## Abstract Title Page

**Title:** Instructional Gaming: Using Technology to Support Early Mathematical Proficiency

**Authors and Affiliations:**

Nancy J. Nelson-Walker
University of Oregon
nnelson3@uoregon.edu

Christian T. Doabler
University of Oregon
cdoabler@uoregon.edu

Hank Fien
University of Oregon
ffien@uoregon.edu

Marshall Gause
Thought Cycle, LLC
mgause@thoughtcycle.net

Scott K. Baker
University of Oregon
sbaker@uoregon.edu

Ben Clarke
University of Oregon
clarkeb@uoregon.edu

## Abstract Body

**Background / Context:**

Widespread concern has been expressed about the persistent low mathematics achievement of students in the US, particularly for students from low-income and minority backgrounds and students with disabilities. For example, results of the 2011 National Assessment for Educational Progress (NAEP) indicate that only 40% of 4th graders score at or above Proficient in math. Difficulties in mathematics achievement are particularly severe for students from low income and minority backgrounds and those with disabilities. For instance, nearly half of all 4th graders identified with a disability scored Below Basic on the 2011 NAEP. Mounting evidence also suggests that students who perform poorly in mathematics in the early grades are likely to continue to struggle throughout elementary school (Bodovski & Farkas, 2007; Morgan et al., 2009). Thus early intervention designed to support the needs of a range of learners is vital.

Instructional gaming technology, when designed and fictionalized well, has the potential to improve the motivation and mathematics achievement of students with or at-risk for mathematics difficulties (MD). Advanced gaming technology can provide a foundation to increase instructional intensity and serve as a motivational component for students who have experienced a long line of failure and frustration (Gersten et al., 2009b). Instructional gaming can, for example, facilitate the instructional interactions that deeply engage at-risk learners in the critical content of mathematics. Technology-based programs are also well suited to serve as targeted or intensive, supplemental interventions within a response-to-intervention framework, because of their capacity to differentiate instruction for a range of learners.

Despite these potential advantages, the research base is scant for efficacious technology tools in early mathematics (Dynarski et al., 2007). The National Education Technology Plan (NETP: Atkins et al., 2010) indicates that technology should be exploited to make learning experiences more meaningful, engaging, and accessible for students struggling to acquire academic proficiency. However, few of the numerous products available on the current market are grounded in research and development efforts that can fully address the agenda of the NETP and adequately meet the instructional needs of students at risk for MD. For example, in the area of early math instruction, the What Works Clearinghouse (WWC) has reviewed a total of 75 elementary programs to date, 22 of which are technology programs. Of these 22 programs, only two products at the elementary level have research studies that meet WWC screening criteria. In other words, less than 10% of the subset of technology programs reviewed have any research that could be used to evaluate their efficacy. Of the two reviewable programs, WWC ratings of impact on student outcomes are "potentially positive" and "mixed."

Moreover, few existing technology products infuse in their design research-based instructional and technological design principles that support students struggling to learn academic content. Many technology-based mathematics programs lack explicit modeling for teaching new and complex concepts, and fail to provide guided practice opportunities to facilitate student learning (Doabler & Nelson-Walker, 2013). In addition, existing technology programs often fail to limit extraneous information, teach key vocabulary, and provide clear instructional examples. A need clearly exists for more efficacious technology tools specific to early mathematics instruction (Dynarski et al., 2007).

**Purpose and Research Questions:**

Project NumberShire supports the development of in-depth knowledge of whole number concepts for students with or at risk for MD in grades K-2, a focus recommended by mathematics education experts (NMAP, 2009; Gersten et al., 2009a). NumberShire is a browser-based, educational video game in which players build an idyllic fairytale village by learning and applying whole number knowledge in three domains of the Common Core State Standards for Mathematics (CCSSO, 2010). This session will describe development and testing of the NumberShire intervention, and discuss results from feasibility studies in kindergarten, first, and second grade classrooms. At the conclusion of this session, participants will be able to identify (a) critical features of technology-based interventions and (b) preliminary results of a study using such an intervention to increase instructional intensity to meet the needs of at-risk learners.

To test the initial feasibility and usability of NumberShire, we aimed to answer four research questions: (1) Is NumberShire reliably efficient and easy for students and teachers to use? (2) Are students able to focus on and benefit from mathematics content in the game, rather than being distracted by other features? (3) Are students operating the game as intended? (4) Are students engaged in NumberShire mini-games and activities?

**Setting:**

Classrooms were located in three elementary schools in Oregon and Massachusetts. School A was a diversely populated, urban, charter school located in Boston, Massachusetts, with access to a math coordinator, two in-building math coaches, and a computer teacher with experience providing technology-based interventions in a well-equipped, up-to-date computer lab. Computers in this school were recent model year PC desktops, running Windows 7.

School B was a Title I school located in a suburban district outside of Eugene, Oregon. More than three-quarters of the student population receives free or reduced price lunch, and approximately 23% of students are identified as English Learners. Computers available in School B were 25 Apple MacBooks, originally marketed in 2005. School B offers access to math interventions and math coaching support one day per week, with no dedicated computer teacher.

School C was a Title I school located in a large, suburban district outside Portland, Oregon. Nearly half of all students in School C are identified as English Learners. School C has a dedicated computer teacher, four computer labs with a variety of technology equipment, and access to math coaching support and a variety of academic interventions.

**Participants:**

In fall 2012, 125 students participated in feasibility testing, 50 of which teachers identified as being at risk for difficulties in mathematics on the basis of student performance on screening measures (e.g., EasyCBM) and other classroom assessments. Participating students were 24 second graders (11 female, 13 male) from 2 second grade classrooms and 101 first graders (48 female, 53 male) from 4 first grade classrooms in Schools A and B. All six participating teachers (2 second grade, 4 first grade) in fall 2012 were female and had varying levels of teaching experience. Demographic information about spring 2013 participants will be collected and summarized in advance of our presentation. The spring feasibility test will be conducted in 4 kindergarten and 6 first grade classrooms in Schools A and C.

**Intervention:**

NumberShire is a fully featured, integrated learning and assessment system designed to support students with or at risk for MD develop proficiency with whole numbers in three domains represented in the K-2 Common Core Standards for Mathematics (CCSSO, 2010): (a) Counting and Cardinality, (b) Number and Operations in Base Ten, and (c) Operations and Algebraic Thinking. Each version of NumberShire (i.e., NumberShire K, NumberShire 1, and Number Shire 2) consists of 12 hours of individualized instructional game play, comprised of 15-minute sessions, delivered 4 times per week for approximately 12 weeks. Players assume the role of a young member of a Renaissance-style village in the fairytale-inspired medieval kingdom of Tally-ho, where the village elder is stepping down and handing over the mantle of leadership to the player. Sim-style game mechanics allow the player to click on village buildings to trigger mini-games targeted at whole number concepts, such as composing and decomposing teen numbers, and word problem solving. Sessions utilize an explicit instructional format and contain three instructional phases: explicit modeling, supported practice, and independent practice. Embedded within each session are four mathematics mini-games, including a Teaching Event (i.e., a mini-lesson targeting a new instructional objective), Assessment Event (i.e., review of a previously mastered objective), Warm-up, and Wrap-up. Mini-games include clear explanations to introduce new material and high quality feedback. A variety of virtual mathematical representations (e.g., number lines, base-10 blocks) and frequent practice opportunities are employed to facilitate procedural fluency and build a robust and enduring conceptual understanding of whole number concepts.

**Research Design:**

This study used design experiment methodology (Brown, 1992; Diamond & Powell, 2011; Shavelson et al., 2003) and iterative end-user testing trials to examine initial feasibility and usability of the NumberShire intervention. Mixed methods research design was also employed to study initial student learning during the feasibility test and guide program revisions in preparation for a formal, rigorous, small scale RCT pilot study to assess intervention promise.

**Data Collection and Analysis:**

Feasibility data were collected in fall 2012 and spring 2013, from a variety of assessment, game, and interview activities. In each school, feasibility testing was conducted across a period of one week. Each week was comprised of four 30-minute research sessions. Within each session, approximately 15 minutes were devoted to student game play and the remaining time was used for assessment and interview activities. Research staff observed game play sessions and conducted interviews with participating students and teachers. To assess students' baseline knowledge of whole number concepts and student mathematics learning during the feasibility test week, research staff administered a (paper-pencil) proximal measure of mathematics achievement at the beginning and end of the feasibility test. In addition, game metrics built into the prototype were used to gather data on student accuracy and latency during game play, across the week's four sessions.

On the second and third days of each testing week, students participated in large and small group interviews conducted by a member of the research staff. Students were asked to recall math activities encountered in the game and describe storyline events to determine whether they could discern and remember targeted, unique features of the prototype. Students were also asked about their game preferences (e.g., favorite characters and activities, reward structure) to gauge general interest and engagement in NumberShire. Daily observations of student game play were also conducted using a standardized observation protocol. Observations targeted student dexterity, ability to navigate the interface with accuracy and ease, and sustained engagement during game sessions. To further assess the feasibility and usability of the NumberShire prototype, research staff conducted interviews with participating teachers. Teachers were asked to describe their perceptions of student interest in NumberShire, the accessibility of the game interface, alignment of NumberShire with essential math content, and the utility of other game features (e.g., teacher data reports, customized instructional recommendations).

In winter 2013, research staff summarized data from all observations and interviews conducted in second grade classrooms. Paired t-tests were used to analyze pre and posttest performance across item sets, game event types (e.g., Teaching Events, Assessment Events), and pre and posttest total score, for second grade students. Data from game play were examined for changes in student performance across similar events from the beginning to the end of the week, for second grade participants. Results for kindergarten and first grade students will be analyzed using the same procedures in summer 2013.

**Results:**

***Student interviews, observations, and teacher interviews.*** Second grade participants enjoyed playing NumberShire, were engaged in game activities, and were able to navigate the interface with ease. Statements from students reinforce our plans to increase the frequency of effort- and performance-related rewards and incentives. Participating teachers indicated students were excited about the game (e.g., teachers said students asked, between sessions, when they would be able to play the game again), believed it targeted important math skills, and expressed interest in obtaining student performance data and customized instructional recommendations.

***Student proximal assessment.*** Preliminary results (see Table 1) suggest that students performed significantly better at posttest in solving subtraction number combinations and using proximal math models to understand place value. Students also performed significantly better at posttest on content targeted in Assessment Events and Warm-up/Wrap-up activities. It appears that the amount of practice included in game activities had a positive impact on student mathematics learning. Total score mean performance increased from pretest to posttest, but the increase was not statistically significant. This finding is not surprising, in part because NumberShire is intended to be a 12-week intervention. We expect students will need more than four days of exposure to the intervention to observe substantial increases in mathematics performance.

***Student performance during game play.*** Data from game play were examined for changes in student performance across similar events from the beginning to the end of the week (see Table 2). For example, during the feasibility test, students began each session with the Magic Slate warm-up. We examined math fact fluency data across sessions to determine whether students demonstrated increased math skill from the first to the last session. Preliminary data indicate students were more fluent and accurate with math fact problems at the end of the week.

**Conclusions:**

This session will describe the NumberShire intervention's instructional design framework, game mechanics, and mathematical content. Video clips of NumberShire game play will be shared along with results from a recent feasibility study conducted in kindergarten, first grade, and second grade classrooms. Results from the first half of the feasibility study suggest second grade students were engaged in and able to use NumberShire with general ease. Second grade teachers perceived the NumberShire intervention as being aligned with important skill objectives and found the features of the intervention useful. Preliminary results suggest NumberShire may support student achievement for second grade students working to demonstrate proficiency with whole number concepts aligned with the Common Core State Standards for Mathematics (CCSSO, 2010). However, results are based on a one-week test of the intervention and the rigor of our study design prohibits causal conclusions. Also, researchers facilitated all research activities during the fall 2012 feasibility test (while researchers supported teachers to implement NumberShire during the spring 2013 feasibility test). Results are presently limited to a small number of second grade students; however, we will share results for a much larger sample in our proposed presentation. Future studies should rigorously test the feasibility and promise of the full-featured NumberShire intervention for improving student math outcomes.

## Appendices

### Appendix A. References

Atkins, D. E., Bennett, J., Brown, J. S., Chopra, A., Dede, C., Fishman, B., . . . Williams, B. (2010). *Transforming American education: Learning powered by technology*. Washington, DC: U.S. Department of Education, Office of Educational Technology. Retrieved from http://www.ed.gov/sites/default/files/netp2010.pdf

Bodovski, K. & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 108*(2), 115-130.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of Learning Sciences,* 2, 141–178. doi: 10.1207/s15327809jls0202_

The Council of Chief State School Officers. (2010). Common core state standards initiative: Designing common state assessment systems. Retrieved from http://www.nga.org/Files/pdf/1004NGACCSSOASSESSMENTS.PDF

Diamond, K. E. & Powell, D. R. (2011). An iterative approach to the development of a professional development intervention for head start teachers. *Journal of Early Intervention, 33*(1), 75-93.

Doabler, C. & Nelson-Walker, N. J. (2013). Evaluation of Technology and Instructional Design and Delivery Principles. Instrument available from the Center on Teaching and Learning at the University of Oregon.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., . . . Sussex, W. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort. Washington, DC: US Department of Education, Institute of Education Sciences.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009a). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools.* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Gersten, R., Chard, D., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009b). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*, 1202–1242.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321.

National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, DC: US Department of Education.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher, 32*, 25–28.

U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Mathematics Assessment.

## Appendix B. Tables and Figures

Table 1. *Student Proximal Assessment Data*

| | | Pretest | | Postest | | | |
|---|---|---|---|---|---|---|---|
| **Measure** | *N* | *M* | *SD* | *M* | *SD* | *t* | *p* |
| Addition fact fluency | 24 | 29.79 | 3.45 | 30.29 | 3.63 | 0.72 | .480 |
| Subtraction fact fluency | 24 | 16.79 | 5.00 | 19.25 | 5.06 | 2.32 | .030** |
| Place value models | 24 | 2.96 | 0.20 | 2.88 | 0.61 | -0.62 | .539 |
| Proximal place value | 24 | 1.13 | 1.48 | 1.92 | 1.44 | 2.22 | .036** |
| Multi-digit addition | 24 | 6.50 | 2.54 | 6.46 | 2.41 | -0.09 | .927 |
| Multi-digit subtraction | 24 | 5.33 | 2.60 | 5.50 | 2.90 | 0.33 | .748 |
| Decomposing numbers | 23 | 6.39 | 3.12 | 6.43 | 3.13 | 0.10 | .922 |
| Comparing numbers | 23 | 3.39 | 1.31 | 3.70 | 1.02 | 1.13 | .272 |
| Word problem type | 24 | 1.58 | 0.65 | 1.63 | 0.71 | 0.27 | .788 |
| Completing a strip diagram | 24 | 4.21 | 1.82 | 4.54 | 1.41 | 1.23 | .224 |
| Drawing a strip diagram | 23 | 4.39 | 2.41 | 4.70 | 1.94 | 0.66 | .519 |
| Word problem equation | 19 | 8.00 | 3.90 | 6.58 | 4.73 | -1.43 | .169 |
| Word problem essential info | 23 | 1.87 | 0.46 | 1.91 | 0.29 | 0.37 | .714 |
| Teaching events | 19 | 17.68 | 5.96 | 17.47 | 5.20 | -0.14 | .888 |
| Assessment events (practice) | 23 | 13.91 | 3.45 | 15.00 | 4.37 | 1.80 | .085* |
| Warm up and wrap up activities | 24 | 46.58 | 7.04 | 49.54 | 8.17 | 2.02 | .056* |
| Multi-digit operations | 24 | 11.83 | 4.72 | 11.96 | 4.79 | -0.16 | .874 |
| Proximal measure total score | 19 | 89.68 | 17.34 | 92.89 | 14.65 | 1.19 | .248 |

**$p < .05$, * $p < .10$

Table 2. *Student Data During Game Play*

| | Latency | | Accuracy | |
|---|---|---|---|---|
| **Activity** | *M* | *SD* | *M* | *SD* |
| Activity 49: Math fact fluency, Day 1[W] | 12193.38 | 5055.30 | 0.38 | 0.14 |
| Activity 50: Word problems +2, Day 1[T] | 8732.18 | 11369.06 | 0.91 | 0.19 |
| Activity 52: Math fact fluency, Day 2[W] | 10672.93 | 6134.62 | 0.35 | 0.16 |
| Activity 53: Word problems +2, Day 2[T] | 6921.99 | 4198.43 | 0.93 | 0.10 |
| Activity 54: Proximal place value models[A] | 69032.44 | 38513.61 | 0.43 | 0.36 |
| Activity 55: Word Problems +1[D] | 7226.09 | 4148.38 | 0.84 | 0.15 |
| Activity 57: Math fact fluency, Day 3[W] | 9953.23 | 4263.78 | 0.41 | 0.21 |
| Activity 59: Compare numbers[A] | 3521.54 | 4759.10 | 0.61 | 0.40 |
| Activity 60: Count up subtraction -1[D] | 11260.28 | 7376.95 | 0.64 | 0.44 |
| Activity 62: Math fact fluency, Day 4[W] | 9203.64 | 4324.81 | 0.46 | 0.31 |
| Activity 62.5: Count up subtraction -2/-3, Day 2[T] | 13498.88 | 8637.27 | 0.90 | 0.23 |
| Activity 63: Word Problems +2[A] | 9333.89 | 19979.93 | 0.89 | 0.18 |
| Activity 64: Count up subtraction -1[D] | 17404.52 | 15888.10 | 0.64 | 0.32 |

*Note.* [W] = Warm-up activity; [T] = Teaching Event; [A] = Assessment Event; [D] = Differentiated Learning Pathway. Student data only includes cases from 20 second grade participants (School A). Latency is the average number of milliseconds elapsed between student responses. Accuracy is the average proportion of items students answered correct.