Planning Prompts Increase and Forecast Course Completion in Massive Open Online Courses

Michael Yeomans
Harvard University
Cambridge, MA
yeomans@fas.harvard.edu

Justin Reich
Massachusetts Institute of Technology
Cambridge, MA
jreich@mit.edu

ABSTRACT

Among all of the learners in Massive Open Online Courses (MOOCs) who intend to complete a course, the majority fail to do so. This intention-action gap is found in many domains of human experience, and research in similar goal pursuit domains suggests that plan-making is a cheap and effective nudge to encourage follow-through. In a natural field experiment in three HarvardX courses, some students received open-ended planning prompts at the beginning of a course. These prompts increased course completion by 29%, and payment for certificates by 40%. This effect was largest for students enrolled in traditional schools. Furthermore, the contents of students' plans could predict which students were least likely to succeed - in particular, students whose plans focused on specific times were unlikely to complete the course. Our results suggest that planning prompts can help learners adopted productive frames of mind at the outset of a learning goal that encourage and forecast student success.

CCS Concepts

• Applied computing~Psychology • Applied computing~Distance learning

General Terms

Decision-Making, Goal Pursuit, Natural Language Processing

Keywords

MOOCS; Learning Analytics; Motivation

Copyright Information

LAK '17, March 13 - 17, 2017, Vancouver, BC, Canada

ACM 978-1-4503-4870-6/17/03$15.00
DOI: http://dx.doi.org/10.1145/3027385.3027416

1. INTRODUCTION

The human mind has an incredible capacity to look into the future, set goals, and plan for action [1]. But in many essential domains, this goal-setting often outpaces goal-achieving. While the link between intention and action is strong, often it is not as strong as we would like.

The intention-action gap is one of the distinctive features of open online education [10]. Since the first Massive Open Online Courses ("MOOCs") were created at Stanford in 2011, over 35 million students have enrolled in one of 4,200 courses offered by over five hundred universities worldwide [42]. These courses offer access to university-level instruction from elite institutions to anyone in the world, free of charge, and they hope to transform the production function of higher education [4]. But despite this growing interest and attention, it is still true that the vast majority of students who have enrolled in MOOCs do not finish.

More importantly, the majority of students with a stated intention to complete a MOOC do not finish. Among students enrolled in HarvardX and MITx courses, who declared at enrollment that they intended to finish their course, only 22% did so [16, 37]. In many domains, research suggests that nudges - simple, inexpensive psychological supports - can be used to help people address their intention-action gaps [46]. To what extent can these same approaches be deployed to help students achieve their stated goals?

In the current research we address this question in a natural field experiment, by prompting some MOOC students to plan their course participation in advance. Planning prompts encourage goal pursuers to elaborate on their implementation strategies while their intentions are still vivid [12, 39]. This intervention has been successful in other domains where the benefits of follow-through lie far in the future, but where goal pursuit is easily derailed in the present [29, 30]. MOOCs are minimally structured by design, so they may be especially vulnerable to obstacles during implementation [6]. This diagnosis implies that MOOCs might be particularly responsive to planning prompts.

Voluntary online courses provide a compelling new setting in which to test theories of planning in goal pursuit. MOOCs require persistence over multiple assignments, months apart, testing the effects of planning over a much longer time scale than in previous research. The online platform also provides exact measurements of planners' subsequent activities, and allows us capture the full text of thousands of student plans, creating a corpus of student natural language that can be parsed and evaluated to determine which kinds of plans were most likely to be successful. These data provide a novel and comprehensive view of how planning can increase follow-through, in online education and in other domains where people struggle with persistence, and firms struggle with retention.

2. BACKGROUND LITERATURE

This research draws from two disparate bodies of work, for which we now provide a brief review. First, we describe how planning prompts have been used to increase follow-through in other domains, and how MOOCs provide a novel extension of that literature. Second, we describe the budding literature on online education, with a particular focus on how MOOCs might serve as a model of goal pursuit, which motivates our formal hypotheses.

2.1. Planning Prompts

Planning is central to social psychological models of goal pursuit, as the path by which current intentions can be translated into future action [1]. Planning works because even when intentions are strong, many long-term goals fail from a lack of implementation [12]. These implementation factors are often overlooked when people forecast their future goal pursuit [25, 28, 36].

The theoretical motivation for planning prompts, then, is to spur attention to these implementation factors while intentions are strong [39]. Elaborating on implementation can help plans last when intentions fade and are overcome by forgetfulness [40] and procrastination [33]. Recent field applications have shown that planning prompts are a low-cost way to encourage follow-through in social beneficial domains like voter turnout, flu shots, and colonoscopies [29, 30, 31].

Online education offers a unique opportunity to study the effect of planning prompts on goal pursuit. MOOCs require sustained effort over months, in contrast to goals pursued within a single lab session, or behaviors (like a doctor's visit) which involve a single plan, enacted once in the future. Additionally, all goal relevant behavior is passively and automatically tracked by the server logs of the MOOC platform, which precludes common measurement problems like attrition, demand effects, social pressure, or self-report bias, which compromise plan-making results from the lab.

Finally, participation in MOOCs is endogenous - students choose to enroll of their own accord. By contrast, lab experiments rely on extrinsic rewards for participation, such as course credit or money. And previous field settings simply cold-called (or cold-mailed) participants, so it was not clear whether plan-making encouraged follow-through, or was simply an effective persuasion technique. All participants in this research enrolled in these courses of their own accord, and students explicitly reported their intentions, so that we could test intentions as a potential moderator [43]. Put simply: we could be reasonably sure our students were there to learn.

2.2. Goal Pursuit in MOOCs

Our research was conducted on EdX, a MOOC platform that has served more than ten million students. MOOCs like these provide new kinds of educational opportunities by making elite college courses widely available for free or minimal cost. But their ultimate value to higher education institutions remains unclear. In particular, the consistently low completion rates - typically less than 10% of enrolled students - call into question their effectiveness [10]. One valid response is to point out that many students only intend to browse, rather than finish the required assignments. Indeed, browsers account for a large percentage of so-called "drop outs", whether assessed using tracking data [2, 19, 27] or by asking students to declare their intentions upfront [16, 37]. But evidence still shows that even when students explicitly intend to finish, follow-through failures are common.

These results have prompted two responses. One response is to consider whether psychologically-informed interventions might be used to align the design of MOOCs with their students' goals and increase course completion. However, this nascent literature has mainly produced null results [20, 23, 50]. One encouraging result shows that participation badges can increase forum participation [2], which is consistent with the hypothesis that a lack of structure is at least partly responsible for drop-outs [6]. These results lead us to the prediction that planning prompts might be a particularly effective way to nudge student follow-through.

Hypothesis 1: Prompting MOOC students to plan their efforts at the start of a course will increase completion rates.

A second response to low MOOC completion rates has focused on modeling the heterogeneity in course completion, to better understand why some students do not follow through on their goals. Forecasting from demographics has been shown previously [21], but that exercise is intrinsically limited - demographics cannot be manipulated by interventions, and provide a sparse understanding of how students' life situation affects their ability to follow through on their goal.

Some understanding into course dynamics can gained from course activity logs [2, 24, 48, 53]. But is unclear whether these results meet the epistemological definition of forecasting - that is, do activity logs anticipate drop-outs in advance, or merely reveal them as they happen? A similar concern arises from forecasts based on in-class discussion boards - posts may simply reveal which students are encountering difficulties, rather than anticipating potential difficulties [52, 54].

In contrast, planning prompts might forecast students' engagement at the moment they start the course. This timing is essential in practice, because most drop outs occur in the first week of the class [10, 16]. This means that targeted interventions will be most successful if they can be deployed early, while students' attention is still piqued. Furthermore, open-ended plans might provide a richer forecast of students' behavior and their expected obstacles. This might also lay the groundwork for personalized learning, allowing the course platform to adapt to students' individual plans for goal pursuit [9, 26, 41].

Hypothesis 2: Course completion can be forecasted from the content of MOOC students' plans at the start the course.

3. METHODS

3.1. Experimental Setting

This research was conducted in three online courses created by HarvardX, teaching Business, Chemistry and Political Science. Each course was taught by a Harvard professor, and paralleled an existing course at the University. The course material consisted of video lectures, assigned readings, and discussion boards, and chapters of the course were doled out in sequence over 2-3 months. Grades were determined by a combination of quizzes, peer-assessed written assignments, and self-assessed participation. Participation was free, though students were given an option to pay for a "verified" certificate, where assessments were remotely proctored and their identity was confirmed.

These are reported as "Study 1" (Business) and "Study 2" (Chemistry and Political Science), because some elements of the survey design in Study 1 were modified before the roll-out in Study 2. Our preregistration describes the methods in Study 2, but was posted before the outcome of Study 1 could be observed, and we follow the same analyses throughout. For completeness, we report the results of analyses in both studies, individually and pooled. We also report how we determined our sample size, all data exclusions, all conditions, and all measures in every study. Additionally, our data, code, materials and preregistered analysis plan are available at https://osf.io/mky8n/.

3.2. Data Collection

The MOOC platform records every action that every student takes on the course platform, including enrollment, verification, and grades. This ensures that course progress can be tracked accurately, exhaustively, and without any effort or awareness on the part of the student. This also produces many possible outcome measures from a MOOC, and we discuss several in this research. However, our pre-registered analysis plan focused on only one primary outcome: whether or not students completed enough of the coursework to "earn a certificate" in their class. This requires earning a grade above a certain threshold (between 70-80%, depending on the class), and if they achieve the threshold before the final date, they are deemed to have "certified" in the class.

3.2.1. Course Grades

To achieve a certificate the class, students completed assignments and quizzes at the end of each chapter in the courses, as well as some peer-grading (Study 1) and some participation credit (Study 2). Students could track some of their progress on the website, but final grades and certificates were not handed out until after the class had closed. The distribution of grades was bimodal - 49% of students in our sample had a grade of zero, while 16% earned a passing grade, and the rest were scattered in between.

3.2.2. Certificate Verification

During the first half of each course, all students had the option of paying ($50-100 USD) for a "verified" certificate, that also involved a formal identity check. By the end of the course, 6% of students in our sample were verified. However, 3.1% verified when they enrolled, before they saw the planning prompts. The other 2.9% upgraded their account during the course, and after the planning prompts. After pre-registration, we determined that verification rates were an outcome of increasing interest to MOOC stakeholders, and investigated this outcome in an exploratory manner. Specifically, we were encouraged by new datasets that distinguished two kinds of verifications in our sample: pre-course enrollments, as a pre-treatment covariate; and in-course upgrades, as a post-treatment outcome.

3.2.3. Pre-Course Survey

Every class run by HarvardX has a pre-course survey embedded as the first chapter of the course. The pre-course survey is optional, and has no effect on grades but it is encouraged as a "way to get to know our students". The majority of enrolled students did not attempt the pre-course survey. However, the vast majority of those students did not complete much else. These results are typical [37], and suggest that the pre-course survey provides good coverage of the students who actually participate in the course.

The pre-course survey has two purposes in this research. First, the survey contained our planning prompt treatments. Second, the pre-course survey also collected information about demographic and behavioral covariates, as part of a standard battery of survey questions included in all HarvardX classes. These answers were used to define our exclusion criteria, and allowed us to test for treatment effect heterogeneity.

3.3. Population of Interest

We decided ex ante to analyze only a subset of the 60,778 students who enrolled in these three courses. All of our primary analyses below follow these pre-registered exclusion criteria.

Most enrolled students were excluded simply because they did not participate in the class after enrollment. Our cut-off rule was to include anyone who completed enough of the pre-course survey to be assigned to a treatment, regardless of whether or not they actually wrote a plan (i.e. intent to treat). We also removed anyone who did not report that they were fluent in written English, because the planning prompt was intended to be a natural language task for students. We also expected that the vast majority of students would start the class in the first month, and planned to drop all late enrollees in our primary analyses.

3.3.1. Pre-course Intentions

Intentions were self-declared in the pre-course survey, as an option in a non-binding multiple-choice question. The exact text was as follows (emphasis added):

People register for HarvardX courses for different reasons. Which of the following best describes you?

Here to browse the materials, but not planning on completing any course activities (watching videos, reading text, answering problems, etc.).

Planning on completing some course activities, but not planning on earning a certificate.

Planning on completing enough course activities to earn a certificate.

Have not decided whether I will complete any course activities.

Students' responses from our experiment are plotted in Figure 1. The majority of students (57%) intended to certify, indicating high interest. And most students who earned a certificate intended to do so (83%). But intentions alone were not enough, as among those students who intended to certify, only a minority completed enough work to achieve that goal (16%).

3.4. Planning Prompt Intervention

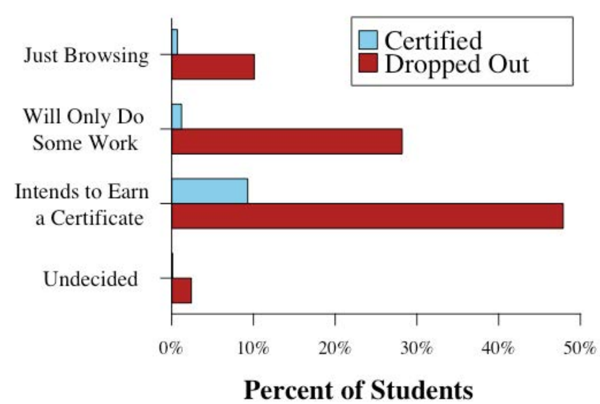

Figure 1. Course Intentions and Completion Rates

All pre-course surveys included a single randomized factor, which was the presence or absence of a planning prompt. However, the protocol varied slightly between the two studies.

3.4.1. Study 1

Students were randomized between two different conditions - *planning* or *control*. The two surveys were exactly the same, except that students in the *planning* condition received a planning prompt. The control condition had no additional materials. This prompt asked students to describe any specific plans they made to engage course content and complete assignments on time. However, students were not explicitly encouraged to make extra plans, or told about the benefits of planning.

Below the prompt, two open-ended text boxes were provided, into which students could type their plans. Students were free to leave the boxes blank if they did not want to engage. In fact, 14% of students left the boxes blank. All of the analyses below estimate intent to treat, which includes students who did not write anything, or who quit the pre-course survey after being assigned to treatment (but not those who quit before, who were excluded).

3.4.2. Study 2

Students were randomly assigned to one of three conditions - *control, simple-planning*, or *planning-plus*. The control condition was identical to Study 1 (i.e. no planning prompt). The *simple-planning* condition was virtually identical to the *planning* condition in Study 1, however, there were some minor wording changes to help clarify the instructions, and the number of text boxes was increased to accommodate additional planning instructions. The exact text of the prompt was as follows:

> *We want to know about what plans you have made to complete this course. In the space below, write down some of your plans to learn. For example, try to specify:*
>
> *a)* ***When and where*** *do you plan to spend time engaging the course content?*
> *b)* ***What specific steps*** *you will take to ensure you complete the required course work?*
>
> *c)* ***How will you respond*** *to obstacles that you might encounter during the course?*
>
> *Please use some of the boxes below to describe the plans you are making for this course. (Note: You don't have to fill every box; just use the different boxes to separate the distinct plans you have).*

The *planning-plus* condition was identical to the *simple-planning* condition*,* with two exceptions. First, at the top of the page, students were explicitly told that planning was a useful strategy to increase follow-through. Second, after typing in their plans, the next page on the survey displayed to students the text of their plans in a list labeled "your plans for this course". Students were encouraged to write their plans down and stick to them during the course.

4. EXPERIMENTAL RESULTS

4.1. Descriptive Statistics

Our samples are far more diverse than any brick-and-mortar school, but this diversity is quite typical for MOOCs. Basic demographics for all three courses are given in Table 1. Students come from a range of ages, educational backgrounds, and current vocations, and a majority are not based in the United States. Students in Study 1 were more likely to be older, working, and already have a degree, while students in Study 2 were younger and more likely to be currently in university. Overall, the vast majority of students (75%) had previously enrolled in a MOOC, suggesting they should be familiar with the domain. We also confirmed that the random assignment was successful across all observables (i.e. p>0.05 for all balance tests in both studies).

Table 1. Descriptive statistics

	Healthcare	Biochemistry	Government
Age	36.5 (11.7)	30.0 (13.5)	35.0 (14.6)
% Female	61.8%	50.6%	55.3%
Lives in USA	31.7%	42.9%	51.6%
Country HDI	.801 (.147)	.824 (.134)	.847 (.120)
Full-Time Employed	64.5%	34.5%	49.8%
Part-Time Employed	17.4%	16.0%	14.8%
Concurrent Student	22.6%	49.7%	33.1%
Bachelor's Degree	84.3%	49.9%	56.0%
Advanced Degree	51.9%	24.2%	26.5%
MOOCs Enrolled	3.9 (3.9)	4.3 (4.4)	4.3 (4.3)
MOOCs Completed	2.6 (3.5)	2.8 (3.9)	3.2 (4.0)
Pre-Course Enrollment	36.9%	59.7%	76.3%
Pre-Course Verification	3.1%	1.8%	4.6%

4.2. Average Treatment Effects

4.2.1. Course Completion.

Across both studies, we find a consistent and robust effect of planning - students prompted to write out their plans at the beginning of the course had a higher follow-through rate (M=17.7%, *95% CI*=[15.6%, 19.8%]) than those who were not prompted write out their plans (M=13.8%, *95% CI*=[11.3%, 16.3%]; $\chi^2(1)=5.2$, p=.023). The robustness of these results are confirmed in a series of logistic regressions in Table 2. The results imply that planning prompts increased course completion by 29% compared to the control condition. For comparison, this effect size

Table 2. Effects of Treatment on Course Completion

	A	B	C	D
All Plans	0.609 (0.353)*	0.252 (0.139)*		0.302 (0.130)**
Simple Plans			0.190 (0.161)	
Plans Plus			0.312 (0.158)**	
Courses	Study 1	Study 2	Study 2	ALL
Course Effects	NO	YES	YES	YES
N	293	1760	1760	2053
pseudo R^2	.013	.008	.008	.009

is similar to the difference between students who have enrolled in (and completed) one MOOC before, and students who had never enrolled in a MOOC.

We conducted two additional exploratory robustness checks, reported in the full paper. First, we expand our sample to include the (surprisingly numerous) people who signed up in the later months of the course and still intended to complete the course. Second, we expand our sample again, to include people who did not intend to complete the course. This also provided a conceptual replication of Sheeran and colleagues [43], who report a moderation of the plan-making effect by initial intentions. In both analyses, we found that (i) the effect of planning was robust in a broader sample; (ii) both intentions and sign-up times had direct main effects on course completion; and (iii) neither of these variables moderated the effect of planning on course completion.

4.2.1. Verification Rates.

We also decided to test whether planning prompts affected students' willingness to pay to upgrade to a "verified" certificate during the course, reported in Table 3. We indeed find that these upgrades were more common among students who saw a planning prompt (*M*=3.6%, *95% CI*=[2.6%, 4.6%]) than those in the control condition (*M*=1.8%, *95% CI*=[0.8%, 2.8%]; $\chi^2(1)=4.41$, *p*=.036). Added to verifications at enrollment, this implies that planning prompts increased the total verification rate by 40%, from 4.8% in control (*95% CI*=[3.3%, 6.3%]) to 6.7% in the planning conditions (*95% CI*=[5.4%, 8.0%]); $\chi^2(1)=2.70$, *p*=.100).

The causal mechanism between course progress and verification is unclear - some students may verify as a proactive commitment device, but others may simply wait to upgrade until after they are sure they will complete the course. The average time between the pre-course surveys and verification upgrades was similar among students who did not receive a planning prompt (*M*=15 days, *SD*=42 days) and those who did (*M*=19 days, *SD*=46 days). Either way, this measure provides new evidence that plan-making has a causal impact on real-stakes commitments to online education.

4.2.1. Simple Planning vs. Planning Plus.

In Study 2, participants were randomly assigned to one of two different planning prompts. Students in the the *planning-plus* condition did not write longer plans (*M*=30.0 words, *95% CI*=[28.8, 31.2]) than students in the *simple-planning* condition (*M*=28.9 words, *95% CI*=[27.7, 30.1]; *t*(1161)=0.6, *p*=.518), or spend longer time writing, on average (plans: *M*=137s, *95% CI*=[131,142]; plans plus: *M*=148s, *95% CI*=[143, 154]; *t*(1161)=1.4, *p*=.141). Furthermore, certification rates in the *planning-plus* condition (*M*=18.8%, *95% CI*=[17.2%, 20.4%]) were only slightly larger than in the *simple-planning* condition (*M*=16.8%, *95% CI*=[15.3%, 18.3%]; $\chi^2(1)=0.7$, *p*=.404), while the difference in verification rates was in the opposite direction (*planning-plus: M*=2.9%, *95% CI*=[2.2%, 3.6%]; *simple-planning*: *M*=5.0%, *95% CI*=[4.1%, 5.9%]; $\chi^2(1)=2.6$, *p*=.104). If there is a true difference between these conditions, it is too small to detect in our data, so we collapse across these two planning conditions throughout our analyses.

Table 3. Effects of Treatment on Verification Upgrades

	E	F	G
All Plans	0.627 (0.319)**		0.627 (0.319)**
Simple Plans		0.877 (0.34)***	
Plans Plus		0.308 (0.373)	
Courses	Study 2	Study 2	ALL
Course Effects	YES	YES	YES
N	1705	1705	1989
pseudo R^2	.010	.017	.045

4.3. Treatment Effect Heterogeneity

MOOCs attract a large and diverse student body, and it is natural to wonder whether the effects of planning prompts are stronger among certain subgroups of students than others. But many pre-treatment covariates could plausibly moderate the treatment effect and we did not preregister any, so our analysis follows a procedure to correct for multiple comparisons [14]. Specifically, we constructed 13 separate logistic regressions, each of which tested a single interaction between the treatment effect and one of the covariates listed in Table 1, after controlling for course fixed effects and covariate main effects. The p-values from those 13 interaction terms were then corrected to account for the expected false discovery rate [5].

This analysis finds that only one of the covariates - current enrollment in a brick-and-mortar school - significantly moderated the effect on course completion. That is, MOOC students who were also at a traditional school were more likely to benefit from the planning prompt (interaction term: β=1.27, *SE*=0.37; *z*(1514)=3.5, raw *p*<.001; corrected *p*=.043). The next strongest moderator, age, is highly correlated with school enrollment, and not significant after this correction. The regression coefficients imply that planning increased completion rates from 13.6% to 19.4% among those not enrolled in school, and from 12.5% to 25.5% among students who were concurrently enrolled in school. Though exploratory, this result is consistent with the diagnosis that follow-through in MOOCs is rare because of a lack of structure. That is, plan-making seems to be more effective when it is supported by a structured learning environment in students' lives.

5. NATURAL LANGUAGE FORECASTING

In this section we explore our second hypothesis. That is, could the planning prompts also be used to forecast student achievement? Like most text data, the content of the plans are unstructured and high-dimensional, which poses problems for traditional analytic approaches [13, 17, 32].

5.1. Length of Course Plans

Although the planning prompts were optional, 87% of the students in our sample wrote sincere plans (i.e. more than two words). Of those who did write something, the average word count was 33.4 words (*SD*=28.4). The length of students' plans was, at best, a weak predictor of their likelihood of completing the course (β=.004, SE=.002, z(1319)=1.6, p=.156). But the content of the plans was rich, and we parsed them to build a more sophisticated forecasting model.

5.2. Forecasting Course Completion

We use standard natural language tools to extract from each student's planning "document" a set of feature counts - essentially, tallies of concepts and phrases that are commonly mentioned.

These counts were then processed by an algorithm to determine the most distinctive features of successful plans.

5.2.1. NLP Forecasting Model

The documents were first processed manually, in two ways. Every text was spell-checked with software assistance. To resolve synonymy, we then created a simple word substitution algorithm - for example, “each day”, “every day”, “per day” and “daily” were all replaced with the word “daily”. These procedures were unsupervised - that is, they were performed without any knowledge of the students’ treatment condition or certificate status.

The resulting documents were then processed automatically, by converting to lowercase; expanding contractions; removing punctuation; removing common function words (“stopwords”); and stemming the remaining words using the standard Porter stemmer. The remaining word stems were then grouped into “ngrams” - groups of two or three sequential word stems. To focus on the most common features, ngrams which appeared in less than 1% of all documents were excluded. This process reduced the documents to a “feature count matrix”, in which each document (i.e. each student) was assigned a row, while each ngram feature was assigned a column, and the value of each cell represented the number of times that ngram appeared in that document. In addition to the ngram counts, we calculated two summary linguistic features: the raw word count, as well as a binary indicator of which prompts were left blank.

This process produces a high-dimensional set of feature counts which must be regularized in some form to avoid over-fitting. We use a common method, the LASSO, implemented using the *glmnet* package [15, 49]. This algorithm estimates a logistic regression with a constraint on the total absolute size of the coefficients. The size of that constraint is determined empirically, by calculating out-of-sample error via cross-validation within the training set.

This algorithm reduces most coefficients in the regression to exactly zero, leaving a smaller set with non-zero coefficients in the model. The model would then be used to predict the likelihood that new documents (not included in training) were written by students who would go on to complete their course. In essence, they were forecasts of the students’ likelihood of success, which could be compared to forecasts based on other data, and forecasts made by the students themselves. All forecasting models also included course fixed effects, so that they would learn differences between students, not between classes.

Forecaster accuracy was primarily measured using the area under the curve metric (AUC). This tests calculates the probability that the prediction for a randomly-chosen certified student will be higher than the prediction for a randomly-chosen drop-out student. This metric is appealing when, as in our case, the outcomes are unbalanced (i.e. more people dropped out than certified), because it captures the *relative* accuracy of predictions across students. However, we are not concerned with the absolute accuracy of forecasting a single student’s outcome correctly. This simulates a common decision-making margin - for example, if a course administrator has to allocate costly interventions among their students.

5.2.2. Study 1 NLP Results

As an initial test of our hypothesis, we use a strict hold-out procedure. The students from Study 1 served as test data (N=156). To enrich the training data for this test, the sample included all students in Study 2 who intended to complete the course (N=1,792). Table 4 shows the selected language features and their assigned coefficients in the model. These forecasts proved successful in anticipating students’ follow-through (AUC=.659, *95% CI*=[.548,.771]; Mann-Whitney U=1186, p=.009). Because no data from Study 1 were included in training, this result implies that the markers of successful plans are not course-specific.

Table 4. NLP Features Selected by the Forecasting Model

Feature	Coefficient	% of Plans
work.cours	-0.0897	2.4
free.time	-0.0496	5.5
plan.studi	-0.0296	4
time.day	-0.0287	1
home.will	-0.0249	2.1
onehour.daili	-0.0193	3.3
discuss.board	0.0004	2.8
plan.engag	0.0006	1.3
will.engag	0.0410	1.2
complet.work	0.0498	1.1
[word count]	0.0642	—-
hour.week	0.0667	2.4
cours.home	0.0834	1.5
watch.lectur	0.2087	4.3

5.2.3. Study 2 NLP Results

To test out-of-sample accuracy using in-course data, we used a nested cross-validation procedure [45, 51]. Specifically, the dataset was split into 20 folds, and predictions for each fold were made using a model trained and tuned on the other 19 folds. We used two techniques to smooth out instability caused by the cross-validation. First, fold assignment was stratified, to balance the course composition and certification rate in every fold. Additionally, the entire cross-validation procedure was repeated 10 times, and the prediction for every student was an average over those 10 out-of-sample predictions. The language features were once again predictive of course completion among the students who had signed up in the first month and intended to complete the course (AUC=.579, *95% CI*=[.537,.622]; Mann-Whitney U=83238, p<.001).

5.2.4. Study 2 Benchmark Predictions

In Study 2, we sought two other responses from students that could be used to benchmark the results of our NLP model. First, we asked students to directly estimate the probability that they will finish enough of the course to earn a certificate. The average student forecast (83.5%) was far more optimistic than the actual completion rate (16.7%). However, these predictions were still a valid signal of course completion (AUC=.597, *95% CI*=[.555,.639]), because students who completed the course had given higher estimates (M=87.7%, *95% CI*=[85.8,89.6]) than those who did not complete (M=82.9%, *95% CI*=[81.9,83.9], $t(1157)$=4.1, p<.001). This prediction provides a useful benchmark, and shows that natural language forecasting can approximate students' own insights into their expected success.

We also asked students about their grit, using an eight-item survey instrument designed to measure resilience in goal pursuit [11]. This metric also produced similar forecasting accuracy to the natural language forecast (AUC=.577, *95% CI*=[.546,.631]).

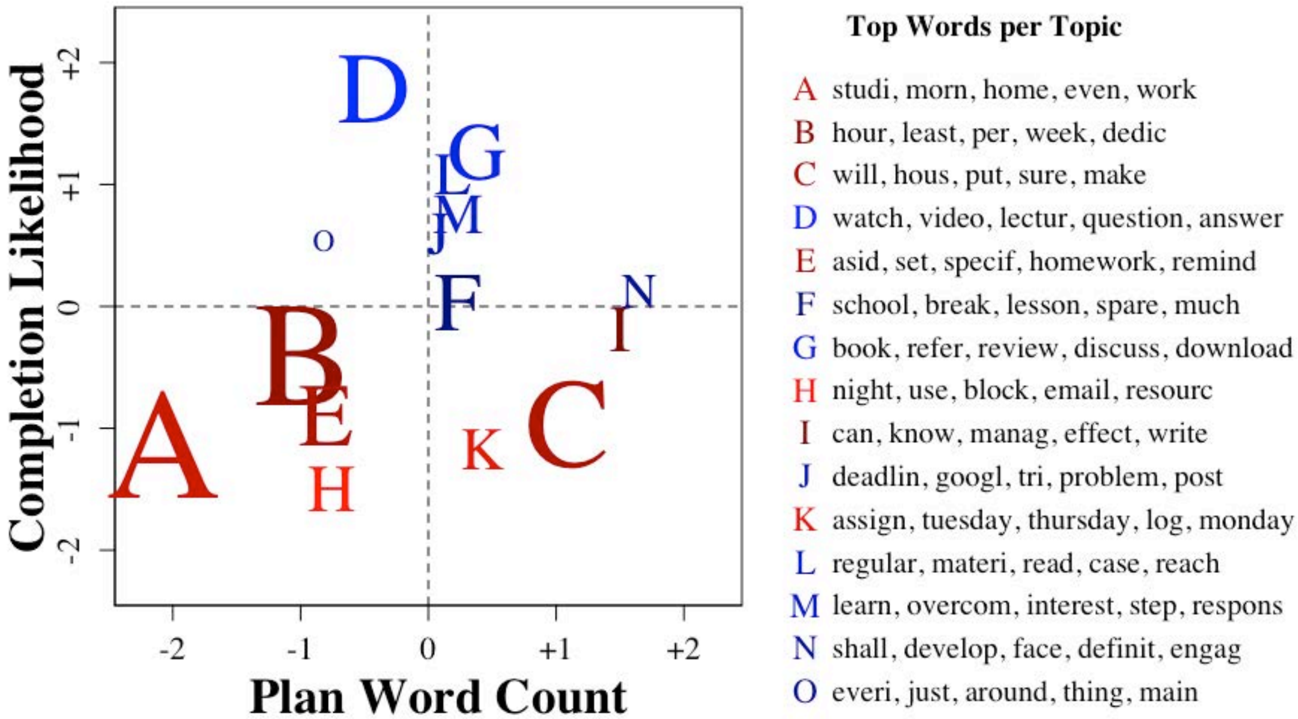

Figure 2. Planning Topics and Course Completion Rates

However, the natural language forecast has two advantages over the grit scale. First, the grit scale forecasts did not improve the fit of the prediction model, beyond a baseline model of students' self-predictions ($\chi^2(1,N=1161)=0.32$, p=.572), whereas the natural language forecasts did explain additional variance in course outcomes ($\chi^2(1,N=1161)=7.46$, p=.006). Second, the grit scale only provides a point estimate of students' expected success. Natural language forecasts, on the other hand, can provide a much richer model of students' follow-through constraints, by revealing the contents of which plans were most (or least) likely to succeed.

5.3. Contents of Course Plans

What, then, were the differences between successful and unsuccessful plans? The feature set in Table 4 is sparse, in part because of the modest sample size for training. Additionally, the coefficients from a lasso regression can be hard to interpret, because the regularization path only selects features that add unique variance as the model becomes more complex. In effect, the selected features are taken out of of their context, by choosing only one feature among many that co-occur together, and which all map onto a common concept.

5.3.1. Topic Modeling

To better understand that mapping, we turn to Latent Dirichlet Allocation as a quantitative model of the planning text [7, 38]. This algorithm clusters words into "topics" based on their co-occurrence, which exploits the very feature that makes the Lasso coefficients opaque. To estimate the topic structure, we pooled all the written plans that were longer than ten words (N=1007). We used a variant of the standard unsupervised LDA algorithm which incorporates a deterministic algorithm for initializing the anchor word of each topic [3]. There are no hard-and-fast rules for choosing the number of topics, so the researcher must choose the an appropriate level of granularity for their own research. Informed by some reasonable guidelines [47], we chose to fit a 15-topic model in this paper, though our basic conclusions are robust across a range of reasonable topic quantities.

After the model was estimated, we calculated the log-normal prevalence of every topic in every document, as well as how that prevalence correlated with both course completion rates and document length (after controlling for differences courses). These two correlations for every topic are plotted against one another in Figure 2 (with units expressed in terms of standardized regression coefficients). The letter labels are sized to scale with total topic prevalence. The legend indicates the top five key words of each topic (by FREX, see [38]). Additionally, the topics that are most distinctive of course completion are presented as word clouds in Figure 3. Qualitatively, the most stark pattern in the topic model was the divide between "context" plans, which focus on the time and location for learning, versus "action" plans, which focus on the materials and methods of learning. We followed up on these observations to quantify them using more structured analyses.

5.3.2. Time Plans

Time was a first-order concern of most students in our experiments - 89% of students who wrote any plan mentioned time at least once, as defined by a pre-written dictionary of time-related words [34]. However, time focus was not necessarily helpful to goal pursuit. In fact, in a logistic regression over all non-blank texts (with course fixed effects), we found that the proportion of time words predicted that a given plan was less likely to succeed ($\beta=-0.223$, SE=0.091, $z(1131)=2.4$, p=.015). That is, plans that focus on time were less likely to succeed than other plans.

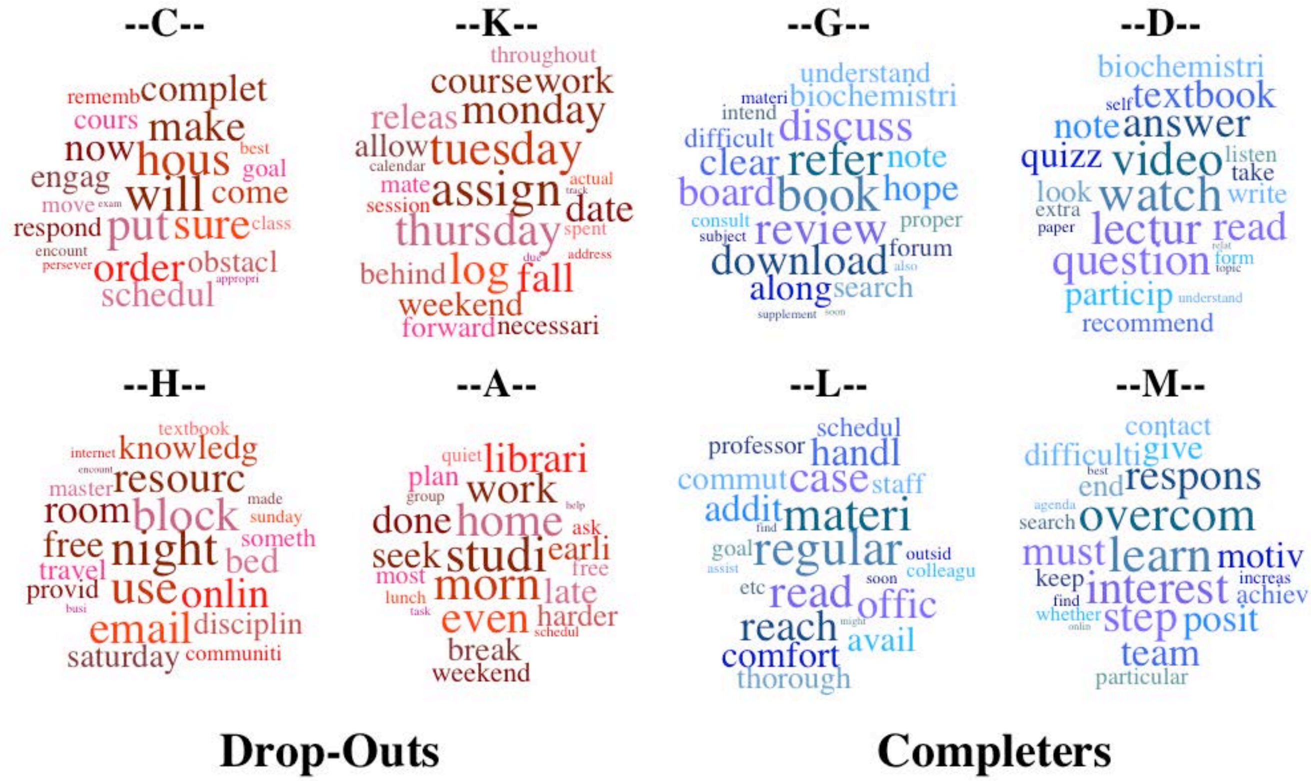

Figure 3. Distinctive words from selected planning topics that predict course completion

Of course, there is a range of possible time plans. In particular, they can also be subdivided using a dictionary of concreteness ratings, along a range from concrete (e.g. “day”, “month”, “afternoon”) to abstract (e.g. “sometime”, “future”, “soon”; see [45]). We calculated the average concreteness of just the words from the time list, following their procedure exactly (i.e. by dropping texts that did not include at least four words from their list). The students who used more concrete time words were less likely to complete the course than students who used more abstract time words (β=-0.168, SE=0.098, z(720)=1.7, p=.086). Concrete time-based plans were less likely to succeed than abstract time-based plans.

These estimates are not causal. Choice of plans is endogenous to contextual factors, like other time pressures, that also affect course completion. But these diagnostic results are still important for understanding goal pursuit. Specific time-and-place plans were most successful in previous field experiments on planning prompts [39]. Our results provide stronger theoretical support for the mechanism behind the causal effects of specific plans. That is, specific plan-making may have been most beneficial for those students who would have focused on time in their open-ended planning prompts. Our work suggests a way to diagnose responsiveness to particular interventions, and future experiments should investigate this hypothesis.

6. GENERAL DISCUSSION

Students enrolling in MOOCs often have ambitious intentions and high expectations that they will follow through on their goal. But most learners who intend to complete a MOOC fail to do so. In this paper we present results from an intervention that is targeted at the follow-through problem. In a field experiment, we asked some students to describe their personal plans for completing the course. They were prompted at the beginning of the course, when intentions were strong, but follow-through was uncertain.

These planning prompts had two benefits. First, planning increased follow-through. We estimate that completion rates among students prompted to make a plan were 29% higher than those who were not prompted to make plans, on average. Additionally, students who planned were 40% more likely to pay for a verified certificate during the course. These effect on completion was almost twice as large among people concurrently enrolled in a traditional school. Planning prompts provided psychological scaffolding for students frame of mind, and paid dividends weeks and months later in terms of greater persistence and completion.

Our results show a second benefit from planning prompts: the text of students’ plans could be used to forecast their success. Natural language processing algorithms could parse the plans and forecast course completion as well as the students’ own predictions, finding predictive features in the text that students do not see for themselves. Students who make plans and succeed are more likely to write about how they will engage with the course, while students who make plans and fail are more likely to write about the concrete steps of when and where they will engage the course.

These results add to mounting evidence from the field for the effect of planning prompts [39]. Ours is the first natural field experiment to show an effect of planning on a long-term goal that requires many actions over time, rather than a single action at one point in the future. The open-ended text data also provide a unique window into the psychology of planning. While previous field experiments in planning have focused on single events where the scope of possible plans is limited, MOOCs are complex goals with many strategies and obstacles. A common task for platforms is plan recognition - that is, anticipating a user's goals from their behavior [8,18]. However, common plan recognition strategies presuppose knowledge of the range of possible plans, whereas in many domains we may want to learn the plans from the data. Our work shows that many MOOC students are willing to report their plans voluntarily at enrollment, and that these plans can be modeled to understand their implementation intentions.

This modeling is important because in an online environment, the mixture - and dosage - of different interventions can be personalized to individual students, based on their needs and likelihood of dropping out. Typically, these prediction problems have been approached using in-course activity data (e.g. [48, 53] though see [21,22]). However, activity data often cannot provide enough lead time for a course designer to intervene before the drop-out is inevitable. Furthermore, MOOC data has shown that most drop-outs occur early in the course [2, 16, 35]. Planning prompts are given to students early, when they have not yet dropped out and their exposure to interventions will be high.

Our results show that planning prompts do not just forecast follow-through failures - they can identify the nature of the impending obstacles. This can provide deeper guidance into the appropriate mid-course intervention. To take one example, Coursera recently introduced a "calendar app" in all of their courses for students to allocate time for future activities. Our results suggest that the treatment effects of this app could be heterogenous, and that an optimal course design policy would push this app into a more prominent place for students whose plans reveal time constraints.

Finally, these results provide context for low completion rates in MOOCs. It is clear that a substantial fraction of students do not achieve the goals they set at the start of the course [2,16,19,37]. However, our research adds to other results that suggest the relative lack of structure in current MOOCs makes it hard for the less diligent students to stick to their goals [6]. As MOOCs become a more established feature in the educational landscape, there will be even more demand for choice architecture to better align students' behavior with their intentions. Planning prompts can be one important response to that demand, and future work should focus on other ways to encourage follow-through. We now have initial evidence that these approaches are effective in MOOCs, and they should be investigated in other forms of online and hybrid learning as well.

7. ACKNOWLEDGEMENTS

This work would not have been possible without the close assistance of Glenn Lopez, Rebecca Petersen, Marshall Thomas, Shilpa Adnani, Zofia Gajdos, Zachary Davis, Heather Sternshein, and Dustin Tingley, and we are grateful for helpful comments from Julia Minson, Todd Rogers, Jacob Whitehill, Joseph Williams and Sendhil Mullainathan.

8. REFERENCES

[1] Ajzen, I. (1991). The theory of planned behavior. *Org Behav Hum Dec, 50(2),* 179-211.

[2] Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In *Proc 23rd Intern'l Conf on World Wide Web,* 687-698.

[3] Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., Wu, Y. & Zhu, M. (2013). A Practical Algorithm for Topic Modeling with Provable Guarantees. In *Proc 30th Inter'l Conf Machine Learning,* 280-288.

[4] Barber, M., Donnelly, K., Rizvi, S., & Summers, L. (2013). An avalanche is coming. *Higher Education and the revolution ahead,* 73.

[5] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B Met,* 289-300.

[6] Banerjee, A. V., & Duflo, E. (2014). (Dis) organization and Success in an Economics MOOC. *Am Econ Rev*, 104(5), 514-518.

[7] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the J Machine Learning Res, 3,* 993-1022.

[8] Carberry, S. (2001). Techniques for plan recognition. *User Modeling and User-Adapted Interaction, 11(1-2),* 31-48.

[9] Chen, C. M. (2008). Intelligent web-based learning system with personalized learning path guidance. *Computers & Education, 51(2),* 787-814.

[10] Clow, D. (2013). MOOCs and the funnel of participation. In *Proc 3rd Inter'l Conf on Learning Analytics and Knowledge,* 185-189.

[11] Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *J Personality Assess, 91(2),* 166-174.

[12] Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Adv Exp Soc Psychol, 38,* 69-119.

[13] Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Polit Anal, 21(3), 298-313.*

[14] Fink, G., McConnell, M., & Vollmer, S. (2014). Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures. *J Dev Effectiveness, 6(1),* 44-57.

[15] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J Stat software, 33(1),* 1.

[16] Ho, A. D., Chuang, I., Reich, J., Coleman, C. A., Whitehill, J., Northcutt, C. G., ... & Petersen, R. (2015). *HarvardX and MITx: Two Years of Open Online Courses Fall 2012-Summer 2014.*

[17] Jurafsky, D. & Martin, J. (2009). *Speech and natural language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* MIT Press.

[18] Kautz, H. A., & Allen, J. F. (1986). Generalized Plan Recognition. *AAAI 86(3237),* 5.

[19] Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proc 3rd Inter'l Conf on Learning Analytics and Knowledge,* 170-179.

[20] Kizilcec, R. F., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational

framings. *Experiences and best practices in and around MOOCs,* 17.

[21] Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *In Proc 2nd ACM Conf on Learning@ Scale,* 57-66.

[22] Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction,22(2),* 6.

[23] Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016, April). Recommending Self-Regulated Learning Strategies Does Not Improve Performance in a MOOC. In *Proc 3rd ACM Conf on Learning@ Scale,* 101-104.

[24] Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. *Empirical Methods on Natural Language Processing*, 60.

[25] Koehler, D. J., & Poon, C. S. (2006). Self-predictions overweight strength of current intentions. *J Exp Soc Psychol, 42(4),* 517-524.

[26] Koller, D. (2011). Death knell for the lecture: Technology as a passport to personalized education. *New York Times, Dec 5,* 5.

[27] Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause Rev, 48(3),* 62-63.

[28] Kruger, J., & Evans, M. (2004). If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *J Exp Soc Psychol, 40(5),* 586-598.

[29] Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proc Natl Acad Sci USA, 108(26),* 10415-10420.

[30] Milkman, K.L., Beshears, J., J.J. Choi, D. Laibson, and B.C. Madrian (2013). Planning prompts as a means of increasing preventive screening rates. *Prev Med, 56,* 92-93.

[31] Nickerson, D.W. & Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychol Sci, 21(2),* 194-199.

[32] O'Connor, B., Bamman, D., & Smith, N. A. (2011). Computational text analysis for social science: Model assumptions and complexity. *Public Health*, *41*(42), 43-50.

[33] O'Donoghue, T. & Rabin, M. (1999). Doing it Now or Later. *Am Econ Rev, 89 (1),* 103-124.

[34] Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC [Computer software]. Austin, TX: liwc. net.

[35] Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving through MOOCs understanding the progression of users in Massive Open Online Courses. *Educ Res, 0013189X14562423.*

[36] Poon, C. S., Koehler, D. J., & Buehler, R. (2014). On the psychology of self-prediction: Consideration of situational barriers to intended actions. *Judgm Decis Making, 9(3),* 207.

[37] Reich, J. (2014). MOOC completion and retention in the context of student intent. *EDUCAUSE Review Online.*

[38] Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *Am J Polit Sci, 58(4),* 1064-1082.

[39] Rogers, T., Milkman, K., John, L., & Norton, M. I. (2016). Making the best-laid plans better: how plan making increases follow-through. *Behav Sci & Pol, In Press.*

[40] Schacter, D.L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *Am Psychol, 54*, 182-203.

[41] Self, J. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *Inter'l J Artif Intell in Educ, 10,* 350-364.

[42] Shah, D. (2015). By The Numbers: MOOCS in 2015 - Class Central's MOOC Report. Retrieved June 05, 2016, from https://www.class-central.com/report/moocs-2015-stats/

[43] Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Pers Soc Psych B, 31(1),* 87-98.

[44] Snefjella, B., & Kuperman, V. (2015). Concreteness and psychological distance in natural language use. *Psychol Sci, 26(9),* 1449-1460.

[45] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J Royal Stat Soc B Met,* 111-147.

[46] Sunstein, C. R., & Thaler, R. (2008). *Nudge. The politics of libertarian paternalism.* New Haven.

[47] Taddy, M. (2012). On Estimation and Selection for Topic Models. In *15th Inter'l Conf on Artif Intel and Stat,* 1184-1193.

[48] Taylor, C., Veeramachaneni, K., & O'Reilly, U. M. (2014). Likely to stop? predicting stop-out in massive open online courses. arXiv preprint arXiv:1408.3382.

[49] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J Royal Stat Soc B Met,* 267-288.

[50] Tomkin, J. H., & Charlevoix, D. (2014, March). Do professors matter?: Using an a/b test to evaluate the impact of instructor involvement on MOOC student outcomes. In *Proc 1st ACM Conf Learning@Scale,* 71-78

[51] Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics, 7(1),* 91.

[52] Wen, M., Yang, D., & Rose, C. (2014). Sentiment Analysis in MOOC Discussion Forums: What does it tell us? In *Proc 7th Intl Conf Educ Data Mining, 130-137*

[53] Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps towards automatic intervention in MOOC student stopout. *Proc 8th Intl Conf Educ Data Mining, 171-179.*

[54] Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015, March). Exploring the effect of confusion in discussion forums of massive open online courses. In *Proc 2nd ACM Conference on Learning@Scale,* 121-130.