The Inevitability of Epic Fail: An Investigation of Implementation Problems Associated with Technology-Rich Research Innovations

Stephen T. Slota, Michael F. Young, Roger Travis, University of Connecticut
stephen.slota@gmail.com, Michael.f.young@uconn.edu, roger.travis@uconn.edu

Abstract: Reviewing large-scale attempts to introduce new technology-based approaches to instruction yields many instances where researchers have used contemporary learning-science research and theory to design innovative, technology-based curricula that show substantial benefits to important instructional variables. Crucially, however, these results arise only when the project originators have close and direct involvement with teacher training and implementation. With evidence for limited long-term success, both educational psychologists and practicing educators must ask: "Why do technology-rich research innovations fail once researchers are no longer directly involved?" The authors aim to address this question in a way helpful to game-based learning researchers by presenting an example of failed implementation of a learning game paired with follow-up research on other instructional tools and methodologies. We propose three reasons why such innovations fail: 1) Fatal Mutation Due to Assimilation; 2) Loss of Fidelity; and 3) Failure to Thrive. Recommendations for addressing these issues follow.

The Inevitability of Epic Fail

Looking back on large-scale attempts to introduce new approaches to instruction, it is not uncommon to identify instances where contemporary learning science research and theory have been instantiated into innovative, technology-based curricula with empirical research findings that evidence substantial benefits to important instructional variables (e.g., student-side variables – test achievement, self-regulation, course completion, motivation, engagement – or teacher-side variables – content knowledge, wait time, professional development, collaboration). Crucially, however, these results only arise when the project originators are closely involved through teacher training and implementation. Once they pull back, the original instructional motivations, objectives, and underlying theories are often lost. Notable among such "new" approaches are Logo (Papert, 1980), *The Adventures of Jasper Woodbury* (Cognition and Technology Group at Vanderbilt [CTGV], 1992), and Apple's Hypercard. Many projects in the realm of game-based learning have experienced a similar fate, including a variety of educational video games (Honey & Hilton, 2011; Young et al., 2012), games for non-academic instruction (e.g., Aronowsky, Sanzenbacher, Thompson, Villanosa, & Drew, 2012; Sylvan, Larsen, Asbell-Clark, & Edwards, 2012), and scholastic tabletop roleplaying games (e.g., Slota, Travis, & Ballestrini, 2012).

With such limited long-term success, both educational psychologists and practicing educators must ask: "Why do technology-rich research innovations fail once project originators are no longer directly involved?" We address this question through a worked example of failed implementation paired with follow-up analysis on other instructional tools and methodologies. More specifically, we propose three reasons why technology-rich educational research innovations fail: 1) Fatal Mutation Due to Assimilation; 2) Loss of Fidelity; and 3) Failure to Thrive. Recommendations for addressing these issues follow.

What Game-Based Learning Researchers Can Learn From History

Technology integration efforts intended to dramatically change American education have a rather checkered implementation history. Indeed, there is still debate over the extent to which even simple technologies like calculators and word processors have meaningfully affected school classrooms, nevermind more advanced tools such as geospatial mapping software, programming languages, 3D avatar-based virtual worlds, scientific modeling environments, and online video series used to "Flip" classrooms. The rise and subsequent fall of these technologies has provided researchers with several notable examples that exhibit how no amount of time, effort, and money has been able to sustain long-term change. The following serve as indicators of how the complexity of innovative technology implementation can override both good ideas and intentions:

Logo

Based on Papert's exploration of constructionism-based learning environments, Logo was developed in 1967 as a programming language that made LISP-like artificial intelligence programming accessible to elementary education students through graphical representations of a small, moving turtle. It received widespread research and acclaim by teachers at the time of its implementation, but rather than acknowledging a new paradigm for student learning, school environments chose to assimilate Logo into their tried-and-true method of education: direct instruction. Once the developers and researchers left the school environment, classroom teachers defaulted to teaching *about* Logo in place of teaching *with* Logo. As a result, students were ultimately led to memorize Logo programming commands (e.g., FD50 RT90) from teacher-generated worksheets. After peaking in the mid-1980s – largely due to the development of the Apple II computer – most of Logo's 197 compilers/interpreters fell out of educational use. After peaking in the mid-1980s – largely due to the development of the Apple II computer – most of Logo's 197 compilers/interpreters fell out of educational use despite numerous research studies showing the benefits of students learning with and about Logo the way Papert envisioned. This led Papert (1993) to apply the Piagetian notion of assimilation to the entirety of American schools and conclude that large scale school reform was largely impossible (Papert, 1997).

Jasper

Vanderbilt University's Jasper Woodbury videodisc series was a professionally filmed video production designed to support middle school mathematics and problem solving. Drawing on contemporary learning theories of situated cognition and anchored instruction, these videos were meant to provide an authentic learning context prior to and in conjunction with other instruction. However, fearing that students were unprepared to handle the series' outward complexity, many teachers chose to educate their classes using the stories as post-instruction word problems rather than a context for inquiry-driven learning. Once the program's researchers ceased their direct interaction with participating educators, the videos became mainly capstone activities, relieving them of any potential they once had for providing an authentic context for learning. The migration of videodisc to DVD coupled with rapid antiquation of the series' content (e.g., the price of gas as part of the calculations) further sealed the program's fate. By the late 1990s, Jasper was almost universally shelved.

HyperCard

HyperCard functioned as the framework from which several other learning technology innovations were based. Teachers and subject matter experts collaborated to create classroom tutorials, including dribble file data for assessments and interactive controls for videodisc players. However, rapid technological advancement during the 1990s rendered HyperCard obsolete, inadvertently categorizing all related instructional innovation as outdated and irrelevant. Work derived from thousands of HyperCard programming hours was lost or abandoned once Apple stopped supporting the system in favor of HTML and Java, and with no simple migration path, HyperCard vanished from the classrooms its creators hoped it would revolutionize. The moral: technology comes and goes so quickly that technology-specific innovative pedagogy may be doomed to fail or eventually to fade.

Operation BIOME

In April 2012, approximately 1,500 educators and administrators were contacted with an offer to utilize a year-long, alternate-reality game (ARG)/roleplaying game (RPG) biology program, *Operation BIOME*, beginning in the Fall 2012 semester. Each teacher received a school-specific email that described the nature of the project, incentives for participating, and other pertinent participation information. Of those who were contacted, 17 expressed interest, 10 responded to additional inquiries regarding their participation, and five followed through with training and implementation. Six more teachers opted into the program after witnessing their colleagues utilizing it at the start of the school year. By the end of August 2012, 11 teachers agreed to begin the program with 600+ students spread across 29 class sections in five suburban and urban schools.

Unlike projects that had simply "gamified" science, *Operation BIOME*'s design began with the objectives/standards and used them as a guide for developing the mechanics and narrative rather than the other way around – a design scheme reflective of the top-down approach often associated with strong curriculum development (Bergmann & Sams, 2012). This placed emphasis on the game's ruleset (i.e., how play happens) in order to bring students closer to doing the things real world scientists do: problem solve, think critically, examine existing literature, generate new questions, and, most importantly, collaborate toward realistic, shared goals (e.g., "cure cancer"). Additionally, because the mechanics and narrative followed the same trajectory as state and national standards, the story missions transparently aligned with the information students needed to successfully complete high stakes tests. Put another way, the mechanics and narrative were designed to carry much of the weight usually given to direct instruction, allowing teachers to use the exploratory prompts as an introduction to the class content and class-time for reflection on and discussion of the concepts and skills students discovered outside the classroom (i.e., scaffolding both successes and 'productive failures' in problem solving, critical thinking, etc.).

In an effort to highlight these instructional goals, all participating teachers were extended the opportunity to share in one or more of five one-hour summer training sessions hosted by the researcher via Google Hangout. Training sessions focused on the underlying rationale for *Operation BIOME*, the use of Edmodo, and the integration of *Operation BIOME* with other web-based tools (e.g., Google Documents, the *Operation BIOME* website). The content of each session was cut into a series of short video tutorials hosted on YouTube in order for teachers to review the material again at their leisure. Bi-weekly support sessions were planned for the Fall 2012 in order to bring together *Operation BIOME* teachers with those involved with similar ARG/RPG programs for Latin and Greek language courses. Troubleshooting for *Operation BIOME*, Edmodo, and all related technologies remained available throughout the Fall 2012 semester.

Fueled with a combination of teacher enthusiasm and administrative interest in increasing student achievement, program rollout began in the early portion of September. However, the project encountered a rapid slowdown after the first three weeks due to unanticipated competition for attention with a variety of other district initiatives (e.g., a school-wide science summer reading program, new safety policies, abrupt changes to the biology curriculum). *Operation BIOME* website use dropped dramatically through the middle of October, and by the time December arrived, all 11 participating teachers had either deviated from the designed approach or stopped utilizing the program altogether. Data collection was temporarily halted while the initial research questions were re-evaluated to determine whether or not they could be reasonably addressed with the available information.

A series of short face-to-face discussions revealed that a number of participants had modified the program without researcher approval (e.g., combining units, skipping sections, altering prompts), ultimately resulting in substantial treatment disparity between the experimental and comparison groups. While the teachers noted the helpfulness of on-going researcher support, none were able to commit to biweekly, online follow-up meetings and made clear that altering their typical instruction to fit the “new” model introduced through *Operation BIOME* was not possible given the other time and resource demands being made by their schools. Additionally, they expressed frustration with tracking online student responses and several commented on the relative efficiency of direct instruction over student facilitation.

Understanding Trajectories of Failure

To help game-based learning researchers escape potential failure, we propose that there are three major trends common among the technology-rich educational research innovations described above: 1) Fatal Mutation Due to Assimilation; 2) Loss of Fidelity; and 3) Failure to Thrive.

1) Fatal Mutation Due to Assimilation

Fatal Mutation Due to Assimilation refers to *teacher-generated changes that are fundamentally in opposition to the theory behind the program.* In the case of Logo, constructivist theory dictated that students needed to discover the program language in context, including having younger students know information that older students did not. Once schools deviated from this approach and began teaching the program language with worksheets, students were taught to parrot back decontextualized commands before being allowed to interact with the computer. For Papert (1980), this was akin to Piaget’s assimilation but at the school rather than individual level: new ideas were forced into existing schemata (e.g., direct instruction) rather than being applied toward the reformation of an already-existing paradigm.

2) Loss of Fidelity

For the purposes of this discussion, we characterize Loss of Fidelity as *participants doing what designers and researchers intended, but not enough, adding materials that are antithetical to project objectives, and/or watering down required activities to a point where they become ineffective.* In other words, Loss of Fidelity occurs when there is a schism between teacher intention and instructional reality. Dusenbury et al. (2003) explored this precise issue in the context of drug abuse prevention programs, identifying five major measures of fidelity: Dosage, Adherence, Program Differentiation, Participant Responsiveness, and Quality of Program Delivery. We believe this framework can encompass technology-based and game-based classroom interventions equally well.

Dosage includes agreed participation in a daily intervention program but, rather than implementing the set treatment each day, only following through with it once per week or month. This may also manifest as a timing issue wherein dosage chronology is miscalculated or deliberately modified (e.g., in the case of medication, consuming the prescribed drug without food or prior to sleep). Teachers often presented the first episode of *The Adventures of Jasper Woodbury* (i.e., *Journey to Cedar Creek*) following direct instruction about distance, rate, and time. While the video was utilized during the appropriate unit, those who placed the implementation too late along the program trajectory broke their prescribed dosage chronology.

Adherence and Program Differentiation refer to the addition of instructional practices or pedagogies that make a unique program more like a particular pre-existing program. For example, while the researcher may wish to implement a purely constructivist program, a participating classroom teacher might choose to add *Classdojo*™ or a similarly designed behaviorism-based program. Such an eclectic approach has the potential to enhance the intervention but also makes innovation less distinguishable from existing programs and may ultimately make it harder to measure the intervention's effects, thereby preventing the researcher from assessing any uniquely added benefit.

Quality of Delivery refers to how well teachers understand the theoretical foundations of a given innovation and dynamically interact with learners in a manner consistent with the design principles. Teaching "in the cracks" (i.e., in a live classroom where interactions cannot be scripted) requires implementers to "fill" non-program activities and discussion with information and responses that are consistent with designer's theoretical framework. This bears a direct relationship to Participant Responsiveness – the way instructional interventions are received by the target audience (i.e., both teachers and students). Because instruction is intended to induce particular learning experiences and interactions, miscommunication or ineffectual implementation may lead the audience to miss the intervention's situated value. When elements like Dosage (e.g., how much Jasper or Logo instruction is needed before measurable changes in math achievement can be expected) conflict with school scheduling or administrative initiatives, Quality of Delivery and Participant Responsiveness tend to suffer dramatic setbacks.

3) Failure to Thrive

The third trajectory, Failure to Thrive, represents a pattern wherein *lack of researcher feedback (i.e., how the program implementation is working) causes instructors to gradually shift away from program goals, theories, and procedures present at the time of initialization.* In part, this appears to involve situations where participating educators "do it for the researcher(s)" or for the status of being part of the research team, or the resources involved in a grant project. Once the project originators leave, the teachers simply move on or revert to prior instructional practices. However, prior research indicates that the problem may be much more complicated than this.

In describing a situated view of naval quartermasters, Hutchins (1995) addressed how success arises from interaction both among and between people and artifacts in the world. From this perspective, failure can occur when any one of these potential interactions is interrupted: teacher-tool, researcher/designer-tool, and teacher-researcher. Each interaction must be functioning and ongoing to provide the feedback necessary for sustaining technology-rich programs over time. While teachers often crave interactions with talented adults and welcome the opportunity to share their insights, debate with researchers, reflect on and explain their own pedagogy, and receive critiques of their teaching from academic peers, the social and cognitive factors arising from broken interaction can obscure progress toward a common objective. Barron (2003) explained this as smart groups capable of generating workable solutions ignoring those solutions as a result of structural social dynamics. Following this line of logic, any interruption of teacher-researcher interaction, intended or not, may allow misunderstanding, lack of investment, social conflict, or other social dynamics to overshadow the project's original goals. These issues eventually consume program implementation and push participating teachers back into their respective comfort zones.

Avoiding the Precipice of Epic Fail

With the advent of the No Child Left Behind Act, sweeping instructional change has become even more difficult to achieve than when many new, innovative instructional projects began during the 1980s and 90s. School devotion to improving test scores has come largely at the expense of innovation, and direct instruction in the form of test preparation has served almost exclusively as the means to contend with the ever-growing gauntlet of standardized testing. This requires designers to devise game-based learning frameworks that meet parent, teacher, district, and researcher needs while simultaneously avoiding the constraints that suffocated Logo, Jasper, HyperCard, and, more recently, *Operation BIOME*.

Because no two school environments are identical, some customization should be expected at every implementation site. However, the authors believe this customization can be accomplished in a way that maintains program fidelity and allows project originators to avoid the three predominant trajectories of failure described herein. Avoiding the precipice of epic fail requires the clear articulation of which program elements – including content and method of implementation – can be entirely altered, changed within a set margin, or not changed at all. This means anticipating which program elements might challenge traditional school instruction and planning ahead to avoid any disruption of the innovation's theoretical integrity. At times, we have referred to this as the "fixin's bar" approach where designers select an array of possible additions or enhancements that could be applied to locally customize the program without ruining the "flavor" of the meat of the innovation.

Fatal Mutation Due to Assimilation can be addressed with planned customization and clear designation of critical components. Project originators should anticipate that each new site will seek to customize the intervention to meet the unique characteristics of the specific context. For game-based learning designers and researchers, specifically, this may mean tailoring the game to fit a pre-existing curriculum rather than attempting to replace all materials that already exist. A 1:1 learning and game objective relationship can ensure overlap between state and national standards (e.g., Young et al., 2012) and decrease the likelihood that participating teachers will simply assimilate the game into direct instruction or other pre-planned activities.

As indicated by *Operation BIOME*, Loss of Fidelity can become a considerable problem when project originators attempt to control and quantitatively evaluate program implementation in a living school environment. Variables tracked through teacher-guided discussion, answers to follow-up questions, and how well each teacher connects new content to prior content can improve the likelihood of measuring the extent to which fidelity has suffered. Similarly, regular follow-up, including audio/video, on-going training, face-to-face focus groups, surveying, student feedback, and other qualitative tools provide a glimpse at teacher understanding and whether or not participants understand how to successfully implement the program *in situ*. All new innovations should be designed with a companion text that explicitly reveals the presumed theoretical framework, going so far as to frame the basis, history, and practical implications of the underlying theory.

Failure to Thrive could be combated through the creation of a self-sustaining, dynamic community of practice (e.g., Lave & Wenger, 1991) that exists alongside the original innovation much like the websites and other metagame materials designed to support many massively multiplayer online games. Any such community must be able to change as needed within the parameters laid out by the project originators and the underlying theory. While this might include the creation of a webpage, forum, YouTube videos, wiki, and/or series of regular meetings, continued success will only come from ongoing facilitation by leading experts (i.e., project originators and trained practitioner-specialists). All teachers hoping to become community practitioner-specialists should be capable of describing the program's underlying theory and show evidence of their ability to adapt the theory to fit within the scope of a living classroom environment. When possible, original practitioners and other expert researchers should return to the community for two-way dialogue concerning information regarding progress in the field, modifications to the theory, and related research projects. Additionally, the application of cost-sharing user fees may increase school buy-in, providing impetus to remain involved with the innovation and assist with the burden of community development. Though teacher-tool and researcher/designer-tool interactions will likely continue regardless of community formation, teacher-researcher communication is the only element that will sustain program fidelity beyond the original scope of the project.

Any game-based learning research model aimed at long-term viability will require researchers capable of on-going school-level involvement, random fidelity checks, and two-way monitoring of the innovation. Project originators should be mindful of ways in which they can maintain an active community role and use on-going educational changes at federal, state, and local levels for guidance on shifting K12 pedagogy toward more meaningfully authentic student learning experiences. While the authors' goals may be lofty, past precedent indicates that even slight breakdowns in program framing and teacher-researcher dialogue can be incredibly detrimental to project longevity. The history of other large-scale attempts to introduce new approaches to instruction have given the game-based learning field a rare opportunity to turn repeated failure into a powerful alternative to the current state of K12 education. In the wise words of *Diablo*'s Cain, let us "stay a while and listen" that we might find the path to educational reformation scattered among the ashes of our predecessors' epic fails.

References

Aronowsky, A., Sanzenbacher, B., Thompson, J., Villanosa, K., & Drew, J. (2012). When simple is not best: Issues that arose using why reef in the conservation connection digital learning program. *GLS 8.0 Conference Proceedings*, 24-29.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307-359.

Bergmann, J. & Sams, A. (2012). Flipping the classroom. *Tech & Learning*, 32(10), 42.

Cognition and Technology Group at Vanderbilt (1992). Technology and the design of generative learning environments. In T.M. Duffy & D. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation*. Hillsdale NJ: Lawrence Erlbaum Associates.

Dusenbury, D., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237-256.

Honey M. A. and Hilton, M. (2011). Learning science through computer games and simulations Committee on Science Learning: Computer Games, Simulations, and Education, National Research Council.

Hutchins, E. (1995). *Cognition In the Wild*. Cambridge, MA: MIT Press.

Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.

Papert, S. (1993). *The Children's Machine*. New York, NY: Basic Books.

Papert, S. (1997). Why school reform is impossible. Review of Tyack and Cuban (1995). Accessed May 31, 2013 from http://www.papert.org/articles/school_reform.html

Slota, S. T., Travis, R., & Ballestrini, K. (2012). Operation BIOME: The design of a situated, social constructivist ARG/RPG for biology education. *GLS 8.0 Conference Proceedings*, 261-267.

Sylvan, E., Larsen, J., Asbell-Clark, J., & Edwards T. (2012). The canary's not dead, it's just resting: The productive failure of a science-based augmented-reality game. *GLS 8.0 Conference Proceedings*, 30-37.

Young, M., Slota, S., Cutter, A., Jalette, G., Lai, B., Mullin, G., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: a review of trends in video gaming for education. *Review of Educational Research*, 82(1), 61-89. doi: 10.3102/0034654312436980