

The Invisible Researcher: Using Educational Technologies as Research Tools for Education

Dwayne E. PARÉ

Psychology, University of Toronto Scarborough
Toronto, Ontario, Canada

and

Steve JOORDENS

Psychology, University of Toronto Scarborough
Toronto, Ontario, Canada

ABSTRACT

As educational technologies become more commonplace, they are often created with the intention of benefiting students through some novel approach, or to fill a perceived educational gap. While these rationales are good ones, it should also be realized that through the use of innovative technologies educators and researchers alike are presented with a unique and powerful opportunity to conduct laboratory-like research in a naturalistic environment. Thus giving the ‘invisible researcher’ the ability to test the desired effectiveness of the tool, and to use the tool as a vehicle to understand learning, all in an unobtrusive manner. This not only ensures that new educational technologies are doing what they were designed to do, but also promises to create pedagogically superior tools and an improved learning environment for both students and educators. To illustrate how this can be successfully implemented, two evidence-based technologies are discussed (the webOption and peerScholar) where research has assisted in tool development and also furthered our understanding of educational theory.

Keywords: Educational Technology, Laboratory Research, webOption, Field Research, peerScholar, Blended Learning

1. INTRODUCTION

As educational institutions embrace technology in hopes of providing new and efficient means of teaching, there is an increasing supply of educational technologies designed to do everything from maximize the efficiency of traditional educational procedures (e.g., electronic document management) to the enhancement of the overall educational experience (e.g., online peer-assessment). It is the design and implementation of these technological tools we wish to focus on here, highlighting the potential that exists for using these tools, not only as powerful educational supplements, but also as

powerful research instruments for better understanding education. With these new technologies educators and researchers have a unique and powerful opportunity to conduct laboratory-like research in a naturalistic environment while remaining unobtrusive – essentially becoming an invisible researcher.

The remainder of this paper will form an argument for why educators, researchers, and technology developers should keep this “tool for basic research” function firmly in mind when designing and implementing their educational technologies. In fact, we believe that the best technologies will intentionally build the capacity for laboratory-like research into their implementation. To begin, we will discuss the roles that theory and research play in educational technologies, followed by details of how the invisibility afforded by these technologies allows for measurement of “real-world” behavior. To support the argument we will highlight the power of using educational technologies as basic research tools in more detail by describing two evidence-based technologies from our work – the webOption and peerScholar.

2. THE ROLE OF THEORY & RESEARCH

When an educational technology is being designed, it naturally should be based on some relevant pedagogical theory [1][2]. Typically though, the theory guides the higher level design (i.e., the purpose of the innovation), but does not necessarily inform all of the lower level decisions (i.e., system and interface configurations), or the specifics of how the learning process will be implemented and supported. Thus, one begins with a tool designed in general to reach some goal, but is implemented via specific features that are often employed because they “seem like good ideas”. For example, an online tool intended to improve reading comprehension might be designed to include pictures with the text because research has shown that pictures-in-text can be beneficial to the learner in technological formats [3]. However, the new tool might also be

designed to display multiple pictures close by the corresponding text even though number of pictures and their proximity to the text has not yet been empirically examined – it just seems to make sense to do it this way.

Therefore, there are two obvious roles that research can serve when implementing a new educational technology. First, it is essential to directly assess whether the tool is reaching the educational goal that it is meant to achieve. That is, is the tool a valid implementation of the educational theory it was informed by. For the reading comprehension example this might be achieved through a series of pre and post tool-use comprehension tests to determine if the educational technology actually improves reading comprehension as intended.

Second, given specific features are sometimes chosen in an intuition-guided manner, it is a good idea to test varied implementations of those features in order to assess the most optimal implementation of the tool. In our picture-in-text example we might want to run studies that vary the number of pictures and their proximity to the text to determine the effectiveness of their inclusion. With this new information the educational technology can be fine-tuned to be optimally effective. In both cases, note that the research goals are focused on confirming the validity and effective use of an educational technology based on existing educational theory.

However, there is also a third important role for research in the context of educational technologies that we wish to focus on here; the role of using the technologies as basic research tools for informing educational theory. That is, these tools can be used to examine learning and educational theories themselves. Going back to the reading comprehension example, one could imagine a scenario where the valid and effective educational tool is now used to examine areas of comprehension theory that have not yet been empirically tested. In fact, it is sometimes the case that testing specific features (e.g., picture number and proximity) might serve the dual purpose of making the tool more effective and informing educational theory.

Details of educational tool designs that allow for the testing of validity, effectiveness, or theory are near impossible to describe in full because of the countless methods and technologies one could use to achieve a goal. Instead, we would like to point out the unique general ability for variable manipulation afforded by educational technologies in a real world context. The specifics of an implementation can be manipulated in a manner typically only possible in experimental laboratories but in a context where behavior has implications for success in a class, and more generally for success in life. Thus, even when research is conducted primarily to assess the validity or effective use of some technological tool, it also has the potential of providing a very powerful context for studying human learning in a mixed-methods way [4] that combines external validity – the extent to which some result holds beyond the laboratory – and internal validity – the extent to which some result is clearly due to the variable manipulated –

all in an invisible manner that is not common in other areas of research.

3. BENEFITS OF BEING ‘INVISIBLE’

The first step in the scientific process is the generation of some hypothesis to be explored and tested. Often these initial hypotheses derive from observational work. That is, behavior is observed in some naturalistic context, and correlations suggestive of potential relations are identified. Once exposed, these potential relations can be investigated more systematically via lab-based studies that directly manipulate variables assumed to play a causal role.

The entire process described above rests on two notions that, 1) when the behavior was originally observed, the act of observing didn’t fundamentally alter the behavior in question, and 2) when the behavior was investigated, the unnatural laboratory setting and research design didn’t bias the experimental results. For example, if an alien creature decided to sit in on a typical university class to see how humans interact, there is a good chance that the mere presence of the alien would strongly alter the way the humans behaved (i.e., the idea of the Hawthorn Effect as originally described by French [5]). In such a case, any follow up work that was based on the assumption that the behavior observed was “natural human behavior” would be fundamentally flawed. Similarity, if the alien creature brought a few humans back to their spaceship in hopes of experimentally validating some hypotheses of human behavior, the mere alienness of the environment and situation could influence the research findings.

The solution to these problems are straightforward; when observing behavior in its natural context, or while investigating some hypothesis, the observer/researcher would prefer to remain as invisible as possible, while maintaining the ability to manipulate controlled variables for empirical testing. The invisible alien who sat quietly in the corner of the lecture would likely receive a very realistic portrayal of natural human behavior in the classroom context, just as would the alien who secretly manipulated a virtual environment to mimic a naturalistic setting. As a result, any subsequent research designed on the basis of these observations would have a higher degree of validity.

Generally there is a trade-off for being invisible though; the more invisible a researcher is the less control they have over the variables intended to affect behavior. A researcher observing video of student interactions is perfectly invisible to the students, but also has absolutely no control to manipulate variables to affect students’ behavior (potentially confounding any interpretations). On the other side of the continuum, a researcher who has students interact in the laboratory has complete control of the manipulated variables, but is also highly visible to all (potentially biasing the findings).

Numerous research designs and experimental settings are used to navigate somewhere in the middle of this continuum in hopes of balancing the costs and

benefits of each location. This is usually done by attempting to bring the field to the laboratory (e.g., [6]) or to bring the laboratory to the field. An example of the latter is when researchers observe participants behind a one-way mirror in a location outside of the laboratory (e.g., in a school). Although, as pointed out by Scott-Miller [7], even when the location of such experiments are not in a lab setting, the unusual looking rooms and the unfamiliar researchers can still lead to various kinds of bias.

Online educational technologies can get around this trade-off and allow researchers to remain invisible while still maintaining laboratory-like control in a naturalistic environment. The fact that these applications have real world consequences (i.e., grades) gives them ecological validity (i.e., external validity) that goes beyond that of what is likely to occur in lab-based studies. To some extent it is the real-world aspect of these applications that allow the experimenter to fade into the background. The behavioral context no longer feels like an experiment. As users become more comfortable with the application context, and more focused on the specific learning goal the application is meant to promote, they will likely begin to respond in more naturalistic ways.

As an example of how educational technologies can be used in this manner, we will now turn to a brief description of two such evidence-based technologies designed and tested at the University of Toronto Scarborough. The first is focused on a blended learning approach (the webOption) and the second is focused on an online peer-assessment system (peerScholar).

4. WEBOPTION: CASE STUDY 1

The first educational technology that we will discuss is one based on blended learning: the webOption. We chose to highlight the webOption because it is a tool that, through invisible research, has both been validated as being effective and valuable in expanding the understanding of online learning. It is a prime example of how an online technology can be utilized as an educational supplement while simultaneously being used as a research tool.

For over five years an increasing number of the courses offered at the University of Toronto Scarborough have been accessible in both traditional and online formats. The webOption process is simple. As professors teach in the traditional manner they are videotaped from the back of the class (with a wireless microphone providing direct audio), and the video and audio are immediately converted to a digital format and uploaded to the web (see Bassili & Joordens [8] for detailed information). Students then have a choice; they can attend lectures in a traditional manner, they can watch the lectures online, or they can do both.

Note that the webOption was implemented with the principal goal of enhancing student satisfaction by providing flexibility in terms of how and when students watched the lectures. That is, students are clearly not

satisfied when scheduling issues prevent them from taking a course they would like, so the purpose of implementing the webOption was to overcome this issue, thereby allowing students to have an educational experience that more closely matched what they hoped for. Initial research confirmed that students were indeed satisfied with the webOption, and that they liked the manner in which specific features were set up [8].

Critical to the current paper, our research did not stop there. Bassili [9] showed that attitudes concerning whether students liked the option of having online lectures were predicted by motivational orientations, whereas the actual choice to attend lectures or watch them online was related to students' cognitive strategies. In addition, Bassili [10] showed that students' perceptions of media richness predicted their tendencies to attend class or watch online, and that students were especially likely to attend classes when they perceived the content to be difficult. Thus, the educational tool is now telling us more about why students might make the various choices they make in a blended learning context.

Perhaps more relevant to the interaction between education and technology, we also examined the degree to which students took control over the lecture pace in the online environment and how that was related to performance in the course. The video player afforded students the option to pause lectures for any reason. Our previous research showed that those students who paused more often performed better in the context of our Introductory Psychology class [8]. However, when the same behavior was examined in the context of our Introductory Calculus courses, the opposite result was found. Those students who paused more frequently performed more poorly. Our subsequent attempts to understand this dissociation (i.e., [11]) revealed that pausing, in an online environment, appears to be indicative of students utilizing a shallow approach to learning [12]. A shallow approach can work in an introductory level "content course" such as Introductory Psychology, but it does not work in an introductory level "skills-based course" such as calculus.

The central point of all this is the following. Had we stopped doing research once we had ascertained that the tool was effective, and that its features seemed well designed, we would have lost an opportunity to learn more about educational issues relevant to online learning. By continuing the research we know more about why students behave the way they do in blended learning contexts, and we have become knowledgeable about how certain features of the online environment can be beneficial or not depending on the course context.

5. PEERSCHOLAR: CASE STUDY 2

The second educational technology that we will discuss is an online tool based on peer-assessment: peerScholar. We chose to highlight peerScholar because like the webOption, it too has been validated and used to extend our current knowledge of learning through research. It is also key to note that the system is one

designed with invisible research explicitly in mind. Various implementations of peerScholar can be run simultaneously in a single class in order to control and manipulate specific variables. Also, students can be randomly assigned to implementations in cases where such assignment is necessary. Although not described in detail, the studies outlined below make use of these features (please refer to corresponding studies for more details).

Initially peerScholar was developed as a way to bring writing assignments back to a large class where all writing had been abandoned in favor of the more efficient and economically feasible multiple-choice assessment. Our major concern was that multiple-choice testing encouraged knowledge-acquisition but did not allow for the knowledge-use promoted by writing, especially critical writing. Therefore the use of peerScholar at the University of Toronto Scarborough had the principal goal of supporting critical thinking and clear written communication. And though the system is capable of supporting many different forms of assignments (e.g., multimedia-targeted peer-assessment), the focus of our research has thus far been primarily on writing.

At its core, peerScholar is based on peer-assessment where students grade the work of fellow students as they learn together how to think critically and write clearly. The entire learning process occurs online, with no time-of-day or location restrictions. An entire peerScholar assignment consists of three phases: a Composition Phase (students write a reaction paper to a one-sided argument), an Assessment Phase (students assess 5 or 6 of their peers' reaction papers), and a Feedback Phase (students get a mark based on the average of their peer-given grades). Each phase leads naturally into the next, with the subsequent phase opening when the previous is complete. Indeed, one of the pedagogical strengths of peerScholar is the fact that students remain focused on the assignment throughout a short period of time, receiving feedback on their own work while everything is still very fresh and in memory.

The peer-assessment aspect is quite different from the traditional method of having an expert judging the quality of the composition. Thus our initial research project within peerScholar was to validate the grades produced by peer graders. Research that directly compared grades acquired within peerScholar to those provided by a more typical expert marker [13]. Specifically, while students were grading each other's work, we also had a pair of expert graders (graduate level teaching assistants) each grade a subset of 120 compositions. This allowed us to compare three indices of agreement in marks and ask: (a) to what extent do two expert markers agree in the marks they give, (b) to what extent do two undergraduate markers agree in the marks they give, and critically (c) to what extent does the average undergraduate mark – the average of all 5 or 6 students who graded a particular composition - agree with the mark given by an expert marker?

Across a number of studies we have found the agreement between experts to provide a Pearson

correlation of about 0.50 – 0.55. This level of agreement is consistent with expert agreement levels observed outside of the peerScholar context [14] [15]. Undergraduates' agreement with each other tended to fall more in the 0.20 to 0.30 range. However, when you average at least 5 undergraduate marks, that average mark agrees with the mark provided by experts at the 0.45 – 0.55 level. That is, it does not differ significantly from the agreement between two experts. The averaging reduces the effects of noise, making the averaged undergraduate mark as valid as the mark provided by a single expert [13].

Validity is clearly a critical thing to establish because, without it, there would be very little willingness on the part of students or professors to accept this new form of assessment. The results of our research suggest that peerScholar provides an efficient way of having written assignments in any lecture context while still producing valid grades. And the researcher remained invisible to the students throughout the process, thereby surmounting problems that other studies are often faced with.

However, peerScholar, and peer-assessment more generally, is not just a way to reinstitute written assignments but, rather, it represents a step forward in how we teach thinking and writing skills. Many positive attributes are associated with seeing and evaluating the work of peers. These include students thinking about their own work more deeply after seeing the work of peers [16], students being exposed to the real world of scientific discourse [17][18], and students learning how to provide and interpret feedback [19]. As well, the timely nature of the feedback can also enhance learning [20].

To ensure peerScholar was effective in reaching its primary educational goal (i.e., utilizing peer-assessment to assist students to better judge the quality of their own work) we empirically tested the pedagogical power of peer-assessment using peerScholar [21]. Students were asked to grade their own written piece as they were submitting it during the Composition Phase. They were then asked to grade their composition again after completing their assessments in the Assessment Phase. Relative to their final mark on the assignment, their self-assessed marks became significantly more accurate after just one exposure to the peer-assessment process.

After finding that peerScholar produced valid grades, and was achieving its intended educational goal, our next step was to look closer at the number of markers used in a peer-assessment assignment. Our intent was twofold; 1) to figure out the right number of peers to achieve an accurate average grade (fine-tuning the system specifics) and, the more educationally relevant question 2) to find out if grading workload offsets the statistical advantages of having more peer assessors [22].

As students are asked to mark more assignments, each composition ends up being graded by more peers, and the average should contain less noise with all else being equal. However, if students are asked

to mark too many compositions, it was hypothesized that their attention and effort may wane and more noise might enter into the marks. Thus, one might imagine a sweet spot reflecting the trade-off between statistical properties and human performance. By randomly assigning students to grade varying numbers of peer assignments (balanced over time) we were able to show that validity measures are highest, sometimes even higher than the agreement between experts, when each student is asked to grade 6 compositions, meaning their composition is in turn graded by 6 peers [22]. Therefore we were able to fine-tune the system while concurrently learning more about mark accuracy and workload in peer-assessment.

Once again, the main point here is had we stopped using peerScholar as a research tool once we discovered the system was effective, we would have missed the opportunity to examine the trade-offs between mark accuracy and workload. In fact, we have found peerScholar so useful in empirical research that we continue to examine many facets of peer-assessment that to-date have had little or no empirical backing (e.g., [23]).

6. CONCLUSIONS

In this paper we have discussed two case studies where educational technologies were successfully used as research tools by “invisible researchers”. It is our hope that we have convinced the reader of the opportunities afforded to them through the use of educational technologies in ways that are not always apparent. This not only ensures that new educational technologies are doing what they were designed to do, as well as creating pedagogically superior tools and an improved learning environment, it also allows for empirical testing of pedagogical and educational theories in a manner that promises to help us better understand the processes of learning.

7. REFERENCES

- [1] R. E. Ferdig, “Assessing technologies for teaching and learning: understanding the importance of technological pedagogical content knowledge”, **British Journal of Educational Technology**, Vol. 37, No 4, 2006, pp. 749-760.
- [2] G. Salomon, “On the nature of pedagogic computer tools: the case of the writing partner”, In S. Lajoie & S. J. Derry (Eds), **Computers as cognitive tools**, pp. 179–196. Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.
- [3] R. N. Carney and J. R. Levin, “Pictorial Illustrations Still Improve Students' Learning from Text”, **Educational Psychology Review**, Vol. 14, No. 1, 2002, pp. 5-26.
- [4] K. Muthu, “Mixed methodology research design in educational technology”, **Alberta Journal of Educational Research**, Vol. 53, No. 1, 2007, 34-44.
- [5] J. R. P. French, “Experiments in field settings”, In L. Festinger, & D. Katz (Eds.), **Research Methods in the Behavioral Sciences**, pp. 98–135. New York: Holt, Rinehart and Winston, 1953.
- [6] I. Straaten, R. C. M. E. Engels, C. Finkenauer, and R. W. Holland, “Sex differences in short-term mate preferences and behavioral mimicry: A semi-naturalistic experiment”, **Archives of Sexual Behavior**, Vol. 37, No. 6, 2008, pp. 902-911.
- [7] S. A. Miller, **Developmental Research Methods (3rd Ed.)**. Englewood Cliffs, NJ: Prentice-Hall, 2007.
- [8] J. N. Bassili and S. Joordens, “Media player tool use, satisfaction with online lectures and examination performance”, **Journal of Distance Education**, Vol. 22, No. 2, 2008, pp. 93-108.
- [9] J. N. Bassili, “Motivation and cognitive strategies in the choice to attend lectures or watch them online”, **Journal of Distance Education**, Vol. 22, No. 3, 2008a, pp. 129-148.
- [10] J. N. Bassili, “Media richness and social norms in the choice to attend lectures or to watch them online”, **Journal of Educational Multimedia and Hypermedia**, Vol. 17, No. 2, 2008b, 453-475.
- [11] S. Joordens, A. Le, R. Grinnell, and S. Chrysostomou, “Eating your lectures and having them too: Is online lecture availability especially helpful in ‘Skills-Based’ courses?”, To appear in the **Proceedings of the 4th International Conference on E-Learning**, 2009, Toronto, Canada.
- [12] J. Biggs, D. Kember, and D. Y. P. Leung, “The revised two-factor study process questionnaire: R-SPQ-2F”, **British Journal of Educational Psychology**, Vol. 71, 2001, pp. 133-149.
- [13] D. E. Paré, & S. Joordens, “Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool”, **Journal of Computer Assisted Learning**, Vol. 24, No. 6, 2008, pp. 526-540.
- [14] H. Blok, “Estimating the reliability, validity, and invalidity of essay ratings”, **Journal of Educational Measurement**, Vol. 22, No. 1, 1985, pp. 41-52.

- [15] R. Miller, "Mark my words, part 1: Teachers", **South African Journal of Higher Education**, Vol. 10, No. 2, 1996, pp. 3-24.
- [16] G. K. Sims, "Student peer review in the classroom: a teaching and grading tool", **Journal of Agronomic Education (JAE)**, Vol. 18, No. 2, 1989, pp. 105-108.
- [17] M. H. Towns, K. Marden, D. Sauder, R. Stout, G. Long, M. Waxman, M. Kahlow, and T. Zielinki, "Interinstitutional peer review on the internet: crossing boundaries electronically in a student-refereed assignment", **Journal of College Science Teaching**, Vol. 30, 2001, pp. 256-260.
- [18] A. Venables and R. Summit, "Enhancing scientific essay writing using peer assessment", **Innovations in Education and Teaching International**, Vol. 40, No. 3, 2003, pp. 281-290.
- [19] J. Heywood, **Assessment in Higher Education**, John Wiley & Sons, London, 1989.
- [20] S. Bloxham and A. West, "Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment", **Assessment & Assessment in Higher Education**, Vol. 29, No. 6, 2004, pp. 721-733.
- [21] D. E. Paré and S. Joordens, "To thine own self be true: self-assessment in the context of peer-assessment", Paper presented at the **Society for Teaching and Learning in Higher Education 28th Annual Meeting**, 2008, Windsor, Ontario.
- [22] D. E. Paré, "Tired of marking?: using peerScholar to explore the change in peer grading reliability as a function of increased number of peer evaluations", Paper presented at the **Society for Computers in Psychology 37th Annual Meeting**, 2007, Long Beach, California.
- [23] D. E. Paré, A. Velji, and S. Joordens, "Mark the marker: using statistical techniques to judge marking quality in peer-assessment" [In preparation].