
38.

Mentira

The Death and Life of an Augmented Reality Curriculum

Christopher L. Holden (University of New Mexico), Julie M. Sykes (University of Oregon), & Steven L. Thorne (Portland State University & University of Groningen)

Abstract

In 2009, Sykes and Holden began working on *Mentira*, a design-based research (DBR) project and game-based curriculum. Years later, *Mentira* is gone. In many ways, we would say the project has failed. Consideration of the details illuminates general truths regarding the potential of DBR to result in innovation at scale. Success and failure are not as simple as we typically understand them. Metrics and mechanisms common to researchers describe only a small facet of projects' lives and deaths, missing most of what looks to be important to sustenance and growth. Without major attention given to broader goals and practices, we overlook likely ways to go from idiosyncratic experiment to meaningful impact. We look at why *Mentira* died and what lives on in its place, with advice for other practitioners and scholars of educational technology.

Introduction – Mentira and its goals

Mentira was created in 2009 as an augmented reality (AR) game-based curriculum for a fourth-semester Spanish course at The University of New Mexico. Sykes and Holden put it together using an internally funded \$10K research grant. It was the first game designed with ARIS, an AR design platform, to be used in a classroom, and a first in many other ways too: the first language learning AR game, first AR curriculum in higher education, and the first AR game to have both field-based, homework, and in-class software components (Holden & Sykes, 2011). A feature of this, and many other DBR projects, is to take on not only the messiness of an existing educational context (Brown, 1992), but also the use of new technologies and alternative pedagogies by participants in those spaces. Consequently, projects like these naturally entertain a wide, perhaps overwhelming, swath of research questions relating to some of the most basic features of game based learning, AR, and mobile technology. This is enough already to encumber a single, small project, but the work also hinges on further issues long debated in relation to educational research generally and about the potential and reality of DBR especially. Generally, “How can educational research proceed and be effective once we agree to a certain level of irreducible complexity in the situation at hand?” More specifically,

- Methods: How can actual evidence about any of the above questions be sought and verified?
- Analysis: Given evidence about outcomes, how can we trace it back to responsible factors?
- Scale: What might this one case imply about any other? How can an intervention scale?

Between our inspiration to make something interesting, and possibly innovative, and all the things we might expect from DBR, there is a lot of unsure ground: using iterative prototypes to simultaneously improve theory and conditions on the ground (Collins, 1992), reconciling issues of epistemology and methodology (Schoenfeld, 1994), creating generalizable knowledge (Barab & Squire, 2004; Bielaczyc, 2006), and doing all of this in a way that is capable of being scaled (Fishman et al., 2004) and sustained (Cole, 2007). The general, though usually unstated, model is: Hope for one of these one-off designs to be proven effective at accomplishing some educational task, where elements of design and implementation can be distilled from the complexity of the original context, both so that the original intervention can be scaled to other similar situations (with *Mentira* we might imagine other Spanish courses at a similar educational level). Then, apply the demonstrated general insights, hopefully replicated through additional studies, published in academic literature reapplied by designers in other circumstances and referenced by policy makers in an effort to drive adoption and other attention system-wide. Understandably, this is a large effort, and any small team can only be expected to do so much. There is a fair amount of skepticism as to whether small scale projects like these can lead to larger scale change. The first part of our *Mentira* story, detailed below, demonstrates ways in which these skeptics may be right.

What Happened? The Failures of *Mentira*

Mentira was, over the course of four years, piloted, implemented in a restricted way, revised, implemented full scale in all sections of the fourth-semester Spanish course it was designed to support, revised again, run again, and then discontinued. Our playthroughs were promising (Holden & Sykes, 2011), but eventually *Mentira* faded away. A full accounting for the project would take too long, but it's worth looking at a couple details to help illustrate a central question: "What might it mean for an educational game DBR project to be successful?" There does seem to be an overall skeptical attitude towards research on "failed projects", but there is a need to better understand and enunciate what divides success and failure. This paper is an attempt to explore these ideas, and simultaneously, make salient the the challenges, and the impending hope, that stems from *Mentira* and many similar projects.

Let us then begin with a simple criterion: can the game actually be played and under what circumstances? As for *Mentira*, it has not been used for 3 years and has since become unplayable through obsolescence. How might an interested party implement, change, test, adopt, or scale *Mentira* if they can't play it? This criterion seems obvious, yet we (education researchers) probably spend too much time debating the merits of games never to be played again—either broken, obsolete, or inaccessible—maybe partly because grants are awarded at the beginning and the fodder for future work are publications and not active user data. There are of course real questions about how many users is worthwhile, *how long* software needs or can be expected to be viable, and what "accessible" means, but these factors seem to receive relatively little attention. We actually did well with *Mentira*, bringing it through many major changes in iOS hardware and software, as well as a major revision of the underlying development platform, ARIS (itself a small-scale DBR project), without further funding. *Mentira* was always free and easily accessible, just as ARIS itself is open source and easy to use. Nevertheless, *Mentira* is now dead. In our particular case too, chances of an outsider adopting the game in the future is even smaller, as the game is strongly set in both a specific level, dialect, and culture of Spanish language, as well in the particular place in which the game is set. Nonlocal parties could emulate the game's mechanics but never truly play it. Not only is our game now obsolete, but, in despite design decisions that would have allowed for non-local play, the game has little potential for being directly scaled.

We also ask whether a program of research can be built around the game, yielding general theoretical insights and proven empirical results. Overall, our experimental design with *Mentira* was open-ended, limiting these contributions to academia through its design and use. We started out with some novel ideas for what data to collect from the play of an AR game (Holden & Sykes, 2011), but like many DBR projects, we collected copious data and only analyzed a limited subset of it. As Anderson & Shattuck worried in their meta-study of DBR projects (2012, p. 24), our data is not accessible to other researchers, and is so contextually bound so as to be nearly useless. When it comes to the traditionally most valuable data, A-B tests of alternate treatments according to short term measures of narrowly identified learning outcomes, we had even less. Worse, we wasted time pretending our exploratory work in poorly known areas (and areas where short term gains on specific learning objectives have only spurious connections to long term educational goals—achievement in a college language course is not a good predictor of lifetime second language use) would yield such data. The high rhetorical value of such data, and hence its high appeal despite its inappropriateness, paralyzed us to some extent. “Is this better than that?” is a good question to answer with an A-B test, but it was not what we were working towards with *Mentira*. We would have gotten more done had we the courage of our convictions, and simply left the A-B tests for later.

Another implicit metric for the success of DBR, perhaps the most important, is iteration. *Mentira* was iterated, and through its iterations, the game and our process and knowledge all improved greatly. But here too we stalled. Important findings never made it back into the game. For example:

1. We combined the heaviest reading with the field trip portion of the game. Both are intense activities and should have been separated for better play and more learning.
2. We wanted students and teachers to co-author *Mentira*. But the teaching assistants were reticent to play the game themselves, limiting their ability to contribute. Involving students in the design took too much time away from the existing course material, and was not something all were prepared or eager for. Instead of hoping all would dig deep into the design, we should have recruited simpler participation generally while retaining openness for deeper involvement.
3. Introducing a game-based learning model within an existing curriculum is not simple from a cultural framing standpoint.

Most students did not play through *Mentira* in the same way people learn to play videogames (Gee, 2003). We wanted them to see a game, but sometimes, they treated it more like a textbook. We saw few replays and instances of getting help *reading* from other players, dictionaries, or in-game resources despite difficult vocabulary and a complex story. Though it was a class and students expected homework, play did emerge in places. We could have capitalized on these more. With these examples and more, responding effectively to clear observations whose certain identification did not require detailed methodology, was beyond us. We failed when we could no longer improve *Mentira*.

What Went Wrong? Roots of Failure

So here we are, with a dead game, a mixed record of use, plenty of unfinished ideas, and a lot of raw “data”. Sounds like a pretty convincing failure. How did we get these things wrong? One big reason is that there was simply too much to do and too few of the right people to do it. People are a critical piece of projects like *Mentira*. In fact, the proximal reason the game is unplayable today is that Sykes moved to another university and was no longer in a position to directly implement *Mentira*. When the expertise

and roles of a very small number of people together cover the many facets of development, use, and research, each of those people is very valuable and hard to replace. For instance, our stalled iteration in the game's story happened when we lacked a motivated graduate student fluent in the appropriate dialect of Spanish. But professors move, students graduate, and teams evolve. We often hear that academics, personality-wise, might not make the best game designers and project managers, but, simultaneously, the structural factors of employment in higher education do much to limit potential collaborations across traditional divides. Even with enough of the right people around, representing multiple stakeholder groups working ideally together as co-researchers on a project like *Mentira*, that work must compete for attention with other more urgent, mandatory activities. Not only is everyone busy with other things, but incentives among stakeholders are misaligned. A successful game is of little value to assistant professor Principal Investigators (PIs) whose time is limited in terms of design and publication. Making games with less money rather than more is similarly skewed in valuation. And what do teaching assistants or part-time instructors care of the experimental model and its data if it tanks their evaluations or makes them uncomfortable? What are the rewards for their fully involved participation on their part?

We had trouble deeply involving students and teachers in certain parts of *Mentira* (there were some real positives too, but this account is intentionally about our difficulties), highlighting the expectations often placed on co-investigators in DBR (Bielaczyc, 2006), and underlies revisions of DBR itself (e.g. Design Based Implementation Research, DBIR, as described in Fishman et al., 2014). Tradition enshrines the idea that everyone involved gets to be more than a subject of the research. Interventions are co-constructed between all participants, not just PIs. But these other stakeholders may have justifiably little to no interest in working to develop the innovation PIs have in mind. Teachers and students play roles whose scripts emphasize enacting rather than reinventing a curriculum. Co-designing a mobile game as we asked students and teachers to do, or even leaving campus for an hour or two to play a mobile game, falls outside those purviews. Alternative pedagogies or new materials, especially if they run contrary to the cultural beliefs of the teachers or students, may not look like progress (e.g. New Math). The inability to overcome these sorts of mismatch as a matter of course or at scale is another source of broad skepticism regarding educational change generally (Tyack & Cuban, 1995). Reinventing education would seem to require the participation of all stakeholders and requires of each a willingness to step outside their defined role within the system. Expecting exception to be the norm is unrealistic. Our shortcomings in *Mentira* involved many proximal factors, but underestimating the power of structural realities is central.

To Build or Not To Build

First and foremost in our minds is a question we have often asked ourselves and our colleagues: "To build or not to build?" That is, is DBR worth the time and trouble? Have we altered, with our example, the impression that games developed by researchers soon end up unplayable and inaccessible? Place-based mobile games, often designed for a local context, only seem to exacerbate the notorious problem of scaling DBR. Our initial team has been scattered to the wind. The empirical results obtained from the experiment were minimal. Moreover, these problems seem to reveal not just individual shortcomings, but the persistence and perniciousness of well-known structural challenges to educational innovation. On the whole, what here could support a future scholar who chooses to use a similar approach?

Now, here is where the story continues. In reality, despite the failures around *Mentira* we describe above, we could not be more optimistic regarding the potential and promise of DBR, especially of place-based mobile game design, to reinvigorate and intensely contextualize diverse educational experiences.

But how can these perspectives be so far apart? The short answer is that so far, we have only seen a narrow point of view: the standard view of the practice, role, and effects of educational research. In fact, we learned a lot about making and implementing AR games for which there was no help in academic literature, and none of this found its way into academic literature either. Even within Games Learning and Society (GLS), an uncommonly open research community, length and format make it difficult to adequately tell deep stories with more than a few words, and a very few pictures. The stories we do tell focus on only our work, where references to others serve as proof of scholarship more than a lived and growing connection. We hide our failures and strain to see successes in the expected form. Our hope is that *Mentira* is a different kind of story.

First, like all who hope to learn anything from video games, we try to remember that failure is productive. It may be that the things we failed most at are themselves not only what will improve going forward, but also the sources of “generalizable knowledge”. Some of the lessons we might try to recognize and adapt to are the structural features that limit innovation: The double-edged nature of small, multifunctional teams; the misalignment of incentives common in education; the siren’s call of naive empiricism; and the difficulty in attaining sustainability in itinerant conditions. None of these are absolute impediments. Even if DBR had no hope of producing scalable interventions, their diagnostic function—to identify root obstructions to innovation—should be reason enough to keep up with the practice. Failure can thus give us insight into what we can do better next time and why what we’re trying to do is difficult in the first place. Recognizing the role of failure is only a small part of our general optimism. In the course of working on *Mentira*, we have witnessed the potential and power of multiple forms of communication, usually across informal channels, to supersede the limitations of official mechanisms to facilitate educational innovation. Place-based mobile game design has grown with us as we worked on this game; researchers and other creators are now in a much better place to begin similar work themselves. As we worked on *Mentira*, we visited colleagues, published websites, ran workshops, staffed email lists to give technical and design assistance, wrote documentation and tutorials for technique, logistics, and design, got involved in the development and outreach of ARIS itself, consulted on new projects, blogged, and worked with other small teams hoping to do better than we did. Our colleagues shared the idea of the *Mentira* project with their students, and encouraged them to take up DBR and place-based game design themselves. We supported those students, colleagues, and others, as they got their hands dirty. Pretty soon, the creators of *Mentira* and the design team of ARIS were a small part of something much bigger: a growing community of explorers from diverse stations and interests: researchers, teachers, museum directors, and technology coordinators. Instead of seeing the goal of *Mentira* being an effective intervention in Spanish 202 at UNM, with a potential to be used in other similar courses, we can look at its impact in helping to create this much larger space. Just as earlier AR experiments (Squire et al., 2007) encouraged us to make our own game in a new area, with the help of many more investigators, we together are finally able to begin directly mapping out some of the total accessible landscape available to this medium: subject areas, educational contexts, locations, instructional paradigms, audiences, design intent, and research methodologies. We would like to share a few examples of what has begun to grow from *Mentira*.

What Grew from *Mentira*’s Death

The biggest effect *Mentira* had was to be the wind in ARIS’ sails. By setting an example for an interesting design that cut to the heart of second language learning pedagogy, by becoming deeply involved with the ARIS project itself, and by seeking to recruit others into the development and use of

place-based mobile games, we were able to provide awareness, motivation, and support to a variety of audiences. Because ARIS centers on storytelling and was easy to use, people who heard about *Mentira* but were not working in the niche of Spanish language learning and higher education in Albuquerque, were able to nonetheless try it themselves. They did not need money, deep technological expertise, or a large, experienced group to get something off the ground. They could be inspired by the designs and programs of other ARIS users, as well as benefit from their experience solving logistical problems. Thousands of people of all ages, backgrounds, and interests gave ARIS a try and became involved in DBR projects of their own. And because of their diverse interests and stations, the breadth of ground covered by the sum of these researches is orders of magnitude beyond what even a large, well-funded research team could ever dream. There has been a proliferation of use cases, design intents, and methodologies, beginning to meaningfully address the myriad possible implications of making and using games to study and improve learning (Holden et al. 2015; Dijkers, Coulter, & Martin, 2012). We succeeded here by joining in this more general effort, at the cost to our own program. The reception of place-based mobile games has been especially strong among those who study language learning. There was already a growing interest in connecting games and language learning, but also the common frustration that empirical work was badly needed but looked to be too expensive and labor intensive. With new means to collaborate across great distances and with new colleagues, this could change. For instance, Sykes leaving UNM was bad for *Mentira*, but a net positive for the cause: she currently directs the Center for Applied Second Language Studies (CASLS) at the University of Oregon. Her team is developing an Android port of ARIS, an open database of place-based language learning experiences (pebll.uoregon.edu), and new games and other DBR projects that iterate on *Mentira*'s mix of local place, culture, language, storytelling, and game design, developing new, deeper content and integrations than would have been possible before. *Ecopod* is a learning game that extends beyond the confines of a single class into college students' residential experiences, and the *Bridging Project* uses place-based game design to establish real connections between intermediate learners of a target language with native speaker communities. Most recently, these efforts are being extended to use augmented reality as a tool to provide relevant, meaningful, and useful language learning for refugee populations. Through CASLS, Sykes can do far more to support better learning than any careful continuation of *Mentira* could have allowed.

Mentira has also helped to recruit scholars, like our third author, to DBR. Steve Thorne is well known for his scholarship on language learning, some of which considers games, and had shared the common skepticism around the academic design of place-based games. But *Mentira* and ARIS, especially as communicated through informal channels, have changed this for him. With John Hellermann, he co-directs the 503 Design Collective at Portland State University, a group of faculty and students who design AR experiences for language learning. Their ARIS game *ChronoOps* is a multilingual introduction to sustainability features of the campus and is the centerpiece of an ethnomethodologically informed research design (Thorne et al., 2015). With a far simpler game than *Mentira*, and multimodal data collection and analysis, this research investigates how semiotic potential is made meaningful and actionable via talk-in-interaction and embodied deixis. His closer connection to place-based mobile game design has spread even further to new people and new areas within the domain of language learning, illustrating how network effects can work towards scale. For instance, he works with Sabine Siekmann and her graduate students at the University of Alaska Fairbanks, looking for new ways that new technologies might further support native language learning in the region. Natalie Cowley, a master's student in this program and a teacher in Kasigluk, AK made a game to connect young people there to the traditional ways of life known by the elders of their community. She takes the level of

community-language-culture integration hinted at in *Mentira* to a much deeper level, actually involving community members in the creation of the game.

We begin to see a wealth of variation as well as convergent evolutions—similar successful traits of diverse origins—in the details of this work. For example, another project concerned with Indigenous language revitalization, Partnerships for Indigenous Knowledge and Digital Literacies (PIKDL) led by Jon Reinhardt and Susan Penfield, also involves community members as game authors as a key part of their process in moving forward collaborations with academics in the development of new place-based mobile games. This not only mirrors Cowley’s design but also the thinking of researchers in other areas like Jim Mathews (2010), who described making mobile games as a way to investigate local issues for Wisconsin high schoolers, as well as many who develop and use mobile game engines like ARIS and Taleblazer as vehicles for student design (Klopfer & Sheldon, 2010; Holden et al. 2015). Variations in individual designs are just as instructive: Bernadette Perry’s *Explorez* is a French game which, while mechanically similar to *Mentira*, uses AR in an inverse sense and takes a different approach to embedding the cultural aspects of language within the game (Perry, 2015). Another French game, Terri Nelson’s *Paris Occupé*, ditches ambulatory exploration altogether, using an ARIS mechanic derived from the design of *Mentira* to work both as field trip and homework, to carry out the entire game virtually in France. The shared practices and stories of this new cohort—built mostly through informal interactions—inform practice and theory far beyond any additional attention we could have given to publishing results from *Mentira*.

Onward to Failure

One-off projects are frequently dismissed among researchers as being incapable of effecting large scale, long term change in education. As Anderson & Shattuck write, DBR’s impact has been limited to “small-scale interventions and in the lives of individual teachers and schools” (2012, p. 24). Founding literature and realities of participants in educational settings seem to explain the consistent difficulty in innovating through DBR. Yet this perspective only accounts for individual interventions and is too narrow in other ways too. Too often, we only see the means of communication and collaboration that take place through articles, conferences, and results obtained through a limited range of methodologies. The story of *Mentira* gives us a glimpse of what we’re missing, both impact beyond academics and how true collaboration in practice means that success and sustainability can be seen on a larger scale than a single intervention. To really impact education, researchers can, and need to, mobilize informal modes of communication. This means much more than tweeting or the adoption of a particular format. It means sharing your struggle more broadly and productively with others who might join in themselves. By sharing the struggle, and directly encouraging and supporting others, a one-off project can do better than scale, it can inspire a movement. Identifying similar examples to *Mentira*, Martin et al. (2014) gives this a name: participatory scaling. Grass roots, bottom up educational change has been argued for in many guises in every era (Papert, 1993). In the case of *Mentira*, we see that extensive sharing and communication allowed a failed individual effort to spark a distributed and growing movement that explores the developmental potential of DBR-informed interventions that highlight agency, collaboration, and learning in places.

References

- Anderson, T., & Shattuck, J. (2012). Design-based research a decade of progress in education research. *Educational researcher*, 41(1), 16-25.
- Cole, Michael. "Sustaining model systems of educational activity: Designing for the long haul." *Children's learning in and out of school: Essays in honor of Ann Brown* (2007): 71-89.
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The journal of the learning sciences*, 13(1), 1-14.
- Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *The Journal of the Learning Sciences*, 15(3), 301-329.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences*, 2(2), 141-178.
- Collins, A. (1992). *Toward a design science of education* (pp. 15-22). Springer Berlin Heidelberg.
- Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National Society for the Study of Education*, 112(2), 136-156.
- Fishman, B., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *JLS*, 13(1), 43-76.
- Holden C., Dikkers S., Martin J., Litts B. (Eds) (2015). *Mobile media learning: Inspiration and innovation*. ETC Press, Pittsburgh
- Holden, C. (2014). The Local Games Lab ABQ: Homegrown Augmented Reality. *TechTrends*, 58(1), 42-48.
- Holden, C. L., & Sykes, J. M. (2011). Leveraging mobile games for place-based language learning. *International Journal of Game-Based Learning (IJGBL)*, 1(2), 1-18.
- Klopfer, E., & Squire, K. (2004, June). Getting your socks wet: Augmented reality environmental science. In *Proceedings of the 6th ICLS* (pp. 614-614). International Society of the Learning Sciences.
- Klopfer, E., & Sheldon, J. (2010). Augmenting your own reality: Student authoring of science-based augmented reality games. *New directions for youth development*, 2010(128), 85-94.
- Martin, J., Dikkers, S., Squire, K., & Gagnon, D. (2014). Participatory scaling through augmented reality learning through local games. *TechTrends*, 58(1), 35-41.
- Mathews, J. (2010). Using a studio-based pedagogy to engage students in the design of mobile-based media. *English Teaching*, 9(1), 87.
- Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. Basic books.

Perry, B. (2015). Gamifying French Language Learning: A Case Study Examining a Quest-based, Augmented Reality Mobile Learning-tool. *Procedia-Social and Behavioral Sciences*, 174, 2308-2315.

Schoenfeld, A. H. (1994). A Discourse on Methods. *Journal for Research in Mathematics Education*, 25(6), 697–710. <http://doi.org/10.2307/749581>

Squire, K., Jan, M., Matthews, J., Wagler, M., Martin, J., DeVane, B., & Holden, C. (2007). Wherever you go, there you are: Place-based augmented reality games for learning. *The Educational Design and Use of Simulation Computer Games*. Rotterdam, The Netherlands: Sense Publishers.

Thorne, S., Hellerman, J., Jones, A., Lester, D. (2015). Interactional practices and artifact orientation in mobile augmented reality game play. *PsychNology Journal*. 13(2-3), 259-286.

Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia*. Harvard University Press.