Reclaiming the Classroom through Design Science Research:

Designing scalability in required college courses.

Abstract In this work, we show how digital innovation can put pressure on the quality of education by creating financial incentives for moving classes online. We posit that design science enables us to create quality courses despite these financial pressures, and we address the design problem of delivering a required in-class introductory college course that can scale to large numbers of students, under resource constraint. The point of departure for our work is the centrality of human interactions in learning environments and we conceptualize a college course as a socio-technical artifact. From intervention theory, we draw meta-requirements that can guide the design of college courses that leverage IT to enable the professor to scale the course while maintaining their role as the course designer and lead. The paper uses the design-build-evaluate cycle of design science research to instantiate the ST artifact and demonstrate its feasibility. Based on the evaluation of the first instantiation, during a full semester course, we refine the original design principles for the class scalable required in-class college courses.

Keywords Digital innovation · Socio-technical artifact · Design science research · Information systems education

1. Introduction

There is little doubt that digital technology has transformed society as we know it. Over the last few decades, digital innovation was instrumental in creating new products and services, new business models and value creation paradigms. More broadly, it helped transform entire industries (Watson et al. 2017; Nambisan et al. 2017). However, digital innovation and digital transformation engender unintended consequences and risks, and these risks often go understudied by Information Systems research (Gregor and Hevner 2013; Silver and Markus 2013). In this paper, we turn our attention to a wicked problem (Buchanan 1992) confronting higher education: the need to balance efficiency with quality education.

Under financial pressure, many universities have responded by increasing classroom size and substituting online courses for campus-based activities. Massive Open Online Courses (MOOC) MOOCs and Virtual Learning Environments (VLE) have proven to be effective alternatives for students in remote locations (Chatterjee and Nath 2014), students who cannot afford traditional education (Dillahunt et al. 2014), or students who just need to refresh their academic background (Hew and Cheung 2014). Yet, there is a growing consensus in the literature on effective college education that points to human relationships between peers and between the instructor and the students as the catalyst for high-quality college education (Kalay, 2004; Bernard et al., 2009; Gebre, Saroyan and Bracewell, 2014).

With respect to class size, recent research converged toward the consensus position that the learning and teaching experience degrade as the student-instructor ration increases. Specifically, at the college level larger classes are associated with lower subject matter mastery, as measured by grades (Kokkelenberg et al. 2008), 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). The most effective teachers carefully motivate student learning by articulating and reinforcing the "payoff" of knowledge acquisition while giving students a sense of control over their achievement (Bain 2004). Larger classes limit the human interaction that teachers traditionally leverage to motivate students, leading to a regression toward using grades as a motivator. The lack of human interaction also fosters strategic learning – which occurs when students focus "primarily on doing well in school, avoiding any challenges that will harm their academic performance and record" (Bain 2004). 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*.

Conceptualizing a semester-long college course is an act of design – "engineering an environment in which [students] learn" (Bain 2004). Thus, our program of research tackles the challenge of designing a *required in-class introductory college course that can scale to large numbers of students, under resource constraint*. We argue that the design science paradigm is best suited to our goal because it is a problem-solving paradigm (Hevner et al. 2004). We conceptualize the crafting of a university course as the design of a Socio-Technical (ST) artifact (Gregor and Hevner 2013; Silver and Markus 2013), substantiated by the interplay of IT, people, processes and organizational structures. For those universities where the pressure on efficiency is irreversible, we submit that design science research enables information systems scholars to uncover and test course designs that optimize the seemingly competing pressures of quality and efficient education.

The plan for this paper is as follow. We first frame the class of problems we are focusing on. We then introduce our kernel theory (Walls et al. 1992): intervention theory. Next, we articulate a set of meta-requirements and design principles drawn from the kernel theory, as applied to our specific class of problems. Following, we discuss the artifact we built and describe the first design iteration. Next, we evaluate the design and show that the artifact can be constructed, that it will work, and that it will be used by the intended audience. Finally, we use the evaluation to draw feedback for a designing a second instantiation, which we articulate in the discussion section of the paper.

2. Problem Definition

Our work focuses on a standard introductory course in a large state business school in the United States of America. The course is titled "Introduction to Management Information Systems" and it is required of all first-year business and economics majors. According to the description, the course covers "the role of information technology in business including the development and use of information systems, hardware and software, the strategic impact of IT for businesses and the nature of the IT career; utilization of management information systems to improve managerial decision-making." Since its inception in the mid-1990s, the course also had a practical component focused on proficiency with individual productivity tools (e.g., Microsoft Excel). Over the last five years, an average of 1,563 students per year enrolled in the course(Author Cited). A similar course is on the books at most major undergraduate business programs.

Over the last decade, the course increasingly migrated from the classroom to an online delivery. Resource constraint, rather than pedagogical superiority, was the primary driver of this decision. After some experimentation, the department settled on a major vendor system with the following features:

- Online Books: All content is delivered online.
- Virtual Microsoft Office environment: No software installation required, a simulated application runs in the Web browser.
- Assignments: The instructor assigns specific skills for the students to practice as graded homework and quizzes.
- Gradebook: Assignments are automatically graded, and the instructor receives a standardized report.
- Search: Users can navigate directly to content referring to a specific skill.
- Videos and interactive "Guide Me" pages: Screen capture videos that demonstrate how to complete individual skills.

The application automatically grades the work students performed in the simulated environment. It can also evaluate "projects" – structured assignments the student completed using the actual software application (e.g., Excel). As documented elsewhere (Author Cited), the introduction of the system resulted in some positive intended consequences and some negative unintended consequences. Specifically, the migration to the online delivery brought increased efficiency – defined as student throughput per section. These efficiencies translate into salary savings exceeding $5.5 million between 2001 and 2016. While it is hard to assess the effectiveness of the course without a formal evaluation, instructors indicated that an in-class delivery "is better" from a pedagogical standpoint, but infeasible with large sections (Author Cited). Amongst the primary limitations, they identified the inability to illustrate particularly difficult concepts, to respond to questions in real time, to convey tacit knowledge (e.g., tips and tricks) that improve students' efficiency and effectiveness. They also lamented the inherent limitations of using a simulated software environment rather than having students practice directly in the software (e.g., Microsoft Excel). More surprisingly, the online delivery engendered some unintended consequence: role reversal, minimization of human interaction, and strategic learning (Author Cited).

Role reversal refers to the "digitization of the professor" whereby the "course solution," over time, disrupts the role of the teacher and takes over fundamental teaching activities. As the department gravitated increasingly toward the online delivery, control of both the content and the pedagogy shifted to the digital "course solution." Practical skills (e.g., Microsoft Office) became predominant over the theoretical concepts listed in the course description and the instructors were relegated to course administration and support.

Minimization of student interaction refers to a course design that steers students toward online resources rather than fostering human interaction with faculty members. In the course, only 6% of enrolled students visited an instructor for at least one face-to-face meeting. 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 (all those who interacted face to face also sent email communications). Further, 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”). These results are based on an analysis of all communication in a section of the online course (enrollment 171 students). However, they are consistent with previous evaluations by instructors involved in the course (Authors citations) and conform to a ‘‘technology-shaping perspective’’ (Markus 2005) whereby even small differences in ST artifact design lead to significant differences in the pattern of use over time (Palen and Grudin 2003).

The lack of human interaction also appeared to foster strategic learning. While strategic learning is not exclusive to online courses, it can be exacerbated when performance goals are set based on extrinsic motivators (e.g., the grade) rather than a mastery orientation.

Note that the department did not plan for role reversal, minimization of human interaction, and strategic learning. These were intended outcomes of the migration to the online learning environment that, we argue, stem from the lack of a “purposeful organization of resources” (Hevner et al. 2004) designed to create an introductory college course that can scale to large numbers of students, under resource constraint. In the remainder of this paper we investigate a competing design that uses technology to enable, rather than substitute, the professor.

3. Designing the Artifact: A Scalable Course

While our context is a college of business required introductory course in Information System, with our design we seek to inform solutions for a class of problems (Sein et al. 2011): the design of in-class required introductory college course that can scale to large numbers of students, under resource constraint. In keeping with design science fundamentals (Walls et al. 1992; Hevner et al. 2004; Sein et al. 2011), we derive requirements from kernel theories and formulate design principles to guide the development of the system. Once the ST artifact is implemented, we leverage observations of usage to evaluate results and, refine the design principles and implementation.

3.1. Kernel Theory: Intervention Theory

In *Intervention Theory and Method*, Chris Argyris posits that “to intervene is to enter in an ongoing system of relationships [...] An intervenor, in this view, assists a system to become more effective in problem solving, decision making and decision implementation in such a way that the system can continue to be increasingly effective in these activities and have a decreasing need for the intervenor” (Argyris 1970). Intervention theory identifies three principles that guide the design of interventions: leveraging valid and useful information, allowing free informed choice by the client, and fostering internal commitment. Valid information is that which can be verified and has been shown to affect the phenomena the intervenor is seeking to influence. Useful information is that which the client would be able to use to “control their destiny” (Argyris 1970). From an intervention theory standpoint, natural aptitude provides valid, but not useful information while study habits provide both valid and useful information. Free informed choice points to the centrality of the client in the implementation of the intervention – and therefore in its design. A free and informed choice is particularly important in situations like college learning, where internal commitment is a precondition to the success of the intervention. 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.

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).

3.2.Meta-Requirements and Design Principles Discovery

We conceptualize the course as an ST artifact. Thus, drawing on intervention theory, we advance meta-requirements for system design that include both IT-dominant and organization-dominant elements (Sein et al. 2011). We organize the meta-requirements along the three principles of intervention theory.

3.2.1. Valid Information through Digital Data Streaming

Since the introduction of computers in business and organizations, information systems theorists have recognized the potential of information technology to create digital representations of processes and activities (Yoo 2010). The computer mediation of everyday activities generates digital representations of events at their inception, a phenomenon known in the literature as digital data genesis (Piccoli and Watson 2008). A Digital Data Stream (DDS) is a "continuous digital encoding and transmission of data describing a related class of events" (Pigni et al. 2016). A DDS channels the digital representation of a class of events at their inception and makes them available for harvesting by organizations (Piccoli and Pigni 2013). In some instances, the DDS is available as a byproduct of existing systems, in others, the firm consciously generates it by deploying the needed technology. We argue that the digital transformation of an introductory college course should focus on tapping into, or generating, the DDS that yield valid and useful data. This meta-requirement calls for the proactive design of real-time data collection of students' behaviors.

MR1: The ST artifact supporting large introductory college courses should record all students' behaviors.

- DP1.1: Generate an attendance DDS for all physical activities (e.g., class session, lab sessions).
- DP1.2: Generate a resources utilization DDS for required and optional resources (e.g., readings, homework).
- DP1.3: Generate a completion and performance DDS for required and optional assignments.
- DP1.4: Generate an interaction DDS for tracking communication.

3.2.2. Free and Informed Choice through Learning Analytics

The advent of Learning Management Systems (LMS) enables the digital data genesis of student activities in and outside of the class. Modern LMS track such variables as login frequency, time spent on the system, download of materials and resources, completion of exercises, communication and the like (Mwalumbwe and Mtebe 2017). This ready availability of digital data spurred the development of learning analytics to implement "analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals" (Van Barneveld et al. 2012).

While early work in learning analytics focused on empowering administrators, intervention theory calls for using learning analytics to empower students' free and informed choice. This meta-requirements calls for placing the locus of decision making with students, once they are armed with valid and useful information, rather than using grades as a mechanism to enforce behavior (Bain 2004).

MR2: The ST artifact supporting large introductory college courses must not conflate behavior with learning

- DP2.1: Activities tracked through DDS have no bearing on the student's learning assessment (i.e., the grade).
- DP2.2: Assignments and homework are a service to students and have no bearing on the student's learning assessment.
- DP2.3: Learning assessment is measured, independently of student behavior, through dedicated ad-hoc evaluations (i.e. exams).

MR2 aims at eliminating external incentives for counterproductive behaviors (e.g., coming to class only to accrue "participation" points, cheating on homework, or engaging in strategic learning). It also fosters learner control, the ability of students to "make their own decisions regarding some aspects of the path, flow, or events of instruction" (Williams 1996). While behaviors and activities are not indicative of performance and skill acquisition, they are important elements of the learning process. One of the risks of learner control is lack of information, whereby students do not realize how well they perform until it is too late to redress the situation (Pistilli and Arnold 2010). Thus, an important characteristic of the class of ST artifacts for introductory university courses is the ability of students to engage with the material and receive appropriate feedback.

MR3: The ST artifact supporting large introductory college courses treats students as self-responsible units and maximizes learner control

- DP3.1: Regular homework and practice assignments are available to students.

- DP3.2: Evaluations and feedback are provided for any assignment students voluntarily submit.
- DP3.3: Assignments are designed to be automatically evaluated and students are directed to physical interactions (e.g., lab hours) to discuss the results if clarification is needed.

An important tenet of intervention theory is that free choice must be *informed choice*, based on valid and useful information. Note that this requirement goes beyond showing students how they performed on homework assignments or examinations. Rather it calls for exposing students to data about all their behaviors and to alert them with decision-making data that provides an interpretation of these behaviors. For example, with access to ever more comprehensive DDS (MR1), it is possible to reliably classify students in risk categories. However, this predictive data is rarely made systematically available to students themselves (Jayaprakash et al. 2014) despite the fact that simple interventions such as alert emails that spur faculty-student interactions lead to better retention rates (Tinto 2012).

MR4: The ST artifact supporting large introductory college courses exposes all behavioral and performance data as soon as it becomes available

- DP4.1: Provide a dashboard for visualizing students' individual behavior and performance.
- DP4.2: Apply learning analytics techniques to identify and alert at-risk students.

MR5: The ST artifact supporting large introductory college courses contextualize behavioral and performance data for students.

- DP5.1: Provide a dashboard for visualizing anonymized aggregated current student cohort behavior and performance
- DP5.2: Provide a dashboard for visualizing anonymized behavior and performance by previous student cohorts.

3.2.3. Internal Commitment through Persuasive Technology

The third principle of intervention theory is internal commitment. Designing technology for maximum influence is the realm of the emerging field of persuasive technology. The computer mediation of everyday activities (Yoo 2010) elevated the role of IT to that of a potential agent of persuasion (Nass 2010). Persuasive technology, defined as "any interactive computing system designed to change people's attitudes or behaviors" (Fogg 2003), can therefore be an agent of influence by delivering persuasive stimuli (Fogg 2009) designed to influence the recipients to form, reinforce or change their attitudes or behaviors (Oinas-Kukkonen 2013). Modern college students are digital natives, comfortable users of personal IT and smartphones. With such audience, the use of triggers – those calls to action "that tells people to perform a behavior now" (Fogg 2003) – can be very promising. This meta-requirement calls for the utilization (MR6) and attentive design (MR7) of IT-enabled signal, spark and facilitator triggers to foster students' internal commitment during the course.

MR6: The ST artifact supporting large introductory college courses proactively triggers appropriate behaviors

- DP6.1: Utilize signal triggers to remind students of deadlines and commitments (e.g., assignment deadlines).
- DP6.2: Utilize spark triggers to alert at-risk students and urge them to action.
- DP6.3: Utilize facilitator triggers to reduce obstacles to performing appropriate behaviors (e.g., prompting a "question of the day" through a conversational interface).

While the notion of triggering is intuitively appealing, the difficulty lies in triggering the behavior at the appropriate place and time to prompt action without frustrating or annoying the recipient (Intille 2004). For example, urging action at times when students are unable to perform it risks causing frustration (Fogg 2009). Thus, triggering activities must leverage individual preferences.

MR7: The ST artifact supporting large introductory college courses encourages sustained use by students by managing triggering risks.

- DP7.1: Signal triggers are contextually aware (e.g., reminders are targeted, rather than unqualified "gentle reminders").

- DP7.2: Students can customize the acceptable triggering window (e.g., time of day, day of week) or suspend triggers (e.g., mute for the day).
- DP7.3: Students can manage the type of triggers they receive (e.g., Requesting a "question of the day").

4. Artifact Implementation

In this section we report on the implementation of the ST artifact and the design iterations over the last twelve months. Note that while we report organizational and IT interventions separately for clarity of exposition, our focus on the ST artifact ensured the reciprocal shaping between artifact building, organizational intervention and evaluation (Sein et al. 2011). We also report preliminary evaluations from the first full implementation (i.e., first in-class section with 37 students in Fall 2017).

4.1. Organizational Intervention

The course was scheduled on Tuesdays and Thursdays for 80 minutes sessions. 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 (see the appendix for the complete list of topics). The session 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. Each assignment followed an ongoing scenario that spanned the duration of the semester and contained links to the official Microsoft documentation (both descriptive and video sources) for the skills it required (see Figure 1 for a sample). The scenario gave continuity and realism to the work so that students could see how their acquisition of increasingly complex skills would translate into their ability to carry out increasingly complex work.

The ST artifact design implemented MR2 by ensuring that no DDS data tracked during class or by the course application (see below) was used to compute students' grade (DP2.1). All assigned activities, such as practice assignments, were assessed if the students completed them, but did not contribute to the final evaluation (DP2.2). Students' mastery was measured by way of two exams covering practical skills, five checkups and a final team project covering theoretical material (DP2.3).

The ST artifact design implemented MR3 by providing nine practice assignments over the course of the semester. As per the design principles in this meta-requirement, none of the assignments was required. Yet, work submitted within the expected one-week deadlines was evaluated (DP3.1). Students received a detailed task-by-task report for each assignment, enabling them to see exactly which skill they had not yet mastered (DP3.2). The report listed, for each required task, the importance of the task (i.e., assigned points value) and the actual points earned by the students (i.e., a percentage of successful completion). Students had access to the practice solutions through the course app. However, reports where not intended to substitute for human interaction. Rather, they directed students to meet with the instructor or the teaching assistants for any questions or clarification students could not answer independently (DP3.3).

A pedagogical design imperative for the practical section of the course is the use by each student of their own personal computer. Unlike the previous online design, which forces everyone to learn the Windows version of Word and Excel by relying on a browser-based simulated environment, our pedagogy requires students to work on their own machines with a full version of the Microsoft Office applications being learned. This approach is realistic, and it mirrors the way students would use the software for school projects or at work. For as well designed as a simulated environment might be, it forces learners to practice individual tasks, resulting in learning that is formulaic and disjointed. Moreover, the simulated environment drastically limits the type of mistakes and "alternative solutions" the student can practice. Thus, students don't develop an overall conceptual understanding of how to use the application to achieve their goals (e.g., an analysis in Excel, standardized and efficient document design in Word).

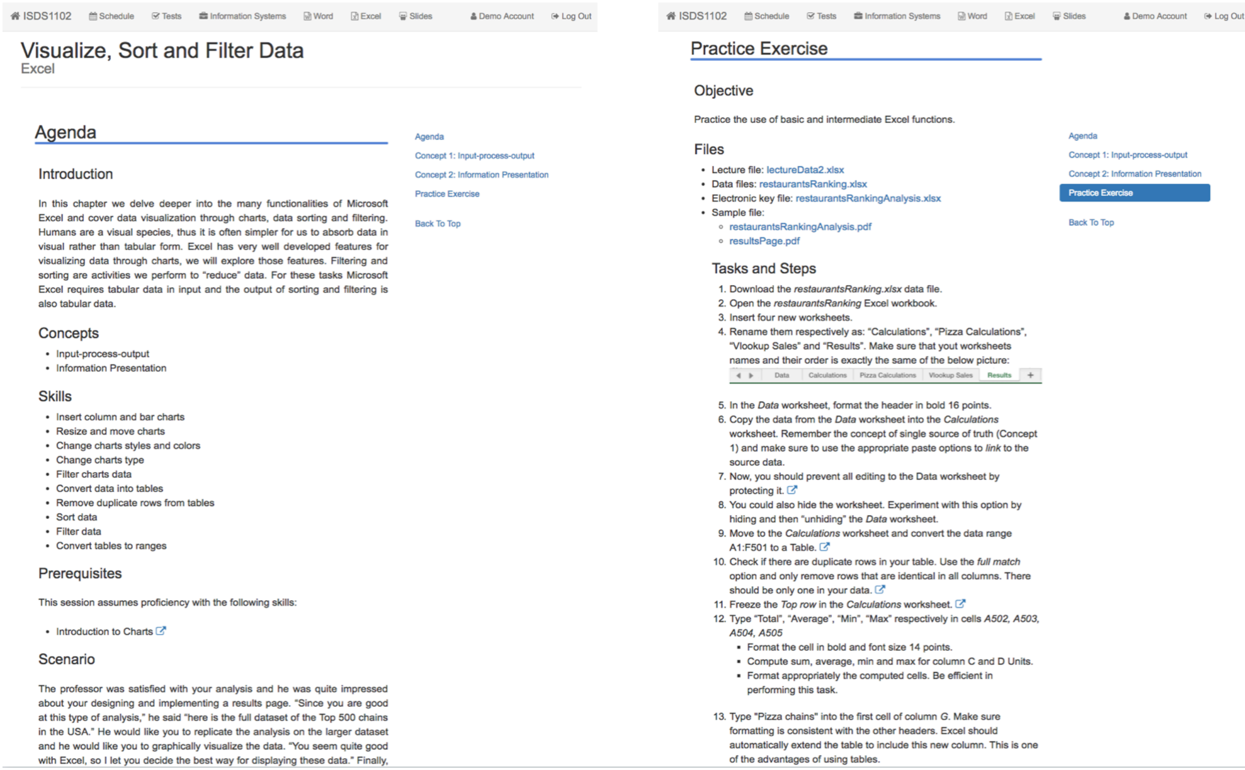

Figure 1: Example of Practice assignment

Given our scalability objective, we developed a custom-made solution to automatically evaluate students' performance in the practice assignments (DP3.3). Developing a custom solution was necessary because there is no available application in the market that would support both Windows and Mac versions of Microsoft Office without relying on a simulated environment. More importantly, we had to be able to create synergy between the structure of the practice assignments and the automatic grader as a prerequisite to the design-build-evaluate cycle (Hevner et al. 2004) required by the DSR approach. Leveraging the fact that Microsoft Office documents are collections of XML files, we used Python to implement grading functions. The software program takes as an input a "key file" created by the instructor and, recursively, each student file. It first extracts the content of the documents by parsing the XML using the *minidom* package. For each practice assignment, a grading sequence is created by calling generic functions that evaluate each skill that comprise the practice assignment using the *numpy* package to perform calculations on arrays and the *prettytable* and *csv* packages to output results (see Figure 2 for sample code). These functions seek a match between the student work and the key, but they are robust to acceptable alternative solutions. For example, if students do not construct a function exactly like the teacher, but necessary elements are present (e.g., the name of the function, the cell range) and results are correct, the students will receive credit for the work. Note however that the automatic grading software is not designed to minimize human interaction by attempting to provide definite answers or pointers to self-help material. Rather it aids students in verifying their own progress, and it stimulates human interaction when needed. In other words, armed with the report from the automatic grading system, the online assignment key and available sample files, the students review their work and evaluate whether they have mastered the skill in the assignment. If they do not understand why a specific task is incorrect, or how to correct errors they have made, they are directed to visit open labs where they received help from the instructor or a TA.

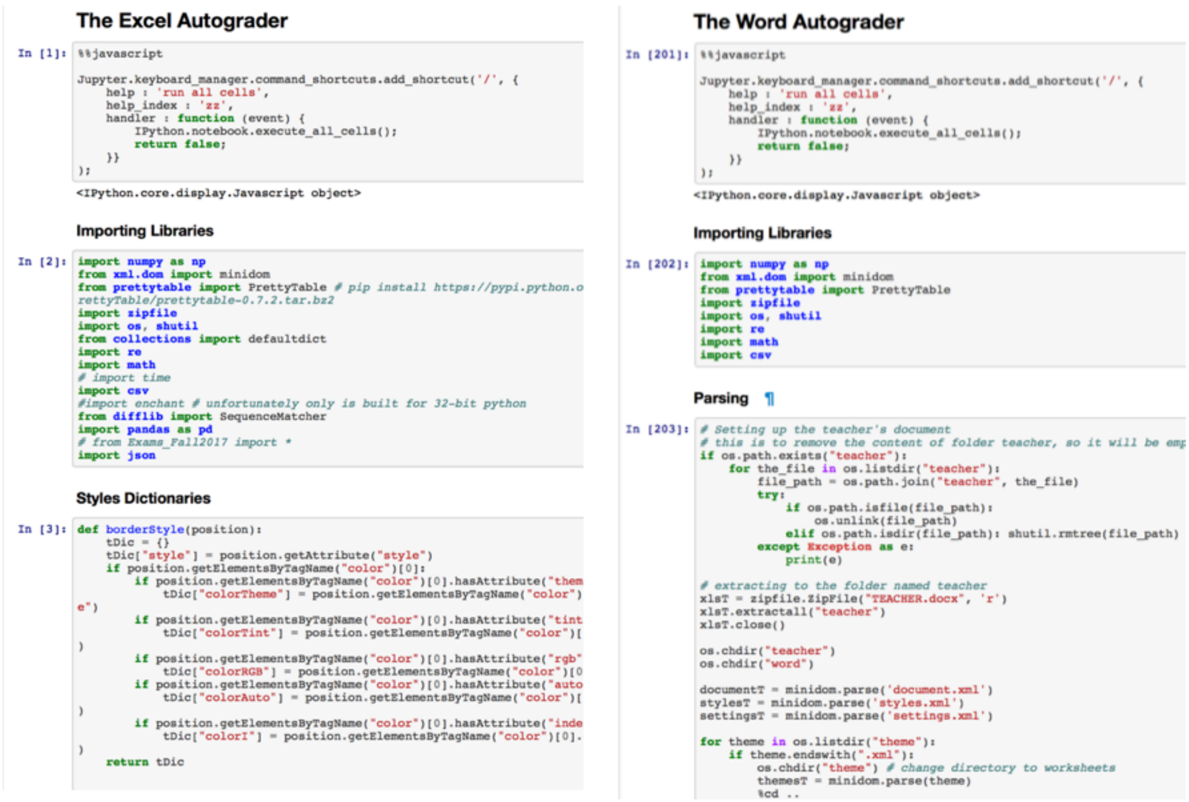

Figure 2: Sample automatic grader code

By controlling the automatic grading system, we have full control of the entire process: from skill identification, to framing of exercises to acquire and practice those skills, to the evaluation approach and the degrees of freedom allowed in the assessment of student work. Moreover, knowing the structure, strengths and limitations of the grading software enables us to design practice assignments that are simultaneously realistic, pedagogically sound and evaluable at scale. For example, if students must learn how to use different types of conditional formatting rules in Microsoft Excel (e.g., "Highlight Cells Rules" and "Top/Bottom Rules") and the automatic grader cannot evaluate multiple conditions on the same column, we structure the assignment so that students apply the first rule on one column and the second rule on a different column. In this way, both skills are tested, yet evaluation can reliably occur at scale.

4.2.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). We chose cloud hosting because it frees us from server maintenance and security responsibilities, while at the same time providing computing and cost scalability proportional to the number of users and calls to the host server. The efficiency thus gained enables our small team to focus on features development and functionality evolution enabled by the reciprocal shaping between artifact building, organizational intervention and evaluation during each semester iteration (Sein et al. 2011). The performance of the current iteration was set to scale reliably up to 45 concurrent users using a virtual machine with one Intel Xenon family CPU with up to 3.3 GHz and 2GB of RAM (an AWS "t2.small" instance).

For the application front-end we adopted the Bootstrap library and the Embedded JavaScript (EJS) template language – ensuring native mobile and desktop compatibility and a responsive app on all platforms. The application was entirely developed with HTML, JavaScript, and Cascading Style Sheets. The code was version controlled using Git, and it was hosted on GitHub to facilitate collaborative coding. The use of free and open-source technologies for developing and hosting the application was imperative given the resource constraints underpinning the project, but also optimal given our design objectives of high scalability and reliability.

The ST artifact design 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 was tracked and monitored using Google Analytics' (GA) free service. GA enabled the collection of students' behavior data at the session, page, and click event level (DP1.2). During the current iteration, the data stream from the Web application was the only one generated on a real-time basis. We manually recorded attendance to physical activities, such as lab and class sessions (DP1.1). The architecture for tracking completion and performance of required and optional assignments (DP1.3) as well as communication outside of physical meeting venues is also in place (DP1.4) but we have yet to leverage the data for artifact evaluation.

The ST artifact design partially implemented MR4 and MR5 in the current iteration. Specifically, in the current instantiation of the ST artifact students do not have access to a dynamic dashboard and we did not implement any learning analytics models. Instead, we experimented 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 can be thought of as a static version of the dashboard we are currently building. Each report exposes to students their data, detailing behaviors enacted (e.g., attendance, access to the materials) and performance on evaluated tasks (DP5.1). 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.

5. Evaluation

At this stage we claim an "improvement" knowledge contribution, focused on "developing new solutions for known problems" (Gregor and Hevner 2013, p. 345). This type of contribution should be evaluated based on its ability to overcome current suboptimal solutions to the class of problems addressed by the ST artifact. A first step in claiming a knowledge contribution of this type is to demonstrate feasibility (Hevner et al. 2004). 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 show that the ST artifact can be built, implemented, and that it will be used by the intended audience. We offer such "proof-by-demonstration" (Nunamaker Jr et al. 1990, p. 98) as the first step in informing the reciprocal shaping between the IT artifact and the design of the course. In this section, we present the evaluation of the ST artifact and we use it to gather essential feedback (Hevner et al. 2004) to feed into the next iteration of artifact construction (see discussion section).

Our evaluation is based on the analysis of the behavioral DDS we collected throughout the semester (MR1), the results of a survey at the end of the semester inquiring about individual app use and seeking feedback about the ST artifact design (see Appendix), as well as the standard course evaluation issued by the administration.

With respect to MR1, we find our design to enable the generation of need DDS. Specifically, only 5.56% of students reported printing all the theoretical materials and 7.43% reported printing all the practice assignments. Thus, we are able to collect accurate resource utilization data (DP1.2) with an approach that is scalable since our app resides on the AWS infrastructure. 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 DDS (DP1.3). All interaction that does not occur during face to face meetings is logged in the communication channels used to support the course - email, and Slack – thus enabling the unobtrusive generation of interaction DDS (DP1.4). The current design relies on manual tracking of attendance to all physical activities (DP1.1). This approach is not scalable, and we are currently working on solving this problem (see discussion).

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%. Of the 30 students who completed the course, four attended every class and all but three attended at least two thirds of the sessions. While attendance did not statistically correlate to performance (see below), it supports our effort to foster human interaction in the learning environment. The formal course evaluation showed that the course was well received, with an overall result on a four-point scale at 3.53 (college total 3.35). Questions influence by the course design included: perception of fairness in evaluation (3.56 vs 3.38), assessment of quality of the feedback (3.44 vs 3.28) and stimulation of interest in the subject matter (3.67 vs 3.32).

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. The results indicate that, while students would start the assignment in class, they rarely completed the work during the allotted class time. As a consequence, they would not submit it for evaluation. A number of students stated in their feedback that when they sat down to complete the assignment at home, the length and complexity of the practice exercise would discourage them. Specifically, 22 out of the 27 respondents agreed with the statement that "shorter practice assignments would be more effective." In other words, in the first iteration of the implementation the design failed to ensure that students would stay on track with the progression of material without using grades to stimulate activity. In the discussion we address how the ST artifact design has evolved based on this result.

Also telling was students timing of preparation for the checkups covering theoretical material. Each checkup covered one or two chapters over two weeks (sessions only on Tuesdays). While session started two weeks before the evaluation, students could sign up for the checkups at the testing center on any day between Monday and Friday during the checkup week (represented by the black horizontal dashed line in Figure 3). Access to the tested material occurred immediately before the exam, with the bulk of it during the testing time span.

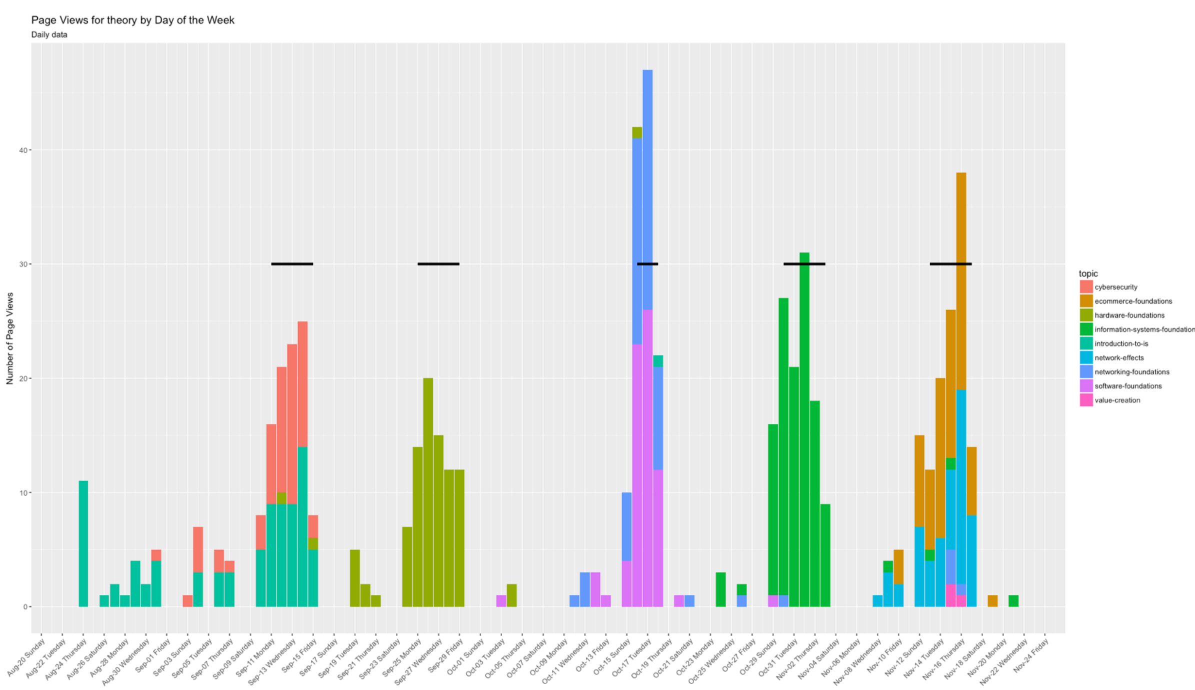

Figure 3: Checkup weeks and content access

More specifically, on average students prepared for the checkup within a two-day window of taking the test. While it would not surprise experienced teachers to find out that students tend to study in the vicinity of examinations, one of the design objectives for our ST artifact is indeed to stimulate a more even and consistent approach to preparation – without using grades as the incentive. The results of the first iteration reiterate the importance of implementing MR6 and MR7 to proactively help students take ownership of the learning experience. We also believe that a faster cycle of feedback (see discussion) would stimulate students' attention toward the course.

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 official Microsoft documentation. 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:

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

Where 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) and the extent of external resources utilization (linkdiv). Linkfr is measured as the total number of external links used, while linkdiv is 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.

6. Discussion

The above evaluation was not designed to study how the ST artifact impacted students' motivation or performance. Rather, the evaluation served 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. Hosting the technology on AWS enables almost infinite scalability as the number of students grows. The implementation of the design principles associated with MR2 and MR3 was successful. Our evaluation shows that the automatic grader is reliable (DP3.3), but the timing of feedback (DP3.2) 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. Moreover, we have rolled out a drop-file feature in the app that streamlines the workflow for the student. Rather than emailing long practice assignments upon completion, the students can drop each completed file and resume work outside of class on the remaining ones. Each chunk provides a data file that is equivalent to the key of the previous chunk, thus reducing carry-over errors, a change that improves both student learning, and the precision of the evaluation produced by the automatic grading software.

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. An added advantage of the current redesign is that the current workflow can be further automated, thus freeing resources from the evaluation and feedback task. The automatic grading software has reached a level of precision that no longer requires manual checking of the output. We are in the process of fully automating it by leveraging compute services that support the running of custom code in the cloud (AWS Lambda). When a student drops a file in the app, the file is stored in an Amazon S3 bucket. Once the grading software is successfully ported to the AWS Lambda service, a message will trigger its execution. The results of the evaluation, a JSON file, is stored in MongoDB and becomes immediately visible to the students inside the app. This evolution of the automatic grader, currently in development, would achieve the fastest cycle of feedback on practice files (i.e., near real-time) in a scalable workflow that could accommodate all the students concurrently taking the course.

Beyond course material changes, such as the "chunking" of large practice exercises, and the app feature changes designed to support them, we have structured the evolution of design principles along two vectors: improved data collection, and persuasive technology triggers. 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 in the 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 widgets (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 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.

The second challenge we face, as captured by our results, is the need to think creatively about scalable systems that aid in influencing students to practice good learning habits and combat strategic learning. In smaller classes, the best teachers are able to motivate students without resorting to the use of requirements and grades (Bain 2004). How can the same objective be achieved in larger required courses? We believe that persuasive technology (Fogg 2009) and digital nudging (Weinmann et al. 2016) hold promise. We are developing two types of triggers, through conversational interfaces, for the next iteration of the course: a)

Facilitator triggers, designed to reduce barriers to accomplishing the behavior (e.g., a "question of the day" trigger designed to stimulate both interest and reflection on the subject matter covered during the week); b) Spark trigger, designed to increase motivation (e.g., alert triggers for ensuring that all students stay on task and to raise awareness of at-risk students who are in danger of falling behind).

7. Conclusions

The higher education "industry" has not been immune to digital innovation and digital transformation. However, there is a growing consensus in the literature that the delivery of high-quality college education hinges on the instructor's ability to engineer a learning environment where students can learn effectively (Bain 2004). Thus, we advocate for an approach to the digital innovation in education that leverages, rather than "digitize," the instructor. Innovation is increasingly the outcome of dynamic problem-solution pairings (von Hippel and von Krogh 2015), and we argue that the design science approach is best suited to identify and advance optimal designs. We consider our socio-technical artifact an early example of what's possible. We continue to iterate the design-build-evaluate cycle of design science research to uncover and solidify the design principles of ST artifacts. Those principles should help us address the wicked design problem of delivering a required in-class introductory college course, one that is efficiently scalable to large numbers of students.

References

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

Author Cited Blinded for peer review

Bain K (2004) What the Best College Teachers Do, 1 edition. Harvard University Press, Cambridge, Mass

Bandiera O, Larcinese V, Rasul I (2010) Heterogeneous Class Size Effects: New Evidence from a Panel of University Students*. Econ J 120:1365–1398. doi: 10.1111/j.1468-0297.2010.02364.x

Bernard RM, Abrami PC, Borokhovski E, et al (2009) A Meta-Analysis of Three Types of Interaction Treatments in Distance Education. Rev Educ Res 79:1243–1289. doi: 10.3102/0034654309333844

Blikstein P, Worsley M (2016) Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. J Learn Anal 3:220–238. doi: 10.18608/jla.2016.32.11

Buchanan R (1992) Wicked Problems in Design Thinking. Des Issues 8:5–21. doi: 10.2307/1511637

Chambliss DF, Takacs CG (2014) How College Works, 1 edition. Harvard University Press, Cambridge, Mass

Chatterjee P, Nath A (2014) Massive open online courses (MOOCs) in higher education—Unleashing the potential in India. In: MOOC, Innovation and Technology in Education (MITE), 2014 IEEE International Conference on. IEEE, pp 256–260

Chickering AW, Gamson ZF (1987) Seven principles for good practice in undergraduate education. AAHE Bull 3:7

Dillahunt TR, Wang BZ, Teasley S (2014) Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. Int Rev Res Open Distrib Learn 15:

Fogg B (2009) A Behavior Model for Persuasive Design. In: Proceedings of the 4th International Conference on Persuasive Technology. ACM, New York, NY, USA, p 40:1–40:7

Fogg BJ (2003) Persuasing Technology. Using Computers to Change What We Think and Do. Morgan Kaufmann Publishers

Gebre E, Saroyan A, Bracewell R (2014) Students' engagement in technology rich classrooms and its relationship to professors' conceptions of effective teaching. Br J Educ Technol 45:83–96. doi: 10.1111/bjet.12001

Gregor S, Hevner A (2013) Positioning and Presenting Design Science Research for Maximum Impact. Manag Inf Syst Q 37:337–355

Hevner AR, March ST, Park J, Ram S (2004) Design Science in Information Systems Research. MIS Q 28:75–105

Hew KF, Cheung WS (2014) Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. Educ Res Rev 12:45–58. doi: 10.1016/j.edurev.2014.05.001

Intille SS (2004) Ubiquitous computing technology for just-in-time motivation of behavior change. Stud Health Technol Inform 107:1434–1437

Jayaprakash SM, Moody EW, Lauría EJM, et al (2014) Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. J Learn Anal 1:6–47. doi: 10.18608/jla.2014.11.3

Kalay YE (2004) Virtual learning environments. J Inf Technol Constr ITcon 9:195–207

Kokkelenberg EC, Dillon M, Christy SM (2008) The effects of class size on student grades at a public university. Econ Educ Rev 27:221–233

Markus ML (2005) Technology-Shaping Effects of E-Collaboration Technologies: Bugs and Features. Int J E-Collab IJeC 1:1–23. doi: 10.4018/jec.2005010101

Mwalumbwe I, Mtebe JS (2017) Using Learning Analytics to Predict Students' Performance in Moodle Learning Management Systems: A case of Mbeya University of Science and Technology. Electron J Inf Syst Dev Ctries 79:

Nambisan S, Lyytinen K, Majchrzak A, Song M (2017) Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. Manag Inf Syst Q 41:223–238

Nass C (2010) The Man Who Lied to His Laptop. What Machines Teach Us About Human Relationships. Current

Nunamaker Jr JF, Chen M, Purdin TDM (1990) Systems Development in Information Systems Research. J Manag Inf Syst 7:89–106. doi: 10.1080/07421222.1990.11517898

Oinas-Kukkonen H (2013) A foundation for the study of behavior change support systems. Pers Ubiquitous Comput 17:1223–1235. doi: 10.1007/s00779-012-0591-5

Palen L, Grudin J (2003) Implementing Collaboration Technologies in Industry. In: Munkvold BE (ed). Springer-Verlag, London, UK, UK, pp 159–179

Piccoli G, Pigni F (2013) Harvesting External Data: The Potential of Digital Data Streams. MIS Q Exec 12:143–154

Piccoli G, Watson RT (2008) Profit from Customer Data by Identifying Strategic Opportunities and Adopting the" Born Digital" Approach. MIS Q Exec 7:

Pigni F, Piccoli G, Watson R (2016) Digital Data Streams. Calif Manage Rev 58:5–25. doi: 10.1525/cmr.2016.58.3.5

Pistilli MD, Arnold KE (2010) In practice: Purdue Signals: Mining real-time academic data to enhance student success. Campus 15:22–24

Sein M, Henfridsson O, Purao S, et al (2011) Action Design Research. Manag Inf Syst Q 35:37–56

Silver MS, Markus ML (2013) Conceptualizing the SocioTechnical (ST) artifact. Syst Signs Actions 7:82–89

Tinto V (2012) Enhancing student success: Taking the classroom success seriously. Int J First Year High Educ 3:1

Van Barneveld A, Arnold KE, Campbell JP (2012) Analytics in higher education: Establishing a common language. Educ Learn Initiat 1:l–ll

von Hippel E, von Krogh G (2015) CROSSROADS—Identifying Viable "Need–Solution Pairs": Problem Solving Without Problem Formulation. Organ Sci 27:207–221. doi: 10.1287/orsc.2015.1023

Walls JG, Widmeyer GR, El Sawy OA (1992) Building an Information System Design Theory for Vigilant EIS. Inf Syst Res 3:36–59. doi: 10.1287/isre.3.1.36

Watson R, Lind M, Haraldson S (2017) Physical and Digital Innovation in Shipping: Seeding, Standardizing, and Sequencing. In: Proceedings of the 50th Hawaii International Conference on System Sciences. pp 4756–4765

Weinmann M, Schneider C, Brocke J vom (2016) Digital Nudging. Bus Inf Syst Eng 58:433–436. doi: 10.1007/s12599-016-0453-1

Williams MD (1996) Learner-Control And Instructional Technologies. In: Handbook of Research for Educational Communications and Techno, D. H. Jonassen. Simon and Schuster Macmillan, New York

Yoo Y (2010) Computing in Everyday Life: A Call for Research on Experiential Computing. Mis Q 34:213–231

Appendix: Course Topics

Table 2 Course topics: Theory and Practice

Tuesdays: Theory Topics	Thursdays: Practice Topics and Number of Tasks	
Introduction to Information Systems	MS Word: Creation of a Simple Document	31
Cybersecurity	MS Word: Template Creation and Personalization	49
Hardware Foundations	MS Word: Creation of a Complex Document	31
Software Foundations	MS Excel: Introduction to Formatting	21
Networking Foundations	MS Excel: Using Functions	29
Information Systems Foundations	MS Excel: Worksheet and Workbook Management	46
Electronic Commerce and IT-enabled Innovation	MS Excel: Visualize, Sort and Filter Data	52
Network Effects and Information Economics	MS Excel: Summarizing Data with Pivot Tables	32
Value Creation with Information Technology	MS Excel: Finalizing your Report	23

Customer Satisfaction Survey

Name ____________________ Computer Platform ______________________________

1. Did you find the design of the app useful?

- ☐ Yes
- ☐ No

1a. Did you print any of the MS Word Projects?

- ☐ Everyone
- ☐ Some
- ☐ Maybe one
- ☐ None

1b. Did you print any of the MS Word Projects?

- ☐ Everyone
- ☐ Some
- ☐ Maybe one
- ☐ None

1c. Did you print any of the Book Chapters?

- ☐ Everyone
- ☐ Some
- ☐ Maybe one
- ☐ None

2. Did you find the design of the Word and Excel practice useful?

- ☐ I did not use them
- ☐ Yes
- ☐ No

2a. Did you find the graded reports useful?

- ☐ I did not get them
- ☐ Yes
- ☐ No, they took too long
- ☐ No, they were unclear

2b. Would shorter projects be more effective? Why (please write comments here)?

- ☐ Yes
- ☐ No

3. Did you use external resources?

- ☐ No
- ☐ Open Lab
- ☐ Checkups SI review
- ☐ Word/Excel SI Reviews

4. Would you be willing to be part of a focus group to make the App better?

- ☐ Yes
- ☐ Sorry, I can't

5. Provide any other suggestions for the class or the resources we use: