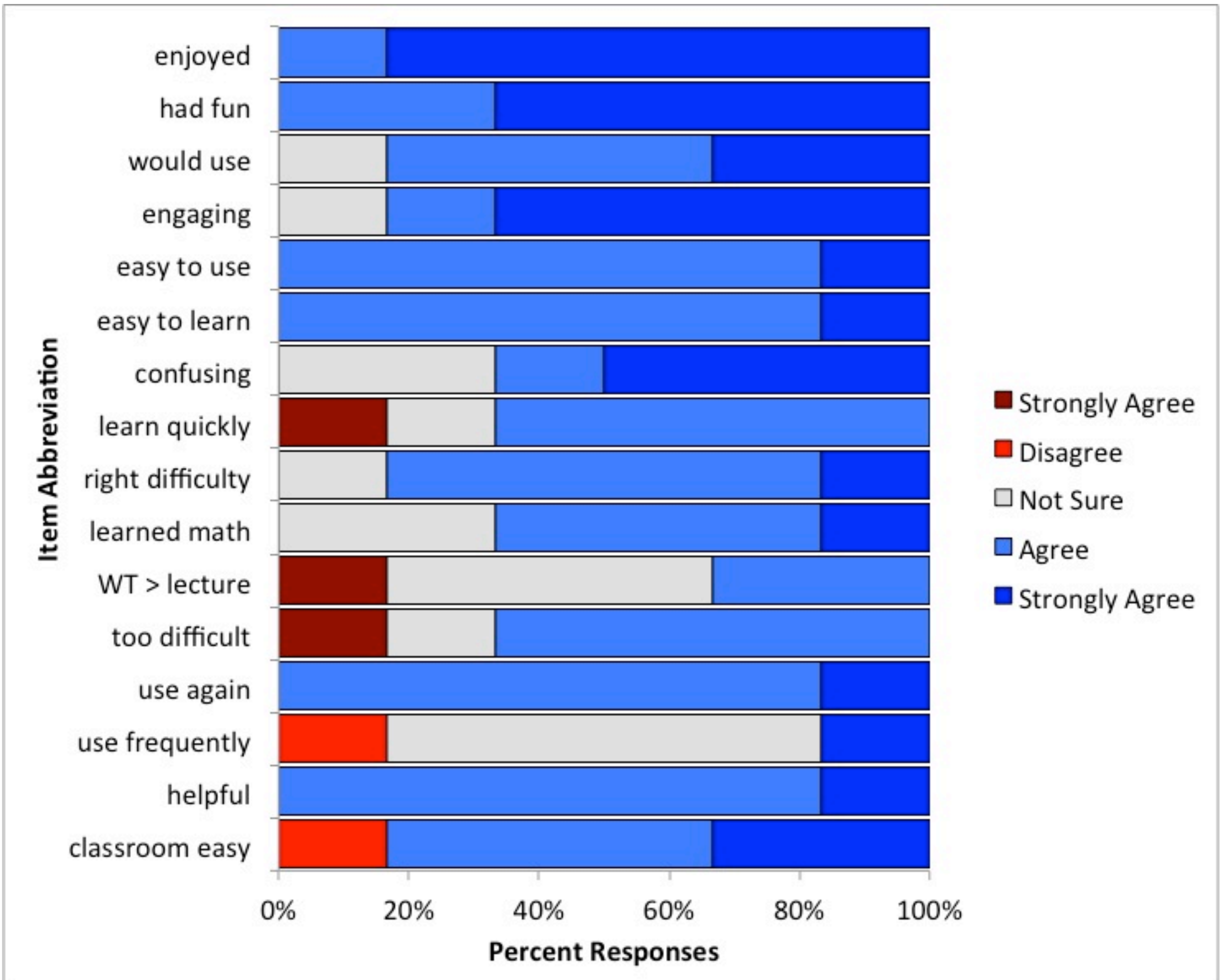
Teachers also completed an analogous survey to the student post-WT survey. Their responses are summarized in Figure 5. Like students, teachers exhibited a strong overall positive response pattern.

As is evident in Figure 4, both the School A and School B teachers believed WT was engaging (questions 1 – 4) and usable (questions 5 – 8). For example, they were positive to questions such as Q2 (had fun): "*My students had fun using Wuzzit Trouble even after using it multiple times*", and Q6 (easy to learn): "*It would be easy for other students to learn how to play Wuzzit Trouble*".

Teachers did provide mixed responses to some questions. For example, three out of the six teachers responded "Not Sure" to Q11 (WT > lecture): "*I feel like my students learned more from using Wuzzit Trouble in the classroom than they would have learned from classroom lecture alone*." Four out of the six teachers responded "Not Sure" to Q14 (use frequently): "*I would like to use Wuzzit Trouble frequently*".

At the same time, there was unanimous agreement to Q13 (use again): "*I would use Wuzzit Trouble again to teach my students more about math*", and strong agreement overall to Q9 (right difficulty): "*I thought the puzzles were at the right level of difficulty for my students*".

Figure 5. Teachers' percent of agreement by question across the five response categories.



Open-Ended Responses

Teachers were also asked open-ended responses about their opinions of WT. The first question asked teachers how they would describe the game to someone else. Most described it as an effective math learning game. One teacher wrote:

> *"Wuzzit Trouble is an enticing math application that students find engaging and easy to use. The students do not realize they are enjoying concepts in math that, for some, may find difficult to attempt."*

One question of interest was whether teachers believed that Wuzzit Trouble provided "*puzzles at the right level of difficulty for* [their] *students*". In response to this question, 5 of the 6 teachers indicated that they believe the puzzles were at the right level of difficulty for their students (one teacher was not sure). For example, one teacher wrote:

> "*Yes, students easily got through the first levels and then tried hard on the harder levels.*"

Teachers were also asked whether there were any obstacles associated with using Wuzzit Trouble. Two of the teachers indicated that there was difficulty for students to log in to the app due to bugs. In our observations, we sometimes observed the log-in section stay present on the screen even during gameplay. This feature was new to WT and specific to the feasibility study.

Teachers were also asked how they would use WT in their classroom again, as well as whether they had any other comments. Teachers provided primarily positive responses, such as:

> *"Wuzzit Trouble would be used in the classroom simply for student engagement and math and logic problem solving. I also appreciated how Wuzzit Trouble provided opportunities for the students to interact with each other verbally on how well they were solving problems or sharing what difficulties they had."*

Despite overall positive responses, one teacher did express concern if the study were to go on for a longer period:

> *"I think that if the study went on for more than two weeks, the students would start to lose interest. I think that the students would need to experience different learning games on a regular basis if you want them to stay engaged."*

We expand on these points in the general discussion.

Discussion: Building Towards Phase II

Multiple assessments in the feasibility study converged on the conclusion that both teachers and students exhibited positive impressions of this version of WT. A consistent finding, both in this feasibility study as well as in our prior usability testing, is that WT is an engaging and user friendly app. Students and teachers alike took away overall positive impressions of WT and exhibited high ratings on usability, feasibility, and engagement. In addition, the adaptive engine assigned levels that wavered between easy and hard puzzles, suggesting it was adapting to students' individual-level performance.

Given WT's feasibility for classroom contexts, we next examine WT in relation to the original research questions that focus on WT's ability to support student learning, motivation, and teacher implementations in the classroom context. We conclude with suggestions for improving WT's classroom feasibility.

Does playing WT increase student learning and attitudes towards mathematics?

Students and teachers from both schools responded positively to statements that asked whether WT supports their learning. This finding suggests that WT has the potential to increase student learning in mathematics. Though the pre and post-surveys did not provide indication of significant learning gains, these surveys suffered from high attrition which may have masked ability to detect the small effect sizes that would be expected from a short, two-week intervention. In addition, there was some evidence that WT helped students to solve equation problems, as indicated by higher post-survey performance on Equation questions relative to other question types, even when controlling for pre-test. WT also showed potential to support students' motivation and attitudes towards mathematics. For example, students from School A - and teachers from both schools - indicated that WT increased student motivation for mathematics. Students at School B as a whole did not show this trend, suggesting that WT's ability to motivate students for mathematics may depend on student characteristics. Nevertheless, we believe these findings suggest that WT has potential to increase student learning and motivation.

What are teachers' impressions of WT and how do they implement WT in their classroom?

Teachers responded overwhelmingly positively to WT, as indicated by responses to survey items as well as their implementation behavior. Almost all teacher responses were in the positive direction for survey questions. In addition, teachers implemented WT approximately double the minimum requirements for participating in the study (almost every day each week, and for approximately 20 minutes each day). Teachers exhibited a variety of implementation models in the classroom. These models varied with regard to how actively they interacted with students during game play, and the extent to which teachers explicitly linked WT to mathematical concepts. Despite this variation, all participating teachers provided positive ratings towards questions of WT's feasibility and implemented WT frequently, suggesting that a variety of implementation models are feasible for the classroom context.

What modifications to WT would maximize its classroom feasibility?

Though WT received high marks from users overall, there are a number of possibilities for improving the feasibility for classroom context. We summarize our suggestions below.

- **Consider developing structured classroom implementation models that are aligned with BrainQuake's theory for how WT helps students learn**. Teachers would benefit from guidance, in the form of implementation options, that can support their use of WT in the classroom. Developing implementation models that are consistent with how WT is expected to help students learn would enhance teachers' ability to use WT effectively. These implementation models might include extension activities that supplement math learning from WT as well as support the linking of WT to mathematical concepts. In addition, implementation models or activities that directly align to common core math standards would be a highly attractive feature for teachers and schools.
- **Build more puzzles that cover wider variety of mathematical topics.** One teacher remarked that for WT to remain a feasible classroom learning tool, it would need to cover a wider range of topics with different levels. Students in the focus group also commented on the need for greater variety of puzzles and levels. A large coverage of mathematical topics would only serve to increase the number of options that teachers have for implementing WT in the classroom, and thereby enhance its classroom feasibility. We also believe adding different kinds of levels in other topics would enhance the student experience.
- **Consider developing a teacher dashboard, so that teachers can view student progress in real-time.** The ability to see students' progress may help teachers actively support students in their WT play. For instance, a summary screen of student performance (e.g., average star counts, star counts from the past five levels, etc.) might motivate teachers to provide individualized support to the struggling students. According to BrainQuake's logic model, these student teacher interactions are a desirable component of WT's implementation, and may lead to improved student learning outcomes.

Overall, this study supports the conclusion that WT has strong potential for being an effective mathematics learning app that can be widely adopted for classroom use. These suggestions may enhance the feasibility of an already highly feasible classroom learning tool.

References

Pope, H., Boaler, & Mangram, C. (in prep). Wuzzit Trouble: The influence of a digital math game on student number sense. Pre-publication draft available at www.BrainQuake.com

APPENDIX. Math Survey Questions

1. Annie likes music and wants to buy as many songs as she can. Each song is priced at 9 dollars. How many songs can she buy with 45 dollars? [please show your work]

A. 8 songs B. 5 songs C. 6 songs D. 13 songs E.10 songs

2. Fill in the blank number:

8 x ___ = 32

A. 5 B. 3 C. 9 D. 4 E. 6

3. Finish the pattern: 55, 49, 43, 37, ___, ___,

A. 31, 25 B. 30, 25 C. 29, 24 D. 32, 26 E. 33, 27

4. Jen had $24 to spend on seven pencils. After buying them, she had $10. How much did each pencil cost? [please show your work]

A. 1 dollar B. 2 dollars C. 3 dollars D. 4 dollars E. 5 dollars

5. Fill in the blank number:

9 x 7 - ___ = 60

A. 3 B. 8 C. 7 D. 2 E. 10

6. Finish the pattern: 3, 6, 12, 24, __, __

A. 30, 36 B. 30, 48 C. 27, 30 D. 48, 98 E. 48, 96