10-1-2019

Implementing online personalized social comparison nudges in a web-enabled coaching system

Michael G. Brown
Iowa State University, brownm@iastate.edu

James J. Schiltz
Iowa State University, jschiltz@iastate.edu

Holly Derry
University of Michigan-Ann Arbor

Caitlin Holman
University of Michigan-Ann Arbor

Implementing online personalized social comparison nudges in a web-enabled coaching system

Abstract

Descriptive norm and opinion leader nudges were tested using a randomized control trial.

Opinion leader nudges have more impact than normative nudges, albeit marginally.

Students with low levels of grade importance responded positively to multiple nudges.

The same group of students who received multiple nudges also ended the course with higher grades.

Disciplines

Curriculum and Instruction | Educational Assessment, Evaluation, and Research | Educational Leadership | Social and Philosophical Foundations of Education

Comments

This accepted article is published as Brown, M., Schlitz, J., Derry, H., Holman, C. (2019). Implementing online personalized social comparison nudging in a web-enabled coaching system. *Internet and Higher Education*. 43; 100691. Doi: 10.1016/j.iheduc.2019.100691. Posted with permission.

1. Introduction

Large face-to-face lecture courses in undergraduate education rarely afford the opportunity for instructors to provide students personalized feedback to guide revision and refinement of their approaches (Brown, 2016; Chamblis & Tackacs, 2014). Instead, students are generally forced to rely upon decontextualized feedback like performance assessments to make changes to their study behavior and coursework strategies. In the event that they struggle, students in large lecture courses are often left without specific direction or ideas for how to improve their approach, and thus performance.

One approach to addressing the lack of personalized feedback in large lecture courses involves helping students see their behaviors in the context of others in the course, for example, with digitally delivered nudges (Fritz, 2017). These tools are designed to help struggling students reflect on and (potentially) make changes to their study strategies (Fritz, 2017). Digital nudges allow instructors to provide personalized feedback at scale in large lecture courses based on the self-reported and observed study behaviors of students in the course (McKay, Miller, & Tritz, 2012).

This approach to feedback builds on what Thaler and Sunstein (2008) refer to as 'choice architecture', where individuals are free to choose among a set of alternatives, but the experience is designed in such a way that it encourages positive decision-making. Choice architecture interventions have been used, for example, to facilitate the course selection process during academic advising appointments by helping students understand the potential impact of co-enrollment in difficult courses on their predicted odds of academic success (Denley, 2014). Such an approach might nudge students towards choices that are more likely to result in academic success without constraining their array of potential choices.

Nudges of this sort are different from direct instruction during a lecture because nudging interventions put information directly in front of the subject and can be personalized to the individual (Thaler & Sunstein, 2008). Very little of the literature on choice architecture interventions focuses on formal education contexts. A recent systematic review of nudges (Szaszi, Palinkas, Palfi, Szollosi, Aczel, 2018) examined 156 empirical studies that included 422 different choice architecture interventions. They found that health interventions (efforts to change eating and drinking behavior) made up nearly half (42%) of the interventions they reviewed. The next most common type of interventions focuses on influencing consumer choices (20%), followed by financial decision interventions and environmental sustainability behaviors, each at 18%. Education related studies represented a mere 4% (about 6 studies) of the reviewed research. As digital and web-enabled tools like LMSs occupy an increasingly central role in undergraduate students' academic work (Selwyn, 2014), there exists a significant opportunity to design, test, and implement effective nudges.

Choice architecture interventions in higher education have focused primarily on out-of-class aspects of student life. For example, there is evidence that simple nudge-based interventions can help students clarify their decision to go to college (Castleman, Arnold & Wartman 2012, Castleman Page and Schooley 2014; Castleman & Page 2015), apply and get accepted into more selective colleges (Hoxby, & Turner 2013), increase rates of tutoring attendance (Pugatch & Wilson, 2016), encourage families to fill out college financial aid forms (Bettinger, Long, Oreopoulos, and Sanbonmatsu, 2011; Castleman and Page, 2015; Castleman and Page, 2016), reduce summer 'melt' (Bird et al., 2017; Castleman et al., 2015), engage with academic advisors (Arnold et al., 2015), and complete more credits during freshman year (Castleman & Meyer, 2016). Many of these interventions have lasting effects on college persistence (e.g. Castleman, Page, & Schooley, 2014; Castleman & Page 2016). Nearly all of these nudges are delivered by an institutional agent, rather than a learning management platform or a digital instructional tool. While all of the research focuses on motivating student behavior, very few leverage peer group comparisons, which have proven effective in other contexts (Szazsi et al., 2018). These out of class focused studies (e.g. Bird et al., 2016; Castleman et al., 2015) use rigorous randomized control trials to identify the positive effects of nudges delivered through web enabled tools on student behavior over time.

In contrast, research on the efficacy of providing feedback through the web as a strategy for changing student study behavior in post-secondary classrooms has produced more mixed results than interventions focused on enrollment, persistence, and college choice. For example, students in three different institutional contexts exhibited only marginal improvement in academic performance from an intervention focused on increasing study time where they were provided a customized study schedule, weekly help tips and text messages support from academic coaches (Oreopoulos, Patterson, Petronijevic, & Pope, 2018). Researchers observed no significant changes in outcomes (credit accrual, grades, or retention) in the wake of the treatment, although they did observe that students increased their study time as a result of participation (Oereopoulos, et al., 2018). They argue that future interventions may be more effective if they focus on a specific study behavior, rather than a collection of strategies.

The existing research on providing students' comparative information about their peers' strategies and behaviors through web-enabled messaging also does not do much to clarify the type of information that might be most effective at encouraging reflection and behavior change. That is because much of the extant research are impact studies focusing on comprehensive use of web-enabled messaging tools. For example, use of individually tailored web dashboards were linked with improved academic outcomes at the end of a course, controlling for different levels of use and non-use (Fritz, 2013; McKay, Miller, & Tritz, 2012). Similarly, students who received performance feedback about a course through Purdue's *Signals* early warning system were retained at higher rates than students in courses who did not use the system (Arnold & Pistilli, 2012). Students in courses that used *Course Signals* also out-performed better-prepared peers who were in courses that did not use the system (Arnold & Pistilli, 2012). Given that attendance and enrollment interventions focus on a singular decision (and related behavior) it may be easier to observe their impact. This study follows from that premise–that by focusing on an individual action within course work we may be better able to identify when, how, and with whom to provide personalized information.

While this evidence suggests that providing students information about their behavior and performance in a course produces performance benefits, what is unclear is the mechanism that results in improved student outcomes. In a typical face-to-face course, as part of student interactions, peers may model effective strategies for coursework, which are adopted by others (Zimmerman & Schunk, 2001). Feedback about peer performance and behavior can help raise a student's awareness that changes in behaviors are needed to overcome performance gaps (Fritz, 2014). In this way, information about successful peers can serve as a model that facilitates the development of study strategies and behavioral engagement, an effect most often observed in smaller courses where peers and instructors have more opportunities to interact (e.g. Azevedo and Hadwin 2005; Schünemann, Spörer, Völlinger, & Brunstein, 2017, Schunk and Zimmerman 2007; Spörer and Brunstein 2009). During reciprocal learning interactions, students observe what their successful peers are doing, and they make changes to their approach (Schünemann, Spörer, Völlinger, & Brunstein, 2017). There may very well be a social component to the development of individual study strategies (Schunk & Zimmerman, 1997).

Research is needed that rigorously evaluates different types of digitally provided informational nudging interventions in large post-secondary courses to determine 1) if they result in behavioral change, 2) if behavioral change differs by the kind of nudge used, 3) if dosage matters for nudging behavioral change, 4) if the effectiveness of a nudge varies by different sub-groups in the course, and 5) if different forms of behavioral change resulting from nudging behavior are related to improved outcomes. By making sense of these questions, instructors and technologists will be better able to design personalized tools that produce their intended impact.

Our objective in this study is to observe how students respond to different types of nudging feedback aimed at specific behavioral strategies for completing homework assignments. We focus on an approach to nudging that has shown promising results in other contexts: providing individuals tailored feedback comparing an individual's behavior to peer behavior as delivered through web-enabled tools (e.g. Szaszi, et al., 2018). We expect that through identifying effective personalized messages, we will be able to develop nudges that can be delivered in large undergraduate courses to help students reflect upon and make changes to their approach to coursework. By identifying effective messages, we can also start to explore the role of timing of the nudge in deploying messages, as prior research suggests that the progression of time in a course influences students' engagement (Author-a).

2. Methods and Materials

This study is part of a large design-based implementation research (DBIR) project (Penuel & Fishman, 2012) focused on the development of the ECoach system, a tailored, web-based student support system. ECoach relies on a personalized messaging system, the Michigan Tailoring System, developed by the University of Michigan that provides individually tailored messages about behavior, based on the student's background, psychosocial characteristics, grades, and behaviors (McKay, Miller, & Tritz, 2012). ECoach aggregates students' self-report and learning analytics data about their behaviors to allow instructors and researchers to deliver personalized interventions that help students direct their energy in a course. In this study, we report the results of a year-long investigation of nudging interventions in an introductory undergraduate statistics course. Below we describe the guiding conceptual framework, detail the design of the data collection process, identify data sources, and provide an overview of our data analysis process.

2.1 Conceptual Framework

The ECoach project broadly has adopted a DBIR lens for our work given the ongoing desire to identify interventions that the literature suggests will improve student learning outcomes, assess their situated success as well as their ability to "cross levels and settings of learning" (Penuel & Fishman, 2012, p. 281), and ultimately craft effective *and* deeply personalized student interventions at scale. Core to the mission of this work is answering the many-layered questions of "what works where, when, and for whom" (Means & Penuel, 2005). ECoach has been intentionally designed to support active research with a robust set of tools including: A/B testing, sampled intervention deployment strategies that enable testing in highly-designed populations (accounting for multiple stratification layers as well as sample size necessary to achieve significance) with ease, and the ability to integrate real-time data to power personalized interfaces that streams from multiple sources.

The nudges we employed in this study are based on Münscher, Vetter, and Scheuerle's (2016) choice architecture taxonomy. The authors identify three crucial points in the decision-making process at which individuals would be responsive to interventions. Specifically, nudging could focus on (1) providing information before a decision is made (such as personalizing information about college costs with financial aid, rather than showing the full price), (2) structuring the decision opportunity (such as making college entrance exams the default for all high school students and requiring people to opt out), and (3) assisting with the execution of the decision (such as sending text messages to remind students and parents of college enrollment tasks).

Individuals often need help with decision-making processes when they have to do so in the context of limited information availability (like a large lecture course with few grading opportunities for students). The nudging interventions that are the focus of this study could, for example, translate feedback about a students' performance into a strategy for concrete behavioral change, thereby making information actionable. For the current study, we have focused on providing social reference points to encourage students to start working earlier on assignments. Münscher and colleagues (2016) argue that researchers and practitioners could capitalize on two powerful forms for providing reference points: descriptive norms and opinion leaders. We focus on this strategy because it reinforces the conventional wisdom that students should start early on their assignments, and because this approach has rarely been tested in educational contexts (Szaszi et al., 2018, p. 359).

2.2. Data Collection

Data collection for this study was instrumented through the Canvas Learning Management System (LMS) and ECoach. Log-records of student behavior, like the date and time that students first opened an assignment, are tracked by the LMS. Students also complete surveys in ECoach at various points throughout the term. The most in-depth survey occurs at the beginning of the academic terms, and includes a variety of validated measures used for research purposes as well as more informal and approachable questions used to collect tailoring data about personality, motivation, preferences, and information about their student experiences. ECoach and the LMS are integrated, and ECoach offers a data export that includes ECoach survey responses as well as ECoach and LMS activity data.

2.3 Experimental Design

Prior to this research we used MTurk to evaluate visualization options and make sense of how users might interpret a given visualization. Survey respondents (n=402) were shown two different visualizations for each nudging condition at random: 1) a visualization of two side-by-side boxes showing how far in advance the student opened the assignment and a comparison box showing the start time for either opinion leaders or the class norm and 2) a horizontal bar chart that showed the number of days before the due date that a student started the assignment compared either opinion leaders or the class norm (figure 1). Respondents were shown each visualization type with sample information for each condition (opinion leader and descriptive norms). Survey respondents were shown four visualizations in total, and asked a series of questions about what the images illustrated, what was being compared, how they might change their study strategies in response, and why they would make those changes. We also asked questions about demographics to compare differences in interpretation by gender identity and highest educational credential obtained.

Figure 1. MTurk Experimental Visualization comparing side by side and bar-chart designs

In our MTurk pilot study, the horizontal bar visualization performed best across both conditions on questions about comprehension and what the individual would do differently in response to the visualization (see figure 2 and figure 3). In the development of our nudges, we also consulted with experts in user interface design and information visualization to ensure that our approach accorded with prior research and best practices in these fields.

In our main study, using a new version of the horizontal bar visualizations (figure 2), we tested both descriptive norm and opinion leader reference points by comparing them to a control group of students who received no message. Students were randomly assigned to one of three conditions: a descriptive norm (Average) message, an 'opinion leader' (BetterThan) message, or a control group who received no message. The Average treatment group saw information about the norm–the average start date of students in the course on the prior homework assignment. The BetterThan treatment group saw information about an opinion leader–the average start date of students who performed as well as or better than themselves on the prior homework assignment. Within the two treatment groups, students were also assigned to dosage groups. Half of the students in each treatment group received a second message.

We conducted a power analysis before engaging in the research each semester to determine the size needed for each potential condition to ensure that our groups were large enough to observe a potential effect of nudging on student behavior. Students were blocked by race, gender, cumulative GPA and year in school and then randomly assigned to the intervention or control conditions. However, this assignment was constrained by students' own self-selection into using the platform during the study's time window.

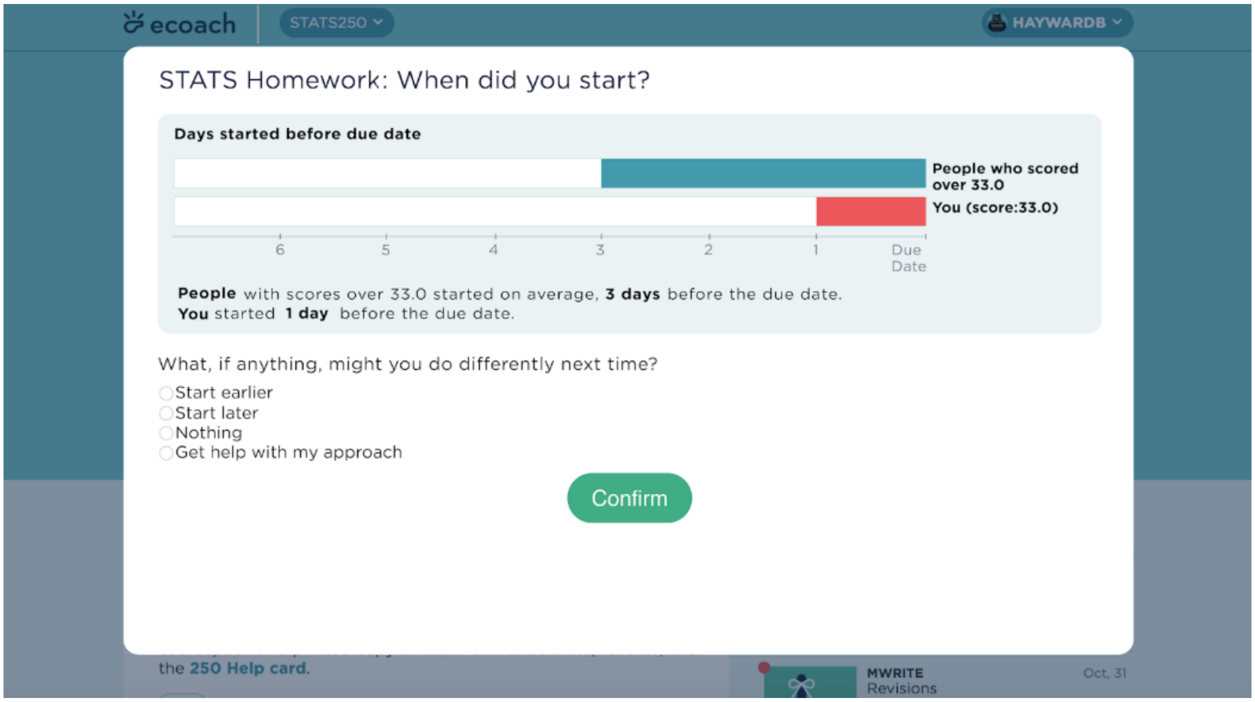

Figure 2. Example Visualization for Opinion Leader nudge

When the nudging intervention was delivered, students viewed the message on the ECoach platform at log-in. This occurred either once before homework 3 (single dose) or before homework 3 and again before homework 6 (two doses). Nudges used student scores and their time before due date measure for the prior homework assignment (so 2 and 5, respectively). The students were delivered information about either the average (Average) or better than (BetterThan) start time on the most recent homework assignment in comparison to their own start time (see figures 2 and 3). Before they could navigate to anything else in the system, students in one of the treatment groups saw a screen entitled "Stats Homework: When did you start?" Students were also asked what they would do differently on the next assignment. They clicked a "Confirm" button to submit their answers.

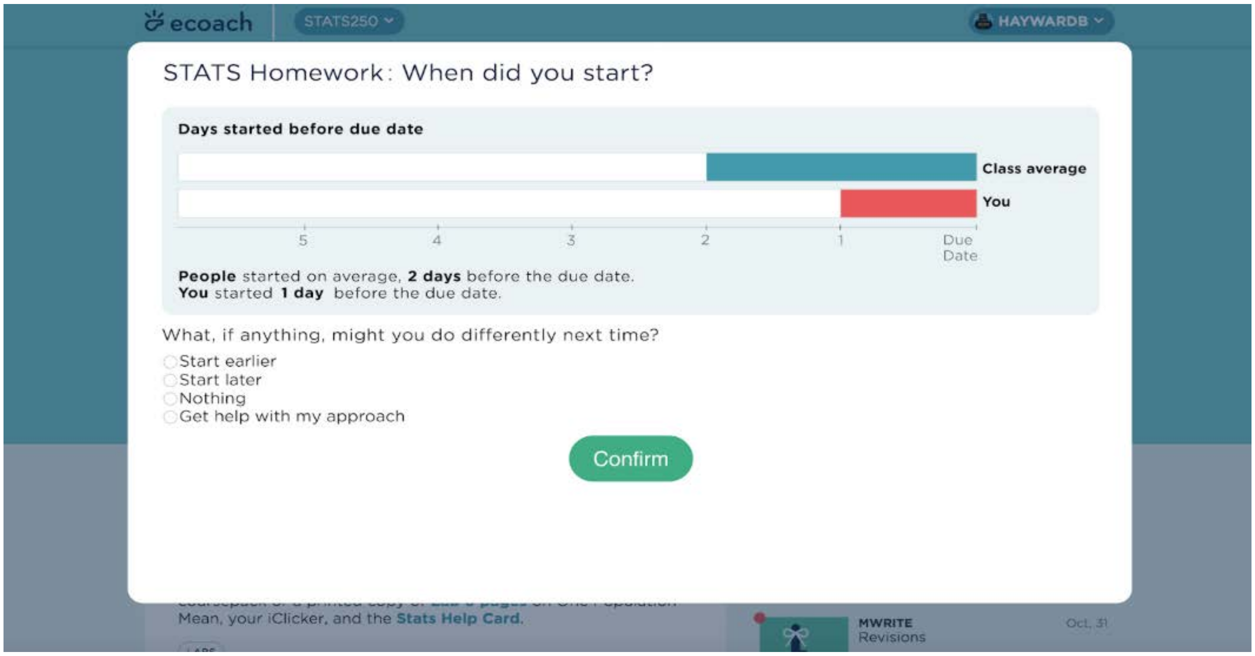

Figure 3. Example Visualization for Descriptive Norm nudge

2.4 Data Sources

We drew data for analysis from four sources. First, a pre-course survey that students completed to access the ECoach system included questions about what grade students wanted to achieve in the course (Grade Expectations), the lowest grade a student would be satisfied with (Grade Satisfaction), and three 10-point scales describing the degree to which it was important they earn their goal grade (Grade Importance), their motivation to earn their goal grade (Grade Motivation), and their confidence in their ability to achieve their goal grade (Grade Confidence). It also included questions about how often they would use the following help resources: make an appointment with the instructor, go to office hours, seek out tutoring, attend supplemental instruction, visit their academic advisor, and attend a study group. Second, user trace data from the LMS was included in our analytical model. Specifically, we used data about students' prior behavior around opening assignment instructions to generate the informational display for the nudge, and to assess the impact of the intervention. Third, data about student demographics (including score on the university math placement exam, their gender identity, and race/ethnicity) and academic preparation were drawn from the institutional student data warehouse. Finally, we incorporated data from the course gradebook in order to investigate the relationship between any behavioral change observed and student outcomes.

2.4.1 Sample

The focal course is a 4-credit introductory statistics course with an enrollment of 1,800-2,000 students each term. Students are introduced to the concepts and applications of statistical methods and data analysis using examples from virtually all academic areas. Exams are worth 80% of the final grade, and the rest is made up of lecture and lab attendance, lab assignments, and homework assignments. Students attended two 1.5-hour lecture sessions each week, where they were introduced to statistical concepts, as well as one 1.5-hour lab session, where they practiced applying the concepts in lab using the software package R.

As part of the course, students completed ten homework assignments. Homework assignments were released through the LMS, so trace data about when a student first accessed the assignment relative to its publication and its due date was available to the researchers. All students in the Average and BetterThan conditions received a nudge after Homework 3. To test the role of timing and dosage on behavior change, half the students in the Average and BetterThan groups also received a second nudge with the same information after Homework 5, making for a total of four treatment groups (Average1Message, Average2Messages, BetterThan1Message, and BetterThan2Messages; see table 1).

Table 1. Count of students by treatment group

	Fall	Winter
Average1Message	269	276
Average2Message	278	231
BetterThan1Message	276	245
BetterThan2Message	250	263
Control	716	791
TOTAL	1789	1806

Data was collected from the Fall 2017 and Winter 2018 semesters. The same instructional team taught the course both semesters. The samples analyzed included students who completed the ECoach beginning of term survey by the time of the first nudging message. Students who had not completed the ECoach survey (which is required to access the system) were not included as they did not meet the intent-to-treat criteria.

2.5 Data Analysis

We report results for change in behavioral engagement in the course, which we operationalized (and refer to) as Time Before Due Date (TBDD). TBDD was calculated as the difference between the date when each student first opened an assignment and the due date of the assignment. Assignments were released to students after their grades for the prior assignment were distributed. Hence, each student had one TBDD score for each assignment. In all cases, the control group was the reference group for the treatment variables, and race, sex, and math placement scores were included in each model as control variables, as these factors have been observed as significant predictors of performance differences in the course, and broadly in this domain of courses, in prior semesters (Matz, et al, 2017).

We conducted three sets of analyses of covariance (ANCOVAs) investigating the effect of the nudges on TBDD, as well as a linear regression model examining final grade in the course. Like analysis of variance (ANOVAs), ANCOVAs reveal whether there are significant differences in mean outcomes across participants at two time points, often pre- and post-measurements, because of different groupings. However, ANCOVAs are different in that they allow for the inclusion of continuous covariates to determine whether/how these variables affect the mean outcomes. When using ANCOVAs with longitudinal data, a common approach is to regress the outcome on a predictor of interest along with a baseline or previous outcome. Including the latter helps account for serial autocorrelation within individuals across different time points.

The first set of ANCOVAs examined the effect of the treatments on mean TBDD scores relative to the control group net of performance on previous assignments and student demographics. We attempted a growth curve analysis to examine change over time in the course, but our analyses did not satisfactorily meet the assumptions for growth curve estimates.

The second set of ANCOVAs explored whether the nudges had differential impacts on TBDD scores based on previous student performance and characteristics. Within this, we investigated whether the nudges had a different influence on TBDD scores according to whether students were below or above the average assignment TBDD prior to the intervention relative to others in their class. For these analyses, the course was split in half, with one group composed of students with below average pre-intervention TBDD scores and the other consisting of students with above average pre-intervention TBDD scores. ANCOVAs were run for both subgroups to determine whether the nudge differentially impacted TBDD scores. This process was repeated for the second set of messages. Significant results would indicate that the nudge had a significant effect on TBDD scores relative to the control group in the particular subgroup.

The third set of ANCOVAs examined whether the nudges had a differential impact on TBDD scores according to students' goals for the course. For these analyses, interactions were included between each self-reported measure and treatment group. A significant interaction would reveal that the impact of the nudge on TBDD scores differed according to student characteristics. ANCOVAs were conducted for both rounds of nudging messages. We also examined the above factors as explanatory influences on the score for the next homework assignment using ANCOVAs, although this was not the primary focus of this study. For the purposes of concision, we do not report those results as we observed no significant relationships. This non-significance made sense given that the homeworks were uniformly skewed toward higher scores, so improvement on scores was less likely than changes to behavior which was more variably distributed across the sample.

Our final analytical model examined the relationships among student course beliefs, treatment group, and end of course grade. We conducted an ordinary least squares regression using final course score out of 13 possible grades (F=1, D-=2, C+=7, A-=11, A=12, A+=13). We controlled for students' score on the university math placement exam, their gender identity, race/ethnicity, and their response to initial survey questions regarding their course beliefs and help-seeking behavior (as described above in section 2.4 Data Sources). We clustered standard errors by term to identify trends across semesters within the data. We created this model to observe if there was a potential latent effect of changes to study strategies that had an end of term benefit for students.

2.6 Limitations

As noted above, we defined our 'intent to treat' group as individuals who were users of the ECoach system. However, there are most likely differences between the students who chose to adopt the ECoach system and those who did not. It may be that ECoach students had higher levels of motivation for goal-oriented tasks in the course. Students who adopted the ECoach system may also, already, have more sophisticated study strategies than non-adopters. As such, the generalizability of our findings are potentially limited to comparable groups of students in comparable contexts. Additionally, we had little information about students' other responsibilities on campus while they took the course (for instance their co-curricular involvement, work, and commute times).

In our effort to develop personalized nudges, we may have provided some students a nudge that encouraged them to be overconfident. For example, students who started early and got a high score might have been provided information that suggested their peers were *starting work later*. Although we do not expect this kind of information to de-motivate students, it may have caused the higher end of student performers to change their behaviors in ways that influenced the aggregate picture post-nudge.

The behavior that we sought to change–the date when a student first opened an assignment within their LMS–is not a direct measure of how long (or with what level of intensity) they engaged with the assignment. A student may have, for example, opened the assignment early based on the nudge and then never reviewed the material again until just before the assignment was submitted. As a proxy for behavioral engagement, our outcome of interest is limited in that it provides insight into how responsive a student was to the nudge, not insight into how their engagement with course material might have changed. Future research might examine the efficacy of different nudging interventions using the time between when a student submitted their first attempt and their final attempt within the auto-grader in the course. We were unable to examine that relationship because of limitations within the data.

3. Results

Table 2. What would you do differently? (n=3484)

	Average		Better Than	
	One Message	Two Messages	One Message	Two Messages
Get Help	2.13%	2.84%	4.40%	2.98%
Nothing	6.11%	5.97%	5.68%	5.40%
Start Earlier	14.49%	16.19%	14.77%	14.49%
Start Later	0.57%	0.57%	0.71%	0.43%

Students generally appeared to understand the purpose of the nudge after they received it. When asked what they might do differently for the next assignment, around 15% of students in each group said that they would start earlier, while around 6% of each group said they would do nothing different (Table 2). When we look at students by our intention to treat–that is, when we compare students who had below average TBDDs–only 7% of students selected a response to what they would do differently that we would not classify as correct. Of this 7%, nearly all (98%) of the students said they would do nothing differently, which suggests a lack of motivation as opposed to a lack of comprehension of the nudge. Among initial demographic variables, across both groups, men opened their assignments later than women in the course, regardless of the nudge type and the assignment.

We found no evidence of a consistent immediate effect on behavior for the first and the second nudges for either treatment group. As math placement score increased, the TBDD also increased, albeit only in the fall semester. There is no evidence that either nudge type has a durable effect on behavior after the next sequential homework assignment. There was also no significant change in grade either between assignments or over the remainder of the semester (see appendix A).

Sub-group analyses examined whether the nudge had a differential impact depending on whether students had spent below or above average TBDD scores preceding the nudge relative to the rest of the class. These analyses did not yield any significant results that were consistent across semesters, suggesting that the nudges did not have a significant effect on TBDD scores regardless of whether students had below or above average TBDD scores on the pre-intervention assignment.

In addition to sub-group analyses, we looked at the interaction of changes in behavioral strategies and initial course beliefs. Among students who reported the lowest levels of Grade Importance, students in the BetterThan2 group opened the sixth assignment 3.12 and 4.08 hours earlier, on average, than students in the control group across the Fall and Winter terms, respectively($p<0.05$).

This interaction suggests that the effect of receiving a second BetterThan nudge on the TBDD score for the next assignment differed according to students' Grade Importance. As illustrated in Figure 4, when Grade Importance was low, receiving a second BetterThan nudge had a positive impact on subsequent TBDD scores relative to the control group. However, as Grade Importance increased, the gap between the two groups gradually narrowed. Only at the highest levels of Grade Importance did the control group have higher TBDD scores than the BetterThan2 group.

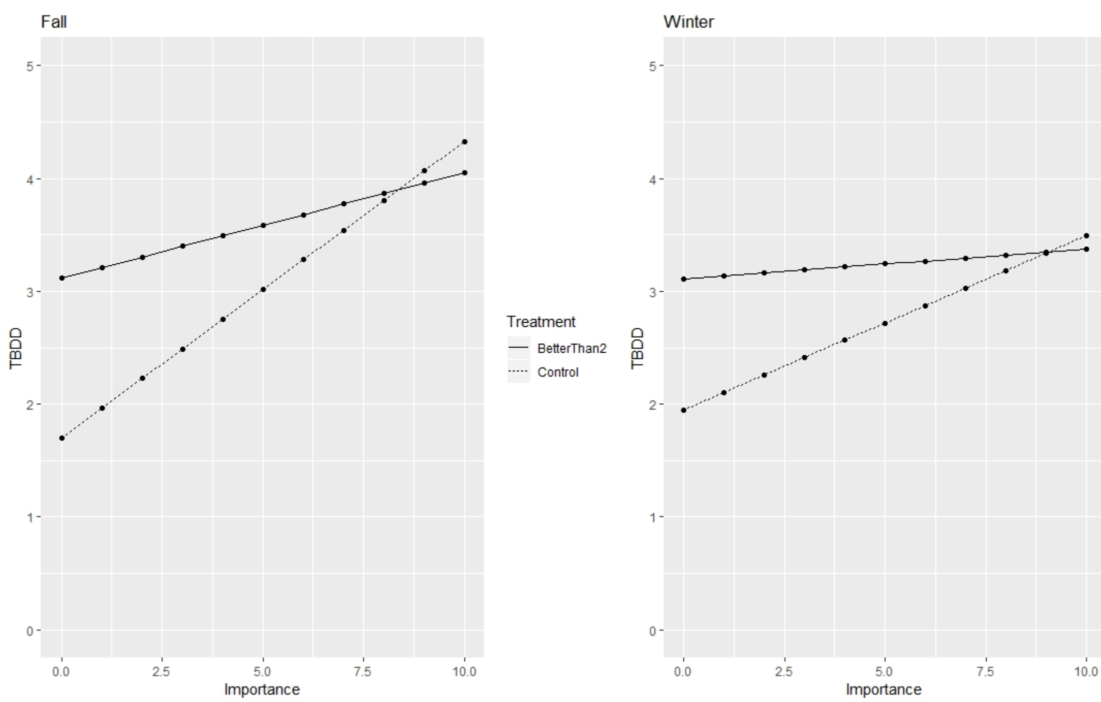

Figure 4. Students with low levels of grade importance are responsive after BetterThan2 nudge

The results of the linear model highlight the importance of initial goals and beliefs about the course on students' end of course performance. Goal grade, expected grade, and self-identifying as "a statistics person" were all positively correlated with higher end of term grades (see table 3). Similar to the TBDD findings, only the interaction of Grade Importance and the second BetterThan nudge was significantly related with a higher grade at the end of term.

Table 3. Regression results grade (1=F, 13=A+, n=3484)

Predictor	*b*	95% CI
(Intercept)	-1.20	[-2.67, 0.28]
Math Placement Exam Score[1]	0.12**	[0.11, 0.14]
Asian, Asian American, and Pacific Islander[2]	0.82**	[0.47, 1.17]
White[2]	0.81**	[0.51, 1.10]
Men[3]	-0.42**	[-0.65, -0.19]
Grade Importance[4]	-0.09	[-0.21, 0.03]
Goal Grade[5]	0.41**	[0.25, 0.57]
Expected Grade[5]	0.19**	[0.10, 0.28]
Grade Motivation[4]	0.13	[-0.00, 0.27]
I am a 'stats' person[6]	0.24**	[0.12, 0.36]
Average1	1.01	[-0.73, 2.74]
Average2	1.66	[-0.06, 3.38]
BetterThan1	0.54	[-1.13, 2.22]
BetterThan2	-0.81	[-2.47, 0.85]
Grade Importance X Average1	-0.12	[-0.34, 0.11]
Grade Importance X Average2	-0.02	[-0.24, 0.19]
Grade Importance X BetterThan1	0.00	[-0.23, 0.24]
Grade Importance X BetterThan2	0.26*	[0.02, 0.49]
Grade Motivated X Average1	-0.00	[-0.23, 0.23]
Grade Motivated X Average2	-0.14	[-0.37, 0.08]
Grade Motivated X BetterThan1	-0.06	[-0.31, 0.19]
Grade Motivated X BetterThan2	-0.15	[-0.39, 0.09]
Second Year[7]	0.33	[-0.19, 0.84]
Third Year[7]	0.69**	[0.18, 1.20]
Fourth Year+[7]	0.63*	[0.08, 1.19]
One on One Help[8]	0.03	[-0.11, 0.17]

Predictor	b	95% CI (continued)
Office Hours[8]	0.24**	[0.10, 0.39]
Academic Advisor[8]	-0.08	[-0.19, 0.03]
Science Learning Center[8]	0.05	[-0.07, 0.17]
Study Group[8]	0.04	[-0.07, 0.15]
Tutoring[8]	-0.23**	[-0.32, -0.14]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively; * indicates p < .05. ** indicates p < .01; R^2 = .289**; 95% CI [.24,.31]; [1]Out of 25; [2]Black, Latinx, Native/Indigenous, and Multi-racial; [3]Women; [4] (1-7); [5] Grade; [6] Strong Disagree/Agree (1-5); [7] First Year; [8] Never, 'If I am failing', 'Probably Not', Maybe, Probably, First Sign, Any Time (1-7)

4. Discussion

This study contributes to the ongoing development and implementation of digitally delivered nudges in higher education and the growing literature on social self-regulation in large undergraduate courses. Across both semesters, we observed no broad trends of behavior change as a result of students receiving digitally delivered nudges with social reference points depicting when their peers accessed assignments relative to their due date. However, when we considered nudge dosage and interactions between the nudge and students' perceived importance in getting a specific grade for the course, we identify important findings that have potential implication for the implementation of nudging interventions in post-secondary courses.

4.1 There does not appear to be a main effect of behavioral change in response to either nudge

Neither the descriptive norm (Average) feedback nor the feedback focused on opinion leaders (BetterThan) resulted in significant differences around access or performance on assignments in this study. Our work only focuses on adopters and it may be that, on average, adopters are students who have already developed relatively effective study strategies. Still, systems like ECoach can only directly impact the behavior of adopters. As such, understanding the mechanisms that produce change among this group is important. Our intention to treat was broad given the behavior we actually hoped to nudge. In developing personalized nudges, our findings suggest that designers should consider the interaction of students' goals and the personalized information that choice architecture interventions provide. We might also consider the delivery platform for implementation. It may be that nudges could be more effective when provided in a platform that all students use (like the learning management system or the online homework system).

4.2 Dosage, message type, and the relative importance of the course for students' goals matter for nudging behavior

Our findings do reliably suggest, however, that when we consider the interaction of grade importance, nudge type, and nudge dosage we can observe a positive influence on behavioral change. The second opinion leader nudge had a positive impact on the TBDD scores of students with low Grade Importance, perhaps motivating them to spend more time on the subsequent assignment. Conversely, students with higher levels of self reported Grade Importance are already motivated to spend more time on assignments in order to achieve a good grade, so the nudge naturally had less of an effect in the interaction term. This suggests that nudges need to be tailored by both need (i.e. could the student benefit from this information) and opportunity (i.e. is the information likely to be salient for the student) in mind.

Additionally, to encourage students to attend to the nudge, more than one message appears to be necessary. It may also be that the additional reinforcement of grade information from prior assignments made students more responsive to the second nudge, as they started to understand their performance in the course and how their strategies could be related to performance. At the time of the first nudge, students may not have had enough information about their performance to reflect on the effectiveness of their strategies.

Students in the two-message BetterThan treatment group also had slightly higher grades at the end of the term–equivalent to about a quarter of the distance between different levels of a grade (i.e. students would move a quarter of the distance from a B- to a B). Further research is needed to see if this impact is affected by more frequent treatment dosage.

The significant difference in both behavior and performance that we observe for students at lower levels of Grade Importance is small. It may be that as the course progresses and the complexity of the material increases, students are already anticipating the need to spend more time on homework assignments. As demands on students' time increase across a semester, students may already be revising their strategies for coursework to accommodate for other in-class and out of class tasks. The significance of the interaction of the second nudge and students' goal grade being linked to behavioral change suggests that there is something important about the exchange of dynamic beliefs, study strategies, and behavioral engagement in the course that matters for developing personalized nudges. We did not fully capture the range of students' other campus engagements because the design of our research focused on students' beliefs about the course and their help-seeking behavior. Future research should test the timing of choice architecture interventions, conditional on students' course beliefs, while accounting for how they spend their time on (and off) campus.

4.3 Implications for implementation of nudging interventions

Our findings suggest that in addition to tailoring based on behavior, effective nudges will simultaneously consider the motivational factors that inform students' study strategies. Nudge designers should closely consider the alignment of the treatment, students' behavior, and their goals when developing models for personalization. This may mean, as we did during our implementation, capturing information about students' goals and motivation related to the course early in the term to aid in personalization and targeting. Additionally, per the prior research from Oreopolous and colleagues (2018) and prior research on the Michigan Tailoring System (McKay et al., 2012), choice architects should consider when and where nudges are situated in the overall system. Given the contrasting results of the McKay and the Oreoplous studies, it is unclear when rich information becomes too much information for students to integrate. The simpler approach highlighted in this study does appear to make a (albeit limited) difference in behavior and performance, so researchers should consider the scope and complexity of informational interventions. A next step, as part of research on the implementation process, might compare platforms for nudge delivery holding constant the type of nudge and the dosage.

Our modest results also suggest that researchers should consider when and where impact matters for developing choice architecture interventions. The gains we observed were the byproduct of a low-cost implementation. We also did not observe significant negative impacts on either students' behavior or outcomes. We should note that one potential outcome of the nudge–the false security risk, where students are provided information that falsely increases their confidence in their strategy–is real. Although we did not observe negative outcomes for students in either the control or treatment group, thought and care should be put into who receives what kind of personalized information nudge in the course.

4.4 Implications for future research

In this study, our investigation of nudge intervention within the ECoach platform highlights the degree to which more research is needed to describe what specific components of the tool impact which students, to characterize what the impact is, and under what conditions they are experienced. Simplistic answers, while satisfying, are unlikely to hold true past the bounds of a single semester, let alone a specific course.

This study does identify students who are possibly part of the "nudgeable"/movable middle: specifically, students who do not initially report perceiving the course outcome as important appear more responsive to nudging than students who identified early on that the course was a high priority. This makes a certain intuitive sense–students who consider the course low priority may not be investing the kind of emotional and cognitive energy as their peers, and so have more potential to change their behavior; they may also not be putting in effort to reflect on the efficacy of their current study strategies. It may be that later in the semester, as they are searching for alternative strategies, students are able to see how the suggested change is easy to do and worth it. Further research is needed to understand how these students approach coursework strategies like when to start homework assignments, what specific issues would provide effective opportunities for nudges, and how students respond to and interpret difference types of nudges.

Further research is also needed that consider the frequency with which nudges need to delivered. Our goal was to develop unobtrusive messages. It may be that amidst the noisiness of a large undergraduate lecture course, an unobtrusive message needs to be repeated relatively frequently to flag a students' attention. Because so much information about the course exists at students' eye level through the LMS, effective nudges need to be delivered at key points throughout the term to be effective. When, where, how, and with what frequency to deliver those nudges is a potential fruitful area for future research. As part of our broader design-based implementation project, our next primary objective is to test the findings of this study across levels of education and across settings. Observing if these results are reproducible in other courses, disciplines, and institutions would provide important evidence for how to scale nudging interventions.

In comparing the two types of nudges we focused on, uniformly using a course average message produced little movement either in student behavior or performance. We expect that researchers will find more fruitful results using the opinion leader approach when personalization is conditioned on course goals. Testing different kinds of messages and different groups of opinion leaders among students by course goals is an important next step. In this study, we did not provide actionable recommendations for how students might approach the next assignment, but in our next projects we aim to pair personalized nudges with personalized action plans, as we expect that students who are less invested in the course may be more responsive if they are provided a pathways to success. Understanding if and how these two informational resources in concert result in a more significant impact on student behavior could further improve the implementation of digital nudging systems.

We used an electronic coaching system to deliver the messages that we tested in this implementation study. While there is extensive adoption of this tool in the course we studied, there may be something significantly different about the kind of student who opts into the system versus the students who choose not to use ECoach. In addition to exploring the timing and dosage of messaging to further refine implementation strategies, a different deployment strategy would enable us to explore how all students in a course respond to the kinds of nudges we outline above. Given the importance of individual beliefs and goals for identifying students who are receptive to the nudging intervention we tested, having a broader diversity of students (including users and non-users of the ECoach system) might uncover more significant impact.

We included all of the ECoach users in our sample, but given the moderate results of this study researchers might also consider testing implementation of nudging systems with message personalization driving sample selection. We could, for example, identify students who would be responsive to the nudge using the Grade Importance item from the survey (or similar measures in other contexts) and deliver a personalized message and a generic message to better understand what factors actually motivate change.

5. Conclusion

This study contributes to the growing literature on nudging student behavior in undergraduate higher education by focusing on a specific academic task across two semesters of a very large undergraduate introductory course. We observe some promising results for personalization that suggest the potential for unobtrusive messaging for encouraging students to make changes to their coursework behavior. Our findings suggest that students who do not enter the course placing a high priority of their end of term grade may be an ideal audience for nudging interventions that help them plan out how to complete some academic tasks. The change in behavior and outcomes we observe are modest, but the intervention we propose is also relatively low cost. We believe with further research and evaluation of implementation the nudging methods we describe here could become part of a host of wrap-around tools and technologies that complement institutions' existing student support infrastructure.

References

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270). ACM.

Arnold, K.C., Castleman, B. L., Chewning, A., & Page, L. C. (2015). Advisor and student experiences of summer support for college-intending, low-income high school graduates. *Journal of College Access, 1(1).*

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition—Implications for the design of computer-based scaffolds. *Instructional Science, 33,* 367–379.

Bettinger, Long, Oreopoulos, & Sanbonmatsu (2011) The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *NBER Working Paper* No. 15361

Bird, K. A. and Castleman, B. L. (2016). Here today, gone tomorrow? Investigating rates and patterns of financial aid renewal among college freshmen. *Research in Higher Education,*
57(4): 395-422.

Bird, K. A., Castleman, B. L., Goodman, J., and Lamberton, C. (2017). *Nudging at a national scale: Experimental evidence from a FAFSA completion campaign*. EdPolicyWorks Working
Paper No. 55. Charlottesville, VA: University of Virginia.

Bleich S, Wolfson J, Jarlenski M. (2015). Indirect Effects From Menu Labeling Can Improve The Public's Health. *Health Affairs Blog*. Retrieved from: http://healthaffairs.org/blog/2015/02/24/indirect-effects-from-menu-labeling-can-improve-the-publics-health/ on November 31, 2018.

Block JP, Roberto CA. Potential benefits of calorie labeling in restaurants. JAMA 2014;312(9),887-888.

Brown, M. G. (2016). Blended instructional practice: A review of the empirical literature on instructors' adoption and use of online tools in face-to-face teaching. *The Internet and Higher Education*, *31*, 1-10.

Camilleri, A., Cam, M. A., & Hoffmann, R. (2018). Nudges and Signposts–The Effect of Smart Defaults and Pictographic Risk Information on Retirement Saving Investment Choices. Retrieved from https://ssrn.com/abstract=3227378

Castleman, B. L., Arnold, K., & Wartman, K. L. (2012) Castleman, B. L., Arnold, K., & Wartman, K. L. (2012). Stemming the tide of summer melt: An experimental study of the effects of post–high school summer intervention on low-income students' college enrollment. *Journal of Research on Educational Effectiveness*, 5, 1-17. Retrieved from: https://eric.ed.gov/?id=EJ952097

Castleman, B. L., & Meyer, K. (2016). Can text message nudges improve academic outcomes in college? Evidence from a West Virginia Initiative. *Center for Education Policy and Workforce Competitiveness Working Paper*, *43*.

Castleman, B. L., Page, L. C., & Schooley, K. (2014) The forgotten summer: Does the offer of college counseling after high school mitigate summer melt among college-intending, low-income high school graduates? *Journal of Policy Analysis and Management*, 33(2), 320–344. doi:10.1002/pam.21743

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, 115, 144-160. doi:10.1016/j.jebo.2014.12.008.

Castleman, B. L., & Page, L. C. (2016). Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence. *Journal of Human Resources*, *51*(2), 389-415.

Castleman, B.L., Owen, L., & Page, L.C. (2015). Stay late or start early? Experimental evidence on the benefits of college matriculation support from high schools versus colleges. *Economics of Education Review*, 47, 168-179. doi: 10.1016/j.econedurev.2015.05.010.

Chambliss, D. F. & Tackacs, P.F. (2014). *How college works*. Harvard University Press.

Davidai S, Gilovich T, Ross LD. (2012). The meaning of default options for potential organ donors. *PNAS*. 109(38):15201–5.

Denley, T. (2014). How Predictive Analytics and Choice Architecture Can Improve Student Success. *Research & Practice in Assessment*, 9, 61-69.

Fritz, J. (2013). *Using analytics at UMBC: Encouraging student responsibility and identifying effective course designs* (Research Bulletin) (p. 11). Louisville, CO: Educause Center for Applied Research. Retrieved from http://www.educause.edu.proxy.lib.iastate.edu/library/resources/using-analytics-umbc-encouraging-student-responsibility-and-identifying-effective-course-designs

Fritz, J. (2017). Using Analytics to Nudge Student Responsibility for Learning. *New Directions for Higher Education*, *2017*(179), 65-75.

Foster GD, Karpyn A, Wojtanowski AC, Davis E, Weiss S, Brensinger C, et al. (2014) Placement and promotion strategies to increase sales of healthier products in supermarkets in low-income, ethnically diverse neighborhoods: a randomized controlled trial. *American Journal of Clinical Nutrition*. 99(6):1359–68.

Gabriele, J. M., Carpenter, B. D., Tate, D. F., & Fisher, E. B. (2010). Directive and nondirective e-coach support for weight loss in overweight adults. *Annals of Behavioral Medicine*, 41(2), 252-263.

Hoxby, C., & Turner, S. 2013. Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014).

Johnson EJ, Goldstein D. Do Defaults Save Lives? Science. 2003 Nov 21;302(5649):1338–9.

Keller PA, Harlam B, Loewenstein G, Volpp KG. Enhanced active choice: A new method to motivate behavior change. Journal of Consumer Psychology. 2011 Oct;21(4):376–83.

Loewenstein G. Confronting reality: pitfalls of calorie posting. Am J Clin Nutr. 2011 Apr;93(4):679–80.

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., … McKay, T. A. (2017). Patterns of Gendered Performance Differences in Large Introductory Courses at Five Research Universities. AERA Open. https://doi.org/10.1177/2332858417743754

McKay, T., Miller, K., & Tritz, J. (2012). What to do with actionable intelligence: E²Coach as an intervention engine. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 88–91). New York, NY: ACM. https://doi-org.proxy.lib.iastate.edu/10.1145/2330601.2330627

Means, B., & Penuel, W. R. (2005). Research to support scaling up technology-based innovations. In C. Dede, J. Honan, & L. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement* (pp. 176–197). New York, NY: Jossey-Bass.

Münscher, R., Vetter, M., & Scheuerle, T. (2016). A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, *29*(5), 511-524.

Oreopoulos, P., Patterson, R. W., Petronijevic, U., & Pope, N. G. (2018). Lack of Study Time is the Problem, but What is the Solution? Unsuccessful Attempts to Help Traditional and Online College Students. *National Bureau of Economic Research*. (No. w25036).

Penuel, W. R., & Fishman, B. J. (2012). Large-scale science education intervention research we can use. *Journal of Research in Science Teaching*, 49(3), 281–304. https://doi.org/10.1002/tea.21001

Pugatch, T., & Wilson, N. (2018). Nudging study habits: A field experiment on peer tutoring in higher education. *Economics of Education Review,* 62, 151-161.

Reeve, J., & Lee, W. (2014). Students' classroom engagement produces longitudinal changes in classroom motivation. *Journal of Educational Psychology*, *106*(2), 527.

Riggs, William, The Role of Behavioral Economics and Social Nudges in Sustainable Travel Behavior (March 22, 2017). Available at SSRN: https://ssrn.com/abstract=2939404 or http://dx.doi.org/10.2139/ssrn.2939404

Schünemann, N., Spörer, N., Völlinger, V. A., & Brunstein, J. C. (2017). Peer feedback mediates the impact of self-regulation procedures on strategy use and reading comprehension in reciprocal teaching groups. *Instructional Science*, *45*(4), 395-415.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational psychologist*, *32*(4), 195-208.

Selwyn, N. (2014). *Digital technology and the contemporary university: Degrees of digitization*. Routledge.

Spörer, N., & Brunstein, J. C. (2009). Fostering the reading comprehension of secondary school students through peer-assisted learning: Effects on strategy knowledge, strategy use, and task performance. *Contemporary Educational Psychology, 34,* 289–297.

Stutzer A, Goette L, Zehnder M. Active Decisions and Prosocial Behaviour: a Field Experiment on Blood Donation. *The Economic Journal*. 2011 Nov 1;121(556):F476–93.

Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, *31*(3), 355-366.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thorndike AN, Riis J, Sonnenberg LM, Levy DE. Traffic-Light Labels and Choice Architecture: Promoting Healthy Food Choices. *American Journal of Preventive Medicine.* 2014 Feb;46(2):143–9

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Erlbaum.

Zywicki, T. J. (2016). Do Americans Really Save Too Little and Should We Nudge Them to Save More: The Ethics of Nudging Retirement Savings. *Georgetown Journal of Law and Public Policy*, *14*, 877.