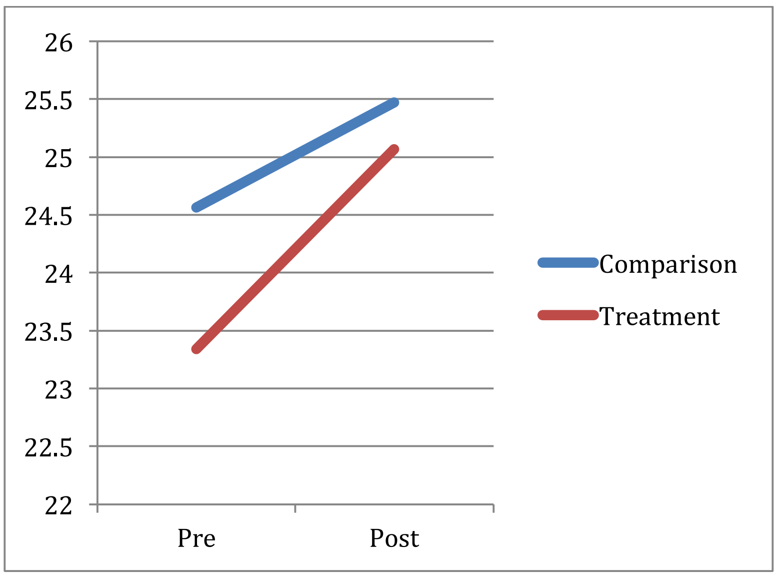
Stanford study shows *Wuzzit Trouble* leads to significant math learning gains

Major new research finding in game-based learning

Dr. Keith Devlin, Chief Scientist at BrainQuake

> "*My 5th period class, which is involved in the study, is an inclusion class with students with learning disabilities. On the last quiz I gave, the percentage of students receiving an A or B grade in this class was [only] one percentage less than those receiving an A or B grade in my Honors class which is filled with students in the gifted an talented program and my schools science magnet program.*"
>
> – Elementary school teacher involved in the *Wuzzit Trouble* study.

A recent Stanford University study showed that just 120 minutes play of the mobile video game *Wuzzit Trouble* led to dramatic math learning results that no one involved had believed were possible. The impact of the finding is even made more dramatic by the fact that the study was done using a comparison group, a protocol that is mandated for medical research but rarely, if ever, used to test learning games.



Graph showing the improvement in results from the written pre-test to the written post-test for the two classes. The treatment class, which the teacher had regarded as the weaker students, had almost caught up with their friends in the comparison class. The only difference between the instruction the two classes received was that treatment class spent 120 minutes of class time (10 minutes per day, 3 days a week, for four weeks) playing *Wuzzit Trouble*.[1]

The graph summarizes the study's key finding. In particular, the treatment group, the class that the teacher described as the weaker of the two, had almost caught up with their friends in the hitherto stronger comparison class. An examination of the numbers behind that graph shows that improvements on five key elements of written test scores averaged a remarkable 16.4%. That is after just two hours of self-guided play, spread over one month.

Since video-game play requires no strong literacy skills, they provide a level playing field for all learners, whatever their background. (Provided only that they have access to the technology

[1] Source: Holly Pope & Charmaine Mangram, *Wuzzit Trouble: The Influence of a Digital Math Game on Student Number Sense*, Stanford University, February 2015.

required — in the case of *Wuzzit Trouble* a tablet or smartphone, with a Web-based version in development for use on a PC or Chromebook). The findings of this study thus indicate significant potential for major impact on the mathematics development of students from disadvantaged backgrounds.

The study

> *"Given the increasing affordability and availability of mobile technology, it makes sense to explore using it in schools. With the type of mathematical engagement that is embedded into* Wuzzit Trouble*, it exemplifies a well-designed video game that could be a suitable context for learning mathematics*." – Holly Pope, M. Ed., lead researcher in the study.[2]

Scientific studies of learning outcomes resulting from new teaching programs, materials, or technologies rarely show improvements beyond 5 to 8 percent. Moreover, it usually requires a study period of several months using the new approach to yield such a result. Significant learning outcomes after a couple of hours using a new approach, spread over a month, are unheard of, and any claims of such a result need to be carefully assessed.

This was the position we at BrainQuake faced when the results came back from a small initial pilot study of learning outcomes for our first mathematics learning app *Wuzzit Trouble*, released in fall 2013. When we asked Stanford to conduct the study, it was not with the expectation of any significant findings. Rather, our goal was to test the research protocol on a small, but scientifically adequate, scale before running a much larger, and longer study. But with such a dramatic result, it was clear we had to look very closely at all aspects of the study, bringing in members of the company's Scientific Advisory Board.[3]

The project was directed by renowned mathematics learning professor Jo Boaler, and carried out by two doctoral students in Stanford's Graduate School of Education, one of whom (Holly Pope) did the bulk of the work. The team followed the strict controls the university requires of its researchers. Since Pope had decided to focus on game-based math learning for her qualifying research paper, the study was subjected to further scrutiny by a Stanford faculty panel.[4]

The study involved two classes of third-graders from an elementary school in California, both taught by the same teacher. Both classes had one hour math lessons daily. One of the two classes, the mathematical weaker of the two, incorporated *Wuzzit Trouble* into their regular math classes on Mondays, Wednesdays, and Fridays, while the other continued with normal classroom instruction.

In statistical study terminology, the *Wuzzit Trouble* class was the *intervention group*, the other class was the *comparison group*. The reason for conducting a study with two groups in this fashion is to be able to say with confidence that improvements are the result of the new approach. With the only difference between the two groups being the use of the new approach, when you compare the two groups in terms of improvements in performance measured by the test (post-test over pre-test), if any, you can conclude that any difference is due to the new approach. (This is why it is important that the two classes be taught by the same teacher.)

For one month, three times a week, the *Wuzzit Trouble* class finished normal instruction ten minutes early and spent the last ten minutes of class-time playing *Wuzzit Trouble*, which they

[2] Source: Conclusion to the paper.

[3] Following standard practice, Advisory Board members are unpaid, but given a nominal, very modest number of shares in the company as a token of appreciation. (Because she engages in BrainQuake learning research, Prof Boaler has no financial stake in company.) Their role is to assist the company by providing impartial expert advice. We chose our board to consist of world famous learning scientists who would not risk damaging their valuable reputations by being associated with anything less than world class science.

[4] I was asked to join the panel as an expert on the game itself, and the learning principles behind it.

downloaded for free from the Apple App Store and played on school-owned iPads. The teacher gave no instruction around the game. That is a total of 120 minutes playing a game at the end of class.

Both groups were tested before the month-long study began, and then tested again at the end, using a written test of mathematics performance with five questions. The test was designed by Holly Pope and Charmaine Mangram, under the supervision of Prof Boaler, who spent several years early in her career in charge of schools' mathematics testing in the UK, and is widely regarded as an expert in testing methods.

The use of comparison groups in this way is mandated for studies of new drugs or other medical treatments. Any study that does not involve the use of a comparison group is automatically suspect. Surprisingly, our *Wuzzit Trouble* study appears to be the first time this approach has been used to determine learning outcomes from a video game. The more common approach is to simply measure improvement in test performance from the start of the intervention (playing the game) and its completion some time later. This is much easier to implement, but it leaves open the possibility that any improvement is due to other factors besides playing the game. The use of a comparison group eliminates that possibility.[5] (This is why comparison groups are always used in medical research.[6])

What is going on?

> *"Educators should be aware of the types of mathematics learning a game promotes. Games that focus on speed and rote learning of skills will increase those skills in students, but may not deepen understanding and may foster misconceptions about math being about speed and rote memory instead of the creative, flexible discipline that it is.* Wuzzit Trouble *is an example of a digital math game that promotes mathematical proficiency and could pave the way for a new generation of mobile math apps*." – Holly Pope, M. Ed., lead researcher in the study.[7]

We were surprised to find such a significant improvement after such a short play-time. The only other instance we were aware of where such a short exposure to a learning video game produced significant learning outcomes was a recent study of the algebra game *DragonBox*, though that study did not involve a comparison group.[8] There are, however, a number of known factors that, taken together, could produce such an outcome. (Further research – which we intend to have done – will be required to unravel the complete story.) *Wuzzit Trouble* was designed to leverage these factors.

The three most significant factors are:

1. *Student engagement*.[9] Engagement is known to be a significant factor in learning. A well-designed video game creates a deep level of engagement rarely generated in a typical mathematics classroom.
2. *Mindset*.[10] Players in a video game quickly learn to adopt an iterative approach involving exploratory trial-and-error, reflection on failure, and subsequent adaptation. This results in a

---

[5] Some studies do set any gains they find against progress of students of the same age at other schools over the same time period, or same-aged students at the same school taught by other teachers. But then you cannot eliminate the possibility that you are measuring the schools or the teachers, not the effect of the intervention. To do that, at the very least you need to compare two classes taught by the same teacher during the period of the study.

[6] Indeed, for medical research there are even more stringent requirements on how the groups are created and who is allowed to know which group is the intervention.

[7] Source: Conclusion to the paper.

[8] Doron Popovic, 2014. "Learning basic Algebra by playing 1.5h". Center for Game Science, University of Washington.

[9] Byron Reeves & J. Leighton Read, 2009: *Total Engagement: Total Engagement: How Games and Virtual Worlds Are Changing the Way People Work and Businesses Compete,* Harvard Press

[10] Carol Dweck, 2007: *Mindset: The New Psychology of Success*, Ballantine Books

positive, "can do" attitude that Stanford researcher Carol Dweck demonstrated has an enormous effect on performance.

3. *Game design*.[11] [12] A well designed video game will lead to rapid, deep acquisition of whatever skills are intrinsically required to succeed in the game. A key word in that sentence is *intrinsic*. As Gee, Devlin, and others have observed, in order for a good video game to yield significant learning of X, the game has to be built tightly around X — essentially, the game mechanic has to be a dynamic representation of X. *DragonBox* does that with symbolic algebra (solving single variable, linear equations); *Wuzzit Trouble* does it with integer arithmetic, general problem solving skills, and algorithmic thinking. Few other video games adopt this approach.

In their paper, Pope, Boaler, and Mangram discuss issues that relate to these three factors, though they do not explicitly describe their observations along those three dimensions.

Analysis of the student scores on the five questions on the pre- and post-test provides particularly significant insight into the crucial learning that *Wuzzit Trouble* produces.[13] Four of the questions were fairly straightforward number problems closely related to the puzzles in the game itself. Both classes showed definite improvement in performance on those questions. It was the fifth question (actually, question number 4 on the sheet given to the students) that produced that dramatic convergence of the two lines in the graph. That question was a non-routine type that the students had been unlikely to have encountered before.

It was an example of what educators call a *complex performance task*. The question admitted multiple solutions, but could not be solved by the application of a learned technique. The students' first challenge was to understand what they were being asked to do. The solution then required that they attend to several constraints at once, and engage in decision-making processes.

Playing *Wuzzit Trouble* helped the intervention class to develop these proficiencies, which then transferred to their performance on the problem task.

The significance of this problem is that, in an era when machines we carry around in our pockets can solve problems amenable to an established technique, it is representative of the kinds of real-world mathematical (or mathematically related) problems that people in many walks of life frequently find themselves faced with. In short, the ability to solve a novel complex performance task is the key (human) mathematical ability in the 21st Century.

Examination of the students' success in meeting the four problem constraints where the difference in performance between the two classes was at a statistically significant level, reveals just how dramatic was the effect of playing *Wuzzit Trouble*. The *Wuzzit Trouble* class displayed increases between pre- and post-assessment of 46%, 22%, 21%, and 4%, whereas the comparison students showed essentially no change apart from an increase of 20% in meeting the first problem constraint.

In the one further problem constraint where there was a measurable difference between pre- and post- test, it was negative. The *Wuzzit Trouble* class showed a decrease of almost 11%. A decrease in performance may seem strange, but reflects the nature of complex performance tasks that require juggling many constraints. The decrease in the *Wuzzit Trouble* class was much less than in the comparison class.[14]

---

[11] James Paul Gee, 2003: *What Video Games Have to Teach Us About Learning and Literacy*, Palgrave

[12] Keith Devlin, 2011: *Mathematics Education for a New Era: Video Games as a Medium for Learning*, AK Peters/CRC Press.

[13] The authors provide the five-question test sheet used as an appendix to their paper.

[14] That figure of a 16.4% improvement cited earlier is the mean of the five figures 46%, 22%, 21%, 4%, and

Other benefits

As the authors note in their report, other schools took part in the study, but because the intervention classes were taught by different teachers than the comparison classes, no reliable quantitative conclusions could be deduced. One of the other teachers did however provide Prof Boaler with some qualitative information of interest. She wrote:[15]

> *Some things I have noticed with the students involved in the study:*
>
> *1. I award my classes with the highest participation average half way through the grading period. While I've done this for several years, this is the first year where I had class participation averages above 90%. These classes were two participating in the study.*
>
> *2. Students made an error while sharing their work in front of the class and as they made corrections they stated, "I'm learning right now."*
>
> *3. The achievement scores for the three classes involved in the study on my class tests and quizzes are higher than those not participating.*
>
> *4. Along the lines of achievement. My 5th period class, which is involved in the study, is an inclusion class with students with learning disabilities. On the last quiz I gave, the percentage of students receiving an A or B grade in this class was [only] one percentage less than those receiving an A or B grade in my Honors class which is filled with students in the gifted an talented program and my schools science magnet program. When I shared the results with my 5th period they attributed their success to how hard they had been working to learn the math. Before the study, these same students had the lowest achievement on a quiz and attribute their low scores to their ability, using phrases like, "We're the dumb class".*

Future research

BrainQuake is committed to developing learning materials based on established scientific learning principles, and to approach leading universities to request that they conduct independent studies, with comparison groups, in order to learn how effective are our products and what changes we need to make. The results of the Boaler–Pope–Mangram study validate that approach.

An educational researcher visiting Stanford for the year from Finland is currently finishing a similar, small scale *Wuzzit Trouble* study at a California middle school. This study differs in two ways from the first Stanford study.

First, the same study is being carried out in parallel at a middle school in Finland, to provide a comparison with a country that scores among the top few in international tests.

Second, the primary pre- and post- tests are in the form of performance in another math learning video game, developed by a Finnish company. That game, *Semideus*, is based on fractions. The study will thus determine the extent to which playing *Wuzzit Trouble* produces learning that transfers from one mathematical domain (integers) to another (fractions), when the testing takes place in a game. We expect the results to be available in May of this year.

Other, larger scale studies are also planned.

–11%. This is not a numerically precise statistical measure, rather a rough indicator of the scale of the difference.

[15] Email to Jo Boaler, December 10, 2014