Go Back to Your Room: Designing a Scalable In-Class Required College Course.

Emergent Research Forum (ERF) Papers

Introduction

In our ongoing program of research, we focus on a wicked problem confronting higher education: the need to balance efficiency and effectiveness. Under financial pressure, many universities have had to accommodate increasing student-teacher ratios. But recent research shows that larger college classes are associated with lower grades, particularly for students who are at the top of the grade distribution (Bandiera et al. 2010). Further, interaction and engagement decrease and students in large classes remain anonymous, leading to a lower motivation for both faculty and students (Chambliss and Takacs 2014). Our work is motivated by the need to scale an introductory information systems course to prepare about 1,000 business college freshmen per year. The point of departure for our work is the centrality of human interactions in learning environments. Thus, rather than looking at MOOCs or online courses, we leverage IT to design a *required in-class introductory college course that can scale to large numbers of students, under resource constraint*.

Conceptualizing a semester-long college course is an act of design, "engineering an environment in which [students] learn" (Bain 2004). Thus, we adopt a design science approach and conceptualize the crafting of a university course as the design of a Socio-Technical (ST) artifact (Gregor and Hevner 2013). In the remainder of the paper we discuss intervention theory and how it guides the design of the course, we discuss the implementation and we report on preliminary evaluation. We conclude with the modified design for next iteration, the result of which will be available for presentation at the conference.

Intervention Theory

The kernel theory underpinning the meta-requirements (MR) for our proposed ST artifact design is intervention theory (Argyris 1970). It identifies three principles that guide the design of interventions: leveraging valid and useful information, allowing free informed choice by the client (in our case the student) and fostering internal commitment. Previous research (citation withheld for blind review), draws on intervention theory to articulate the meta-requirements (Walls et al. 1992) upon which to build and scale a learning environment for required introductory college courses under resource constraint. In this section, we briefly sketch both the theory and the meta-requirements stemming from it.

Valid information is that which can be publicly verified and shown to affect the phenomenon of interest. Useful information is that which the client would be able to use to "control their destiny" (Argyris 1970). For example, while students' natural aptitude will reliably predispose them to master specific subject matter (e.g., analytics), they cannot modify natural ability. From an intervention theory standpoint, natural aptitude provides valid, but not useful information. Study habits are also reliably shown to affect subject matter mastery. An understanding of study habits provides information that is both valid and useful.

MR1: An in-class required introductory college course that can scale to large numbers of students should record students' behaviors both in class (e.g., attendance) and outside (e.g., study patterns).

Free informed choice points to the centrality of the client in the implementation of the intervention – and therefore, in its design. Free and informed choice is particularly important in college learning, as opposed to surgery for example. While a surgery patient has no control over the outcome of the intervention, students exert great control over their effort and commitment to learning.

MR2: An in-class required introductory college course that can scale to large numbers of students must not conflate behavior with learning. In other words, grades should never be used to enforce behaviors (e.g., attendance points).

MR3: An in-class required introductory college course that can scale to large numbers of students treats students as self-responsible and maximizes learner control.

MR4: An in-class required introductory college course that can scale to large numbers of students exposes all behavioral and performance data as soon as they become available (e.g., providing alerts to at-risk students based on tracked behaviors and performance).

MR5: An in-class required introductory college course that can scale to large numbers of students contextualize behavioral and performance data for students (e.g., in the form of contextualized dashboards enabling students to compare their behavior to the class, previous cohorts, or their own targets).

Internal commitment refers to the degree of ownership and responsibility the client feels with respect to the intervention. The power of internal commitment comes from individuals' sense of purpose for the initiative and their beliefs about the control they exert over their action and the outcome.

MR6: An in-class required introductory college course that can scale to large numbers of students proactively triggers appropriate behaviors (e.g., by engaging students outside of class using conversational interfaces and alerts).

MR7: An in-class required introductory college course that can scale to large numbers of students encourages sustained use by managing triggering risks (e.g., personalizing triggers based on student preferences or characteristics).

The three principles of intervention theory are interdependent. The availability of valid and useful information is necessary for the client to make decisions that are free and informed. At the same time, the outcome of these decisions provides information that contributes to the stock of valid and useful information available to the client and the intervenor. Moreover, to the extent that the results of choices being made by the client are positive, those choices should strengthen internal commitment (Argyris 1970).

The Course

While our aim is to contribute a general design for scalable in-class required college courses our intervention takes place in the "Introduction to Management Information Systems" at a large state business school in the USA. The course is required of all first-year business majors and a similar course is on the books at most major undergraduate business programs. Over the last five years, an average of 1,563 students per year enrolled in the course with all sections taught 100% online. The migration to online delivery was incremental, following a very successful pioneering effort in the late 1990s (citation withheld). The online effort created considerable efficiencies, saving the college over $5.5M between 2001 and 2016, but human interaction between students and faculty has all but disappeared. An analysis of all communication in one section of the online course (Fall 2017, enrollment 171 students) corroborates previous evaluations by instructors (Authors citations). Only 6% of enrolled students visited an instructor for at least one face-to-face meeting during the semester. All other communication occurred via email, with 54 out of 163 students sending at least one message. Thus, 66.9% of students never interacted with a faculty member at all. 71.2% of messages pertained to administrative and procedural questions (e.g., "I was wondering [about] the difference between the $135 price and the $180. What comes with each price?"), 26% of the messages pertained to software issues (e.g., "I uploaded my file and I got 100% on it, and when I went to press submit it would not let me but it showed up at the bottom that I did get a 100%"), and only the remaining 2.3% was devoted to content questions (e.g., "I just finished the project but I'm confused on how to print preview the workbook since there is no file tab").

A Scalable In-Class Design

The first full semester instantiation of our designed consisted of two 80 minutes sessions per week. On Tuesdays, the instructor held interactive lectures covering theoretical material on IT foundations (e.g., networking) and IS foundations (e.g., value creation with IT) topics. On Thursdays, using a flipped-classroom pedagogy, the class acquired intermediate skills in Microsoft Word and Excel. While our design encompasses both the theoretical and practical elements of the course, we focus this paper on the first two elements of intervention theory (MR1-MR5) and the skills components of the course. The practical sessions started with the instructor showing how to perform some of the more difficult or conceptually challenging tasks using a purposely designed data file that students could download to follow along. After this mini-lecture of 15-30 minutes, the students worked on practice assignments using their personal computer (Mac or Windows) with a flipped-classroom approach.

We implemented MR2 by ensuring that no Digital Data Stream (DDS) tracked during class or by the course application was used to compute students' grade. All assigned activities, such as practice assignments, were assessed if the students completed them, but did not contribute to the final evaluation. Students' mastery was measured by way of two exams covering practical skills, five checkups and a final team project covering theoretical material. We implemented MR3 by providing nine practice assignments over the course of the semester. Work submitted within the expected one-week deadlines was evaluated. And students received a detailed task-by-task report for each assignment. These reports where not intended to substitute for human interaction, rather, they directed students to meet with the instructor or the teaching assistants for any doubts.

Given our scalability objective, we developed a custom-made solution to automatically evaluate students' performance in the practice assignments – the autograder. Microsoft Office documents are collections of XML files. Written in Python, the autograder has functions to compare student work to a "key file" made by the instructor. Those functions are robust to acceptable alternative solutions. Given space limitations we do not describe the system, but we will have it available for examination and use at the conference. By controlling the automatic grading system, we have full control of the entire process: from skill identification, to framing of exercises students use to acquire and practice those skills, to the evaluation approach and the degrees of freedom allowed in the assessment of student work.

Enabling Technology: The Custom Designed Course Application

We custom-developed the software application enabling the ST artifact on a back-end using the MEAN (MongoDB, Express.js, AngularJS, and Node.js) free and open-source JavaScript stack. It runs on a Linux OS instance from Amazon Web Services (AWS). For the application front-end we adopted the Bootstrap library and the Embedded JavaScript (EJS) template language – ensuring native mobile and desktop compatibility with HTML, JavaScript, and CSS and a responsive app on all platforms.

We implemented MR1 by requiring students to sign-in to access course material: chapters, course slides, practice assignments, topic and test schedules, and performance results. The application traffic is tracked and monitored using Google Analytics' (GA). During the current iteration, we manually recorded attendance to physical activities, such as lab and class sessions. In the current instantiation of the ST artifact, we implemented MR4 and MR5 with a reporting system delivering two types of custom pdf reports via email. The first type, a detailed task-by-task report contains results from practice assignments. We implemented an R script that, using the CSV output of the automatic grading software, automatically creates and emails to the student a pdf report with the performance feedback detailed in the previous section. The second type of reports expose valid and useful information to students, detailing behaviors enacted (e.g., attendance, access to the materials) and performance on evaluated tasks. For example, students received a chart comparing their lab attendance data to the class average. We generated the report for each student twice during the semester, at midterm and prior to the final exam. The reports offered a comprehensive overview of individual behaviors, in addition to a comparison of individual activities with an aggregate (average) of the class.

Evaluation

At this stage we claim an "improvement" knowledge contribution, focused on "developing new solutions for known problems" (Gregor and Hevner 2013, p. 345). Because we are designing a required in-class introductory college course that can scale to large numbers of students under resource constraint, the first step in the evaluation is to provide "proof-by-demonstration" (Nunamaker Jr et al. 1990, p. 98). In this section, we present the evaluation of the ST artifact and how we use it to gather essential feedback to feed into the next iteration of artifact construction (Sein, et al. 2011).

With respect to MR1, we find our design to enable the generation of needed DDS because we are able to collect accurate resource utilization data with an approach that is scalable since our app resides on the AWS infrastructure. Specifically, only 5.56% of students reported printing all the theoretical materials and 7.43% reported printing all the practice assignments. We also evaluate all performance activities, required or optional, using electronic means (e.g., the automatic grading software). As a consequence, we can scale our collection of performance data. As with any other behavior in the course, class attendance was not required (MR2). Despite the prevailing rhetoric at the school suggesting that unless students are forced they will not go to class, attendance was consistently above 75% and the formal course evaluation showed that the course was well received 3.53/4 (college total 3.35/4).

While they were present and actively engaged during the sessions, in aggregate students did not consistently complete the assignments. On time completion was the only requirement to receive feedback, and completion steadily declined during the semester – from 69% on the first assignment to 7% on the ninth assignment. In the end of semester survey 22 out of the 27 respondents agreed with the statement that "shorter practice assignments would be more effective." We conclude that in our design implementation the design failed to ensure that students would stay on track with the progression of the course, without using grades to stimulate activity.

Finally, our results show that performance on practical skill development – as measured by the score on the exams testing Word and Excel competency – is correlated with the use of external links to the Microsoft documentation and other beneficial learning resources. However, when separately regressing the use of external resources on exam performance, we find that results hold only for Excel (Table 1). We tested four models with the following general specification:

$$\text{Mastery (Word|Excel)} = \beta 0 + \beta 1 \times (\text{linkfr } |\text{linkdiv}) + \beta 2 \times (\text{attendance}) + \beta 3 \times (\text{completion}) + \varepsilon$$

Mastery is tested via a one-hour comprehensive exam. Control variables include attendance and assignments completion rate. Independent variables of interest are the frequency of external resources utilization (linkfr) measured as the total number of clicks to external resources, and the extent of external resources utilization (linkdiv), measured as the total number of external resources used at least once.

	Model 1: Word	Model 2: Word	Model 3: Excel	Model 4: Excel
(Intercept)	86.9388**	85.7489**	27.7601*	25.0508
Link Frequency	0.0460		0.5050*	
Link Diversity		0.1228		0.8969**
Attendance	-21.8330	-21.2051	34.1738*	35.5872*
Assignment Completion	6.7653**	6.4808*	0.6804	-0.1653
Resid.SE	11.84	11.82	15.47	14.95
F	4.093	4.136	6.216	7.243
Prob>F	0.0171	0.0164	0.0027	0.0012
Adj. R-Squared	0.2489	0.2515	0.3585	0.4008
Note: * denotes significance at α=0.05, and ** denotes significance at α=0.01.				

Table 1. Regression Analysis Results

The tests are independent because there is no overlap of content or external resources in the two exams. Resource usage had no effect above and beyond the amount of practice by the students (completion) in Word. Conversely, for Excel, each incremental visit to a documentation page or video results in a half-point increase in final score ($p = 0.0103$) and each incremental visit to a new resource results in an almost one-point increase (out of 100) in final score ($p = 0.0041$). We ascribe the difference to the fact that most students have some familiarity with features in Word but find Excel more difficult both conceptually and syntactically. Thus, ready access to explanatory material has a stronger impact on their learning and performance in the latter. These results provide important feedback for iterating the ST artifact design along two dimensions: a) providing appropriate links to needed material within the flow of student activity and b) designing digital nudges that can motivate students to take advantage of the resources, without relying on grades or other performance incentives.

Design Evolution and Second Iteration

The above evaluation serves to inform the design process (Sein, et al. 2011) and point to needed ST artifact changes. Based on the evaluation of the first full ST artifact implementation, we iterated the design and introduced the following changes ahead of the Spring 2018 semester implementation.

The first consideration is that the system is ready to scale to large numbers. All behavioral DDS are seamlessly tracked and can be used for analysis with the exception of attendance. The implementation of the design principles associated with MR2 and MR3 was successful. Our evaluation shows that the automatic grader is reliable, but the timing of feedback was reevaluated based on our observations and student feedback. Specifically, we "chunked" the assignments to overcome the length of practice obstacle and we eliminated deadlines to enable students to receive feedback at any time rather than only once a week. Each of the nine practice assignments is subdivided into 4-5 pedagogically consistent chunks. Each chunk provides a data file, thus reducing carry-over errors, and a drop-file feature in the app streamlines the workflow for the student. In the new design, the results are not emailed as pdf reports, but they are accessible privately by each student in the app through a JSON file produced by the automatic grader. These new features have enabled us to grade files every day and lift the lockstep constraint. In other words, any file dropped by the student is evaluated and made available in the app, regardless of sequencing of completion. The result is a faster cycle of feedback that on-average completes under 24 hours and we expect will increase the completion rate of assignments – without motivating via grades.

Our evaluation of actual student use patterns shows that they tended not to print material and instead work within the app. The use of embedded links significantly impacted students learning on the most difficult practical skills component of the course. These early results suggest that the design of the content can simultaneously improve the objective of valid and useful data collection through DDS generation and improved pedagogical value. In the next iteration we are focusing on embedding valuable multimedia elements and links in the theoretical content. Early examples are videos, links to external resources, and embedded widget for students to improve their understanding (e.g., a password strength evaluation widget in the cybersecurity chapter). As the design of content moves away from a book format and closer to an interactive knowledge app, we expect students to further limit printing while improving their understanding of the material.

While at this point we have relied heavily on log data to capture student behaviors, we intend to leverage emerging multimodal learning analytics techniques (Blikstein and Worsley 2016) to improve our collection of valid and useful information at scale. Specifically, we are currently developing an attendance system that relies on face recognition. The system is to be used to collect attendance in all co-located learning activities: class sessions, lab sessions, review sessions and office hours. Keeping with our design principles, such system can be used to record student behaviors in an effort to expose them through dashboards and will not be used for grading. We also expect the system to become the basis for real-time student recognition during the class, once augmented reality solutions become viable. We deem such a system as an important instrument to help reduce the feelings of anonymity that pervade students in large classes.

REFERENCES

Argyris C. 1970. *Intervention Theory and Method: A Behavioral Science View*. Reading, Mass: Addison-Wesley.

Bain K. 2004 *What the Best College Teachers Do*. Cambridge, Mass: Harvard University Press.

Bandiera O., Larcinese V. and Rasul I. 2010. "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students," *The Economic Journal* (120: 12), pp. 1365–1398.

Blikstein P. and Worsley M. 2016. "Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks," *Journal of Learning Analytics* (3: 2), pp. 220–238.

Chambliss D.F. and Takacs C.G. 2014. *How College Works*. Cambridge, Mass: Harvard University Press.

Gregor S. and Hevner A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37: 2), pp. 337–355.

Nunamaker J.F., Chen M. and Purdin T.D. 1990. "Systems Development in Information Systems Research," *Journal of Management Information Systems* (7: 3), pp. 89–106.

Sein M., Henfridsson O., Purao S., et al 2011. "Action Design Research," *MIS Quarterly* (35: 1), pp. 37–56.

Walls J.G., Widmeyer G.R. and El Sawy O.A. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3: 1), pp. 36–59.