"IT'S JUST LIKE LEARNING, ONLY FUN" – A TEACHER'S PERSPECTIVE OF EMPIRICALLY VALIDATING EFFECTIVENESS OF A MATH APP

Ka Rene Grimes
Momentous Institute, 106 E 10th St, Dallas, TX 75203

ABSTRACT

The purpose of this paper is to provide a narrative of work in progress to validate a math app designed for number sense. To date I have conducted classroom research and pilot studies across ten early childhood classrooms in two schools and will begin an empirical study at the beginning of the 2014-2015 school year. Through my work I believe the fields of neuroscience, education, and digital science offer robust and unique ways to address at least two barriers I encountered: identifying instructional computer adaptive software containing embedded assessments and designed explicitly with cognitive models of learning; and developing ongoing collaborative research networks to validate this software. In an attempt to inform the work of those working in the fields of digital science, cognitive science and education, my reflection includes the background, content, context and observations of my studies to date, as well as insights and emerging hypotheses for consideration.

1. INTRODUCTION

The quote in the title of this paper is from a former student; a testimony to being engaged while using a math app. Engagement is an important component of instruction to measure, but what methods exist for measuring learning and understanding of instruction delivered via apps? Does instructional software exist that allows for learning and assessment to occur simultaneously? How can classroom teachers undertake investigating these questions through formal research? Computer adaptive software offers great potential to solve these questions; collaborative networks offer the potential to research these questions. In this paper I offer a narrative of my work in progress to validate *Native Numbers©,* a math app, through collaboration with a network of researchers and educators. My purpose in sharing this narrative is to provide insights and observations that may inform future work of those in the fields of cognition, education, and computer science. This narrative adds to the limited body of research conducted explicitly through collaborations between educators, researchers and other stakeholders, in an attempt to validate learning effectiveness of educational software designed intentionally with cognitive models of learning.

2. BACKGROUND

As a teacher I must know if the academic content of any instructional resource is effective and as a practitioner-researcher I am bound to investigate effectiveness thorough formal or informal study. This study began during the last semester of my master's graduate work. As background, I highlight here three watershed events as to why I chose to examine the educational effectiveness of *Native Numbers©*, a math app for early number sense. First, moving from teaching prekindergarten to first grade I perceived many of my former students had not progressed in some of the math skills I considered they had mastered a full year previously. I had ultimate confidence in their kindergarten teachers' instruction, so I was greatly puzzled. Second, my graduate work in the Mind, Brain and Education (MBE) program at the University of Texas,

Arlington (UTA) prepared me to take on the role of practitioner-researcher. Core to the program design is a goal of enabling teachers to become teacher-researchers, trained and ready to bridge the divide between neuroscience and education (Schwartz & Gerlach, 2011).Outside of my required course work, I read research and attended professional development related to number sense and experimented with different instructional methods. Third, while attending a workshop for *Native Numbers©*, I recognized both the cognitive models and the number sense research this app had included in its design. I asked my director for permission to conduct a study of the app as an elective independent study. We contacted the developers and initiated a collaboration to study the educational effectiveness of *Native Numbers ©* for early number sense.

Validation of assessments and instructional resources for math is critical as the long term effect of mathematical difficulties, and the need for intense and explicit assessment and intervention, cannot be overstated. Morgan, Farkas, and Wu (2009) conducted a study using data from the U.S. Department of Education's National Center for Educational Statistics to investigate five year growth trajectories of students with mathematical difficulties (MD). They found students who manifested MD in kindergarten were likely to continue showing patterns of these struggles throughout their elementary years, with the clear implication being need for additional assistance prior to the end of kindergarten (p. 319). In *Screening for Mathematical Difficulties in K-3 Students*, Gersten, Clarke, Haymond, and Jordan (2011) illuminate the serious need for early math screening and note that specific components of efficient assessments remain unclear, with one problem being how can we gain the maximum amount of information in the minimum amount of time (p.15). The authors suggest "any assessment instrument must be guided by findings from developmental and cognitive psychology…and by mathematics educators' expertise" (p. 4).

Native Numbers© is adaptive and mastery-based and in this sense is both a curriculum and an assessment. Tasks are designed specifically around the concept of a mental number line on which representations of numeric quantities and magnitude are manipulated (Berch, 2005; Dehaene, 2011; Griffin, Case, and Siegler, 1994). Order, presentation and mastery are based on number sense development (Griffin, Case, and Siegler, 1994), as well as the cognitive models of Perceptual Control Theory (Powers, 1998), Skill Theory (Schwartz & Fisher, 2004) and Flow (Nakamura & Csikszentmihalyi, 2009). *Native Numbers'©* sequencing of skills, with unlimited attempts to progress through proficiency to mastery, fits the model Pellegrino, Chudowsky and Glaser (2001) describe as "intelligent tutoring systems (that) have a strong cognitive research base and offer opportunities for integrating formative and summative assessments, as well as measuring growth" (p. 257). Shepard, Daro, and Stancavage (2013) contend that learning progressions are "an advancement beyond traditional scope and sequence schema…document how learning typically unfolds…ensure the close connections between assessment and instruction…(and) hold promise for the deepening of student learning" (p. 137). At face value, *Native Numbers©* holds promise to meet all the aforementioned current needs.

3. INSIGHTS FROM ACTION RESEARCH AND PILOT STUDIES

Goals of my independent study were to: provide the developers feedback from teachers and students on the usability of the app and web-based dashboard; and investigate our hypothesis that raw scores taken from the dashboard would correlate with scores of identical tasks on a validated math screener. The study included a pilot in a first grade classroom and a formal study in three kindergarten classrooms. My district agreed I could provide anecdotal informal feedback, but ultimately denied a formal study. The following year, at a different location, the second iteration of this study included: formal study of two kindergarten classrooms, a pilot study of two first grade classrooms and informal observations of two second grade classrooms. Complications encountered during the pilot study postponed the formal study until the next school year; the study currently in progress. Insights from all studies to date are combined under the headings of: The Possibilities of Adaptive Software with Embedded Assessment; Design and Analyses of Computer Software Models; and Collaborative Research Networks Support All Stakeholders.

3.1 The Possibilities of Adaptive Software with Embedded Assessment

In each of the studies specific observations held true regardless of whether I was the observer or a fellow teacher observed her own students. These observations warrant further study; study, which because of the number of observations required, could be difficult to conduct through traditional paper and pencil

assessment. Students who had historically indicated difficulty with math skills all made significant progress, but sometimes this progress required multitudes of tasks; in one case over three thousand to complete just one of the 25 activities. The reality of a classroom setting makes it impossible for any one teacher to provide three thousand attempts to show proficiency or fluency in one concept. Pellegrino et al (2001) describe this exact scenario: "the most useful kinds of assessment…support a process of individualized instruction, allow for student interaction, collect rich diagnostic data, and provide timely feedback….(S)ignificant amounts of information must be collected, interpreted, and reported. No individual… could realistically be expected to handle the information flow, analysis demands, and decision-making burdens involved without technological support" (p. 272). Across the three grade levels and schools, the progression of learning was extremely varied. No two students learn in the same way over the same period of time. Use of computer adaptive software has the ability to make this differentiation feasible. In the five kindergarten classrooms, a total of 92 students used the app for approximately three weeks for approximately the same amount of time. Of those 92 students 13 students completed the 25 skill sets, reaching the level of fluency; 22 completed 13 of the skill sets with at least proficiency; and 10 students were able to complete only three or fewer skill sets. Combining the number of tasks of the 13 students who complete the app resulted in over 53000 tasks. It would be unrealistic to believe a classroom teacher could provide even just 13 students with 53000 drills or problems in three weeks just a few minutes per day, much less assess the results.

Another salient observation is that many of the students had significant differences in the number of tasks it took to reach fluency in counting up by one and counting down by one. Comparing just these two skills across the classrooms, 15 out of 92 kindergarten students completed these with fluency. The difference between the number of tasks required to reach fluency in counting down over counting up ranged from 0 to 75 with a mean of 36; 36 out of 52 first grade students completed skills with a mean of 32 and a range between 0 and 120; and 26 out of 32 second grade students completed both skill sets with a mean of 20 and a range from 0 to 125. This data highlights the significant difference in learning progression of students. Of particular interest were the students whose previous classroom performance indicated a high level of mathematical understanding, yet the number of tasks to reach fluency in counting down was much higher than expected; this was particularly true for the subset of students who also had recognized difficulties of attention. Scant few studies exist on the cognitive process of subtraction in elementary students and these small findings warrant further investigation. (See Barrouillet, Mignon & Thevenot, 2008, for literature regarding the cognitive process of subtraction, working memory and discussion of competing theories).

3.2 Design and Analyses of Computer Software Models

The pilot study of the first grade classrooms at the second school included a cross over design whereby all students would be assessed with a validated screener utilizing pre-, mid-, and post-tests. This design eliminates classroom variables such as ongoing instruction. One of the difficulties encountered in the pilot study was the initial analyses of data from the validated screener showed most students were at ceiling. Furthermore, at the mid-test, little statistical significance was found. Why students were at ceiling on the validated screener yet needed multiple tasks to reach proficiency is puzzling and suggests the need for a closer look at micro-skill development. The number of variables is great; however one possibility could be that the screeners are validated using means of groups. It may be that computer adaptive assessment counters floor and ceiling effects of paper and pencil or human administered tests (de Beer, 2010, p. 243). One methodology researchers are exploring is using meta-analyses of single subjects; each student is their own case and then multiple cases are analyzed. The *Journal of School Psychology* recently dedicated a special issue to this subject (2014). Byiers, Reichle and Symons (2012) offer an explanation of how single subject designs can be used for evidence based practice. Forbes, Ross and Chesser (2011) remind us we must keep the individual at the heart of each study, "(e)ven if group statistics do not detect individual learning gains, or even suggest no gain through lack of statistical significance, individual gains are not insignificant at a practical level" (p. 171). What Works Clearing House has issued standards for determining what constitutes a single case design as well as standards for visual analyses (Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, & Shadish, 2010). Heyvaert and Onghena (2014) explain how to design various intervention methods for measuring effectiveness of an intervention and also recommend combining both visual and statistical analyses of the data. Additionally, there is a rich amount of literature addressing Bayesian analytic methods for single subject designs. For a brief idea of how Bayesian models can inform analyses of single

subject designs see: de Vries and Morey (2013), Ferron, Farmer and Owens (2010), and Rindskopf (2014). Vos (2007) provides a thorough explanation of how to use Bayesian methods to determine concept learning, in particular in determining mastery. Meta-analyses of single subject design studies, analyzed using both visual and statistical methods for validating computer adaptive software are warranted, in particular those using various Bayes models. This type of analyses will most likely require a collaborative relationship with a statistician well versed in Bayesian models. Computer technology that is constructed with cognitive models is complex and may very well require complex methods of validation and collaboration.

3.3 Collaborative Research Networks Benefit All Stakeholders

One means of incorporating pedagogy in analyses is to include teachers in collaborative research networks. Educators *do* conduct research and the field of action research is rich with data. The research that typically occurs in classrooms is considered action research which generally is not designed in the way that experimental research is conducted, there may or may not be control groups, and the participant (or sample) sizes are not normally large enough to provide statistical power (Sigler, 2009, p.23). The collaborations within my research occurred through networks established by the MBE program at UTA; in particular the collaborative partnership of the Research Schools Network (See Schwartz & Gerlach, 2011). Many collaborations similar to the MBE framework exist and studies from these educator and researcher collaborations are emerging: some completed while the educator is enrolled in a graduate program, fewer from classroom educators not in one of these programs (Cornelissen, van Swet, Beijaard, & Bergen, 2013). It isn't always the researcher who contacts a school district; I contacted the developers and asked to research their app. Regardless of how the collaborations are established, every stakeholder has an equally vested interest in researching educational tools; there is no room for hierarchy. Connell presents a model (Figure 3.1), a "gold standard," of collaborative networks, the Connell Adaptation Loop. The model is a complex adaptive system: each element provides feedback and energy, allowing for growth and emergence. Researching educational practice is complex, iterative, and requires time.
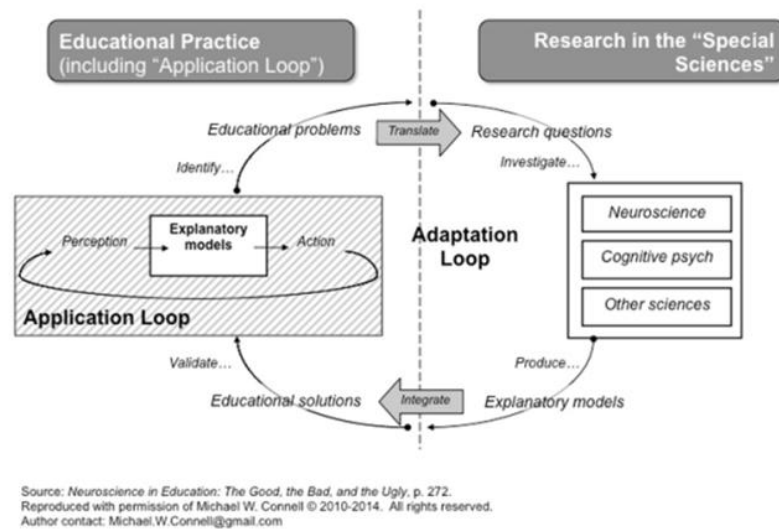
Figure 1. Connell Adaptation Loop

4. CONCLUSION

In my work to date I have seen what I believe is a model of computer adaptive software with embedded assessment that provides students with individualized instruction. What remains is validation of these claims. The observations in this paper are from a study in progress and are meant to further dialogue between researchers and teachers in an effort to better understand the complexity of designing and researching innovative technologies for education. The literature presented here is not exhaustive and the challenges of developing and sustaining collaborations has not been included. I believe the fields of neuroscience, the learning sciences and digital science offer robust and unique ways to design, research and validate innovative, educational technologies by developing ongoing networks of collaborative relationships. I look forward to the day when my students no longer say, "It's just like learning, only fun" but "I learned. It was fun!"

ACKNOWLEDGEMENT

REFERENCES

Analysis and Meta-Analysis of Single-Case Designs [Special Issue]. (2014). *Journal of School Psychology*, 52, 109-248.

Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, *99*(4), 233-251.

Berch, D. B. (2005). Making sense of number sense implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38*(4), 333-339.

Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, *21*(4), 397-414.

Cornelissen, F., van Swet, J., Beijaard, D., & Bergen, T. (2013). Exploring knowledge processes based on teacher research in a school–university research network of a master's program. *Journal of Educational Change*, *14*(2), 139-176.

De Beer, M. (2010). A modern assessment psychometric approach to dynamic assessment. *Journal of Psychology in Africa*, *20*(2), 241-246.

de Vries, R. M., & Morey, R. D. (2013). Bayesian hypothesis testing for single-subject designs. *Psychological methods*, *18*(2), 165.

Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior research methods*, *42*(4), 930-943.

Fischer, K. W. (2009). Mind, brain, and education: building a scientific groundwork for learning and teaching. *Mind, Brain, and Education*, *3*(1), 3-16.

Forbes, S. A., Ross, M. E., & Chesser, S. S. (2011). Single-subject designs and action research in the K-12 setting. *Educational Research and Evaluation*,*17*(3), 161-173.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). *Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure*. The MIT Press.

Heyvaert, M., & Onghena, P. (2014). Randomization tests for single-case experiments: State of the art, state of the science, and state of the application. *Journal of Contextual Behavioral Science*, *3*(1), 51-64.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities: Research and Practice*, *21*(1), 37–47.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

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

Nakamura, J., & Csikszentmihalyi, M. (2009). 18 Flow Theory and Research. *Oxford handbook of positive psychology*, 195.

Native Numbers©, iPad© app. Connell, M.W., & Durso (2012). Via http://www.nativebrain.com/apps

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Powers, W. T. (1973). Behavior: The control of perception. *New York, NY: Hawthorne*.

Rindskopf, D. (2014). Nonlinear Bayesian analysis for single case designs.*Journal of school psychology*, 52(2), 179-189.

Schwartz, M., & Fischer, K. W. (2004). Building general knowledge and skill: Cognition and microdevelopment in science learning. *Cognitive developmental change: Theories, models, and measurement*, 157-185

Schwartz, M., & Gerlach, J., (2011). The birth of a field and the rebirth of the laboratory school. *Educational Philosophy and Theory, 43 (1),* 67-74.

Schwartz, M. S., & Gerlach, J. (2011). Guiding principles for a research schools network: Successes and challenges. *Mind, Brain, and Education*, 5(4), 172-179.

Shepard, L., Daro, P., & Stancavage, F. B. (2013). The Relevance of Learning Progressions for NAEP.

Sigler, E. (2009). Action Research in Public Schools: Is it Research? Should it be Reviewed?. *Journal of Empirical Research on Human Research Ethics: An International Journal*, 4(2), 17-25.

Vos, H. J. (2007). A Bayesian sequential procedure for determining the optimal number of interrogatory examples for concept-learning. *Computers in human behavior*, 23(1), 609-627.