Guest Editorial:
More Questions than Answers:
Responding to the Reading and Mathematics Software Effectiveness Study

[Kimberly M. Fitzer](), [Joseph R. Freidhoff](), [Anny Fritzen](), [Anne Heintz](), [Matthew J. Koehler](), [Punya Mishra](), [Jim Ratcliffe](), [Tianyi Zhang](), [Jinjie Zheng](), and [Wenying Zhou[1]]()
Michigan State University

Problems worthy of attack, prove their worth by hitting back—Piet Hein

There have been few large-scale empirical studies of the effectiveness of educational software in improving student learning, even though educational technology has become a ubiquitous tool for learning both in and out of the classroom. The recently released report to Congress, "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort," (Dynarski et al., 2007) produced by the National Center for Education Evaluation and Regional Assistance aims to fill this gap. (The complete report can be downloaded from http://ies.ed.gov/ncee/pdf/20074005.pdf).

The study is impressive in many regards. The sample included 439 teachers and 9,424 students in 33 districts. Sixteen different software products for reading in first grade and in fourth grade, mathematics in sixth grade, and high school algebra in ninth grade were selected. Classrooms within participating schools were randomly assigned to treatment or control conditions. Treatment classrooms used one of 16 preselected software titles, while not interfering with instruction in control classrooms.

Teachers in the treatment group received software and training on one of the software products and were expected to use it in their teaching. It is worth noting that teachers in control classrooms were also permitted to use educational software products and even introduce new technology products at their discretion. Assessments of effectiveness cast a wide net by including standardized tests and classroom observations. In all grades, students were given reading and mathematics tests both at the beginning and end of the school year. Teachers were interviewed to determine their attitudes toward the technology products and to assess how these tools were being used.

The study's key finding is that the use of preselected educational software products did not make a statistically significant difference in student test scores between the control and the treatment groups, though the report suggests that there was substantial variation between schools regarding the effects on student achievement. The media were quick to respond to these findings (e.g., Paley, 2007; Trotter, 2007; *ZD Net* Editorial, 2007a, 2007b; among others). Our concern, however, is that readers might misinterpret the nuances of the study, and the subsequent media reports as being an indictment of *all* educational technology. There is some evidence that this is already happening. For

instance, a group opposing a recent technology bond initiative in a Michigan school district, quoted the study and other media reports as evidence for supporting its case (see the Common Sense for Okemos Web site at http://commonsenseforokemos.org).

As scholars of educational technology it is vital that we engage in a serious dialogue regarding the implications of these findings for research, scholarship, and policy. This guest editorial is just one small step in this direction. We would like to thank the editors of CITE Journal for providing us with the opportunity to write this first response, as it were, allowing us to initiate the conversation. We expect that this dialogue will continue online, through the mentoring blog sponsored by the Society for Information Technology and Teacher Education and other venues.

There is much to admire in this study. First, the focus of the study is particularly relevant given the financial resources that many schools, districts, and local governments are investing in educational technology. Administrators and policy makers need to know if technology-based instructional interventions really improve what matters most—student achievement. Second, the selected focus on low-income schools is consistent with the *No Child Left Behind* (NCLB) goal of providing achievement-driven interventions to all students, particularly traditionally underserved populations. The emphasis on low-income schools is laudatory, but it is complicated somewhat by the fact that most of the software programs selected for this study were tutorials. This approach is consistent with previous surveys showing that low-income and otherwise disadvantaged schools tend to use such software programs, rather than using more open-ended programs emphasizing higher order thinking skills.

Third, investigating technology use in first-, fourth-, sixth-, and ninth-grade reading or mathematics classrooms provides data that would help us understand how students and teachers interact with educational software at different levels and in different content areas.

Fourth, by carefully selecting software products that showed some prior evidence of effectiveness, the researchers reduced the possibility that nonsignificant outcomes could be attributed to inferior software products.

Finally, given the difficulties of measuring immediate results within real-world environments, the longitudinal design of the study is a definite strength. Regardless of whether the researchers find significant results in Year 2, we admire the foresight of a longitudinal design that carefully disaggregates the treatment data.

In addressing the question of whether or not software-based interventions are successful, however, this report raises many more questions for the research community than it answers. A primary question is how broadly do the results of the study apply? Despite its merits, the study displays certain fundamental limitations, only some of which are discussed in the report. For example, the authors point out that the study "was not designed to assess the effectiveness of educational technology across its entire spectrum of uses, and the study's findings do not support conclusions about technology's effectiveness beyond the study's context, such as other subjects areas" (p. xiv). Most media reports of the study, however, fail to convey these nuances of interpretation.

Another question is how much software use is sufficient to produce a difference in learning outcomes? For example, the selected software programs were used, on average, for only about 10-11% of the instructional time.[2] More importantly, what is not clear is the degree to which the other 90% of instructional time was meaningfully coordinated

with computer use. If the use of educational technologies is regarded as a separate classroom activity, rather than an integral one, its effectiveness will clearly be limited. Some evidence in the report supports this position, particularly for fourth grade, where effects were larger when teachers in the treatment condition reported higher levels of use of the selected software programs.

There are several questions about the training the teachers received, how it was implemented in the classroom, and how student learning was assessed. Was the length of the intervention adequate to observe an effect? Furthermore, how much training do teachers need to effectively use educational software? For example, the fact that teachers' confidence in using the software dropped from 95% after training to around 60% after they began to use the software suggests that perhaps teachers were not adequately prepared to use these technologies or that they realized its limitations once they actually started using it. One can also question whether standardized test scores are the only appropriate measures of student learning.

Although these questions are all important, the fundamental issue lies in understanding the study's limited scope and generalizability. Considering the larger potential of technology to influence learning requires conceptualizing the relationships between technology, the student, the teacher, and the classroom context. Important factors in this broader conceptualization are understanding theories of pedagogy and classroom discourse, the roles played by teachers, and concomitantly, the kind of teacher training or professional development required to integrate technology into classroom practice. In contrast, this study oversimplifies the case by pushing aside these complicated relationships, and treating all the software programs as members of the same generic set of "mathematics software" (or "reading software").

Educational software programs, however, are not monolithic entities. They encompass a diverse range of products with different strengths and weaknesses, each instantiating a different perspective or approach toward learning. The study fails to consider what is unique about the new educational technologies—their diverse and varied potentials, their demands for new forms of participation, and their essential connectivity with life outside the classroom. The research ignores this diversity and instead focuses its attention on what may be considered the lowest common denominator for educational software— tutorial programs.[3]

The one-size-fits-all approach overlooks the fact that different software programs embed in their very design different philosophies of teaching. Tutorial software programs (such as the majority of those selected in this project) embed within them a certain model of learning—what has traditionally been called a transmission model. This approach runs contrary to most of what we know today about effective learning.

Although tutorial software programs may be easier to study, they underrepresent the diverse instructional approaches to learning. For example, no Internet-based approaches or student-directed collaborative and constructivist-oriented technologies that provide different models of learning are represented in this study. For a study published in 2007, such a lapse is surprising.

It is also important to realize that the pedagogies inherent in specific software designs are refracted through the intentions of the teacher, as well as the goals, motivations, and prior experiences of students. Research suggests that a given technology is less likely to be adopted if it deviates too greatly from prevailing values, pedagogical beliefs, and practices of the teachers (Zhao, Pugh, Sheldon, & Byers, 2002). Ignoring these factors will lead to situations in which the pedagogical philosophies of the teacher and the pedagogies

embedded in the software program could be in conflict. The design of this study appears to assume that software choice is the only variable that could affect student learning, revealing a lack of sensitivity to the situational complexity of technology integration.

At issue is why we are still framing the role of technologies within a "transmission model" of learning. Were teachers in this study regarded as secondary to the effectiveness of the selected software programs? How else can we explain why some of the classrooms in the experimental group tended to have more "student" time than teacher-led time. For example, in the algebra treatment classrooms, 88% of the activity time was focused on independent practice, as opposed to 34% in the control classrooms. Why did teachers' roles change so dramatically and why did they all but disappear from the software-based approach? The selection bias toward tutorial programs may have been the cause of this shift.[4]

The puzzling absence of the teacher can be inferred from the report's title, which indicates that the study measured the "effectiveness of … software products" on the "first student cohort." This implies a direct relationship between software programs and student achievement. This research approach overlooks a large body of scholarship on the critical role that teacher beliefs play in how pedagogies and curricula are constructed through the very act of teaching (see Ertmer, 2005, for a good review).

Given that the design did not restrict teachers in the control classrooms from using technology, the study may have produced a paradoxical situation in which teachers in the control classroom may have integrated technology in a more effective manner than their experimental counterparts who replaced their instructional approach with a tutorial program.

Research studies are typically critiqued for lacking ecological validity. In some way, the questions we have asked speak to the possibility that this study may be too ecologically valid. One can argue that the selection of software by committee and its imposition on teachers who receive minimum training provides a picture of real-world practice. This is not an approach, however, that we, as technology researchers and educators, endorse.

By reducing teacher autonomy and overemphasizing the technology through the software programs selected, this study deliberately or inadvertently provided a valuable, un-retouched snapshot of how instructional technology is used in our schools. It provides little guidance, however, for researchers, educators, developers, and policymakers who wish to develop better ways to take advantage of educational software.

Although we have raised many questions about the current study, we do believe that it informs the current debate about the role of technology in schools. In fact, educational technologists and researchers in educational technology should take heart from these findings.

If the results had shown that the use of educational technologies selected for the study improved student scores, tutorial-based programs would have been touted as the way forward. What these findings indicate, not surprisingly, to those of us who deal with these issues, is that the integration of technology in teaching is a complex and wicked problem (Rittel & Webber, 1973). Simple solutions will just not work.

This study should be a rallying cry for the next generation of scholars and researchers to develop newer and better research designs that address some of these questions. We hope the next round of studies will help us develop a better understanding of the complex and contextual interplay between technology, pedagogy, and content (Koehler & Mishra, in press; Mishra & Koehler, 2006).

Notes

[1] Authorship has been presented alphabetically to reflect the equal contribution by all the authors. This paper builds on a nonevaluated course assignment for a doctoral seminar led by Dr. Mishra. The authors would like to thank Glen Bull and Lynn Bell for their support, comments, and feedback.

[2] This prompted one colleague to retort, somewhat facetiously, that a first grader probably spends more time finding his or her crayons over the course of a school year!

[3] Although specific descriptions are lacking, the report indicates that almost all but one of the software programs were tutorial in nature. No rationale is provided for this apparent bias. It should be noted that generating experimental control is much easier with tutorial programs than with more participatory educational software products.

[4] The fact that the classroom using algebra software programs did not significantly differ in scores from the control classrooms, even while spending more time in individual work, raises an intriguing possibility regarding the role of teachers and the nature of the assessments used. The report does not provide enough evidence for us to reach any firm conclusions, but this is clearly an area worthy of further attention.

References

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort. (Publication No. 2007-4005). Retrieved April 27, 2007, from Institute of Education Sciences, U.S. Department of Education, Web site: http://ies.ed.gov/ncee/pdf/20074005.pdf

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration. *Educational Technology, Research and Development, 53*(4), 25-39.

Koehler, M. J., & Mishra, P. (in press). Introducing Technological Pedagogical Content Knowledge. In AACTE Technology & Innovation Committee (Eds.), *The handbook of technological pedagogical content knowledge for teaching and teacher educators.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mishra, P., & Koehler, M.J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teacher knowledge. *Teachers College Record, 108*(6), 1017-1054.

Paley, A. R. (2007). Software's benefits on tests in doubt: Study says tools don't raise scores. *Washington Post.* Retrieved May 15, 2007, from http://www.washingtonpost.com /wp-dyn/content/article/2007/04/04/AR2007040402715.html

Rittel. H., & Webber, M., (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155-169.

Trotter, A. (2007, April 4). Federal study finds no edge for students using technology-based reading and math products. *Education Week, 26.* Retrieved May 23, 2007, from http://www.edweek.org/

ZDnet Editorial. (2007a, April 5). Is education software worth anything? Retrieved May 15, 2007, from http://education.zdnet.com/?p=969

ZDnet Editorial. (2007b, April 11). Another hit on educational software. Retrieved May 15, 2007, from http://education.zdnet.com/?p=987

Zhao, Y., Pugh, K., Sheldon, S., & Byers, J. L. (2002). Conditions for classroom technology innovations. *Teachers College Record, 104*(3), 482-515.