

Using Design-Based Research in Higher Education Innovation

Cristi Ford, Darragh McNally and Kate Ford
University of Maryland University College

Abstract

This paper discusses the design-based research approach used by the Center for Innovation in Learning and Student Success (CILSS) at the University of Maryland, University College (UMUC). CILSS is a laboratory for conducting applied research that focuses on continuous improvements to the university's instruction of curriculum, learning models, and student support to identify promising innovations for underserved populations in adult higher education; to drive adoption of next-generation transformational online learning; to develop new educational models based on learning science, cutting edge technology, and improved instructional methods; to help more UMUC adult students succeed by increasing retention and graduating more students in shorter time frames (thus reducing their costs). As such, leveraging technology and pedagogy in innovative ways is key to the Center's work. CILSS serves as the research and development arm for the university, promoting innovative ideas and breakthroughs in learning.

In this paper, we detail one interpretation of design-based research (DBR) and how it can be applied by an innovation center working within a university for program evaluation. We also posit that the conceptual framework and assumptions of andragogy (Knowles, 1984) have applicable relevance to the instructional shifts that include adaptive learning in the curriculum. A review of the literature on DBR explores the central features of this approach. A review of andragogy as the conceptual framework for this paper highlights what we believe to be the central features of the evaluation approach of adaptive learning software. We then present the model used by CILSS when designing and testing a pilot project. To illustrate the approach, we provide the example of a recent pilot that uses the adaptive learning software RealizeIt in UMUC's Principles of Accounting I course, a course that traditionally has lower than average success rates.

Keywords: Design Based Research; Innovation; Continuous Improvement; Adaptive Learning; Evaluation; Accounting; Adult Learners

Ford, C., McNally, D. & Ford, K. (2017). Using design-based research in higher education innovation. *Online Learning, 21*(3), 50-67. doi: 10.24059/olj.v%vi%i.1232

Introduction

DBR is not so much a precise research methodology as it is a collaborative approach that engages both researchers and practitioners in the iterative process of systematically analyzing, designing, and evaluating educational innovations and interventions aimed at solving complex, real-world educational problems. Whereas traditional educational research methods are aimed at examining what works (i.e., efficacy), often *in a controlled laboratory setting*, DBR is concerned with understanding and documenting how and why the designed intervention or innovation works *in practice* (Anderson & Shattuck, 2012; The Design-Based Research Collective, 2003; Nieveen & Folmer, 2013; Plomp, 2013).

Literature Review

Central Features of DBR

DBR is frequently described in the literature as being pragmatic, interventionist, and collaborative. Similar to action research, DBR involves problem identification, assessment, and analysis in an applied educational setting, along with the implementation and evaluation of some type of change or intervention to address a problem. Although action research and DBR are grounded by theoretical and empirical evidence, they also privilege practical evidence, knowledge, and solutions (Anderson & Shattuck, 2012; Lewis, 2015; McKenney & Reeves, 2014; Plomp, 2013). Where these two methods typically diverge is around the premium placed on collaboration. Both Anderson and Shattuck's and Plomp's work have asserted that action research is typically performed by teaching professionals with the goal of improving their own practice rather than as a collaborative partnership with a research and design team.

Starting with the initial assessment of the problem and the specific context in which it occurs and continuing throughout the iterative design, implementation, and evaluation process, DBR relies on the collaboration of a multidisciplinary team comprised of researchers, practitioners, subject matter experts, designers, and others, including administrators, trainers, or technologists, whose expertise may be crucial to the project (McKenney & Reeves, 2014). DBR also draws from multiple theories to inform design, as illustrated by the Carnegie Foundation's Pathway program. The design was informed by theories related to student engagement, mathematics learning, learning strategies, and non-cognitive learning factors, including perseverance and growth mindsets (Russell, Jackson, Krumm, & Frank, 2013).

DBR is an iterative approach involving multiple cycles of design and in-situ testing of the design. The knowledge generated during each phase of the DBR process is used to refine the design and implementation of the intervention, which is why DBR is also considered adaptive (Anderson & Shattuck, 2012; McKenney & Reeves, 2014). This differentiates DBR from other types of educational research (Bannan, 2013), which typically involve a single cycle of data collection and analysis focused on producing knowledge.

Implementing and evaluating a high-fidelity representation of an intervention in-situ can involve a substantial commitment of funding, time, and resources. Effective planning and the use of low-fidelity rapid prototyping during the early stages of a DBR project enable the team to test their assumptions and quickly reject bad designs or to modify the design prior to implementation or summative evaluation of the intervention's effectiveness (Easterday, Rees Lewis, & Gerber, 2014).

For practitioners, administrators, and policymakers, the contextual relevance of the intervention is just as important as the methodological rigor and efficacy (Fishman, Penuel, Allen, Cheng, & Sabelli, 2013). DBR integrates design research, evaluation research, and validation research. Consequently, a variety of quantitative and qualitative research methods and design techniques are required to develop, test, and refine an intervention while generating knowledge and design principles that address the relationship between teaching, learning, and context variables (Anderson & Shattuck, 2012; Bannan, 2013; Reimann, 2016).

Challenges Associated with DBR

It is beneficial to first consider and classify the object of research to determine whether DBR is the right approach. For example, Kelly (2013) indicated that design research may not be cost-effective for simple or closed problems. DBR may be more effective in cases in which previous solutions or interventions failed or the specifics of the problem require assessment, clarification, and solution design. According to Kelly, DBR is indicated when one or more of the following conditions are present:

- When the content knowledge to be learned is new or [is] being discovered even by the experts.
- When how to teach the content is unclear: pedagogical content knowledge is poor.
- When the instructional materials are poor or not available.
- When the teachers' knowledge and skills are unsatisfactory.
- When the educational researchers' knowledge of the content and instructional strategies or instructional materials are poor.
- When complex societal, policy or political factors may negatively affect progress (p. 138).

DBR entails multiple cycles of design and implementation refinements that can span multiple semesters or even years, during which collaborative partnerships, resources, and funding may be constrained or overtaken by competing priorities (Anderson & Shattuck, 2012; The Design-Based Research Collective, 2003). DBR considers not just design efficacy but also the conditions that impact the effectiveness of implementation in practice. Yet without a plan for actively managing project scope during these iterations, the DBR team runs the risk of gold plating an intervention to account for every possible permutation in the implementation environment or pursuing additional incremental improvements that exceed the purpose, goals, and requirements of the project. Criteria must be established to guide decision-making about whether or when to abandon, adapt, or expand a design (Dede, 2004). Generally, CILSS abandons a pilot project when outcomes appear harmful to students, for example, by harming learning outcomes or grades. An iteration with mixed results is usually not cause to abandon the project; rather, it is an opportunity to refine and repeat the iteration before moving on to the next stage of the pilot.

At UMUC we have created our own process flow and iteration process. CILSS generally plans on three to five iterations, beginning with one section and scaling up to a full randomized control trial (RCT) with all sections in a given term. CILSS uses a multi-method approach, including interviews, focus groups, surveys, and analytics. Ultimately, any research project culminates with randomized control trial, testing the effect of an intervention that has been developed over several iterations.

Addressing Implementation at Scale

Implementation at scale requires greater consideration of the extent to which the intervention may interact or conflict with other variables in the learning environment, including existing policies, curriculum, assessment methods, and instructor willingness and ability to implement the intervention or change (Lewis, 2015). Interventions that worked in controlled settings or on a small scale have often failed as they are scaled up, due to variations and adaptations at the system and classroom levels (Fishman, Penuel, Allen, Cheng, & Sabelli, 2013; Penuel, Fishman, Cheng, & Sabelli, 2011). These issues can be addressed by DBR. As an extension of

DBR, Design-Based Implementation Research (DBIR) is focused on building organizational or system capacity for implementing, scaling, and sustaining educational innovations. DBIR's research focus extends to the identification and design of organizational routines and processes that support collaborative design and productive adaptation of core design principles across settings (Fishman et al., 2013; Penuel et al., 2011).

Conceptual Framework: Andragogy

Andragogy encompasses a set of core assumptions about adult learners intended to inform the design and delivery of adult education (Knowles, Holton, & Swanson, 2014). These assumptions should be viewed along a pedagogical-andragogical continuum to the extent that an adult learner may differ from a child learner. According to McAuliffe, Hargreaves, Winter, and Chadwick (2009), andragogical learning design draws from theories of transaction, which focus on the context-dependent and pragmatic needs of learners.

Andragogy is a learner-centric process model. Underlying andragogy's process model is a competency model associated with a level of performance. The competency model is designed to reflect the values and learning expectations of the learner, faculty, the institution and society. An adult learner originating from previous learning environments that emphasized passive, teacher-centric learning approaches will likely require additional real-time help in developing his or her ability to engage effectively as self-directed learners (Blondy, 2007; Cercone, 2008; Knowles, 1973; Merriam, 2001).

At UMUC, performing learner and contextual analyses based on andragogical assumptions help inform the development of these competency models and the corresponding instructional design and planning at a macro-level. However, DBR is concerned with addressing persistent problems of practice. Therefore, we must also consider the variances course instructors may encounter in each learner's self-directedness, preparedness, and motivation. Knowles recognized, both conceptually and practically, that an adaptive, flexible approach was needed to address the variability of individual adult learner needs and behavior across learning situations and contexts (Holton et al., 2001). Through diagnostic experiences, self-assessment, and the immediacy and accuracy of feedback, self-directed adult learners can also monitor their own learning and development against the underlying competency model (Knowles, 1996).

Nonetheless, online asynchronous learning platforms present a challenge in terms of the lag between the revelation of an individual difference or need related to our andragogical assumptions and the individual instructor's ability to adapt the learning process or provide help or guidance in real-time—at the teachable moment. Given Holton et al.'s (2001) assertion that the primary focus of andragogy is on *how* rather than *why* adult learning transactions occur, it is reasonable for administrators, designers, and instructors to question the extent to which embedded andragogy design considerations can be executed reliably in practice at the micro-level of the individual learner and to work collaboratively to develop solutions that support both the instructor and the learner.

Researchers have indicated that learning is improved when we can personalize the learning and adapt for the student's ability by identifying problem areas and addressing them immediately (Murray & Pérez, 2015). While our DBR process is undergirded by andragogy assumptions and principles, our adaptive learning design recognizes individual adult learners' differences at the

learning transaction level to facilitate learning and provide help or guidance when mistakes are made. Among the andragogy process elements specified by Knowles that are reflected in technology-enabled adaptive learning design are diagnosis of learning needs, development of objectives, or more specifically a learning pathway comprised of content and learning activities oriented to learners' specific needs, and the evaluation of learning through the re-diagnosis of learners' needs.

Adaptive Learning at UMUC

RealizeIt is an adaptive learning software that provides the availability for many learning paths to a final destination—the interaction of which alters the educational environment from a fixed setting to a flexible (adaptive) context. The core elements of adaptive learning include incremental learning; an opportunity for continual feedback for learners given regular assessment, benchmarking, indexing growth; and offers potential advantages over current online learning pedagogical approaches. RealizeIt assumes that students are not forced to learn at the average speed of the class; rather, each student can take the time individually needed to learn. This means that completion can be accomplished in a shorter time for some, while extended time to fill in gaps of learning for others will be needed. Although adaptive courseware has been successful in other institutional contexts, it was imperative for adaptive learning to be tested with the UMUC student population.

The CILSS DBR Model

The Problem Statement and Research Design

Courses with high enrollment and low success rates (or lower than average success rates) are referred to as Obstacle Courses at UMUC. While the success rates for many of these courses are in fact higher than the national average, the university would, nonetheless, like to see these success rates improve. High enrollment and low success is common nationally for courses, such as Introduction to Accounting, which are required by more than one major but in which students struggle. Implementing RealizeIt was proposed as a possible way to ameliorate the low success rates in several courses. Adaptive courseware has been shown to allow students in an online environment to have their needs assessed individually with data about their abilities being collected in real time. To test whether this is the case, a piloting process that would take place over several terms was designed. This process drew on DBR research to design and iteratively improve courseware for UMUC's Principles of Accounting I to test the effectiveness of this platform (the specific adaptive learning software, with content designed and embedded by UMUC) on course outcomes in the online environment.

The Team

While CILSS is a research and innovation center, implementing a pilot requires multiple stakeholders to work together—both researchers and practitioners. As UMUC's classes are, for the most part, taught partly or wholly online, the Learning Design and Solutions department (LD&S) is a vital part of any team that aims to test the effectiveness of courseware. UMUC's LD&S is made up of cutting edge designers who are fully engaged in bringing innovation to bear on issues in higher education.

The collegiate faculty is fully involved in any piloting within their programs. The accounting department was an essential part of the RealizeIt pilot team. This was especially the case because the existing Principles of Accounting I course needed to be mapped into the RealizeIt system. As well as collegiate faculty, several subject matter experts (SMEs) were also required to validate the mapping of the course and to ensure that the existing syllabus, readings, and other class materials were embedded in RealizeIt as well as possible. It was essential that the process of embedding the course in RealizeIt was done well to ensure that the pilot was testing the effectiveness of the courseware and not held back by issues with material being improperly embedded.

The Iterations

To ensure that students are not harmed by a pilot that does not benefit them and that pilots do not fail in a way that causes harm to the students or the university, several iterations of a pilot are planned in advance. At UMUC there are four separate sessions in an academic term. In the case of using RealizeIt, this meant that the platform was used initially in one course for the entirety of the eight-week session. This allowed the LD&S team and SMEs to test the prototype they created on a smaller unit of analysis and to test how well it worked, highlight any issues, and decide what could be done better in the future. After this had been accomplished, the pilot was expanded to encompass several sections in a semester. Again, problems and challenges were noted so that the platform and any supports could be improved for the next term. Next, the platform was used for several sections of a term, using different instructors. Finally, the platform was used as part of a randomized control trial, in which students (and instructors) were randomly assigned to either a treatment group (a section using the RealizeIt system) or a control group (a section using the traditional platform). Several methods were used to determine what advantages, if any, RealizeIt gave to students.

Scaling Up and Knowing When to Stop

One criticism that has been made of DBR is that because the research process is iterative, it is not clear which iteration is the final iteration (Dede, 2004). Iterations can potentially carry on forever. This may be the case in some settings; however, the final iteration is built into the original research design, and the iterations culminate with a RCT and full intensive evaluation.

The problem of interventions that work well in controlled settings but not when scaled up has received much attention in the education literature (e.g., Duffy & Kirkley 2004; Sternberg et al., 2011). CILSS took several steps to increase the likelihood that results found in the pilot would also be found in the real world. One such step was randomly assigning instructors to teach using the RealizeIt platform. Most instructors had not used RealizeIt before. Although more favorable results may have been more likely using instructors who volunteered to teach using RealizeIt, this would be stacking the deck in favor of positive evaluation results. Instructors who volunteer to teach using RealizeIt may be more comfortable with and enthusiastic about the software than the average instructor, resulting in selection bias.

In keeping with Brunswik's (1956) theory of representative design, we recognize that it is the average instructor who will have to use RealizeIt if it is fully scaled up within the university, and so the results of the pilot evaluation must reflect this. This again highlights the importance of the collegiate faculty being fully invested members of the pilot team: UMUC collegiate faculty appreciate the importance of well-researched innovations and so are as interested as the researchers in representative and robust results.

As mentioned, the number of iterations is built into the research design from the beginning. Generally, three iterations are required, with the third iteration being a large-scale RCT. The first iteration is usually carried out by a faculty member who is invested in the innovation and may be part of a session/term for one section or the entire session/term for one section. The second iteration addresses any issues uncovered in the first. This iteration is for several sections and lasts the duration of the term/session. The second iteration uses several different instructors for a plurality of viewpoints on how well the intervention works. The third iteration again addresses any issues uncovered during the second and is a full scale RCT in which half the sections in a given term are randomly assigned to treatment (in effect, randomly assigning the instructors also). This allows CILSS to statistically analyze the effect of the intervention on course outcomes and student satisfaction and perceptions.

One or two iterations may be added at any point in the cycle. For example, if the first iteration goes poorly for a reason that can be identified, it may be best to repeat this iteration rather than move to the second stage. If the results of the RCT are mixed or not significant, it is necessary to repeat this iteration before deciding whether to scale up the pilot.

The Evaluation

Although data are collected and analyzed while the pilot is ongoing, the final iteration of the pilot is the most intensive regarding data collection. As final grades alone are often a poor measure of success, data are gathered on student interaction with the platform, student quiz and exam grades, student discussion posts (qualitative and quantitative), and student impressions and experience with the platform. As the final iteration of the pilot is a randomized control trial, the same data are collected for both the treatment and control groups.

ACCT 220

Principles of Accounting I (ACCT 220) is required for several majors at UMUC, including Business and Finance. Like many introductory courses nationally, it typically has a high enrollment and a lower rate of success. UMUC uses data analytics to monitor the performance of such courses that can be obstacles for students. Adaptive learning software has shown to be promising in similar contexts, increasing success rates by creating individual learning paths for students.

ACCT 220 went through four complete iterations of the RealizeIt system (three planned iterations and one supplemental iteration). The first iteration was in Spring 2016. RealizeIt software was used for one pilot section in a fully online section for all eight weeks, the entire length of the course with selected faculty who were engaged in building the pilot. The instructor in this first iteration was not randomly assigned. She was a faculty member who was a member of the pilot team. In future iterations, the instructors would be randomized to better judge how the project would perform at scale.

Iteration 1 Results

Results were analyzed to test whether RealizeIt had an effect on course success rates and final grades. Data from UMUC's data warehouse allowed us to control for variables that might have an effect on outcomes of interest, such as age, cumulative credits, and course success rate. The analysis showed a significant positive effect on course success rates (the likelihood of a student achieving a final grade of C or higher). The analysis also showed a significant positive effect on grade. The average grade was 2.6 for students in the control sections and 3.0 for those in

the RealizeIt sections. An Ordinal Least Squares (OLS) regression model (which controlled for demographics and student historical academic performance) estimated that being in a RealizeIt section increased final grades by .55 points on average and holding all else equal ($p=.02$). This means that a student in a control section with a C+ (2.3) would be expected to have a grade of B- (2.7) in a RealizeIt section. Of course, given the small sample size ($n=55$), these results were promising but not definitive.

Interviews with the instructor and feedback from students indicated a number of areas that could be improved. It was evident with the first iteration of the course that the adaptive treatment sections needed to be recalibrated with the amount of technical support required. We also identified instances in which the RealizeIt system was not appropriately displaying figures or calculations. The time calculated on the nodes was automatically set at 20 minutes per node—feedback from students highlighted this as a point of frustration as the nodes rarely required only 20 minutes. It was evident that we needed to address the technical issues and reset the predicted times in order to set a realistic expectation for students.

Iteration 2

The second iteration of ACCT 220 was in Summer 2016. This time, three pilot sections were used. Again, the sections were fully online, and the RealizeIt system was used for all eight weeks. Sections were randomly chosen and students were given the option to opt out and be enrolled in a traditional online classroom. The three pilot sections and three control sections resulted in a sample size of 169 students, 82 of whom used RealizeIt. Instructors were assigned to teach the sections before the sections were randomized, effectively randomizing the instructors. This controlled for any bias introduced by instructors who may have been more interested in technology or who were more enthusiastic about this approach to teaching.

Iteration 2 Results and Adaptations

Quantitative results from the second iteration were not as encouraging as those from the first iteration. The analysis showed that there was no significant effect of the treatment on course success, controlling for demographic and other student variables ($p=.64$). There was also no significant effect of the treatment on final grade ($p=.90$).

Interviews with the instructors and feedback from students once again indicated a number of areas that could be improved. One area of insight was around the faculty. Adaptive learning requires faculty to shift their mindset regarding the ways in which they engage with students in the course. Our findings suggested that we needed to better prepare faculty to communicate the shift that happens in utilizing adaptive technology in tandem with learning analytics. Instructor training became an area of greater focus. As a result, we created a faculty mentor program so that faculty who had used the platform and felt comfortable with the technology could help new instructors, encouraging them to engage with the technology and answering any questions they may have. This allowed us to test our hypothesis that if faculty were better prepared, the student experience would improve.

Iteration 3

The third iteration of ACCT 220 was once again a randomized control trial. This trial involved 15 treatment sections and 16 control sections. The sample size was 797 students, 412 of whom were in RealizeIt sections. In this iteration, all students were asked to complete a baseline survey and an end of semester survey that asked for information not available through the data

warehouse (such as hours of employment, previous use of adaptive software, etc.) and for detailed feedback on perceptions of RealizeIt. User data from the RealizeIt system was also collected for this iteration, allowing us to see at which points in the RealizeIt system students were experiencing difficulty.

Iteration 3 Results and Adaptations

The analysis of the data showed that students in the RealizeIt sections were more likely to successfully complete the course with a final grade of C or higher than those in the control sections, controlling for demographic variables and a measure of how many hours the student works in paid employment ($p=.08$). That is, the effect of the treatment on course success was positive and significant.

The average grade was 1.8 for students in a control section, and 2.1 for students in a RealizeIt section. An OLS model estimated that the effect of being in a RealizeIt section was an average increase of .24 grade points for students' final grades, holding all else equal ($p=.00$). This model once again controlled for student demographics and historical academic performance. This result is robust to the addition of the survey variables, such as the student's level of confidence with technology, whether they had previously used adaptive software, and how many hours they work in paid employment.

These results mirror the results of the first iteration (in which only one section used RealizeIt). However, the third iteration has several advantages over the first. The sample size is much larger in this iteration (about 14 times larger). This means that we can be more confident in the results of our statistical analysis. The instructors were assigned to sections before the sections were randomized, effectively randomizing the instructors. And all online sections were part of the pilot and were randomized to treatment or control (each online term has several sessions, which begin at different times). This means that the sections that ran later in the term were as likely to be chosen for RealizeIt as those that ran earlier in the term. This is important, as there may be unmeasured differences between the students who take courses in the first session and those who take classes in the last session.

Beyond the final grades of the students, it was important to determine at what point RealizeIt was having an effect on student learning and to ensure that the aggregation of final grades into grade points was not creating the illusion of significant difference. To this end, we examined the effect of being a member of the treatment group on the constituent parts of the final grade. RealizeIt students had higher grades in all but one of the outcomes examined. However, the results are significant for only three outcomes, as can be seen in Table 1. The gains from Homeworks outweighs the loss seen in Quiz 2 however, as Homeworks (which is a combination of all homework assignments over the term) is worth 20% of the final grade, while Quiz 2 is worth 10% of the final grade.

Dependent Variable		RealizeIt Coefficient	P-value
Quiz 1		2.24 (1.67)	.18
Quiz 2		-6.92 (1.56)	.000
Quiz 3		.56 (1.76)	.75
Homeworks		4.03 (1.73)	.02
Final Exam		.49 (2.07)	.81
Final Grade		3.17 (1.58)	.05

Table 1. Fall 2016 Grades. All models are OLS regression and include controls for Age, Gender, Credits Earned, Current Session Workload, Campus, Pell, Cumulative GPA, and whether the student was Active Duty. Final Grade is measured in percentage points, not grade points.

The end of semester surveys also provided data on any difficulties the students had with the course, their impressions of their instructors, and the material covered. Broadly, there were few statistically significant differences between the two groups on these measures. Of course, it is worth noting that the sample size for these end of semester analyses is smaller because of the response rate. Two hundred and one students out of 797 participated in the end of semester survey (25% response rate).

When asked to rate their instructors on responsiveness, students in the control group rated their instructors 4.26 on average, while students in the RealizeIt group rated their instructors 4.42 on average. An OLS model estimated that being in the treatment group meant rating the instructor .3 points higher on the 5-point scale, on average and holding all else equal ($p=.04$). Students in the treatment section also rated their instructors higher on whether they provided helpful feedback. The average was 4.24 for control sections and 4.35 for RealizeIt sections. An OLS model estimated that being in the RealizeIt section meant rating the instructor .5 points higher, on average and holding all else equal ($p=.04$).

Students were also asked whether they thought this course was less rigorous, equally rigorous, or more rigorous than other UMUC courses they had taken. Half (50%) indicated that it was more rigorous, and 46% indicated it was equally rigorous. An ordered logistic regression

controlling for demographic and other variables showed that being in the treatment group had no effect on perceptions of course rigor for this question ($p=.20$).

In addition to being asked how the course compared to other UMUC courses taken, students were asked how rigorous the course was compared to non-online courses taken in the past. Almost half (47%) indicated that the course was more rigorous than non-online courses they had taken, while 48% indicated it was equally rigorous. An ordered logistic regression controlling for demographic and other variables showed that being in the treatment group had no effect on perceptions of course rigor for this question ($p=.49$).

The final section of the end of semester survey questionnaire asked students who were in the treatment sections about their impressions of RealizeIt. Students were asked the extent to which they agreed with statements about RealizeIt, using a 5-point Likert scale from 1 = Strongly Agree to 5 = Strongly Disagree. Table 2 presents these results (Fall 2016 columns), which have been reordered here for ease of interpretation, with higher scores being better.

	Fall 2016			Spring 2017		
	N	Mean	SD	N	Mean	SD
RealizeIt helped me learn subject better	103	3.71	1.41	118	3.82	1.38
RealizeIt provided feedback to stay on track	106	3.65	1.38	118	3.74	1.36
RealizeIt helped me better learn course material	105	3.76	1.38	118	3.79	1.42
RealizeIt's grading procedure was effective and logical	105	3.32	1.46	118	3.69	1.45
RealizeIt benchmarks helped my learning	106	3.75	1.35	117	3.87	1.38
RealizeIt feedback helped me learn	106	3.58	1.32	118	3.75	1.42
Effective compared to non-RealizeIt classes	105	3.60	1.33	116	3.79	1.38
RealizeIt assessments effectively measured my learning	105	3.49	1.41	118	3.79	1.37
RealizeIt increased engagement with course content	104	3.68	1.33	117	3.81	1.41
RealizeIt was easy to use	105	4.05	1.34	117	3.92	1.37
I was well prepared for using RealizeIt	105	3.74	1.30	117	3.79	1.26
RealizeIt instructions were clear and effective	106	3.92	1.32	115	3.89	1.37
Tech support helped me solve any RealizeIt issues	105	3.68	1.27	113	3.60	1.25
I would take another course using RealizeIt	106	3.58	1.48	117	3.79	1.44
Time spent in RealizeIt was valuable	105	3.76	1.29	117	3.86	1.34

Table 2. Students' Impressions of RealizeIt

As there is no comparison to the control group for these questions, the mean result for each question is reported. However, it is worth noting that the survey instrument used here was based on Dziuban, Moskal, Cassisi, and Fawcett (2016) to allow researchers to compare across institutions. The favorable results here are comparable to those reported by Dziuban et al. (Table 3).

Item	n	X	SD
RealizeIt helped me learn the course material	241	4.02	0.92
RealizeIt's assessment exercises were effective	235	3.82	0.78
Difficulty of the "learning path" sequence	240	3.35	0.81
Difficulty of the learning material	241	3.24	0.78
Difficulty of the questions asked	239	2.99	0.80
RealizeIt increased my engagement	233	3.92	0.87
Grading accurately reflected my knowledge	229	3.81	0.86
Ability levels reported by RealizeIt were accurate	235	3.79	0.84
I would take another course using RealizeIt	234	4.09	0.99
RealizeIt system became personalized to me	228	3.67	0.87
I followed recommended "next steps"	239	3.51	1.11
Time spent in RealizeIt	229	3.31	1.15
RealizeIt was easy for me to use	234	4.24	0.77
The instructions in RealizeIt were clear	241	4.12	0.80
"Learning Path" was easy to use	184	4.01	1.00
"Guidance panel" was easy to use	211	3.91	0.89
RealizeIt provided me with the necessary feedback	237	3.86	0.78
"Guidance panel" was helpful	182	3.81	0.67

Table 3. Dziuban et al.'s (2016) Study of RealizeIt Effectiveness at University of Central Florida: Student Reactions to Survey Items. Differing n's represent missing data.

Students were also asked about the pace of RealizeIt, whether they ever ignored RealizeIt's suggestions for completing content, and how much time they spent in RealizeIt relative to non-adaptive learning courses. The majority of students (67%) indicated that the pace of RealizeIt was just right, while 19% indicated it was somewhat fast. Forty-three percent indicated that they rarely or very rarely ignored RealizeIt's suggestions for completing content, while 18% indicated they did so often or somewhat often. Fifty-three percent indicated they spent more time or much more time in RealizeIt than non-adaptive courses, and 34% indicated they spent the same amount of time as in non-adaptive courses. Again, the sample size for these responses is quite small, and so results should be interpreted cautiously.

Finally, the end of semester survey allowed students to give qualitative responses to questions regarding technical issues and what could be improved with the system. These data points were analyzed and used to make recommendations to instructional designers at the institution designers, as well as to the vendor engineers.

These qualitative data were combined with the RealizeIt usage data to identify the points at which students had difficulty or were dropping out of the system. The data showed several questions that were queried by students at high rates. These questions were investigated by the designers and rephrased to ensure clarity. The data also showed several objectives that a high number of students began but did not finish the objective. Designers used this information to revisit problematic objectives to determine if the material was unclear or not well aligned with the learning objective. These improvements aim to enhance the usability of any aspects of the design that are less than optimal.

Iteration 4

Although CILSS had reached the end of the proposed cycle of iterations at the third iteration, university administration requested one more iteration to gather additional data and determine whether the pilot was suitable for upscaling. The same data points were gathered for this iteration as for the third iteration and results were similar. The sample size for Spring 2017 was 29 sections (14 RealizeIt sections and 15 control sections), which amounted to 808 students, 413 of whom used RealizeIt. Once again, RealizeIt students were more likely to successfully complete the course ($p=.01$). There was also a significant effect on final grade. RealizeIt students completed with a final grade that was .32 points higher than non-RealizeIt students, on average ($p=.00$). The average grade was 2.22 for students in the control sections and 2.48 for students in the RealizeIt sections.

Table 4 shows that, for Spring 2017, RealizeIt students also had significantly higher results in Quiz 1 (6.5 points), Quiz 3 (3.4 points), Homeworks (5.7 points), and in the Final Grade (5.1 points). Unlike Fall 2016, there were no measures for which RealizeIt students received lower grades.

		RealizeIt Coefficient	P-value
Dependent Variable	Quiz 1	6.52 (1.36)	.00
	Quiz 2	.15 (1.28)	.91
	Quiz 3	3.42 (1.33)	.01
	Homeworks	5.74 (1.55)	.00
	Final Exam	1.50 (1.81)	.41
	Final Grade	5.10 (1.31)	.00

Table 4. Spring 2017 Grades. Final Grade is measured in percentage points, not grade points.

As with Fall 2016, there were few differences between the two groups on measures of any difficulties the students had with the course, their impressions of their instructors, and the material covered. Once again, students in the treatment group rated their instructors .3 points higher on responsiveness, on the 5-point scale, on average and holding all else equal ($p=.00$). RealizeIt students also rated instructors .4 points higher on responsiveness ($p=.00$), .35 points higher on course knowledge ($p=.00$), .4 points higher on maintaining accurate course records ($p=.00$), and .3 points higher on the quality of their grading ($p=.01$), on average and holding all else equal.

Descriptive statistics for students' impressions of the RealizeIt system and its effectiveness and usefulness were also strikingly similar to the results from the third iteration (See Table 2: Spring 2017 columns) (and once again similar to results reported in Dziuban et al. [2016]—Table 3).

Results and Lessons Learned from RealizeIt Pilot

The researchers have presented the approach to DBR taken by CILSS at UMUC and the results of a project that illustrates this approach. Like the RealizeIt pilot discussed, the DBR approach itself is also subject to continual improvement. The approach detailed here allowed us to learn about issues related to technical problems, faculty training, and design. These problems were then tackled for subsequent iterations in order to improve the intervention and retest. However, the DBR approach itself can also be improved. More formalized focus groups would have been an asset in the initial iterations. Focus groups would have allowed for more qualitative data on student perceptions and could have potentially pinpointed problem areas sooner.

Students were randomly assigned to either treatment or control sections. Once students were assigned, they were given the option to opt out of the pilot. This approach sought to overcome problems of selection bias, as students who are more interested or motivated are likely to be the ones who sign up for a pilot study (Campbell & Stanley, 1971). However, as the section chosen for RealizeIt was the first online section, and therefore the first section to fill, it is possible that the students who were enrolled were also students who were somewhat more motivated or organized than the average student and are therefore not representative of the population. Nonetheless, as this was the first iteration and focused on design issues and technical problems, this does not pose a problem for our research design. The two sections (one treatment and one control) resulted in a sample size of 55 students, 26 of whom used RealizeIt.

While CILSS had not planned the final iteration involving the hybrid sections, this was added during the third iteration. The reason for this was the positive feedback that instructors were getting from students and the desire on the part of the university administration to gather more data to better determine whether the pilot should be scaled up. As discussed in the literature review section, this ability to add iterations to a research design is both a blessing and a curse. On one hand, it allows for flexibility in the research design. On the other hand, it means that an evaluation can continue indefinitely, with invested researchers always needing one more iteration. We do not believe that to be the case here. Once the current iteration is completed, further iterations will only be carried out at the behest of university administrators should they feel that more data are needed to make a decision to scale up the pilot.

Discussion

There is a growing demand in the field of education for providing educational technology evaluations that are systematic and measure the efficacy of educational technology solutions. In their review of the DBR literature, Anderson and Shattuck (2012) found that 68% of the interventions involved online and mobile technologies. However, the majority of studies focused on K-12 student populations rather than the higher education sector. This revealed the current gap in DBR research studies focused on the iterative design and implementation educational technology interventions in the higher education sector. The approach using DBR provides researchers the opportunity to utilize a collaborative framework with practitioners at all levels in the field. The intent of this paper is to provide some additional insights of using DBR as a framework to move to platforms that are adaptive in nature. The RealizeIt platform afforded us with the opportunity to advance self-directed learning consistent with andragogically-informed design and to improve student outcomes. Our use of DBR in this scenario is consistent with previously recommended and applied uses of DBR to address the question of how education should leverage technology to address complex open problems and the related questions around learning, teaching, and assessment (Bannan, 2013; Kelly, 2013). Further, our use of iterative design and evaluation cycles enabled us to surface important methodological issues associated with studying learning in what Kelly (2013) described as a “complex and nested learning environment” (p. 140) within the cyberinfrastructure. Experience that includes mistakes can provide the basis for rich learning. For the first time, we had comprehensive and robust data to measure the learning occurring in the online environment.

While this study allowed the use of a mixed method approach, we know that future studies are required. Consistent with current thinking on DBR, assessment targets surface during the unfolding design and implementation cycles, for which appropriate measures must be developed. Likewise, the validity and reliability of those measures must be actively considered throughout the project (Kelly, 2013) so that the evidentiary methods and claims are properly aligned to subsequent iterations and implementations of design prototypes. Here, in this evaluation, we took the lens and philosophy of a qualitative researcher and, in that sense, knowing what students believe matters. If a student believes s/he learned, it is likely that the student’s next action will be based on that belief, for example, signing up for an additional class. However, student self-reports only create one narrow view of the evaluation of this new learning paradigm. Upon completion of this study, longitudinal impacts of students and their academic careers should be observed as a result of their participation in adaptive learning in core foundational courses for their major.

In hindsight, the evaluation also was challenging, given the rapid pace of the cycle of semesters and gathering the data. It should be noted that although there were opportunities in the course cycle to improve the course, it was not consistently possible to make improvement on the very next rollout of the course, given the overlap of the course sessions. Consistent with Bannan’s (2013) Integrative Learning Design Framework, we plan to include additional targeted focus groups, observation/modeling, and interviews at the end of the final iteration cycle to validate that we accurately identified all levels of feedback about the innovation pilot prior to making recommendations about a full-scale implementation. While randomized control trial was used to test the final product that had been developed through earlier iterations, this provides a culminating evaluation of the whole cycle, giving us a holistic view and harnessing the power of the DBR approach.

References

- Anderson, T., & Shattuck, J. (2012). Design-based research a decade of progress in education research? *Educational Researcher*, 41(1), 16–25. Retrieved from <http://dx.doi.org/10.3102/0013189X11428813>
- Bannan, B. (2013) The Integrative Learning Design Framework: An illustrated example from the domain of instructional technology. In T. Plomp & N. Nieveen (Eds.), *Educational design research - Part A: An introduction* (pp. 114–133). Enschede, The Netherlands: Netherlands Institute for Curriculum Development. Retrieved from <http://international.slo.nl/publications/edr/>
- Blondy, L. C. (2007). Evaluation and application of andragogical assumptions to the adult online learning environment. *Journal of Interactive Online Learning*, 6(2), 116-130. Retrieved from <http://www.ncolr.org/jiol/issues/pdf/6.2.3.pdf>
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.
- Campbell, D. T., & Stanley, J. C. (1971). *Experimental and quasi-experimental designs for research* (Vol. 4). Rand McNally.
- Cercone, K. (2008). Characteristics of adult learners with implications for online learning design. *AACE Journal*, 16(2), 137-159. Retrieved from https://www.editlib.org/index.cfm/files/paper_24286.pdf?fuseaction=Reader.DownloadFullText&%3Bpaper_id=24286&origin=publication_detail
- Dede, C. (2004). If design-based research is the answer, what is the question? A commentary on Collins, Joseph, and Bielaczyc; diSessa and Cobb; and Fishman, Marx, Blumenthal, Krajcik, and Soloway in the JLS special issue on design-based research. *The Journal of the Learning Sciences*, 13(1), 105–114. Retrieved from http://dx.doi.org/10.1207/s15327809jls1301_5
- The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8. Retrieved from <http://www.jstor.org/stable/3699927>
- Duffy, T. M., & Kirkley, J. R. (2004). Learning theory and pedagogy applied in distance learning: The case of Cardean University. In T. M. Duffy & J. R. Kirkley (Eds.), *Learner-Centered Theory and Practice in Distance Education: Cases from Higher Education* (pp. 107-141). New York, NY: Routledge.
- Dziuban, C., Moskal, P., Cassisi, J., & Fawcett, A. (2016). Adaptive learning in psychology: Wayfinding in the digital age. *Online Learning*, 20(3). Retrieved from <https://olj.onlinelearningconsortium.org/index.php/olj/article/view/972>

- Easterday, M., Rees Lewis, D., & Gerber, E. (2014). Design-based research process: Problems, phases, and applications. In Proc. of International Conference of Learning Sciences (Vol. 14). Retrieved from http://egerber.mech.northwestern.edu/wp-content/uploads/2014/10/DesignResearch_Methodology_ICLS_2014.pdf
- Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *Yearbook of the National Society for the Study of Education*, 112(2), 136-156. Retrieved from https://www.sri.com/sites/default/files/publications/fishman_penuel_allen_cheng_sabelli_2013.pdf
- Holton, E. F., Swanson, R. A., & Naquin, S. S. (2001). Andragogy in practice: Clarifying the andragogical model of adult learning. *Performance Improvement Quarterly*, 14(1), 118-143. Retrieved from <http://dx.doi.org/10.1111/j.1937-8327.2001.tb00204.x>
- Kelly, A. E. (2013). When is design research appropriate? In T. Plomp & N. Nieveen (Eds.), *Educational design research - Part A: An introduction* (pp. 134-151). Enschede, The Netherlands: Netherlands Institute for Curriculum Development. Retrieved from <http://international.slo.nl/publications/edr/>
- Knowles, M. (1973). *The adult learner: A neglected species*. Houston, TX: Gulf. Retrieved from <http://files.eric.ed.gov/fulltext/ED084368.pdf>
- Knowles, M. (1984). *Andragogy in action*. San Francisco, CA: Jossey-Bass.
- Knowles, M. (1996). *Andragogy: an emerging technology for adult learning*. London: Routledge. Retrieved from <https://www.nationalcollege.org.uk/cm-andragogy.pdf>
- Knowles, M. S., Holton, E. F., III, & Swanson, R. A. (2014). *The adult learner: The definitive classic in adult education and human resource development*. New York, NY: Routledge.
- Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, 44(1), 54–61. Retrieved from <http://dx.doi.org/10.3102/0013189X15570388>
- McAuliffe, M., Hargreaves, D., Winter, A., & Chadwick, G. (2009). Does pedagogy still rule?. *Australasian Journal of Engineering Education*, 15(1), 13-18. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/22054952.2009.11464018>
- McKenney, S., & Reeves, T. C. (2014). Educational design research. In J. M. Spector, M.D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 131–140). New York, NY: Springer. Retrieved from http://dx.doi.org/10.1007/978-1-4614-3185-5_11

- Murray, M. C., & Pérez, J. (2015). Informing and performing: A study comparing adaptive learning to traditional learning. *Informing Science: the International Journal of an Emerging Transdiscipline*, 18, 111-125. Retrieved from <http://www.inform.nu/Articles/Vol18/ISJv18p111-125Murray1572.pdf>
- Nieveen, N., & Folmer, E. (2013). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), *Educational design research - Part A: An introduction* (pp. 152-169). Enschede, Netherlands: Netherlands Institute for Curriculum Development. Retrieved from <http://international.slo.nl/publications/edr/>
- Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40(7), 331–337. Retrieved from <http://dx.doi.org/10.3102/0013189X11421826>
- Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research - Part A: An introduction* (pp. 10-51). Enschede, Netherlands: Netherlands Institute for Curriculum Development. Retrieved from <http://international.slo.nl/publications/edr/>
- Reimann, P. (2016). Connecting learning analytics with learning research: the role of design-based research. *Learning: Research and Practice*, 2(2), 130–142. Retrieved from <http://dx.doi.org/10.1080/23735082.2016.1210198>
- Russell, J. L., Jackson, K., Krumm, A. E. & Frank, K. A. (2013). Theories and Research Methodologies for Design-Based Implementation Research: Examples from Four Cases. *Yearbook of the National Society for the Study of Education*, 112(2), 157–191. Retrieved from https://education.uw.edu/sites/default/files/u735/NSSE_2013_Russell_Jackson_Krumm_Frank.pdf
- Sternberg, R. J., Birney, D., Jarvin, L., Kirlik, A., Stemler, S., & Grigorenko, E. L. (2006). From molehill to mountain: The process of scaling up educational interventions (firsthand experience upscaling the theory of successful intelligence). In M. A. Conchas & R. J. Sternberg (Eds.), *Translating theory and research into educational practice—Developments in content domains, large-scale reform, and intellectual capacity* (pp. 205-222). Mahwah, NJ: Erlbaum.
- Wademan, M.R. (2005). *Utilizing development research to guide people-capability maturity model adoption considerations*. Doctoral dissertation. Syracuse: Syracuse University. Dissertation Abstracts International, 67-01A, 434. (UMI No. 3205587)