Title
Using Smartpens to Examine and Influence the Relationship between Homework Habits and Academic Achievement in Introductory Engineering Courses

Author
Rawson, Kevin

Publication Date
2019

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Using Smartpens to Examine and Influence the Relationship between Homework Habits and Academic Achievement in Introductory Engineering Courses

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Mechanical Engineering

by

Kevin Christopher Rawson

December 2019

Dissertation Committee:
Dr. Thomas Stahovich, Chairperson
Dr. Mona Eskandari
Dr. Vagelis Hristidis

The Dissertation of Kevin Christopher Rawson is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I'd like to thank my advisor, Dr. Stahovich, for his guidance, especially in pushing me to always improve the clarity in my writing and in the presentation of my research. To my lab mates in the Smart Tools Lab, thanks for the friendship and help along the journey: Jim Herold, Hank Lin, Jack Lee, Reyna Garcia, Tim Van Arsdale, Matt Ung, Justin Gyllen, Teresa Wright, Josh Frear, Amir Darvishzadeh, Negin Entezari, Adam Koehler, Vashishtha Bhatt, and Ryan Mercer.

I'd like to thank my family: Mom, Dad, Amy and Robin (Mitch, Rick, and Gemma too) for your support and encouragement throughout.

The text of Chapter 2 in this dissertation is a reprint of the material as it appears in the Journal of Educational Psychology, February 2017. The co-authors, Dr. Thomas Stahovich and Dr. Richard Mayer, listed in that publication directed and supervised the research which forms the basis of this chapter.

This work was additionally supported by the National Science Foundation under award numbers 0935239, 1432820 and 1612511.

ABSTRACT OF THE DISSERTATION

Using Smartpens to Examine and Influence the Relationship between Homework Habits and Academic Achievement in Introductory Engineering Courses

by

Kevin Christopher Rawson

Doctor of Philosophy, Graduate Program in Mechanical Engineering
University of California, Riverside, December 2019
Dr. Thomas Stahovich, Chairperson

This dissertation examines students' homework behaviors and their relationship to academic achievement in introductory engineering courses. Much of the prior work examining the relationship between homework and achievement has relied on student self-reports of homework effort. Our results demonstrate that such self-reports are problematic. Instead, we avoid this methodological shortcoming by using smartpens to objectively measure students' learning activities in an unobtrusive manner and with a high level of fidelity. This dissertation examines how much, how frequently, and when students work on their homework assignments, and if these factors are related to

achievement. This dissertation also examines if informing students of their homework behavior influences them to change that behavior and improve achievement.

This work makes four major contributions. First, we developed quantitative measures of student homework behavior that are related to academic achievement. Second, we demonstrate that self-reported measures of student homework effort are problematic. Third, we show that measures of homework effort early in a course are nearly as effective at predicting achievement as measures from the entire course. This result suggests that student behavior does not change significantly over a course. Finally, we show that informing students of their homework behaviors, and providing suggestions for improving those behaviors, is an insufficient motivator to change behaviors and improve achievement. This result suggests a two-stage model of metacognition for study behaviors, requiring both monitoring (i.e., being aware of how one is studying) and regulation (i.e., adjusting how one studies based on feedback) to affect changes in behavior.

This work makes both applied and methodological contributions to educational research. In contrast to existing research, our results demonstrate a strong and consistent relationship between students' homework behaviors and academic achievement. Additionally, this work shows that students' homework behaviors are established early in a course, and tend to remain relatively constant throughout a course.

This work highlights the potential of educational data mining and smartpen technology for educational research. Our results confirm the unreliability of studies employing self-reports. Our studies also speak to the value of replication in education research.

Table of Contents

Acknowledgements....................iv
Abstract....................v
Tables of Contents....................viii
List of Figures....................ix
List of Tables....................x
Chapter 1: Introduction....................1
Chapter 2: Homework and Achievement: Using Smartpen Technology to Find the Connection....................4
Chapter 3: Homework Habits that Predict Academic Success....................49
Chapter 4: Effects of Weekly Reports of Study Behavior in a College Course: Evidence for a Two-Stage Model of Metacognition....................101
Chapter 5: Conclusions....................159
References....................165
Appendix....................175

List of Figures

Figure 1. A model of academic learning. 9
Figure 2. A typical statics problem. 19
Figure 3. A solution to the statics problem from Figure 2. 20
Figure 4. Correlation between *Early Work Fraction* (E) and course grade. 72
Figure 5. Correlation between *Problems Attempted* (P) and course grade. 74
Figure 6. Correlation between *Stroke Count* (S) and course grade. 75
Figure 7. Correlation between *Length of Ink Written* (L) and course grade. 76
Figure 8. Correlations between each of the four measures of conscientiousness homework habits and course grade averaged over the seven courses. Error bars represent one standard deviation. 78
Figure 9. Correlation of stepwise regression models relating all four measures of conscientiousness homework habits to course grade. 80
Figure 10. Correlation between *Combined Measure of Conscientiousness* (C) in homework habits and course grade for each of the seven courses. 82
Figure 11. Correlations between *Combined Measure of Conscientiousness* (C) in homework habits and course grade averaged over the seven courses. Error bars represent one standard deviation. 83
Figure 12. Correlation of stepwise regression models relating the sequence of *Early Work Fraction* (E) values to course grade. 85
Figure 13. Correlation of stepwise regression models relating the sequence of *Problems Attempted* (P) values to course grade. 87
Figure 14. Correlation of stepwise regression models relating the sequence of *Stroke Count* (S) values to course grade. 88
Figure 15. Correlation of stepwise regression models relating the sequence of *Length of Ink Written* (L) values to course grade. 89
Figure 16. An example of the weekly progress report. 105
Figure 17. A two-stage model of metacognition for study behaviors. 113
Figure 18. A model of how weekly reports affect metacognition for study behaviors. 114
Figure 19. A typical statics problem. In this example, the student must solve for the minimum and maximum force *F* that will keep the two blocks with mass *m* and 2*m* from moving, given the blocks are attached together by a rope and the coefficient of static friction between the blocks and the inclined plane is μ_s. 116

List of Tables

Table 1 *Thirteen Measures Derived Through Smartpen Technology* 25
Table 2 *Correlation Between Course Grade and Each of 13 Smartpen Measures for all Students and Each Cohort Separately* 29
Table 3 *Means and Standard Deviations for Each of 13 Smartpen Measures for all Students and Each Cohort Separately* 31
Table 4 *Three Core Features of Conscientiousness* 51
Table 5 *Summary of Added Instructional Treatments in Each Course Offering* 63
Table 6 *Means and Standard Deviations for Conscientiousness Measures for Each Cohort over the Entire Quarter* 70
Table 7 *Two Metacognitive Processes Related to Study Behavior* 112
Table 8 *Feedback Related to Total Time Spent on Homework* 120
Table 9 *Feedback Related to Problems Attempted on Homework* 121
Table 10 *Feedback Related to Early Work on Homework* 122
Table 11 *Feedback Related to Lecture Notes* 124
Table 12 *Feedback Related to Total Time Spent Reading* 127
Table 13 *Feedback Related to Assigned Pages of Reading* 128
Table 14 *Feedback Related to Midterm Exam Scores* 128
Table 15 *List of Weekly Reports Provided to Students in the Experimental Group* 132
Table 16 *Weekly Reports Read and Course Statistics* 133
Table 17 *Comparison of Proportion Started Early Between the Experimental and Control Group* 135
Table 18 *Comparison of Problems Attempted (Normalized) Between the Experimental and Control Group* 137
Table 19 *Comparison of Total Time (Normalized) Spent on Homework Between the Experimental and Control Group* 138
Table 20 *Comparison of Stroke Count (Normalized) Between the Experimental and Control Group* 139
Table 21 *Comparison of Length of Ink Written (Normalized) Between the Experimental and Control Group* 140
Table 22 *Comparison of Percentage of Lectures Submitted Between the Experimental and Control Group* 141
Table 23 *Comparison of Total Time Reading (in Minutes) Between the Experimental and Control Group* 143
Table 24 *Comparison of Explanatory Text Reading Time (in Minutes) Between the Experimental and Control Group* 144
Table 25 *Comparison of Cumulative Percentage of Assigned Pages Read Between the Experimental and Control Group* 146
Table 26 *Correlation Between the Number of Read Weekly Reports and Measures of Success in the Course* 147
Table 27 *Correlation Between Early Work Fraction* (E) *and Course Grade* 175
Table 28 *Correlation Between Problems Attempted* (P) *and Course Grade* 176
Table 29 *Correlation Between Stroke Count* (S) *and Course Grade* 177
Table 30 *Correlation Between Length of Ink Written* (L) *and Course Grade* 178

Table 31 *Correlation of Stepwise Regression Models Relating All Four Measures of Conscientiousness Homework Habits to Course Grade* 179
Table 32 *Correlation Between Combined Measure of Conscientiousness* (C) *in Homework Habits and Course Grade* 180
Table 33 *Correlation of Stepwise Regression Models Relating the Sequence of Early Work Fraction* (E) *Values to Course Grade* 181
Table 34 *Correlation of Stepwise Regression Models Relating the Sequence of Problems Attempted* (P) *Values to Course Grade* 182
Table 35 *Correlation of Stepwise Regression Models Relating the Sequence of Stroke Count* (S) *Values to Course Grade* 183
Table 36 *Correlation of Stepwise Regression Models Relating the Sequence of Length of Ink Written* (L) *Values to Course Grade* 184

Chapter 1

Introduction

Homework is a component of nearly every curriculum, from kindergarten to graduate school. Because of the ubiquitous nature of homework, the relationship between homework activity and achievement has been explored extensively. However, the results are mixed. Some studies have found a positive relationship between homework and achievement (Ebbinghaus, 1885), some have found a negative relationship (Jones & Ruch, 1928), and yet others have found no relationship (Schuman, Walsh, Olson, & Ethridge, 1985). Most of this research has relied upon self-reported measures of homework effort, and these disparate results brings into question whether this is a valid method to examine this issue. To avoid these pitfalls, we seek to objectively measure students' homework behavior using digital smartpens that capture student homework activity in a natural and unobtrusive manner. From the smartpen data, we are able to extract temporal, spatial, and sequential measures of the students' behavior as they complete homework, with a level of detail and fidelity not previously possible.

We conducted a sequence of three studies exploring the relationship between homework behaviors (or habits) and achievement. In the first study (Chapter 2), we used smartpens to record students' homework behavior over an entire course and identify behavioral measures that correlate with academic achievement. In the second study (Chapter 3), we examined if these objective measures can be used for early prediction of academic success. In our third study (Chapter 4), we examined if informing students of their homework behaviors, and providing them with recommendations for changing their behaviors, lead to improved academic achievement.

This dissertation makes several important contributions:

1. We identify objective measures of student homework behavior that are significantly and strongly correlated with achievement.
2. We show the quantity of time spent on homework as well as the quality of how the homework time is used are related to achievement.
3. We show that self-reported measures of homework effort are problematic, casting doubt upon the results of much of the prior work examining the relationship between homework and achievement.
4. We show measures of conscientious habits (i.e., being on time, being on task, and producing high effort) in solving homework problems are strongly and consistently correlated with achievement.
5. We show that measures of conscientious habits early in a course are nearly as predictive of achievement as those same measures taken over an entire course, laying the groundwork for a practical and scalable early warning system to detect students at-risk of poor performance in a class.
6. We show that informing students of their homework behaviors is an insufficient motivator to influence changes in behavior.

7. We find evidence for a two-stage model of metacognition. More specifically, we find that monitoring one's study behavior is insufficient to improve learning outcomes. Instead, improving learning outcomes requires that monitoring be coupled with regulating one's study behavior, that is, adjusting how one studies based on feedback.

This dissertation consists of three main studies. These studies are presented in journal paper format. One of these studies (Chapter 2) has already been published (doi: 10.1037/edu0000130), while the other two studies (Chapter 3, Chapter 4) are not yet published.

Chapter 2

Homework and Achievement: Using Smartpen Technology to Find the Connection

Abstract

There is a long history of research efforts aimed at understanding the relationship between homework activity and academic achievement. While some self-report inventories involving homework activity have been useful for predicting academic performance, self-reported measures may be limited or even problematic. Here, we employ a novel method for accurately measuring students' homework activity using smartpen technology. Three cohorts of engineering students in an undergraduate statics course used smartpens to complete their homework problems, thus producing records of their work in the form of timestamped digitized pen strokes. Consistent with the time-on-task hypothesis, there was a strong and consistent positive correlation between course grade and time doing homework as measured by smartpen technology (r = .44), but not between course grade and self-reported time doing homework (r = -.16). Consistent with an updated version of the time-on-task hypothesis, there was a strong correlation between measures of the quality of time spent on homework problems (such as the proportion of ink produced for homework within 24 hr of the deadline) and course grade (r = -.32), and between writing activity (such as the total number of pen

strokes on homework) and course grade (r = .49). Overall, smartpen technology allowed a fine-grained test of the idea that productive use of homework time is related to course grade.

Introduction

Homework is defined as "tasks assigned to students by school teachers that are meant to be carried out during non-school hours" (Cooper, 1989, p. 7). Homework has the potential to improve academic learning, perhaps by extending time to learn beyond the classroom and priming active cognitive processing for learning (Cooper, 1989, 2001; Mayer, 2011). Assigning homework problems to be solved by students outside of class time is a common practice in college courses in engineering, mathematics, and science. The goal of the present study is to determine how students' problem-solving activity on homework is related to their course grade in introductory-level engineering courses.

Smartpen Technology

Suppose a teacher assigns homework problems for students to work on each week. How can we know the degree to which students engage with the homework assignment? We could ask them to report how much time (or effort) they put into the homework assignments, but self-reported measures can be problematic. Instead, imagine that a teacher could assign homework problems to students and be able to monitor the student's homework activity at any time and any place, even outside of class. In short, suppose we had a way to know when a student was working on a homework assignment and we were able to record every pen stroke a student made while working on a handwritten assignment. This level of rich data mining of student handwritten homework activity employed in the current study is enabled by the use of newly developed smartpen technology that accomplishes this goal (Herold, Stahovich, Lin, & Calfee, 2011).

Rationale

Researchers have long sought to understand the role of study activities (including homework activities) in academic achievement. For example, Jones and Ruch (1928) examined the relationship between the amount of time spent studying and first semester grade point average. More recently, Credé and Kuncel (2008) conducted a meta-analysis of 10 study habit, skill, and attitude inventories and found that they had incremental validity in predicting academic performance.

Much of this work relies on surveys and students' self-reports of study habits, which may limit the reliability. For example, Schuman, Walsh, Olson, and Etheridge (1985) found little relation between study time and grades, and attributed this to "the possible invalidity of student reports of their own studying" (p. 961). Blumner and Richards (1997) found that a study habit inventory was useful for differentiating between high- and low-performing students. However, the authors concluded that: "It will be necessary to directly observe students in the act of studying. Only in this manner can it be determined that students actually do what they say in response to such an inventory" (p. 132).

In our present work, we take up this challenge, and use Livescribe Smartpens™ to measure students' homework activity. These devices have an integrated camera and are used with dot-patterned paper. They serve the same function as a traditional ink pen and also record the work as timestamped pen strokes. We conducted studies in three offerings of a sophomore-level undergraduate engineering course in statics. Students in these courses completed their homework assignments using the smartpens, thus producing records of the work in the form of timestamped digitized pen strokes.

Homework

There is encouraging evidence—much dating from the 1980s (Keith, 1982)—for the educational value of homework (Cooper, Robinson, & Patall, 2006; Hattie, 2009; Xu, 2013). At the grossest level, Hattie (2009) reported an average effect size of $d = .29$ favoring homework, based on five meta-analyses involving 295 experimental tests and over 100,000 students. In another review of research on the relation between homework and achievement, Cooper, Robinson, and Patall (2006) found a weighted average correlation of $r = .24$ based on 69 separate correlations. Importantly, the research team found the positive correlation between homework and achievement was greater for older students (e.g., high school students) than for younger students (e.g., elementary school students).

Although early research focused on the quantity of homework activity (such as the reported time spent on homework), Xu (2013) has proposed that the next step in research on homework is to more carefully examine the quality of homework activity—including the learner's effort and activity. A methodological obstacle to determining the relation between homework and achievement is that much of the existing research is based on students' self-reported time (or effort) on homework rather than on their actual activity. A related methodological obstacle is that the focus is on what homework is assigned by teachers rather than on what is done by students as they work on their homework.

The present study overcomes these challenges by employing a computer-based technology for tracking the details of students' homework activity in real time using smartpens. This technology provides a level of detail about what students are doing and when they are doing it that is not possible in classic research on homework. Thus, this technology-enhanced system provides data for an updated examination of the connection between homework and achievement.

Theory and Predictions

The amount of time that students choose to give to a task can be considered a measure of student engagement (Hattie, 2009; van Gog, 2013). Student engagement during learning is at the heart of theories of meaningful learning such as cognitive load theory (Sweller, Ayres, & Kalyuga, 2011) and the cognitive theory of multimedia learning

(Mayer, 2009, 2014), and theories of academic motivation such as self-efficacy theory (Schunk & Pajares, 2009) and attribution theory (Graham & Williams, 2009). Figure 1 shows the proposed causes and consequences of student engagement during learning. In terms of what causes students to exert effort, the left side of Figure 1 proposes that instructional features (such as interactivity and personalization) and student characteristics (such as self-efficacy and interest) can prime the level of student effort during learning. A major task of research on instructional design is to identify instructional features that cause the learner to exert effort to learn, and a major task of research on academic motivation is to identify motivational beliefs that cause the

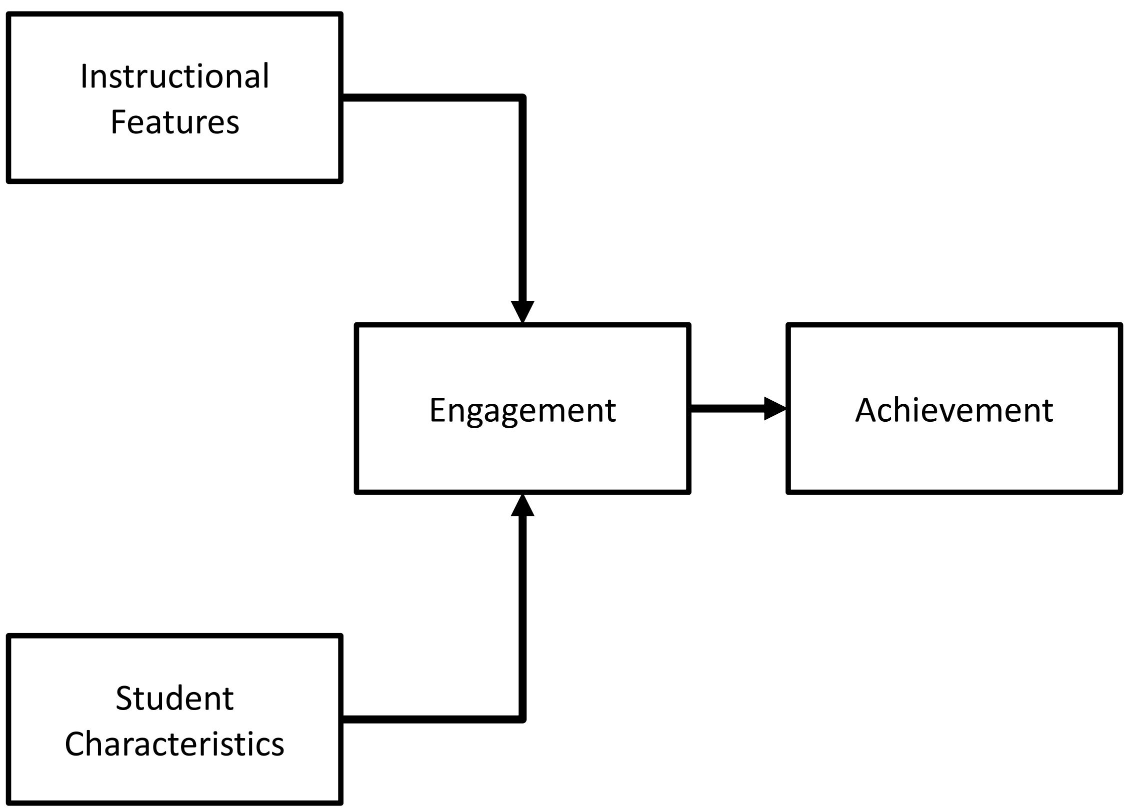


Figure 1. A model of academic learning.

learner to exert effort to learn. In terms of what are the consequences of students engagement, the right side of Figure 1 shows that effort to learn can lead to better learning outcomes as indicated by measures of achievement.

According to this basic model of academic learning, engagement (as indicated by the amount of time that students allocate to a task) is a mechanism affecting learning outcomes (as indicated by achievement). Our focus in the current study is on the relation between time on a study task (i.e., doing homework assignments) and grades in a college course. Thus, our focus is on a crucial link in a model of academic learning. A major new contribution is a more detailed measurement of student engagement on a study activity (i.e., doing handwritten homework assignments) than has been previously available.

Our predictions are based on the time-on-task hypothesis (Hattie, 2009; van Gog, 2013), which holds that learning new material is related to the amount of time a student is effortfully engaged in a productive learning activity. Productive learning activities are those that cause the student to attend to relevant material, mentally organize it, and relate it with relevant prior knowledge (Mayer, 2009, 2014). Spending time on homework is one way to increase productive learning time beyond the school day.

Time-on-task—defined as the amount of time a student spends engaged in an academic task—can be “counted among the most important factors affecting student learning and achievement” (van Gog, 2013, p. 432). Rooted in Ebbinghaus’ (1885/1964) classic studies on verbal learning which showed that time spent studying a word list is related to the amount learned, time-on-task has been recognized as a potentially important variable in academic learning since the 1960s (Berliner, 1991; Carroll, 1963; van Gog, 2013). In a review of meta-analyses, Hattie (2009) found an average effect size of d = .38 for time-on-task based on four meta-analyses examining 136 experimental comparisons.

Over the years, the concept of time-on-task has evolved to reflect a focus on *engaged learning time*—time in which the learner is actively exerting effort on a task—rather than *allocated learning time*—time in which the instructor provides opportunities for learning (Berliner, 1984; Karweit, 1984). Within engaged learning time, furthermore, researchers have come to focus on *productive learning time*—time in which the learner is exerting effort to learn on an appropriate academic task (Berliner, 1984). For example, van Gog, Ericsson, Rikers, and Paas (2005) point to the role of *deliberate practice*—spending extended periods of time effortfully engaged in tasks at an appropriate level of challenge that allow for continual improvement. Early work by Anderson (1993) provides an exemplary demonstration of the role of practice time in learning with computer-based cognitive tutors, and current work continues to demonstrate the positive impact of solving practice problems in e-learning (Clark & Mayer, 2011). Although learning mechanisms were not highlighted in the early conceptions of time-on-task, the updated version of the time-on-task hypothesis is consistent with the idea that meaningful learning requires active cognitive processing in working memory during learning such as attending to relevant information, mentally organizing it into a coherent structure, and relating it to relevant prior knowledge activated from long-term memory (Mayer, 2011). What turns learning time into productive learning time is that the learner is engaged in appropriate cognitive processing on appropriate tasks during learning—processing that leads to constructing new knowledge and skills. Based on these revisions in the classic concept of time-on-task, we expand the time-on-task hypothesis to focus also on the quality of time spent on homework. Overall, we examine three predictions about the relation between homework and achievement concerning the quantity of time (i.e., how much) and the quality of time (i.e., when).

1. How much: The most straightforward prediction of the time-on-task hypothesis is that time spent solving homework problems is related to course grade. However, a problem with traditional research on homework is that some studies use self-reported estimates of time spent doing homework. An important improvement in the current technology-supported study is that we have access to the actual time that students were working on their homework problems, including when they started and ended each session.

2. When: In addition to focusing solely on time spent on homework, a more sophisticated approach is to measure the quality of the time, such as the degree to which the homework activity was performed in advance of the deadline for submission. Although traditional research

on homework generally does not include measures of when the homework was done, our technology-supported environment allows us to test the prediction that doing homework farther in advance of the deadline is related to course grade.

3. How many: In addition to focusing solely on time spent on homework, a more sophisticated approach is to measure how the time was spent. This challenge is problematic with traditional research on homework that does not involve in-process measures of homework activity. However, in our technology-supported environment, a straightforward way to measure the amount of effort put into doing homework is to count the number of strokes performed in solving homework problems. This allows us to test a more focused version of the time-on-task hypothesis: number of strokes performed while solving homework problems is related to course grade.

We examine these three predictions, and related predictions, across three cohorts of engineering students enrolled in an introductory course in statics.

Related Research on Data Mining in Education

Educational data mining with computer-based instructional systems has a rich history dating back to large-scale studies of computer-assisted instruction (CAI) in schools in the 1960s (e.g., Atkinson, 1968), extensive use of log files for modeling student learning with computer-based cognitive tutors (Anderson, 1993), and the subsequent use of log files with intelligent tutoring systems (Koedinger, D'Mello, McLaughlin, Pardos, & Rosé, 2015). In recent years, researchers have made significant progress in educational data mining or EDM (Koedinger, D'Mello, McLaughlin, Pardos, & Rosé, 2015; Romero, Romero, Luna, & Ventura, 2010). Much of the data used in this work is extracted from log files of intelligent tutoring systems (Beal & Cohen, 2008; Li, Cohen, Koedinger, & Matsuda, 2011; Mostow, González-Brenes, & Tan, 2011; Shanabrook, Cooper, Woolf, & Arroyo, 2010; Stevens, Johnson, & Soller, 2005; Trivedi, Pardos, Sráközy, & Heffernan, 2011) and learning management systems such as Moodle and Blackboard (Krüger, Merceron, & Wolf, 2010; Romero, Ventura, Vasilyeva, & Pechenizkiy, 2010). This work relies on a variety of data mining techniques including clustering (Antonenko, Toy, & Niederhauser, 2012; Stevens et al., 2005; Trivedi et al., 2011), model prediction (Li et al., 2011; Mostow et al., 2011; Stevens et al., 2005), and sequence analysis (Beal & Cohen, 2008; Kruger et al., 2010; Romero, Romero, et al., 2010; Shanabrook et al., 2010).

Our work differs from this in that we record and mine data from learning activities involving writing on paper, rather than activities involving typing on a computer keyboard. The work of Oviatt, Arthur, and Cohen (2006) suggests that natural work environments are critical to student performance. In their examinations of computer interfaces for completing geometry problems, they found that "as interfaces departed more from familiar work practice..., students would experience greater cognitive load such that performance would deteriorate in speed, attentional focus, metacognitive control, correctness of problem solutions, and memory” (p. 191). Similarly, Anthony, Yang, and Koedinger (2008) found that handwriting interfaces were more beneficial than keyboard interfaces for math tutoring systems. Mueller and Oppenheimer (2014) made a similar finding in relation to note-taking. They examined student note-taking using both longhand and laptops, and found that the latter can lead to shallower processing. Lectures were shown on a screen, with students taking notes, followed by distractor tasks. Using a model including both word count and verbatim overlap (three-word chunks from student notes matching the lecture transcript), they were able to predict performance on a test of the lecture material with a correlation coefficient of $r = .41$.

Macfadyen and Dawson (2010) mined data from a learning management system (LMS) to predict final course grade. Their best model was able to explain 33% of the variance in grade utilizing three features: the number of mail messages sent, the number of assessments finished, and the total number of discussion messages posted. This provides some insights about the relationship between studying and course performance. However, the type of data available from a LMS—such as records of downloading course materials and submitting electronic assignments—does not provide a direct measurement of students’ homework activity. We use smartpens to capture a fine-grained record of students’ handwritten homework.

Researchers have used video recording to analyze students' problem-solving activities (Blanc, 1999; Hall, 2000). While this approach provides a detailed record of student work, the analysis is time-consuming. For example, Blanc (1999) made 75 recordings of students solving mathematics problems, but analyzed only two of the recordings. This sort of video analysis would be intractable in our studies, which involve hundreds of students completing homework throughout a quarter-long course. For our studies, smartpens provide a convenient and scalable approach for capturing high-resolution, timestamped records of problem-solving work.

There have been prior studies examining learning activities in statics. For example, work by Steif and Dollár (2009) examined usage patterns of a web-based statics tutoring system and found that learning gains increased with the number of tutorial elements completed. Similarly, work by Steif, Lobue, Kara, and Fay (2010) examined whether students can be induced to talk about the bodies in a statics problem, and if doing so can increase a student's performance. They used tablet PCs to record the students' spoken explanations and their handwritten solutions, but the written work was left mostly unanalyzed.

Researchers have only recently begun using smartpens for assessment. For example, Herold and Stahovich (2012) used smartpens to examine the homework of students who were asked to provide self-explanations for their solutions to statics problems. The study found that students who generated self-explanations were more likely to complete homework problems in the order assigned (i.e., complete one problem before beginning the next) than were students who did not generate self-explanations.

Our work builds on that of Rawson and Stahovich (2013) who used smartpens as part of a technique for making early predictions of student success or failure in a statics course. They used smartpens to record students' work on one homework assignment and a corresponding quiz given early in the course. They computed a number of features from this digital ink data including, for example, the total time spent on the homework and the amount of ink written. By themselves, these features were only weakly predictive of a student's course performance. However, when combined with a concept inventory score (Steif & Dantzler, 2005), these features produced useful early predictions.

In our work, we employ many of the ink features they developed. However, our goals are different. While their goal was to use data collected at the beginning of a course to make early predictions of success and failure, ours is to understand the relationship between homework habits and course performance. Our analysis considers homework behavior over the entire duration of a course, while they considered work from only a single assignment and quiz.

Recently, Van Arsdale and Stahovich (2012) demonstrated that the spatial and temporal organization of a student's solution to an engineering problem is indicative of the correctness of that solution. They recorded students' work on exam problems using smartpens and characterized the problem-solving activity in terms of the sequence of problem-solving steps and the arrangement of the work on the page. While they focused on a microscale analysis of problem-solving behavior on individual exam problems, we consider a macroscale analysis of homework habits over the duration of a course.

Herold, Stahovich, and Rawson (2013) used smartpens to examine the correlation between effort on a homework assignment and grade on that assignment. They characterized effort in terms of the amount of time spent and the amount of ink written. They also examined transfer from homework problems to subsequent homework, quiz, and exam problems. They characterized problem solving work by the amount of time the pen was in contact with the paper, which is only a fraction of the time spent on the problem. They found that this “writing time” was correlated with performance on subsequent problems. Our work is similar in that we also examine the relationship between homework activity and success. However, we consider a longer time scale and our focus is understanding how homework habits over an entire course relate to success in that course.

Method

Participants and Course Setting

The participants were three cohorts of undergraduate engineering students at the University of California, Riverside who were enrolled in an entry-level course in statics—92 students in the winter quarter of 2010 (Year 1), 109 students in the winter quarter of 2011 (Year 2), and 127 students in the winter quarter of 2012 (Year 3). The winter term is the first offering of the statics course for the academic year. The majority of the students in the course are from mechanical engineering, although students from several other engineering majors, including materials science and environmental engineering, also take the course. Mechanical engineering students typically take the course in the sophomore year. The course includes two 80-min lecture periods per week. Students also attend a 50-min discussion section each week. The course employs a traditional lecture format.

Statics is the part of engineering mechanics focused on the equilibrium of objects subject to forces. The solution to a statics problem typically includes free body diagrams and equilibrium equations. The former represent the forces acting on a system, while the latter are the application of Newton's Second Law. Figure 2 shows a typical homework problem from the course and Figure 3 shows the sort of solution a student might generate for this problem. This image was constructed from digitized pen strokes captured with a smartpen.

The coefficient of static friction between block A and its incline is 0.25. What must the minimum coefficient of static friction between block B and its incline be, if the blocks are in equilibrium? Neglect friction in the pulley.

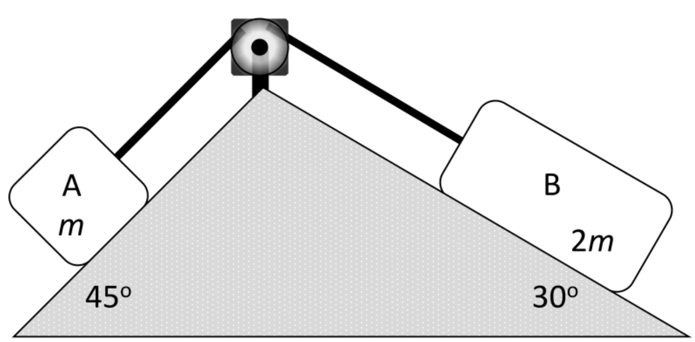


Figure 2. A typical statics problem.

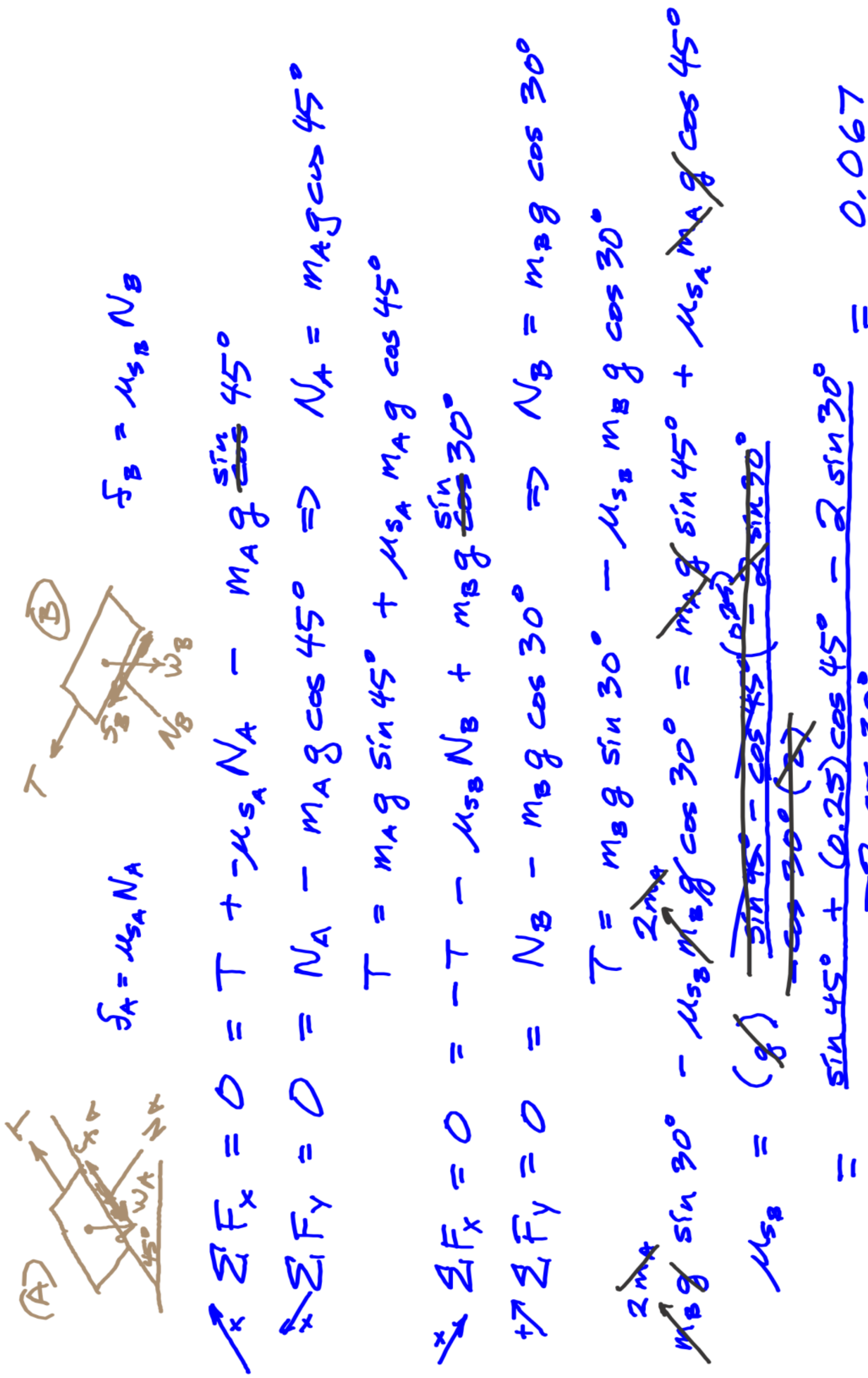


Figure 3. A solution to the statics problem from Figure 2.

In Year 1 students used Newton's Pen, an intelligent tutoring system for statics (Lee, Stahovich, & Calfee, 2011). This system was utilized during several discussion periods.

In Year 2 there were four separate discussion sections, each of which was provided with one of three different experimental treatments. Students from two discussion sections were asked to provide self-explanations for the problem-solving steps for six of the homework assignments. These students were provided with self-explanation prompts for these assignments. Students in a third discussion section used Newton's Pen during some of the discussion periods. The fourth discussion section served as the control. Students in this section did not provide self-explanation, nor did they use Newton's Pen.

In Year 3 students were randomly assigned to one of six experimental groups. Four of the groups were asked to provide self-explanation for the problem-solving steps on their homework. Each of these four groups was provided with varying amounts of scaffolding for self-explanation. Students in a fifth group used Newton's Pen during some discussion periods. Students in the sixth group served as the control. For the final homework assignment, all students were prompted to provide self-explanation without scaffolding. Also, in some discussion periods, students were given problems to solve. They began the problems in discussion, and if necessary completed them later. They submitted these solutions with their homework.

Course grade for all three cohorts was based on the following weighting: 10% for the homework score, 10% for the quiz score, 10% for the project score, 20% for the first midterm exam score, 20% for the second midterm exam score, and 30% for the final exam score. The exams and quizzes were not identical across cohorts but the content and format were similar. For example, for all 3 years the first midterm included one problem requiring students to compute a moment, one problem involving equilibrium analysis of a two-dimensional system, and one problem involving equilibrium of a three-dimensional system. All problems, except an ethics problem on the final exam, required free-form solutions, which typically required one or more free body diagrams and equilibrium equations. Problems were graded using a rubric that examined the correctness of the major elements of the solution. For example, an equilibrium problem might include a free body diagram, geometric calculations, and equilibrium equations. The credit for the problem would be divided over these elements according to their complexity, with more points being assigned to the more challenging elements. If an element was missing, the student would receive no credit for that element. Points were deduced from each element for various types of errors such as sign errors, missing terms (e.g., a missing force in a force equilibrium equation), incorrect terms (e.g., using "sine" instead of "cosine"), and so forth.

Procedure

Beginning in the third week of the course, students used smartpens to complete all homework assignments, quizzes, and exams. Students were instructed to use their smartpen instead of a pencil. We did not collect data from the first two homework assignments and quizzes. In Years 1 and 2, there were a total of nine homework assignments, and we collected data from the last seven. In Year 3, there was a total of eight assignments and we collected data from the last six. In all years we collected data from five quizzes (all quizzes except the first two), two midterm exams, and one comprehensive final exam. In Year 1, the seven homework assignments comprised a total of 41 problems, in Year 2 there were 44 problems, and in Year 3 there were 40. The instructor was aware of the general goal of the study—to capture student problem-solving data from the homework that could be related to course performance—but the data were not analyzed until after each cohort completed the course and received their final grades, thus eliminating the possibility of bias in assigning grades.

Livescribe smartpens create two records: ink on paper and timestamped digitized pen strokes. In Year 1, students submitted both the paper copy of each assignment and their smartpens. We extracted the data from the smartpens and returned them to the students so they could complete their next assignment. For Year 2, we developed software to enable students to submit their assignments electronically. To do this, a student docked the smartpen to a PC using a USB cable. Our software then extracted the ink data and submitted it to a server for grading. We graded the homework electronically and returned it as a PDF. In Year 2, electronic submission was optional. Students could still submit the paper copy of an assignment, in which case we extracted the ink data form the smartpen at the end of the quarter. To encourage students to submit their work electronically, for some assignments the due date for electronic submission was several hours later than for paper submission. In Year 3, all students were required to submit their work electronically. However, if a student had technical difficulties submitting a particular assignment, he or she could still submit it on paper and we extracted the ink data at the end of the quarter.

In Years 2 and 3, some students provided self-explanation with their homework. As self-explanation was not the focus of this project, and to maintain consistency across all students, we excluded the ink for the explanations from our analysis. However, we did include the self-explanation ink from the last assignment in Year 3, as all students provided self-explanation for that assignment. In Year 3, some homework submissions included problems that were solved in part during a discussion period. We excluded these problems from our analysis as they are not typical homework problems.

The Livescribe smartpens have two clocks. One is used to display the current time of day, while the other is used to create timestamps for the pen strokes. The former can be adjusted, while the latter cannot. Having a nonadjustable clock for time stamps ensures that the time of the pen strokes is correct, even when there is a change to or from daylight saving time, for example. We used the time of an exam to determine the offset between the timestamp clock and the actual time of day. For Year 3, we also directly measured the offset before distributing the pens to the students. With this calibration approach, the offset of the timestamp clock is accurate to within about 5 min, which is adequate for our purposes.

For all 3 years, we conducted a survey at the end of the course with questions about demographics, study habits, and perceptions about the course and instructional technology used.

Table 1

Thirteen Measures Derived Through Smartpen Technology

Measure	Description
Total homework time	Total time to complete all assignments
Due date ink fraction	Proportion of pen strokes written within 24 hr of due date
Late night ink fraction	Proportion of pen strokes written between midnight and 4 a.m.
Number of homework sessions	Number of sessions used to complete the assignments
Total strokes	Number of pen strokes written to complete the assignments
Equation strokes	Number of equation pen strokes written to complete the assignments
Diagram strokes	Number of diagram pen strokes written to complete the assignments
Cross-out strokes	Number of cross-out pen strokes written to complete the assignments
Total ink length	Total distance (in inches) the pen travels on paper for all assignments
Problems attempted	Number of problems for which the student wrote at least 50 pen strokes
Average time per problem	Total homework time divided by problems attempted
Average pen speed	Total ink length divided by total homework time
Out of order	Number of times a student transitions to a problem other than the next one in the assignment

Data Mining With Smartpen Technology

We developed software to enable us to manually partition students' ink data into the individual problems comprising each assignment, quiz, and exam. The software renders the ink data on a computer display, enabling one to navigate through the pages of writing. A mouse is used to select the ink for an individual problem and assign a problem number to it. We then use software (Lin, Stahovich, & Herold, 2012) to automatically label each pen stroke as either an equation, free body diagram, or cross-out stroke (see Figure 3).

Once the digital ink has been partitioned into problems and labeled, we computed 13 quantitative measures to characterize a student's homework activity, as summarized in Table 1. Our first measure, *total homework time*, is the total time spent to complete all of the homework assignments. We define the time to complete one assignment as the time from the first pen stroke of the assignment to the last, excluding any periods of inactivity longer than 10 min. Any long inactivity periods partition the homework effort into *sessions*. Consecutive pen strokes within a session are never more than 10 min apart, while strokes from different sessions are always at least 10 min apart.

We use three measures to characterize the time effort over the assignment period. *Due date ink fraction*, computed as the fraction of the pen strokes written within 24 hr of the due date, measures the extent to which students wait until the "last minute" to complete an assignment. Similarly, *late night ink fraction*, computed as the fraction of the pen strokes written between midnight and 4 a.m., measures the fraction of work done late at night. Finally, *number of homework sessions* is simply the total number of sessions required to complete the assignments, with a new session counted when there is at least a 10-min break from the previous pen stroke.

In addition to considering the amount of time spent on homework, we also consider the amount of writing. As the name suggests, *total strokes* is the total number of pen strokes written to complete the assignments. We also count the number of *equation strokes*, the number of *diagram strokes*, and the number of *crossout strokes*. These measures are computed using the auto-labeler from Lin, Stahovich, and Herold (2012). In addition to stroke count, we also consider the length of the pen strokes. *Total ink length*, which is computed in units of inches, is the total distance the pen tip travels on the paper.

We use three measures to characterize effort on individual homework problems. *Problems attempted* is the number of problems for which the student wrote at least 50 pen strokes. It is unlikely that a student made significant progress on a problem if he or she wrote fewer strokes than this. For example, simply writing “Problem 1” takes at least eight strokes. *Average time per problem* is the ratio of *total homework time and problems attempted*. This provides a means of comparing the effort of students even if they did not complete the same number of problems. *Average pen speed* is the ratio of *total ink length and total homework time*. This measure characterizes the pace of the work. Finally, the *out of order* measure describes the frequency with which a student works nonsequentially. Prior work has found that expert students often solve problems in the order assigned, while novice students may begin one problem and then move on to another before completing the former (Herold, Stahovich, Lin, & Calfee, 2011). The *out of order* measure is the number of times a student transitions to a problem other than the next one. For example, the sequence of problems 1, 3, 1, 2 has an *out of order* value of two. The transitions from 1 to 3 and 3 to 1 are nonsequential.

When computing these measures, we exclude any ink that was written more than 5 min prior to the time the homework assignment was posted. This tolerance compensates for the 5-min uncertainty in our timestamp clock calibration. We also exclude any ink written more than an hour after an assignment due date. As our electronic submission system did not prevent late submissions, some students did submit their homework late. We include any pen strokes written during this past-due hour in the *due date ink fraction*.

One of the questions on the end-of-class survey asked students to report the amount of time it took on average to complete a homework assignment, which we used to compute self-reported time on homework. For the first 2 years, the available choices for answering the question were: less than 2 hr, 2–4 hr, 4–6 hr, 6–8 hr, 8–10 hr, and more than 10 hr. In Year 3, the choices were reduced by one so that the last choice was “more than 8 hr.” When computing the total time spent on homework, we consider a student’s average assignment time to be the midpoint of the selected interval. However, if the student selected the largest choice, we use the lower bound (i.e., either 8 or 10 hr). For example, if a student in Year 1 reported “2–4 hr,” we would compute the total self-reported time over the seven homework assignments to be 21 hr. Similarly, if they reported “more than 10 hr” we would compute the value to be 70 hr.

Table 2

Correlation Between Course Grade and Each of 13 Smartpen Measures for all Students and Each Cohort Separately

Measure	All students	Cohort 1	Cohort 2	Cohort 3
Total homework time	.44*	.42*	.59*	.31*
Due date ink fraction	-.32*	-.38*	-.48*	-.20*
Late night ink fraction	-.06	-.08	-.15	-.04
Number of homework sessions	.33*	.05	.58*	.27*
Total strokes	.49*	.55*	.60*	.40*
Equation strokes	.49*	.54*	.61*	.40*
Diagram strokes	.41*	.46*	.51*	.34*
Cross-out strokes	.32*	.33*	.34*	.33*
Total ink length	.42*	.44*	.50*	.39*
Problems attempted	.45*	.35*	.68*	.27*
Average time per problem	.33*	.32*	.39*	.29*
Average pen speed	-.02	.02	-.11	.07
Out of order	.10	-.17	.27*	.09

*$p < .05$.

Results

Data Set

Our dataset includes data on 13 measures from a total of 328 students: 92 from Year 1, 109 from Year 2, and 127 from Year 3. All of these students completed the course and received a final course grade. We excluded data from one student in Year 2 and four from Year 3 because their digital ink data was corrupted.

As described in the Method section, some students in Years 2 and 3 were asked to write self-explanations and some others used an intelligent tutoring system. We wanted to determine whether the same pattern of results could be obtained in different contexts. We performed one-way analysis of variance (ANOVA) to determine if these treatments led to any significant differences in final grades between the experimental and control groups. In both cases, the differences were not significant (p = .706 for Year 2 and p = .957 for Year 3) and thus, in our analysis, we ignore these distinctions between students.

Table 2 shows the correlation between each of the 13 measures and course grade for all students, and for each cohort separately, with significant correlations at p < .050 denoted with an asterisk. We focus on the results for all students, and view the cohort data as a form of replication. Table 3 shows the means and standard deviations of each of the 13 measures for all students, and for each cohort separately.

Some of our measures are sensitive to the number of problems assigned. As the number of homework problems varied between the three cohorts, we performed another analysis in which we normalized the features by the number of problems assigned to the cohort. Four features—*due date ink fraction*, *late night ink fraction*, *average time per problem*, and *average pen speed*—did not require normalizing as they are insensitive to the number of problems. Normalizing the measures produced only a negligible change in the correlation with course grade. The correlations changed by less than .01 (and p by less than .003) for all measures.

We also investigated whether gender is significant to course performance. For Cohort 1 the average score for male students was .71 (n = 78), while the average score for female students was .65 (n = 12). However, this difference in means was nonsignificant, with $p = .212$. Similarly for Cohort 2, the average score for male students was .66 ($n = 87$), while the average score for female students was .63 ($n = 16$). This difference in means was again nonsignificant, with $p = .530$. For Cohort 3, the average score for male students was .68 ($n = 105$), while the average score for females was .61 ($n = 19$). This difference between means was significant, with $p = .028$.

Table 3
Means and Standard Deviations for Each of 13 Smartpen Measures for all Students and Each Cohort Separately

Measure	All students		Cohort 1		Cohort 2		Cohort 3	
	μ	σ	μ	σ	μ	σ	μ	σ
Total homework time (hr)	17.1	8.5	17.7	6.4	16.9	9.1	17	9.2
Due date ink fraction	.7	.3	.7	.2	.7	.2	.6	.3
Late night ink fraction	.1	.1	.2	.2	.1	.1	.1	.1
Number of homework sessions	36.0	19.6	38.8	16.7	32.0	18.1	37.3	22.3
Total strokes	20189.5	9186.6	18944	6880.2	21162.0	10002.3	20257.0	9855.3
Equation strokes	14858.8	6939.1	13975.2	5411.3	15411.0	7330.4	15025.1	7542.9
Diagram strokes	4925.9	2373.6	4572.3	1708.0	5318.8	2767.8	4844.9	2390.9
Cross-out strokes	404.7	278.7	396.5	207.6	432.2	359.9	387.0	241.8
Total ink length (inches)	5936.2	3030.1	5568.8	2308.0	5979.0	3252.0	6165.7	3280.8
Problems attempted	34.4	8.9	35.7	5.9	35.5	10.2	32.4	9.2
Average time per problem (min)	29.1	11.3	29.5	9.8	27.6	10.8	30.0	12.6
Average pen speed (inches/second)	.104	.045	.093	.037	.105	.038	.112	.053
Out of order	20.2	15.3	17.7	11.1	22.6	19.6	20.1	13.5

We also examined the correlation between measures of prior knowledge and course performance. Here we use two measures to quantify prior knowledge: the student's SAT score (based on combined verbal, quantitative, and writing scores) and their high school GPA. For Cohort 1 (r = .534, p < .001) and Cohort 3 (r = .284, p = .003), there was a significant correlation between SAT score and final course grade, but not for Cohort 2 (r = .091, p = .378). The correlation between high school GPA and final course grade was significant for Cohort 1 (r = .317, p = .003) and Cohort 2 (r = .285, p = .004) but not for Cohort 3 (r = .184, p = .052).

We performed a stepwise linear regression to examine the predictive ability of our entire set of measures. In computing a stepwise model we required the probability of $F \leq .05$ to enter a measure, and the probability of $F \geq .10$ to remove a measure. We initialized the model by including three measures: *total strokes*, *total homework time*, and *problems attempted*. For all students, *total strokes*, *problems attempted*, *out of order* and *due date ink fraction* were selected with r = .57, p < .001. For Cohort 1, *total strokes* and *out of order* were selected with r = .67, p < .001. For Cohort 2, *total strokes, problems attempted*, and *due date ink fraction* were selected with r = .72, p < .001. For Cohort 3, only *total strokes* was selected with r = .40, p < .001. Thus, in the analysis with the best statistical power (i.e., the combined data from all students), there is evidence that each of four smartpen measures (i.e., *total strokes, problems attempted, out of order,* and *due date ink fraction*) makes a unique contribution to predicting course grade.

As a follow-up we conducted another stepwise linear regression identical to the one described previously, but with SAT score entered as the first variable and the smartpen variables entered in order of their correlation. For all students, *SAT score*, *text strokes*, *problems attempted*, *out of order* and *due date ink fraction* were selected with $r = .63$, $p < .001$. For Cohort 1, *SAT score*, *total strokes*, *out of order* and *late night ink fraction* were selected with $r = .76$, $p < .001$. For Cohort 2, *SAT score, problems attempted*, and *due date ink fraction* were selected with $r = .71$, $p < .001$. For Cohort 3, *SAT score* and *text strokes* were selected with $r = .63$, $p < .001$. Thus, in the analysis with the best statistical power (i.e., the combined data from all students), there is evidence that each of four smartpen measures (i.e., *text strokes, problems attempted, out of order,* and *due date ink fraction*) contribute uniquely to predicting course grade, even when the effects of prior knowledge are controlled (i.e., smartpen variables predict course grade beyond the effects of SAT score). Overall, although construction of a factor-analyzed measurement instrument based on smartpen variables is beyond the scope of this study, there are indications that course grade is uniquely predicted by a collection of smartpen measures.

The final course grade includes homework score with a weight of 10%. To examine if this artificially increased the correlations between our measures of homework activity and course grade, we recomputed the course grade excluding homework score and recomputed the correlations. This resulted in only a negligible change in the correlations, and no change in the factors chosen in the regression analyses. More specifically, for Cohorts 1 and 2, the changes in correlations were no greater than .04. For Cohort 3, the changes were no greater than .07.

How Much: Is Homework Time Related to Course Grade?

According to the basic version of the time-on-task hypothesis, students who spend more time working on their homework should get better grades in the course. The first line of Table 2 shows the correlation between total time spent on the homework problems and course grade for all students combined, and for each cohort separately. As the table illustrates, there is a significant correlation for each cohort and for all students combined, consistent with predictions. Overall, there is strong and consistent support for the time-on-task hypothesis, based on data collected through smartpen technology.

What happens when we look at students' self-reported time on homework per week as reported on a postquestionnaire? In contrast to the significant correlation between course grade and the actual time on homework recorded through smartpens, the correlation between course grade and self-reported time on homework is not positively significant for all students combined (r = -.16) nor for each of the three cohorts (r = -.29 for Cohort 1, r = -.14 for Cohort 2, and r = -.13 for Cohort 3). Instead, the correlation is negative for all three cohorts, and the negative correlation for Cohort 1 is statistically significant.

Furthermore, is the students' self-reported time consistent with the actual measured time to complete homework assignments? For two of the three cohorts, there was only a weak correlation between self-reported time and *total homework time*: For Cohort 1, $r = .21$, $p = .052$; for Cohort 2, $r = .16$, $p = .139$; and for Cohort 3, $r = .35$, $p < .001$. Additionally, nearly all students overreported their homework time. For Cohort 1, 88.5% of students overreported homework time with an average overestimation of 19.0 hr. For Cohort 2, 85.5% of students overreported homework time with an average overestimation of 23.5 hr. For Cohort 3, 85.5% of students overreported homework time with an average overestimation of 13.4 hr.

This pattern of differences between actual time and self-reported time points to the value of technology-supported measures of homework activity in testing the time-on-task hypothesis. This set of contrasting findings constitutes a major contribution of this study.

When: Is the Timeliness of Homework Activity Related to Course Grade?

According to the updated version of the time-on-task hypothesis, which considers the quality of the time spent on homework, students who commonly wait until the last minute to do homework (i.e., within 24 hr of the due date) or who commonly do homework late at night (i.e., midnight to 4 a.m.) should get worse grades in the course. Consistent with this prediction, the second line of Table 2 shows a significant negative correlation between *due date ink fraction* and course grade for all students, and for each cohort. In contrast, the third line of Table 2 does not show a significant correlation between *late night fraction* and course grade for any of the cohorts, suggesting perhaps that working late at night is not necessarily an indication of lower quality time. Overall, a major empirical contribution is strong and consistent evidence that the quality of how homework time is spent (as measured by the proportion of homework time done within 24 hr of the deadline) is related to course grade. The smartpen technology allows us to address this prediction of an updated version of the time-on-task hypothesis.

Similarly, breaking an assignment up into multiple sessions may be a way to enable distributed practice—spreading practice over multiple sessions—which has been shown to improve learning (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Accordingly, time-on-task should be most efficient when it is spread over multiple sessions. Consistent with this prediction, the fourth line of Table 2 shows that *number of homework sessions* correlates significantly with course grade for all students combined and for two of the three cohorts. Again, smartpen technology allows us to address a prediction about the time-course of homework activity using data that is not otherwise available.

How Many: Is the Amount of Writing Activity Related to Course Grade?

According to the updated version of the time-on-task hypothesis, which considers the amount of productive activity, students who create more pen strokes while working on homework assignments should get better grades in the course. Consistent with this prediction, the fifth line of Table 2 shows a significant correlation between *total strokes* and course grade for all students, and for each cohort. Also consistent with predictions, the next lines in Table 2 show the same pattern of significant correlations (for all cohorts) between grades and *equation strokes, diagram strokes, cross-out strokes*, and *total ink length*, respectively. Overall, there is consistent evidence that higher achievement is related to the level of effort exerted by students as indicated by their pen strokes. Similarly, Table 2 shows that for all students and for each cohort separately, there is a significant correlation between the *number of problems attempted* and course grade (line 10) and between *average time per problem* and course grade (line 11).

What is not related? Two variables did not correlate consistently with course grade—*average pen speed* and *out of order*—perhaps because they are not appropriate measures of the amount of productive activity. Writing faster or slower does not necessarily indicate more or less effort, and trying problems out of order can be attributed to several causes other than effort, including lack of concentration.

Discussion

Empirical Implications

Concerning issues about time-on-task, actual time spent working on homework problems was positively correlated with course grades, but self-reported time spent on homework problems was not. Concerning the time course of homework activity, the amount of homework time spent within 24 hr of the deadline was negatively correlated with course grades, the amount of homework time spent between midnight and 4 a.m. was not correlated with course grades, and breaking homework time into more sessions was positively correlated with course grades. Concerning actual behavior and effort on homework problems, course grades were positively correlated with the total number of pen strokes, equation strokes, diagram strokes, cross-out strokes, total ink produced, total problems attempted, and time per problem. Course grades were not consistently correlated with average pen speed or solving problems out of order.

Theoretical Implications

This study investigates a crucial link in a model of academic learning, the link between engagement or effort to learn, as measured by the amount and time students allocate to a learning task, and performance, as measured by learning outcome in a college course. In particular, the present study examines the idea that the amount of time that students spend in productive learning is related to academic achievement in a course. Although no causal conclusions can be drawn, the work draws attention to a potential causal mechanism leading to learning—namely, amount of productive learning activity. Importantly, both the quantity of time spent on homework and the quality of how homework time is used are related to achievement. Higher quality use of time is reflected in doing homework long before it is due and breaking assignments into smaller sessions. Effortful activity on homework is reflected in the number of pen strokes, the total ink produced, and the number of problems attempted. A major contribution of this project is to enable more detailed measures of student effort or engagement— which is proposed to be the mechanism underlying academic learning.

Practical Implications

This is a correlational study that examines actual performance in a real college course, so no causal conclusions can be drawn. However, this study offers preliminary evidence for the potential of homework as an aid to student achievement, particularly when students work on their homework in a timely and effortful way. The role of productive time on task has long been recognized as a critical issue in intelligent tutoring systems (Anderson, Corbett, Koedinger, & Pelletier, 1995).

Methodological Implications

This study highlights the potential of educational data mining techniques in general (i.e., techniques for measuring and summarizing learner activity during learning), and smartpen technology in particular, for educational research. Smartpen technology allows for assessing student study activity at a level of detail that was not previously possible, and thereby offers a new and powerful methodology for testing implications of educational theories.

Our results confirm what other researchers have proposed (Blumner & Richards, 1997; Schuman et al., 1985): Students' self-reports of study effort are often unreliable. Finally, this study points to the useful role of replication in educational research (as noted by Shavelson & Towne, 2002) by showing the same pattern of results across three independent cohorts of students.

Limitations and Future Directions

This work is a first step at building techniques that can provide automated assessment of performance from an analysis of handwritten homework. Our present analysis examines the relationship between the amount of effort on homework and performance. In future work, we plan to examine how patterns of homework activity contribute to success. For example, our current analysis suggests that doing homework just before it is due may not be a successful strategy, but causal claims cannot be drawn based on the correlational relationship identified in this study. Thus experimental work is needed to test causal claims. Experimental research should be designed to explicitly test hypotheses suggested by this study concerning the possible causal role of productive time on task by directly manipulating this factor and examining the effects on learning outcome. Within the context of experimental research, future work is also needed to examine potential moderating variables such as the learner's prior knowledge and metacognitive skills.

The correlations involving self-reported homework time and actual homework time should be interpreted in light of the fact that our self-report measure of homework time involved asking the learner to check a category that was converted to a number for analyses (such as “two to four hours per week” being recorded as 3, “four to six hours per week” being recorded as 5, etc.), whereas our smartpen measure of homework time is based on a continuous scale.

Another concern is whether the act of being asked to use smartpens, and the ensuing awareness of being observed, could cause students to be more careful about how they deal with homework problems than they would otherwise, could make them want to use scratchpaper before solving homework problems with a smartpen, and could create discomfort or distraction that affect homework behavior. In short, it is important to ensure that students do their homework in their usual way but with use of their smartpens and nothing else. In the present study, students were instructed to show all their work using their smartpen, and a postexperimental questionnaire indicated reasonable compliance. On a survey from Cohort 3 asking students to rate smartpen use on a scale of 1 (“doing all homework elsewhere”) and 7 (“using the pen to do everything”) the mean rating was 5.1 (SD = 1.7). Future work should involve more evidence concerning fidelity, such as poststudy interviews. Similarly, the total time measure was not based on any activity before the first pen stroke so it would not include time to initially read and think about the problem before starting to answer. Another thorny issue concerns whether course grade is an adequate measure of learning outcome. In the present study, course grade was based on tests that involved concepts related to the homework problems, but no detailed method of alignment was implemented.

The present work is based on the idea that a deeper analysis of the sequencing of homework activities can provide additional insights about successful and unsuccessful study strategies. Identifying such strategies will lead to experimental studies that ultimately enable automated coaching systems to examine students' study habits and recommend interventions aimed at increasing academic success. Additionally, we plan to extend the smartpen technology to the study of note-taking during classroom lectures in order to identify classroom learning strategies that are related to course grade.

This work is a step in applying educational data mining techniques to learning activities in traditional, rather than online, environments. Our current studies have focused on one course (i.e., statics), and more work is needed to determine how our techniques will generalize to other domains for which homework assignments comprise handwritten problem solving. We anticipate that our techniques will be applicable to assessing homework habits in a variety of math, science, and engineering subjects.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4,* 167–207. doi:10.1207/s15327809jls0402_2

Anthony, L., Yang, J., & Koedinger, K. R. (2008). Toward next-generation intelligent tutors: Adding natural handwriting input. *IEEE MultiMedia, 15,* 64–68. doi:10.1109/MMUL.2008.73

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60,* 383–398. doi:10.1007/s11423-012-9235-8

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist, 23,* 225–239. doi:10.1037/h0020791

Beal, C. R., & Cohen, P. R. (2008). Temporal data mining for educational applications. *Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence, 5351,* 66–77.

Berliner, D. (1984). The half-full glass: A review of research on teaching. In P. L. Hosford (Ed.), *Using what we know about teaching* (pp. 51–84). Alexandria, VA: Association for Supervision and Curriculum Development.

Berliner, D. (1991). What's all this fuss about instructional time? In M. Ben-Peretz & R. Bromme (Eds.), *The nature of time in schools*. New York, NY: Teachers College Press.

Blanc, P. (1999). Solving a non-routine problem: What helps, what hinders. *Proceedings of the British Society for Research into Learning Mathematics, 19,* 1–6.

Blumner, H. N., & Richards, H. C. (1997). Study habits and academic achievement of engineering students. *The Journal of Engineering Education, 86,* 125–132. doi:10.1002/j.2168-9830.1997.tb00275.x

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723–733.

Clark, R. C., & Mayer, R. E. (2011). *e-Learning and the science of instruction* (3rd ed.). San Francisco, CA: Pfeiffer.

Cooper, H. (1989). *Homework*. White Plains, NY: Longman. doi:10.1037/11578-000

Cooper, H. (2001). *The battle over homework* (2nd ed.). Thousand Oaks, CA: Corwin.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research 1987–2003. *Review of Educational Research, 76,* 1–62. doi:10.3102/00346543076001001

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3,* 425–453. doi:10.1111/j.1745-6924.2008.00089.x

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. doi:10.1177/1529100612453266

Ebbinghaus, H. (1964). *Memory*. New York, NY: Dover. (Original work published 1885)

Graham, S., & Williams, C. (2009). An attributional approach to motivation in school. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 11–34). New York, NY: Routledge.

Hall, R. (2000). Video recording as theory. In D. Lesh & A. Kelley (Eds.), *Handbook of research design in mathematics and science education* (pp. 647–664). Mahwah, NJ: Lawrence Erlbaum.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 metaanalyses relating to achievement*. New York, NY: Routledge.

Herold, J., & Stahovich, T. (2012). Characterizing students handwritten self-explanations. *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

Herold, J., Stahovich, T., Lin, H.-L., & Calfee, R. C. (2011, June). The effectiveness of "pencasts" as an instructional medium. *Proceedings of the American Society for Engineering Education 118th Annual Conference and Exposition*, Vancouver, British Columbia, Canada.

Herold, J., Stahovich, T. F., & Rawson, K. (2013). Using educational data mining to identify correlations between homework effort and performance. *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Jones, L., & Ruch, G. M. (1928). The twenty-seventh yearbook of the national society for the study of education: Nature and nurture, part ii-their influence upon achievement. In G. M. Whipple (Ed.), *Achievement as affected by amount of time spent in study* (pp. 131–134). Bloomington, IL: Public School Publishing Co.

Karweit, N. (1984). Timed-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership, 41,* 32–35.

Keith, T. Z. (1982). Time spent on homework and high school grades: A large-scale path analysis. *Journal of Educational Psychology, 74,* 248–253. doi: 10.1037/0022-0663.74.2.248

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science, 6,* 333–353. doi:10.1002/wcs.1350

Krüger, A., Merceron, A., & Wolf, B. (2010). A data model to ease analysis and mining of educational data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 131–140). Pittsburgh, PA.

Lee, C.-K., Stahovich, T., & Calfee, R. C. (2011). A pen-based statics tutoring system. *Proceedings of the American Society for Engineering Education 118th Annual Conference and Exposition*, Vancouver, British Columbia, Canada.

Li, N., Cohen, W. W., Koedinger, K. R., & Matsuda, N. (2011). A machine learning approach for automatic student model discovery. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 31–40). Eindhoven, the Netherlands.

Lin, H., Stahovich, T. F., & Herold, J. (2012). Automatic handwritten statics solution classification and its applications in predicting student performance. *Proceedings of the 119th American Society for Engineering Education Annual Conference and Exposition*, San Antonio, TX.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education, 54,* 588–599. doi:10.1016/j.compedu.2009.09.008

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511811678

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139547369.005

Mostow, J., González-Brenes, J. P., & Tan, B. H. L. (2011). Learning classifiers from a relational database of tutor logs. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 149–158). Eindhoven, Netherlands.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25,* 1159 –1168. doi:10.1177/0956797614524581

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology - UIST ’06* (pp. 191–200). New York, NY: ACM Press.

Rawson, K., & Stahovich, T. F. (2013, June). Predicting course performance from homework habits. *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Romero, C., Romero, J., Luna, J., & Ventura, S. (2010, June 11-13). Mining rare association rules from e-learning data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 171–180). Pittsburgh, PA.

Romero, C., Ventura, S., Vasilyeva, E., & Pechenizkiy, M. (2010). Class association rules mining from students’ test data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 131–140). Pittsburgh, PA.

Schuman, H., Walsh, E., Olson, C., & Etheridge, B. (1985). Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces, 63,* 945–966. doi:10.1093/sf/63.4.945

Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35–54). New York, NY: Routledge.

Shanabrook, D. H., Cooper, D. G., Woolf, B. P., & Arroyo, I. (2010). Identifying high-level student behavior using sequence-based motif discovery. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 191–200). Pittsburgh, PA.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Steif, P. S., & Dantzler, J. A. (2005). A Statics Concept Inventory: Development and psychometric analysis. *The Journal of Engineering Education, 94,* 363–371. doi:10.1002/j.2168-9830.2005.tb00864.x

Steif, P. S., & Dollár, A. (2009). Study of usage patterns and learning gains in a web-based interactive static course. *The Journal of Engineering Education, 98,* 321–333. doi:10.1002/j.2168-9830.2009.tb01030.x

Steif, P. S., Lobue, J. M., Kara, L. B., & Fay, A. L. (2010). Improving problem solving performance by inducing talk about salient problem features. *The Journal of Engineering Education, 99,* 135–142. doi:10.1002/j.2168-9830.2010.tb01050.x

Stevens, R., Johnson, D. F., & Soller, A. (2005). Probabilities and predictions: Modeling the development of scientific problem-solving skills. *Cell Biology Education, 4,* 42–57. doi:10.1187/cbe.04-03-0036

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. doi:10.1007/978-1-4419-8126-4

Trivedi, S., Pardos, Z. A., Sárközy, G. N., & Heffernan, N. T. (2011). Spectral clustering in educational data mining. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 129–138). Eindhoven, Netherlands.

Van Arsdale, T., & Stahovich, T. (2012, June). Does neatness count? What the organization of student work says about understanding. *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

van Gog, T. (2013). Time on task. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 432–433). New York, NY: Routledge.

van Gog, T., Ericsson, K. A., Rikers, R. M. J. P., & Paas, F. (2005). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educational Technology Research and Development, 53,* 73–81. doi:10.1007/BF02504799

Xu, J. (2013). Homework and academic achievement. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 199–201). New York, NY: Routledge.

Chapter 3

Homework Habits that Predict Academic Success

Abstract

Can homework behaviors during the first few weeks of a college course help identify students who are at risk of earning poor grades in a course? According to the early-warning hypothesis, measures of conscientious habits for solving homework problems during the first few weeks of a course predict course grade. Conscientious homework habits include working on time (as measured by the percentage of homework done at least 24 hours before the due date), working on task (as measured by the percentage of assigned problems that were attempted), and working with high effort (as measured by number of pen strokes and length of ink written). We examined these measures on weekly homework assignments for 659 college students from seven offerings of two introductory-level engineering courses taught by the same instructor over an eight-year period. The four homework habit measures were computed using smartpen technology that recorded timestamped pen strokes on homework assignments during 10-week quarters. Across the seven course offerings, by the third week of the course, the four measures of conscientious homework habits correlated significantly and positively with course grade. Results support the early warning hypothesis and point to the relation between conscientious work habits exhibited early in a course and academic success.

Objective and Rationale

Imagine an introductory-level college course in which an instructor could examine a student's studying behavior over the first few weeks of the term and from this determine the degree to which the student was on a trajectory for success or failure in the course. More specifically, suppose that an instructor could predict a student's final course grade based on the student's behavior in completing homework during the first few weeks of a course. Such "early warning" predictions could alert the student and instructor when interventions are needed.

The present study seeks to enable this vision of an early warning system. To address this goal, the present study examines the early warning hypothesis: Objective measures of homework habits early in an introductory college engineering course are predictive of eventual achievement in the course. Specifically, this study examines the correlation between measures of homework behavior during each week and course grade across seven college engineering courses.

Homework problems are a common element of college courses in science, technology, engineering, and mathematics (STEM). Recent work has shown that objective measures of students' total homework effort in an introductory-level engineering course are positively and significantly correlated with achievement in the course (Rawson, Stahovich, & Mayer, 2017). The present study employs Rawson et al.'s (2017) technology for recording digital records of students' behavior during homework problem solving. The digital record was obtained using smartpens that captured student's writing as timestamped, digitized pen strokes. The goal of the present study is to determine if objective measures of effort on homework completed early in an introductory-level engineering course are predictive of the eventual achievement in the course. If so, this could enable an "early warning" to alert both the student and the instructor when interventions are needed for a student to be successful in a course.

Table 4
Three Core Features of Conscientiousness

Feature	At work	In homework
On time	Showing up for work on time	Percentage of homework done 24 hours before due date (*Early Work Fraction*)
On task	Doing what you are asked to do	Percentage of assigned problems that were attempted (*Problems Attempted*)
High effort	Giving each task a high level of effort	Number of pen strokes (*Stroke Count*) and length of ink written *(Length of Ink Written)*

Theoretical Framework

Conscientiousness. Conscientiousness, which involves commitment to work diligently on assigned tasks, has been proposed as a 21st century skill necessary for success in school, work, and life (Pellegrino & Hilton, 2012). Although cognitive and social skills such as creativity, communication, and collaboration often get more attention as 21st century skills, there is encouraging evidence that personal skills such as conscientiousness may also be important predictors of academic and job success. Conscientiousness is a personal competency that sometimes is called work ethic, grit, responsibility, initiative, perseverance, productivity, citizenship, or career orientation (Pellegrino & Hilton, 2012). In short, there may be personal aspects of academic success that go beyond the competencies of cognitive ability and content knowledge.

In this study, we examine what we call *homework habits* — behaviors that students exhibit in completing their homework assignments that reflect core features of conscientiousness. Table 4 lists three core features of conscientiousness and the homework metrics we use to measure them: (1) showing up to work on time which is measured by the percentage of homework completed more than 24 hours before the due date (i.e., on time feature), (2) doing what is asked of you which is measured by the proportion of problems attempted (i.e., on task feature), and (3) giving your work a high level of effort which is measured by the total number of pen strokes and total amount of ink written (i.e., high-effort feature).

Homework. An increasing body of evidence illustrates the educational value of homework (Keith, 1982; Hattie, 2009; Cooper, Robinson, & Patall, 2006; Xu, 2013; Rawson et al., 2017). Hattie (2009) performed a meta-analysis involving 295 experimental tests and over 100,000 students, and found an average effect size of d = .29 favoring homework. Cooper et al. (2006) found a weighted average correlation of r = .24 based on 69 separate correlations. In addition, they found stronger achievement correlations amongst older students (e.g. Grades 7-12) versus younger students (e.g. Grades K-6), and when parents reported time on homework. Xu (2013) proposed studying the quality of homework, including the learner's effort and activity, as much of the existing research is dependent upon students' self-reported time (or effort) on homework activities. Work by Rawson et al. (2017) illustrated the weakness of using self-reported measures for quantitative analysis. They observed a negative correlation between self-reported time spent on homework and achievement, but saw a positive correlation between objective measures of homework time, as measured with smartpens, and achievement. Additionally, they found no significant correlation between subjective self-reports of time spent on homework and the objective recorded time from smartpens for two of the three cohorts examined.

Conscientiousness in completing homework assignments may improve academic learning through two classic study principles: the practice principle and the self-testing principle. First, practice has been shown to be an effective study strategy since the early days of both psychology and education, as reflected in the foundational work of Ebbinghaus (1885/1964) and Thorndike (1910). Research continues to show that time on task is an important determinant of learning (van Gog, 2013). Thus, spending time on solving homework problems effectively increases time on task.

Second, the self-testing principle (or testing effect) is a phenomenon in which completing practice questions about previously learned material results in improved learning compared to engaging in other study strategies, such as restudy (Brown, Roediger, & McDaniel, 2014; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Karpicke & Grimaldi, 2012; Miyatsu, Nguyen, & McDaniel, 2018; Roediger & Karpicke, 2006). Solving homework problems causes learners to practice on the kind of task that they will be asked to perform on course exams and quizzes (Herold, Stahovich, & Rawson, 2013; Fiorella & Mayer, 2015, 2016; Gyllen, Stahovich, & Mayer, 2018). Thus, solving homework problems can be seen as a form of self-testing.

Literature Review

Research on predicting academic success generally does not involve course-specific measures that can serve as an early warning system. For example, global measures of student aptitude, behaviors, and habits have long been used for prediction of student achievement. High school grade-point average (GPA) and scores on college entrance exams such as the ACT and SAT are frequently used as predictors of overall student performance in college (Bridgeman, McCamley-Jenkins & Ervin, 2000; Crooks, 1980; Geiser & Santelices, 2007; Hills, Klock, & Bush, 1965; Passons, 1967; Sawyer & Maxey, 1979; Wilson & Shrock 2001; Wollman & Lawrenz, 1984).

Concerning student study habits, Robyak and Downey (1979) analyzed students one year after they took a study skills course and observed significant increases in academic performance. Blumner and Richards (1997) examined the study habits of engineering students by measuring distractibility, inquisitiveness, and compulsiveness, and determined that more successful students tended to be less distractible and more inquisitive. Dunlosky et al. (2013) performed a detailed literature review of 10 learning techniques, and found distributed practice and practice testing having the highest relative utility across learning conditions, student characteristics, course material, and criterion tasks. Zollanvari, Kizilirmak, Kho, and Hernandez-Torrano (2017) used measures of self-regulatory learning behaviors to predict low and high GPA students, with their model able to predict with greater accuracy than using prior student performance measures.

Personality traits have also been explored as predictors of student performance. For example, Lounsbury, Sunstrom, Loveland, and Gibson (2002) examined work drive, general intelligence (measured via the Otis-Lennon test), and the “Big Five” personality traits, and found that they could explain up to 26.8% of the variance in a student’s course grade, but the Big Five personality traits did not add significantly to the prediction of course grade above and beyond work drive.

Researchers also have used pretests of skills and knowledge related to a course as a means to predicting student outcomes (Crooks, 1980; Wollman & Lawrenz, 1984; Wilson & Shrock, 2001). For example, Crooks (1980) used a pretest of physics and mathematics ability to predict course grade in an introductory physics class, and found that the pretests were better predictors of grade than more general indicators such as GPA or SAT score.

Educational data mining with computer-based instructional systems has a rich history dating back to large-scale studies of computer assisted instruction (CAI) in schools in the 1960s (Atkinson, 1968). Log files have been used extensively for modeling student learning from computer-based cognitive tutors (Anderson, 1993) and intelligent tutoring systems (Koedinger, D’Mello, McLaughlin, Pardos, & Rosé, 2015). In recent years, researchers have made significant progress in educational data mining or EDM (Koedinger et al., 2015; Romero, Romero, Luna, & Ventura, 2010). Much of the data used in this work is extracted from log files of intelligent tutoring systems (Beal & Cohen, 2008; Li, Cohen, Koedinger, & Matsuda, 2011; Mostow, González-Brenes, & Tan, 2011; Shanabrook, Cooper, Woolf, & Arroyo, 2010; Stevens, Johnson, & Soller, 2005; Trivedi, Pardos, Sráközy, & Heffernan, 2011) and learning management systems such as Moodle and Blackboard (Minaei-Bidgoli, Kashy, Kortenmeyer, & Punch, 2003; Krüger, Merceron, & Wolf, 2010; Picciano, 2012; Romero, Ventura, Vasilyeva, & Pechenizkiy, 2010). This work relies on a variety of data mining techniques including clustering (Antonenko, Toy, & Niederhauser, 2012; Stevens et al., 2005; Trivedi et al., 2011), model prediction (Li et al., 2011; Mostow et al., 2011; Stevens et al., 2005), and sequence analysis (Beal & Cohen, 2008; Krüger et al., 2010; Romero, Romero, et al., 2010; Shanabrook et al., 2010).

Our work differs from this line of data-mining research as we extract data from the students' learning activities written on paper, rather than activities involving typing on a computer keyboard. Prior research by Oviatt, Arthur, and Cohen (2006) shows that natural work environments are critical to student performance. In their examinations of computer interfaces for completing geometry problems, they found that "as the interfaces departed more from familiar work practice…, students would experience greater cognitive load such that performance would deteriorate in speed, attentional focus, metacognitive control, correctness of problem solutions, and memory." Similarly, Anthony, Yang, and Koedinger (2008) found that handwriting interfaces were more beneficial than keyboard interfaces for math tutoring systems. Mueller and Oppenheimer (2014) made a similar finding in relation to note taking. They examined student note taking using both longhand and laptops, and found that the latter can lead to shallower processing. Lectures were shown on a screen, with students taking notes, followed by distractor tasks. Using a model including both word count and verbatim overlap (three-word chunks from student notes matching the lecture transcript), they were able to predict performance on a test of the lecture material with a correlation coefficient of $r = .41$.

Macfadyen and Dawson (2010) mined data from a learning management system (LMS) to predict final course grade. Their best model was able to explain 33% of the variance in grade utilizing three features: the number of mail messages sent, the number of assessments finished, and the total number of discussion messages posted. This provides some insights about the relationship between studying and course performance. However, the type of data available from a LMS—such as records of downloading course materials and submitting electronic assignments—does not provide a direct measurement of students' homework activity. In the present study, we use smartpens to capture a fine-grained record of students' handwritten homework.

Researchers have used video recording to analyze students' problem-solving activities (Blanc, 1999; Hall, 2000). While this approach provides a detailed record of student work, the analysis is time-consuming. For example, Blanc (1999) made 75 recordings of students solving mathematics problems, but analyzed only two of the recordings. This sort of video analysis would be intractable in our studies, which involve hundreds of students completing homework throughout a quarter-long course. For our studies, smartpens provide a convenient and scalable approach for capturing high-resolution, timestamped records of problem-solving work.

Researchers have only recently begun using smartpens for assessment. For example, Herold and Stahovich (2012) used smartpens to examine the homework of students who were asked to provide self-explanations for their solutions to statics problems. The study found that students who generated self-explanations were more likely to complete homework problems in the order assigned (i.e., complete one problem before beginning the next) than were students who did not generate self-explanations.

Our work builds on that of Rawson and Stahovich (2013) who used smartpens as part of a technique for making early predictions of student success or failure in a statics course. They used the smartpens to record students' work on one homework assignment and a corresponding quiz given early in the course. They constructed a set of features that included manually generated scores on a homework assignment and a quiz, self-reported measures of how much written work was produced directly versus being copied from scratch work, and features extracted from the homework and quiz digital ink data (e.g. the total time spent on the homework and the amount of ink written). By themselves, the ink features (other than the amount of ink written) were only weakly predictive of a student's course performance. However, when combined with a concept inventory score (Steif & Dantzler, 2005), meaningful correlations with grade were observed. The features we use for characterizing homework habits, with the exception of *Length of Ink Written*, differ from those used by Rawson and Stahovich (2013). Furthermore, none of our features require manual grading of student work. While they examined homework habits from a single assignment, we examine homework habits over the successive assignments. Additionally, they considered only a single course offering, while we consider two different courses and a total of seven course offerings.

Van Arsdale and Stahovich (2012) demonstrated that the spatial and temporal organization of a student's solution to an engineering problem is indicative of the correctness of that solution. They recorded students' work on exam problems using smartpens and characterized the problem-solving activity in terms of the sequence of problem-solving steps and the arrangement of the work on the page. While they focused on a microscale analysis of problem-solving behavior on individual exam problems (taking minutes), we consider a macroscale analysis of homework habits (over days and weeks).

Herold et al. (2013) used smartpens to examine the correlation between effort on a homework assignment and grade on that particular assignment. They measured effort on a per-problem basis and its relationship with performance on subsequent, related homework, quiz, and exam problems. They characterized effort by the amount of time the pen was in contact with the paper, which is only a small fraction of the time spent on a problem. They found that this “writing time” was correlated with performance on subsequent problems. Our work is similar in that we also examine the relationship between homework effort and success. However, we consider a longer time scale, and overall success in the course, rather than success on individual assignments or subsequent individual problems.

Predictions

According to the early warning hypothesis, we make the following prediction: Measures of conscientiousness in completing homework (*Early Work Fraction*, *Problems Attempted*, *Stroke Count*, and *Length of Ink Written*) during the first few weeks of a course will predict eventual achievement in the course (course grade) just as well as when those measures are applied to homework habits over the entire course. In particular, according to the early warning hypothesis, by the third week of the course, course grade should correlate positively and significantly with *Early Work Fraction* (prediction 1), *Problems Attempted* (prediction 2), *Stroke Count* (prediction 3), *Length of Ink Written* (prediction 4), and a composite based on all four measures (prediction 5). Finally, we expect that data from homework habits from later weeks in the quarter will not results in predictions that are substantially better than those based on data from the first few weeks (prediction 6).

Method

Subjects and Design

The participants were 659 students enrolled in seven offerings of two lower-division courses taught by the same instructor over an eight-year period at a university in southern California: two offerings of Introduction to Mechanical Engineering and five offerings of Statics. There were a total of 170 students from Introduction to Mechanical Engineering (Intro1: 60, Intro2: 110) and 489 from Statics (Statics1: 92, Statics2: 109, Statics3: 127, Statics4: 65, Statics5: 96). These numbers exclude six participants who were excluded from our analysis because of hardware (smartpen) malfunctions. Most students were Mechanical Engineering majors (Intro1: 73.7%, Intro2: 87.1%, Statics1: 84.4%, Statics2: 52.4%, Statics3: 62.1%, Statics4: 50.8%, Statics5: 68.8%) or other engineering majors (Intro1: 7.0%, Intro2: 3.0%, Statics1: 3.3%, Statics2: 26.2%, Statics3: 29.8%, Statics4: 35.6%, Statics5: 23.8%). Most students were male (Intro1: 80.7%, Intro2: 83.2%, Statics1: 86.7%, Statics2: 84.4%, Statics3: 84.7%, Statics4: 78.0%, Statics5: 80.0%).

Course Settings and Procedure

The two offerings of Introduction to Mechanical Engineering were scheduled in the spring quarter of 2015 (Intro1) and the winter quarter of 2016 (Intro2). The Statics offerings were scheduled in the winter quarter of 2010 (Statics1), the winter quarter of 2011 (Statics2), the winter quarter of 2012 (Statics3), the spring quarter of 2016

(Statics4), and the winter quarter of 2017 (Statics5). Introduction to Mechanical Engineering provides students with an overview of the topics that will be studied in more depth in the subsequent Mechanical Engineering curriculum. Topics include forces in structures and fluids, materials and stresses, and thermal and energy systems. Statics is a part of engineering mechanics and is focused on the equilibrium of objects subject to forces. Topics include force systems, equilibrium of two- and three-dimensional systems, and equilibrium of frames and machines.

All course offerings employed a traditional lecture format, and all but one was scheduled on a Tuesday-Thursday schedule with two 80-min lectures per week. The one exception was an offering of Introduction to Mechanical Engineering (Intro1) that was scheduled on a Monday-Wednesday-Friday schedule with three 50-min lectures per week. All courses included a weekly 50-min discussion section.

All courses had weekly homework assignments in which students solved problems. For Introduction to Mechanical Engineering, typical homework problems required a student to apply engineering formulas to compute properties of engineered systems. For example, a student might be asked to compute the stress in a bolt or the flow rate of oil in a pipe. For Statics, problems typically involved drawing free body diagrams and constructing equilibrium equations to determine the magnitudes of forces. Students in both courses used smartpens to complete the homework assignments, as well as quizzes and exams. In some course offerings, students also used smartpens to take lecture notes (Stahovich & Pilegard, 2018). Smartpens are ink pens that digitize the writing as it occurs. These devices are used with a special dot-patterned paper and create two records, ink on paper and timestamped digitized pen strokes. We use the digitized ink data to analyze student homework activities.

Both courses are required for the Mechanical Engineering major, but other majors do take them. Mechanical Engineering students who follow the recommended course plan take Introduction to Mechanical Engineering in the winter quarter of their freshman year, and Statics in the winter quarter of the sophomore year. Both courses are also offered in the spring quarter, but that term is considered off-track.

The courses were all taught by the same instructor and the course content remained mostly uniform from one offering to the next. However, there were some

Table 5

Summary of Added Instructional Treatments in Each Course Offering

	Intro1	Intro2	Statics1	Statics2	Statics3	Statics4	Statics5
DocViewer	X	X				X	X
Lecture Notes	X	X				X	X
Design Project			X	X	X		
Preparatory Assignments		X					
Weekly Reports						X	
Newton's Pen 2			X	X	X		
Self-Explanation				X	X		
Discussion Problems					X		

variations between courses as summarized in Table 5. In four course offerings — Intro1, Intro2, Statics4, and Statics5 — students used an instrumented document viewing program called DocViewer to read course documents, including the digital textbook. The students in these four courses also took lecture notes with a smartpen and submitted their notes electronically to receive course credit. Students in three of the Statics offerings — Statics1, Statics2, and Statics3 — completed a design project in addition to the usual homework assignments. This project required students to apply principles of statics to determine suitable dimensions for the components of a mechanical system. Students in Intro1 completed several preparatory quizzes, in addition to the usual homework. The preparatory quizzes were designed to engage students in reading course materials prior to lecture (Gyllen, Stahovich, Mayer, Entezari, & Darvishzadeh, 2018). Students in Statics4 were given weekly reports providing feedback on their effort on previous homework assignments. The feedback included the amount of time spent on the assignment, the number of problems attempted, and how much of their work was done within 24 hours of the due date. The feedback also included how much time they spent reading and how much of the assigned reading they completed as measured by DocViewer. Students in Statics1 used an intelligent tutoring system, called Newton's Pen 2 (Lee, Stahovich, & Calfee, 2011), during several discussion periods. This system scaffolded students in solving statics problems. Students in Statics3 were given problems to solve during some of the discussion sections. They began the problems in a discussion section and submitted the solutions along with their homework.

In most of the course offerings, all students in the course completed the same work. However, students in Statics2 were split into four separate groups (by discussion section) that received different experimental treatments. Students in two of the groups were asked to answer self-explanation prompts requiring explanations for the steps in their solutions for problems on six of the homework assignments. Students in the third group used the Newton's Pen in some of the discussion sections. Students in the fourth group served as the control, and did not provide self-explanation or use Newton's Pen 2. A one-way analysis of variance (ANOVA) revealed no significant differences in the final course grades between the four groups (p = .706).

Students in Statics3 were randomly split into one of six experimental groups. Four of the groups were asked to answer self-explanation prompts with varying amounts of scaffolding (Wood, Bruner, & Ross, 1976; Azevedo & Hadwin, 2005; Bielaczyc, Pirolli, & Brown, 1995; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Hall & Vance, 2010) in the construction of written explanations. Students in the fifth group used Newton's Pen 2 during discussion sections. Students in the sixth group served as the control. For the final homework assignment, all students were prompted to provide self-explanation without scaffolding. An ANOVA revealed no significant differences in the final course grades between the six groups (p = .957).

In our analysis, we consider each course offering individually. Thus, differences between courses create no issues for our analysis. For Statics2 and Statics3, there were differences within the course in the treatments students received. However, as ANOVA revealed no significant differences in learning outcomes as measured by final course grade, we ignore these within-course differences.

Final course grades for both offerings of Introduction to Mechanical Engineering were computed in similar ways, except that for Intro2, the preparatory assignments counted for 5% of the final grades. For both course offerings, taking and submitting lectures notes counted for 5% of the final grade. For our analysis, we compute final grade for both courses using a modified scheme based on the following weighting: 10% for the homework score, 10% for the quiz score, 20% for the first midterm exam score, 20% for the second midterm exam score, and 30% for the final exam score. We normalize the resulting score to produce a 100-point scale. We exclude the preparatory quizzes from this grade so that a consistent grading scheme is used for both courses. We exclude the lecture notes as they represent class participation rather than achievement.

Final course grades for the various offerings of Statics were also computed in similar ways, with some variations. For offerings with a design project, the project contributed 10% of the final course grade. Also, in some offerings, taking and submitting lecture notes counted for 5% of the final grade. For our analysis, we exclude both the design project and the lecture notes and compute the final course grade with the same scheme we use for the Introduction to Mechanical Engineering courses. We exclude the design project from the grade so that a consistent grading scheme is used for all courses. We exclude the lecture notes as they represent class participation rather than achievement.

For Statics1, Statics2, and Statics3, students received their smartpens during the second week of the quarter and began using them with the third homework assignment. For the other course offerings, students received their smartpens at the beginning of the term and used them for all of the homework assignments.

The method by which students submitted the digital ink for their homework evolved over the span of the study. Students in Statics1 (which is the earliest course in the study) submitted both a paper copy of their homework and their smartpen each week. We graded the paper copy of the homework and extracted the ink data from the smartpen. Students in Statics2 (which is the second earliest course in the study) had the option of submitting a paper copy or a digital copy of their work, with digital submission encouraged by having a due date that was extended by a few hours on some assignments. The students used software we developed called InkViewer to submit the digital copies. If students submitted homework on paper, we extracted the digital ink data from their smartpens at the end of the quarter. In the other five offerings, students used InkViewer to submit all of their homework.

Intro1 and Intro2 each had nine homework assignments with 70 and 72 problems, respectively. We collected digital ink data from all of these assignments. Statics1 and Statics2 also had nine assignments, but we collected data from only the last seven, as students received their smartpens after the second homework was completed. Likewise, Statics3 had eight assignments and we collected data from only the last six. In total, we collected data from 41 problems for Statics1, 44 problems for Statics2, and 40 problems for Statics3. Statics4 and Statics5 each had 10 homework assignments, but as the last assignment was not graded and students were not required to submit their work, we exclude it from our study. In total, we collected data from 60 problems for Statics4 and 60 problems for Statics5. We followed guidelines for research with human subjects and obtained IRB approval.

Materials and Apparatus

Students completed their homework assignments using Livescribe smartpens. They wrote their homework solutions in Livescribe single-subject, spiral-bound notebooks containing 8.5 in. by 11 in. college-ruled paper. As described above, most students used the InkViewer software to submit their homework assignments electronically.

Measures

Building on the work in Rawson et al. (2017), we focus our analysis on four measures of homework habits: *Early Work Fraction* (E) as a measure of completing work on time, number of *Problems Attempted* (P) as a measure of being on task, and *Stroke Count* (S) and *Length of Ink Written* (L) as measures of effort in working. The *Early Work Fraction* is computed as the fraction of the pen strokes written at least 24 hours prior to the homework submission due date. This measure quantitatively describes the extent to which students begin their work early, rather than waiting until the proverbial last minute. The *Early Work Fraction* is the complement of the due date ink fraction measure defined in Rawson et al. (2017) (i.e., *Early Work Fraction* and due date ink fraction sum to one). In this work we use *Early Work Fraction*, rather than due date ink fraction, to obtain a measure that has a positive correlation with performance. (For Statics2, we compute the *Early Work Fraction* using the due date for electronic homework submission.) *Problems Attempted* is measured as the number of problems for which a student wrote at least 50 pen strokes. If a student wrote fewer than 50 strokes for a particular problem, it is unlikely that he or she made significant progress on it. For example, simply writing “Problem 1” takes at least eight strokes. The *Stroke Count* is the number of pen strokes written. The *Length of Ink Written*, which is measured in units of inches, is the distance the pen tip travels on the paper.

We use subscripts to indicate the cumulative measures through specific homework assignments. For example, S_4 is the number of strokes written on the first four assignments cumulatively. For Statics1, Statics2, and Statics3, there is no data for the first two homework assignments. For these courses, for example, S_4 is the number of strokes written for assignments three and four.

Results

Table 6 shows the mean and standard deviations of the four measures of homework habits for each cohort. Note that for Statics1, Statics2, and Statics3, the data does not include the first two homework assignments, as students had not yet received their smartpens. The cohort average for *Early Work Fraction* ranged from .23 for Intro2 to .34 for Statics5. The cohort average fraction of *Problems Attempted* ranged from .80 for Intro1 to .89 for Statics5. The average number of strokes written (*Stroke Count*) ranged from 15,056 for Intro2 to 26,219 for Statics5. Likewise, the cohort average *Length of Ink Written* ranged from 4,320 in. for Intro2 to 8,327 in. for Statics5. As a reference, 4,320 in. of ink is equivalent to drawing 665 lines across a page of letter paper (8.5 in. by 11 in.) from the left margin to the right margin with one in. margins. Similarly, 8,327 in. of ink is equivalent to drawing 1,281 lines across a page.

Table 6
Means and Standard Deviations for Conscientiousness Measures for Each Cohort over the Entire Quarter

Course		*Early Work Fraction*	*Problems Attempted* (% Assigned)	*Stroke Count*	*Length of Ink Written*
Intro1	**M**	**.27**	**56.3 (80%)**	**16,129**	**5,005**
	SD	.25	17.4 (25%)	8,616	2,624
Intro2	**M**	**.23**	**63.2 (88%)**	**15,056**	**4,320**
	SD	.24	9.9 **(14%)**	4,841	1,782
Statics1	**M**	**.26**	**35.7 (87%)**	**18,944**	**5,569**
	SD	.22	5.9 **(14%)**	6,880	2,308
Statics2	**M**	**.26**	**35.5 (81%)**	**21,162**	**5,979**
	SD	.24	10.2 **(23%)**	10,002	3,252
Statics3	**M**	**.42**	**32.4 (81%)**	**20,257**	**6,166**
	SD	.26	9.2 **(23%)**	9,855	3,281
Statics4	**M**	**.29**	**49.7 (83%)**	**23,728**	**6,626**
	SD	.24	14.0 **(23%)**	10,382	3,271
Statics5	**M**	**.34**	**53.7 (89%)**	**26,219**	**8,327**
	SD	.29	10.4 **(17%)**	10,491	3,994

Are homework habits early in the quarter related to course performance?

According to the early warning hypothesis, scores on conscientiousness measures of homework habits exhibited in the first few weeks of the quarter will predict course grade. Work in Rawson et al. (2017) demonstrated that the cumulative homework effort over an entire term was strongly and significantly correlated with course grade. As our goal in the present study is to enable an early warning for at-risk students, we examine how the predictive capability of the four measures of homework habits varies over the term. More specifically, we compute the cumulative measures after each assignment and compute their correlation with final course grade.

Figure 4 shows the correlation between the cumulative *Early Work Fraction* and course grade for all seven courses. (Correspondingly, Table 27 in the appendix lists the correlation coefficients.) Each data point represents the cumulative *Early Work Fraction* up to and including a particular assignment. For example, the data points for "E_3" represent the cumulative *Early Work Fraction* for all assignments up to and including homework assignment 3. Filled data points represent correlations that are significant ($p < .05$) while unfilled data points represent correlations that are not significant. The correlation is significant by homework 3 with the exception of Statics1 and Statics4 (Statics4 approaches significance with $p = .060$). Recall that for Statics1, Statics2, and Statics3, no data was available for the first two assignments, as students had not yet received their smartpens. The ratio of the correlation coefficient at homework 3 to that at the final homework, excluding Statics1 and Statics4, averages 1.00. Consistent with

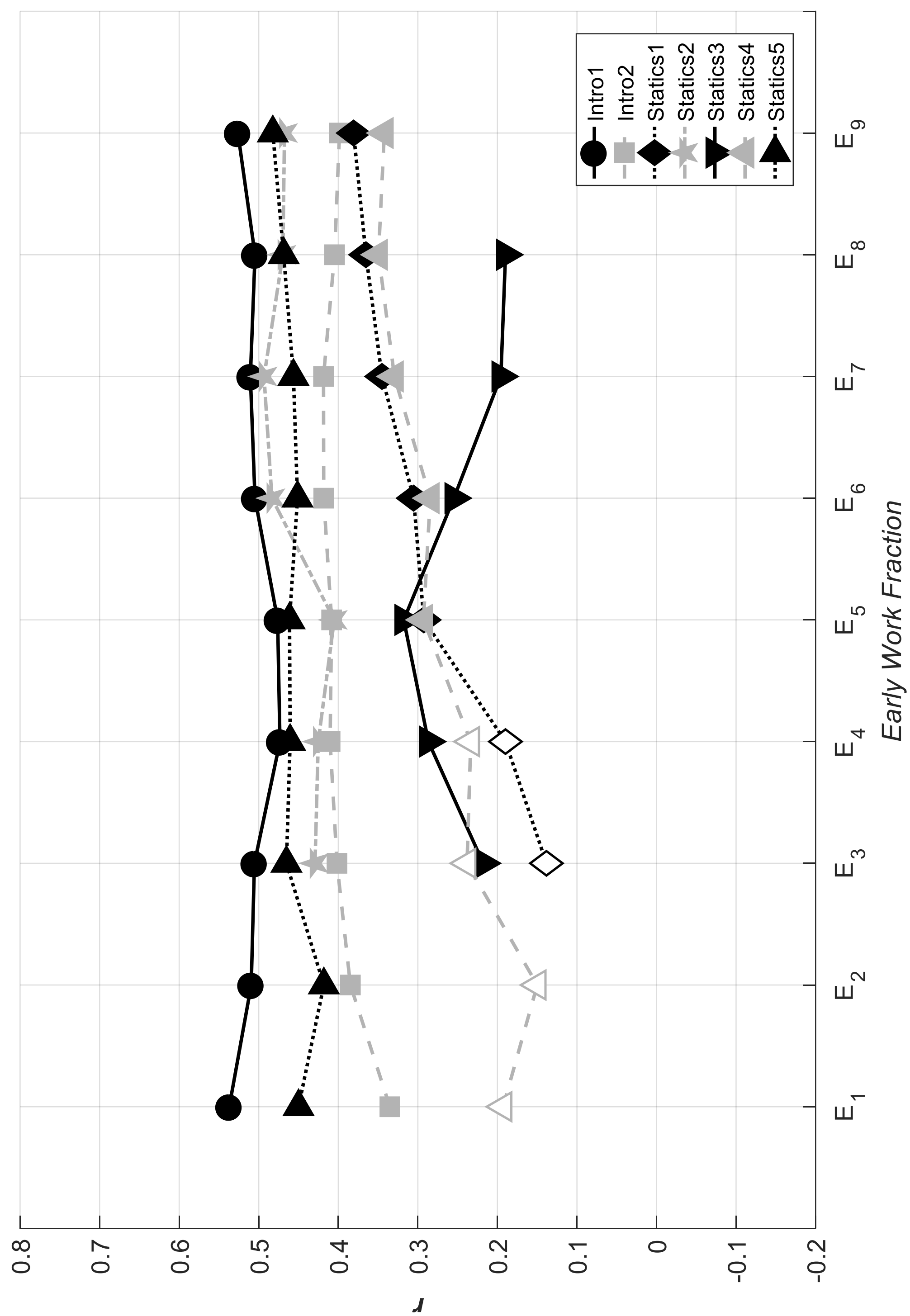


Figure 4. Correlation between *Early Work Fraction* (E) and course grade.

prediction 1, we conclude that a measure of completing work on time (*Early Work Fraction*) on early homework assignments correlates positively and significantly with course grade.

Figure 5 shows the correlation between the cumulative number of *Problems Attempted* and course grade for all seven courses. (Correspondingly, Table 28 in the appendix lists the correlation coefficients.) The correlation is significant for all seven courses by homework 3. The ratio of the correlation coefficient at homework 3 to that at the final homework averages .75. Consistent with prediction 2, we conclude that a measure of being on task (*Problems Attempted*) on early homework assignments correlates positively and significantly with course grade.

Figure 6 shows the correlation between the cumulative *Stroke Count* and course grade for all seven courses. (Correspondingly, Table 29 in the appendix lists the correlation coefficients.) For all courses, except Intro2, the correlation between cumulative *Stroke Count* and grade is significant by the third assignment. For Intro2, the correlation is not significant until homework 4. The ratio of the correlation coefficient at homework 3 to that at the final homework, excluding Intro2, averages .82.

Figure 7 shows the correlation between the cumulative *Length of Ink Written* and course grade for all seven courses. (Correspondingly, Table 30 in the appendix lists the correlation coefficients.) With the exception of Intro2 and Statics4, the correlation is significant by homework 3. For Statics4, the correlation is not significant until homework 4 (p = .136 at homework 3). Intro2 is an outlier: the correlation is non-significant (p >

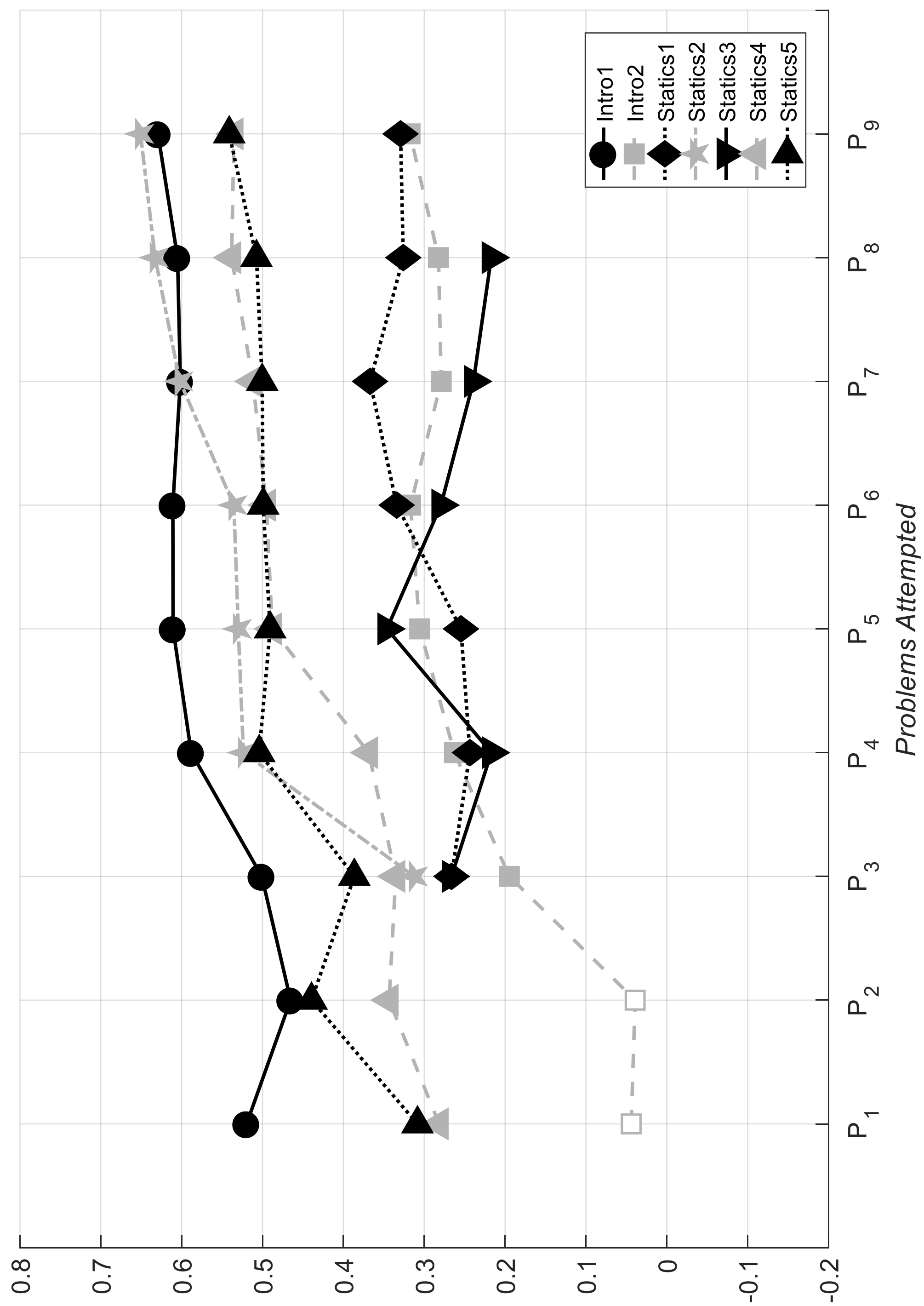


Figure 5. Correlation between *Problems Attempted* (P) and course grade.

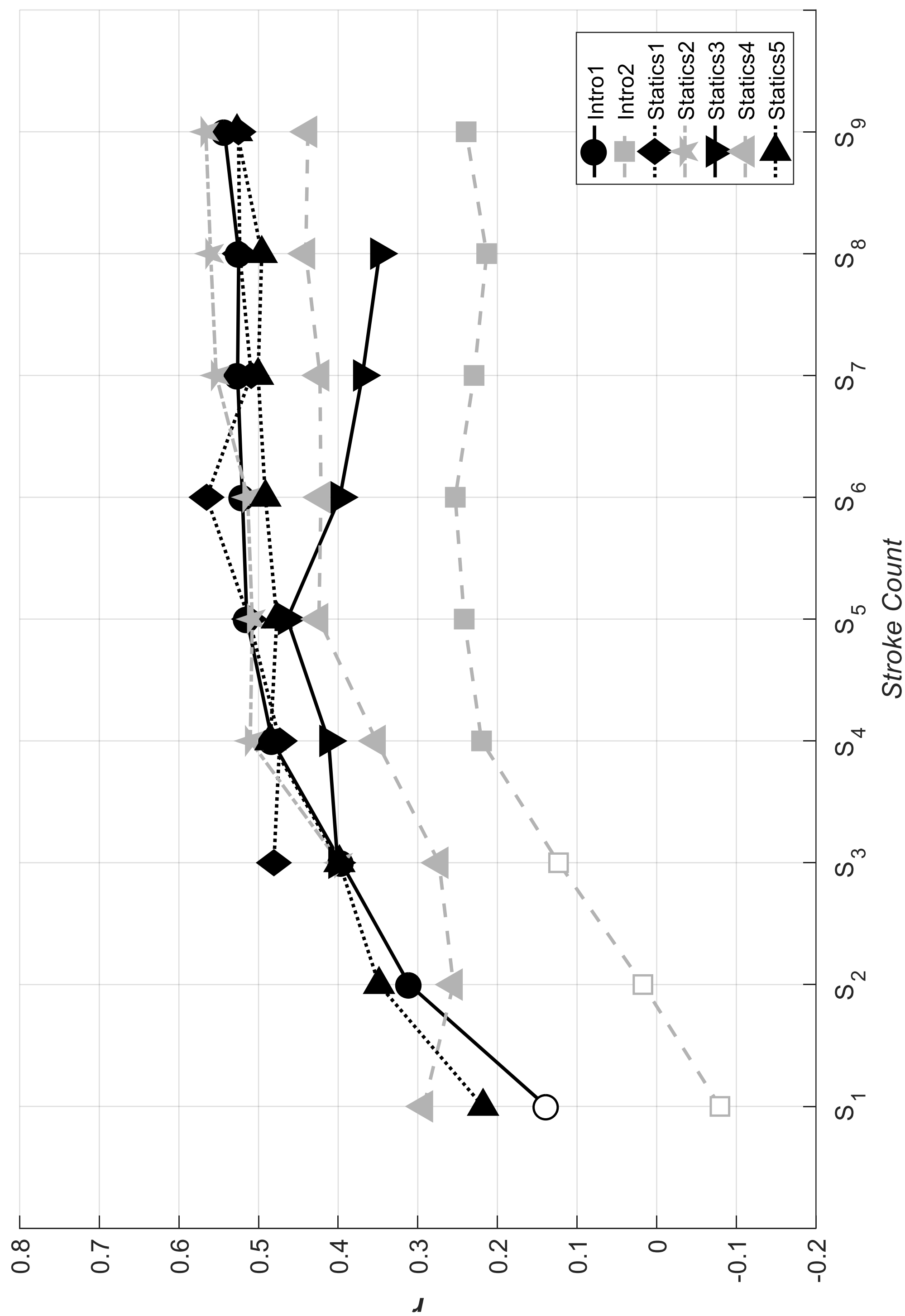


Figure 6. Correlation between *Stroke Count* (S) and course grade.

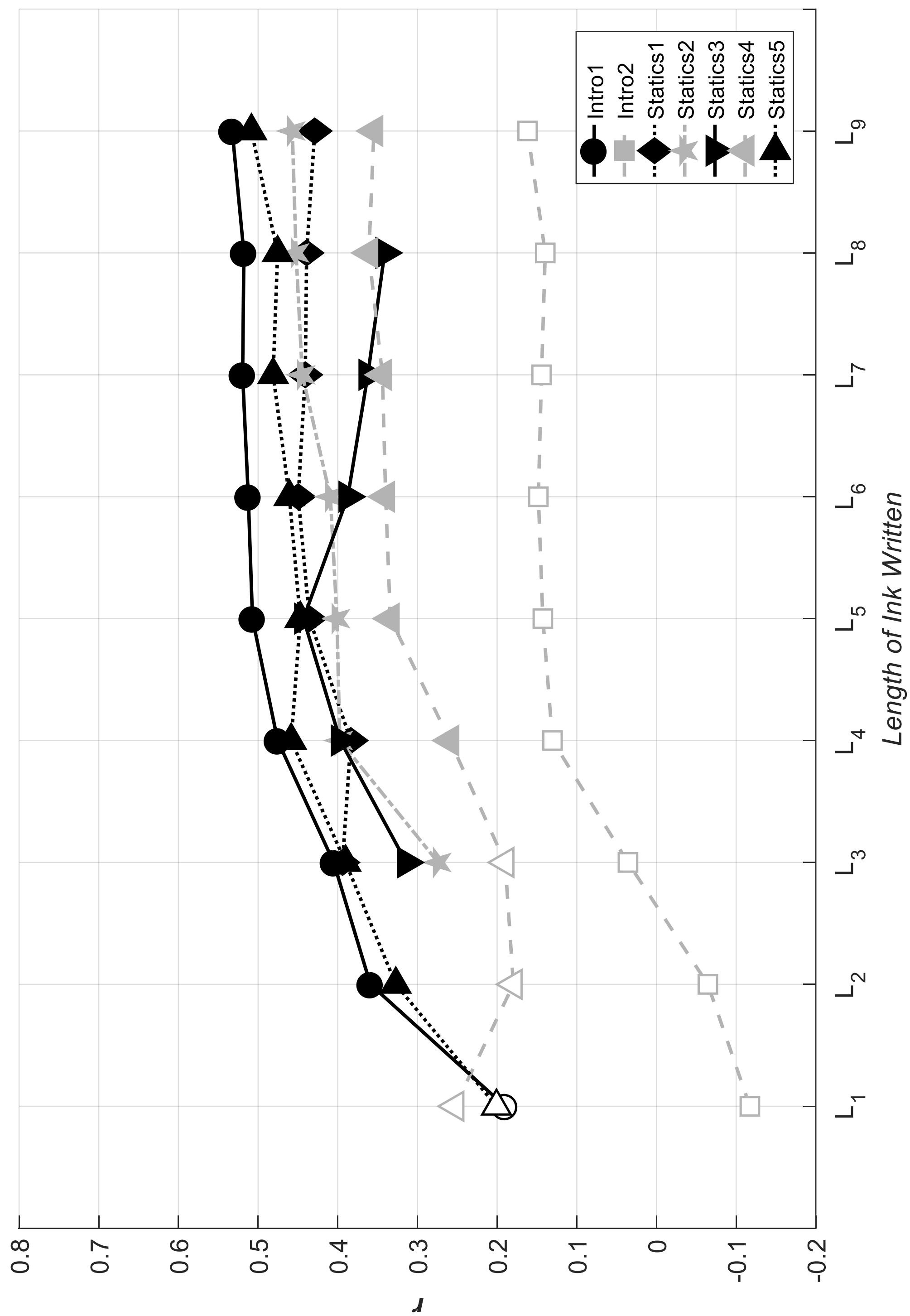


Figure 7. Correlation between *Length of Ink Written* (L) and course grade.

.100 until homework 9, with p =.091), and for the first two homework assignments, the correlation is negative. For the other five courses, the ratio of the correlation coefficient at homework 3 to that at the final homework assignment averages .85. Consistent with predictions 3 and 4, we conclude that measures of working with high effort (*Stroke Count* and *Length of Ink Written*) on early homework assignments correlates positively and significantly with course grade.

Figure 8 shows the correlations between each of the four conscientiousness measures of homework habits and course grade averaged over the seven courses. On average, the *Early Work Fraction* produces early predictions that are nearly as good as those based on the entire homework record. The average correlations for the other three measures — *Problems Attempted*, *Stroke Count*, and *Length of Ink Written* — strengthen until about the fifth homework assignment and then become roughly constant as the number of assignments increases. Consistent with predictions 1-4, when we average over all courses, it becomes even clearer that measures of conscientious homework habits in the first few weeks of a course predicts course grade.

Generally, consistent with the early warning hypothesis, performance on each of the four homework habit measures correlated positively and significantly with course grade by the third week of the quarter. This means that by the third week of the quarter it was possible to determine who needed extra help in the course.

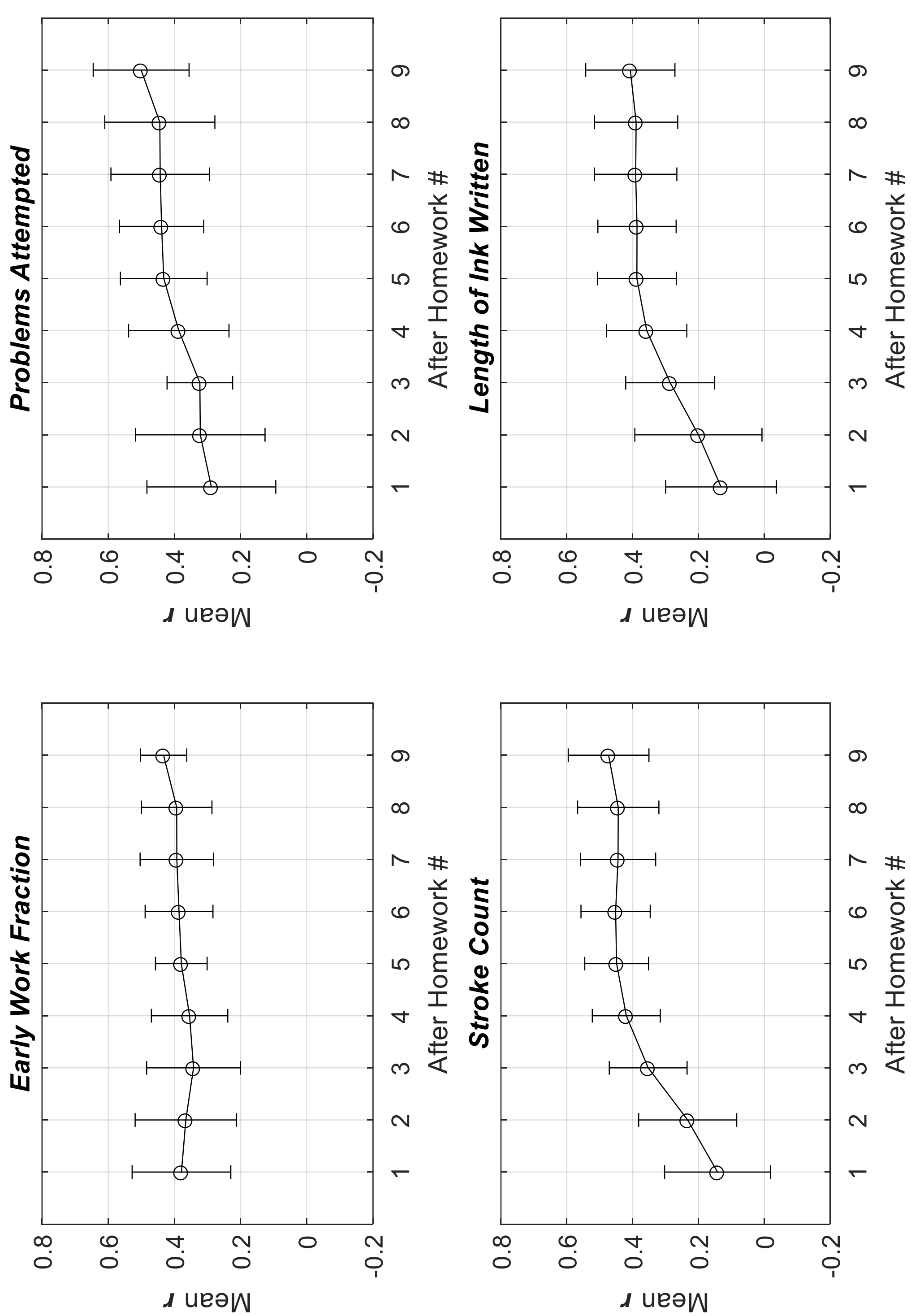


Figure 8. Correlations between each of the four measures of conscientiousness homework habits and course grade averaged over the seven courses. Error bars represent one standard deviation.

Does combining the four measures increase the predictive ability?

We performed stepwise multiple linear regression to examine the predictive ability of the four measures of conscientiousness in homework habits, taken together, over the term. In computing a stepwise model, we require the probability of $F \leq .05$ to enter a measure, and the probability of $F \geq .10$ to remove a measure. Figure 9 shows the results of this analysis. (Table 31 in the appendix lists the correlation coefficients.) For all courses, the correlation of the stepwise models built with the four features is positive and significant after all homework assignments. For the four courses with homework data for the first assignment (Intro1, Intro2, Statics4, and Statics5), the correlation with grade after the first assignment is comparable to the correlation achieved with the benefit of the full set of homework data. For these four courses, the ratio of the correlation coefficient at homework 1 to that after the final assignment averages .78. For all courses, the ratio of the correlation coefficient at homework 3 to that at the final homework assignment averages .88. Thus, taken together, the four features produce a strong early correlate of final course grade.

As a second means of examining the collective power of the four measures of conscientiousness in homework habits, we computed a weighted combination of them. As the measures have different scales, we first converted each measure to a z-score. More specifically we computed the *Combined Measure of Conscientiousness*, C, as:

$$\mathrm{C} = \frac{1}{3}\left(\frac{\mathrm{E} - \mu_{\mathrm{E}}}{\sigma_{\mathrm{E}}}\right) + \frac{1}{3}\left(\frac{\mathrm{P} - \mu_{\mathrm{P}}}{\sigma_{\mathrm{P}}}\right) + \frac{1}{6}\left(\frac{\mathrm{S} - \mu_{\mathrm{S}}}{\sigma_{\mathrm{S}}}\right) + \frac{1}{6}\left(\frac{\mathrm{L} - \mu_{\mathrm{L}}}{\sigma_{\mathrm{L}}}\right)$$

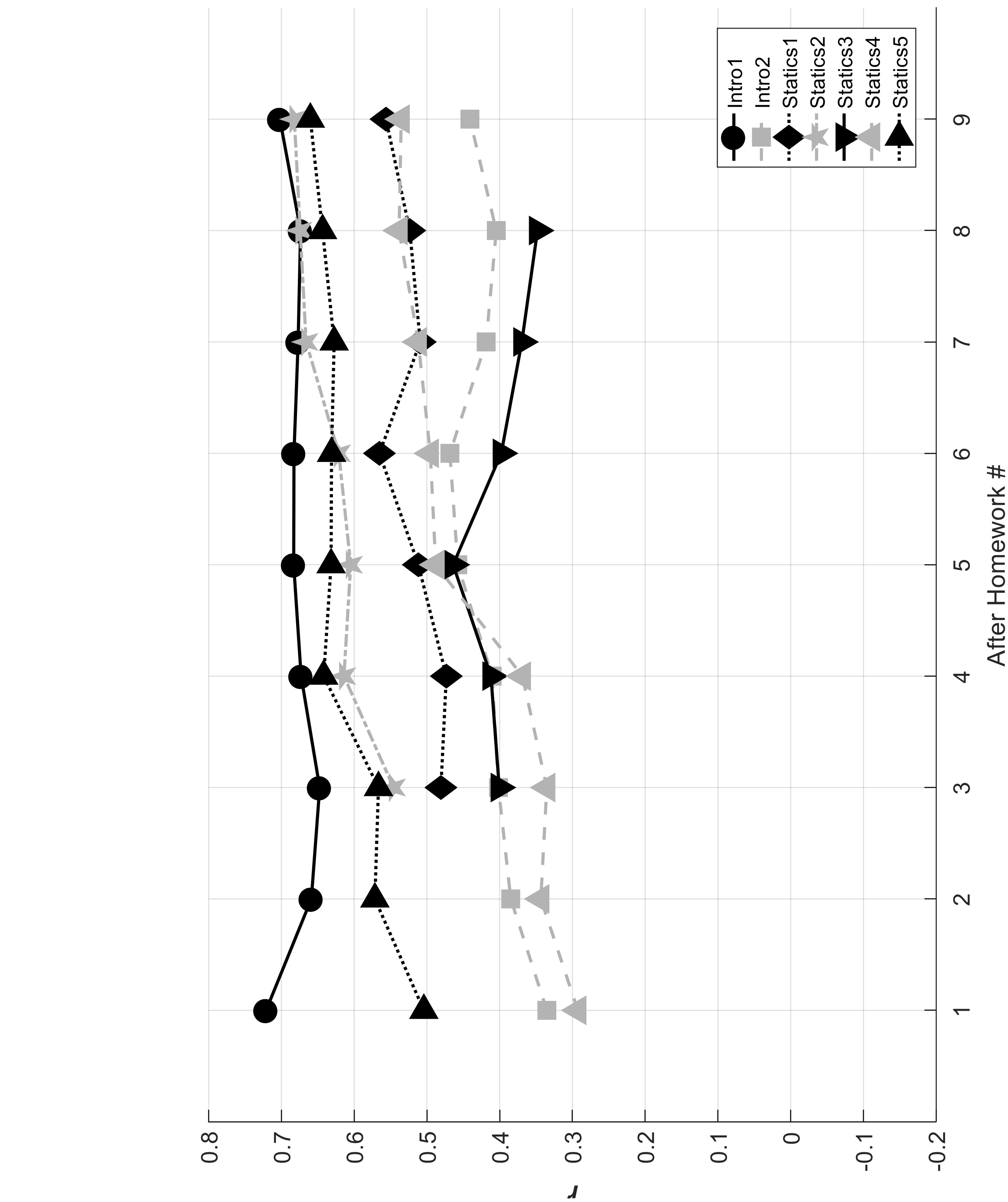


Figure 9. Correlation of stepwise regression models relating all four measures of conscientiousness homework habits to course grade.

where μ_E and σ_E are the mean and standard deviation of the *Early Work Fraction*, μ_P and σ_P are the mean and standard deviation of *Problems Attempted*, μ_S and σ_S are the mean and standard deviation of *Stroke Count*, and μ_L and σ_L are the mean and standard deviation of *Length of Ink Written*. The means and standard deviations are computed separately for each course. In this calculation, the three features of conscientiousness — being on time, being on task, and producing high effort — are all weighted equally. *Early Work Fraction*, which is a measure of being on time, has a weight of 1/3. *Problems Attempted*, which is a measure of being on task, also has a weight of 1/3. *Stroke Count* and *Length of Ink Written*, which are both measures of producing high effort, each have a weight of 1/6 so that the high-effort feature as a whole has a weight of 1/3.

Figure 10 shows the correlation between the *Combined Measure of Conscientiousness* and course grade for each of the seven courses. (Table 32 in the appendix lists the correlation coefficients.) For all courses, the correlation between the *Combined Measure of Conscientiousness* and course grade was significant by the third homework assignment. Furthermore, for all courses, the ratio of the correlation coefficient at homework 3 to that at the final homework assignment averages .86. Thus, the *Combined Measure of Conscientiousness* produces a strong early correlate of final course grade.

Figure 11 shows the correlation between the *Combined Measure of Conscientiousness* and course grade averaged over the seven courses. After homework 3, the *Combined Measure of Conscientiousness* produces an average correlation with

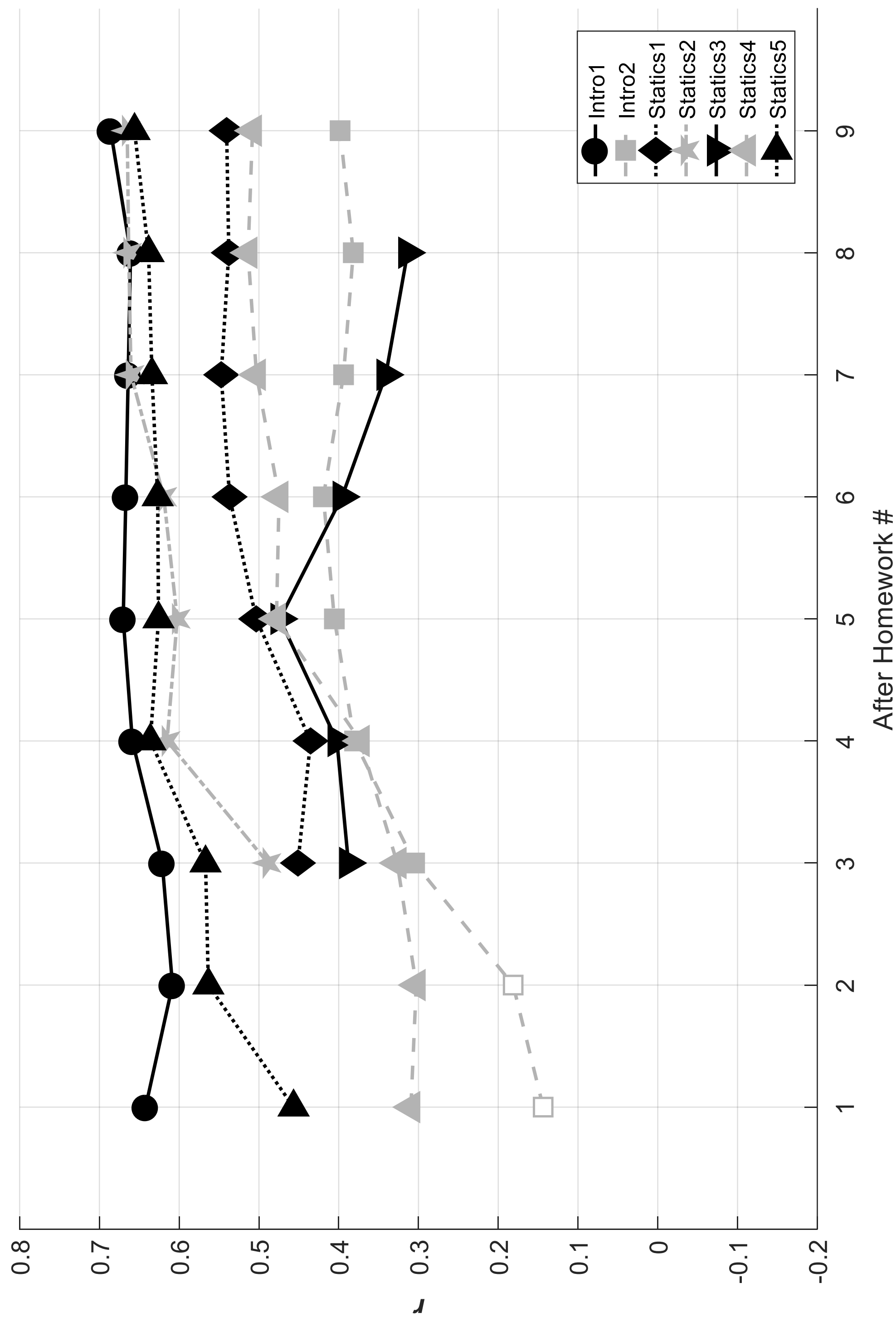


Figure 10. Correlation between *Combined Measure of Conscientiousness* (C) in homework habits and course grade for each of the seven courses.

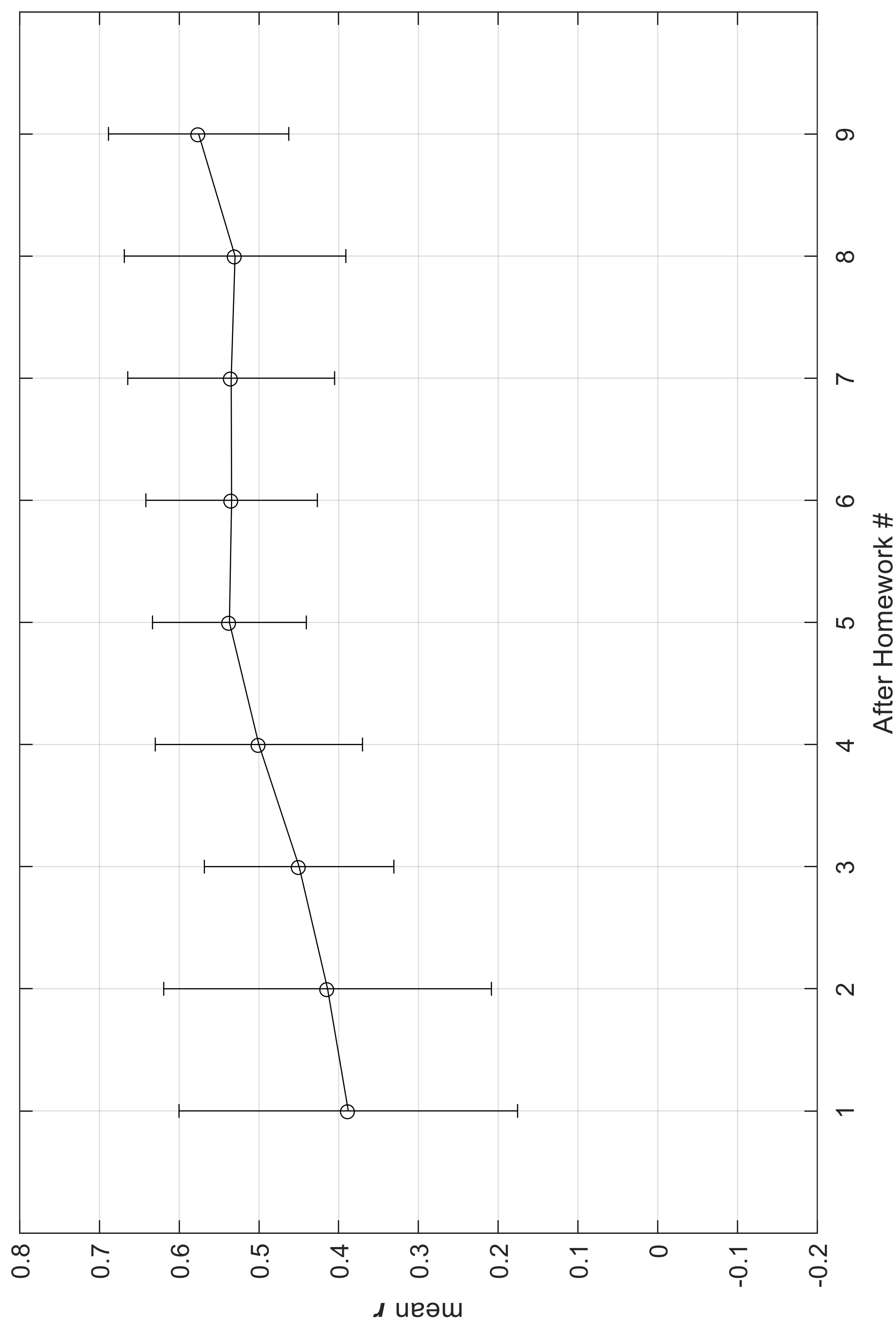


Figure 11. Correlations between *Combined Measure of Conscientiousness* (C) in homework habits and course grade averaged over the seven courses. Error bars represent one standard deviation.

course grade of .45, while the average correlations for *Early Work Fraction*, *Problems Attempted*, *Stroke Count*, and *Length of Ink Written* are .34, .32, .35, and .29, respectively. Consistent with prediction 5, we conclude that a composite of conscientious measures based on early homework assignments correlates positively and significantly with course grade.

Does consideration of the progression of homework habits increase the predictive capability?

Our results have shown that measures of conscientious homework habits on early assignments correlate positively and significantly with course grade. Here we examine if consideration of the progression of study behavior over the term provides additional predictive power. To this end, we construct stepwise regression models that relate the sequence of measures to final course grade. (In computing stepwise models, we again require the probability of $F \leq .05$ to enter a measure, and the probability of $F \geq .10$ to remove a measure.)

Consider the graph for Intro1 in Figure 12. (Table 33 in the appendix lists the correlation coefficients.) The n^{th} data point on the graph represents a stepwise regression model computed from n independent variables comprising the n sequential values of the *Early Work Fraction* measure: E_1, E_2, …, E_n. For example, the first data point represents a stepwise model constructed from one independent variable (E_1); this model has a correlation of $r = .54$ ($p < .05$). Likewise, the 9^{th} data point represents a

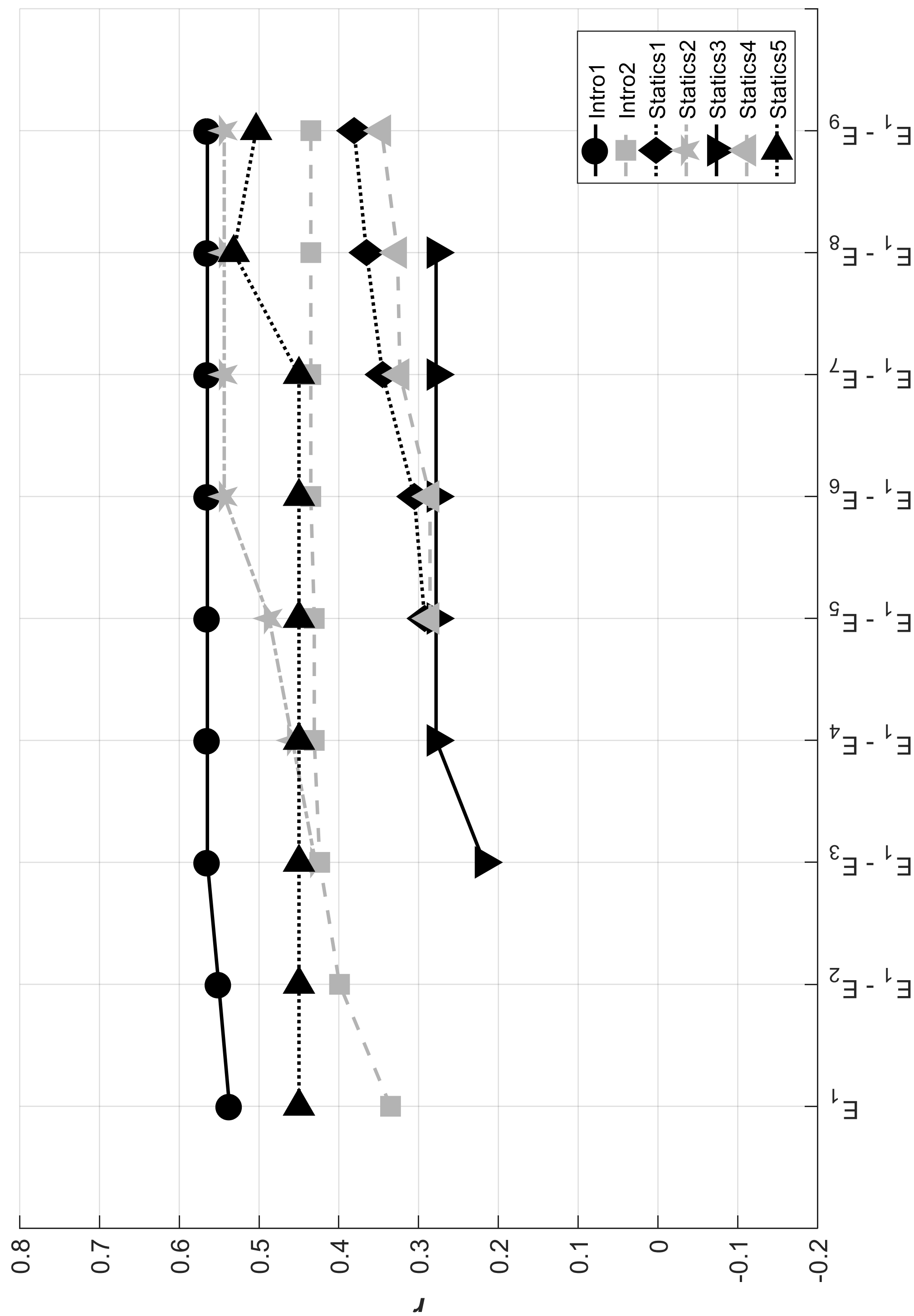


Figure 12. Correlation of stepwise regression models relating the sequence of *Early Work Fraction* (E) values to course grade.

stepwise model constructed from nine independent variables (E_1, E_2, …, E_9); this model has a correlation of $r = .56$ ($p < .05$). By considering the sequence of measures, these models explicitly examine how changes in student habits over the term relate to course performance.

Comparison of the results in Figure 12 with those in Figure 4 reveals that considering the sequence of *Early Work Fraction* values does generally result in a small increase in the predictive power compared to considering a single value of the measure, with an average increase in the correlation coefficient by 5.8%.

Figure 13 shows the results of the same stepwise regression analysis, but now utilizing *Problems Attempted*. (Table 34 in the appendix lists the correlation coefficients.) The correlation achieved using the sequence of *Problems Attempted* values was similar to that achieved using single values: on average, the correlation coefficient of the stepwise models was only 0.3% greater.

Figure 14 shows the results of the same stepwise regression analysis, but now utilizing *Stroke Count*. (Table 35 in the appendix lists the correlation coefficients.) The correlation achieved using the sequence of *Stroke Count* values does generally result in a small increase in the predictive power compared to considering a single value of the measure, with an average increase in the correlation coefficient of 8.8%.

Figure 15 shows the results of the same stepwise regression analysis, but now utilizing *Length of Ink Written*. (Table 36 in the appendix lists the correlation coefficients.) Here again, the correlation achieved using the sequence of *Length of Ink*

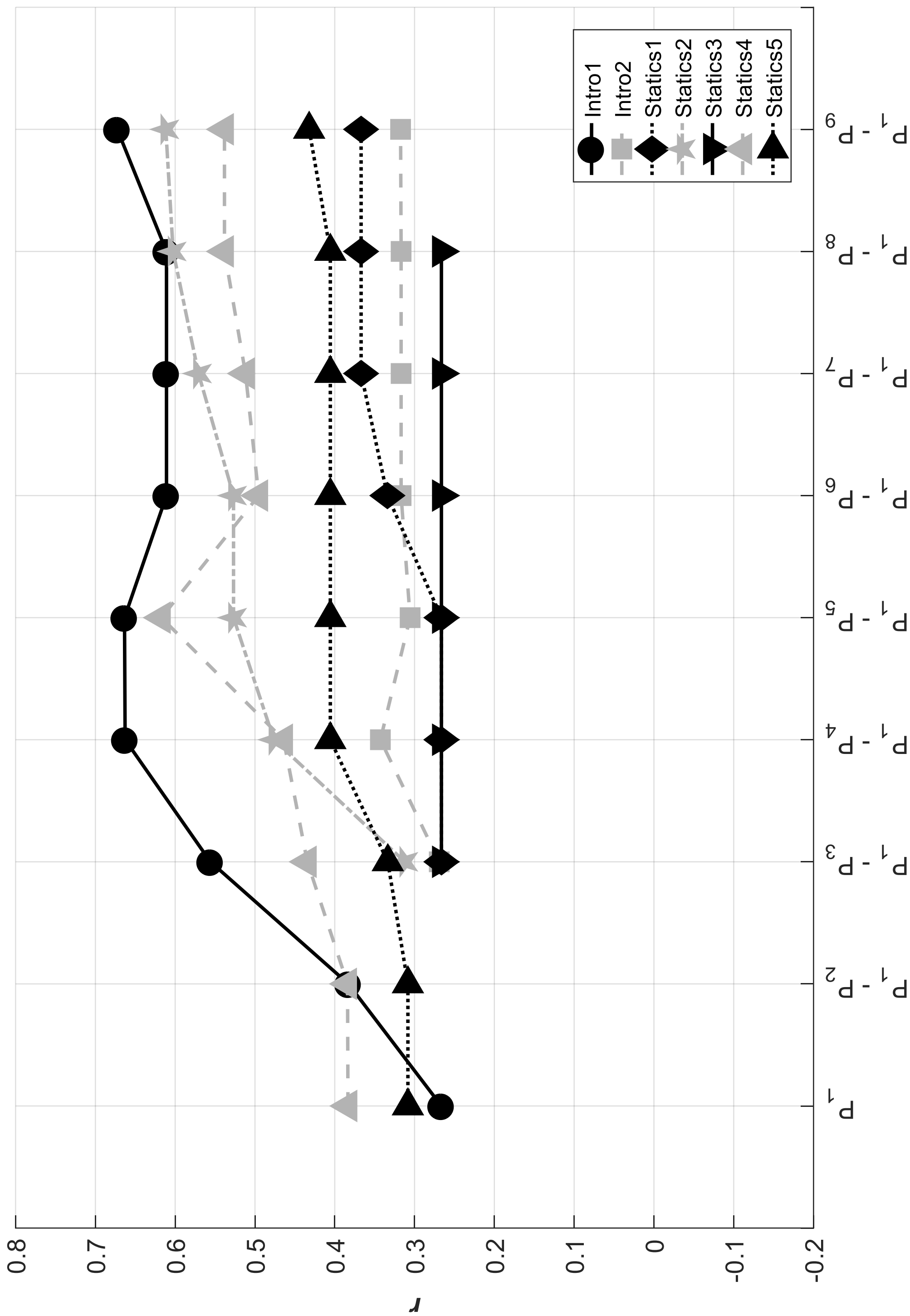


Figure 13. Correlation of stepwise regression models relating the sequence of *Problems Attempted* (P) values to course grade.

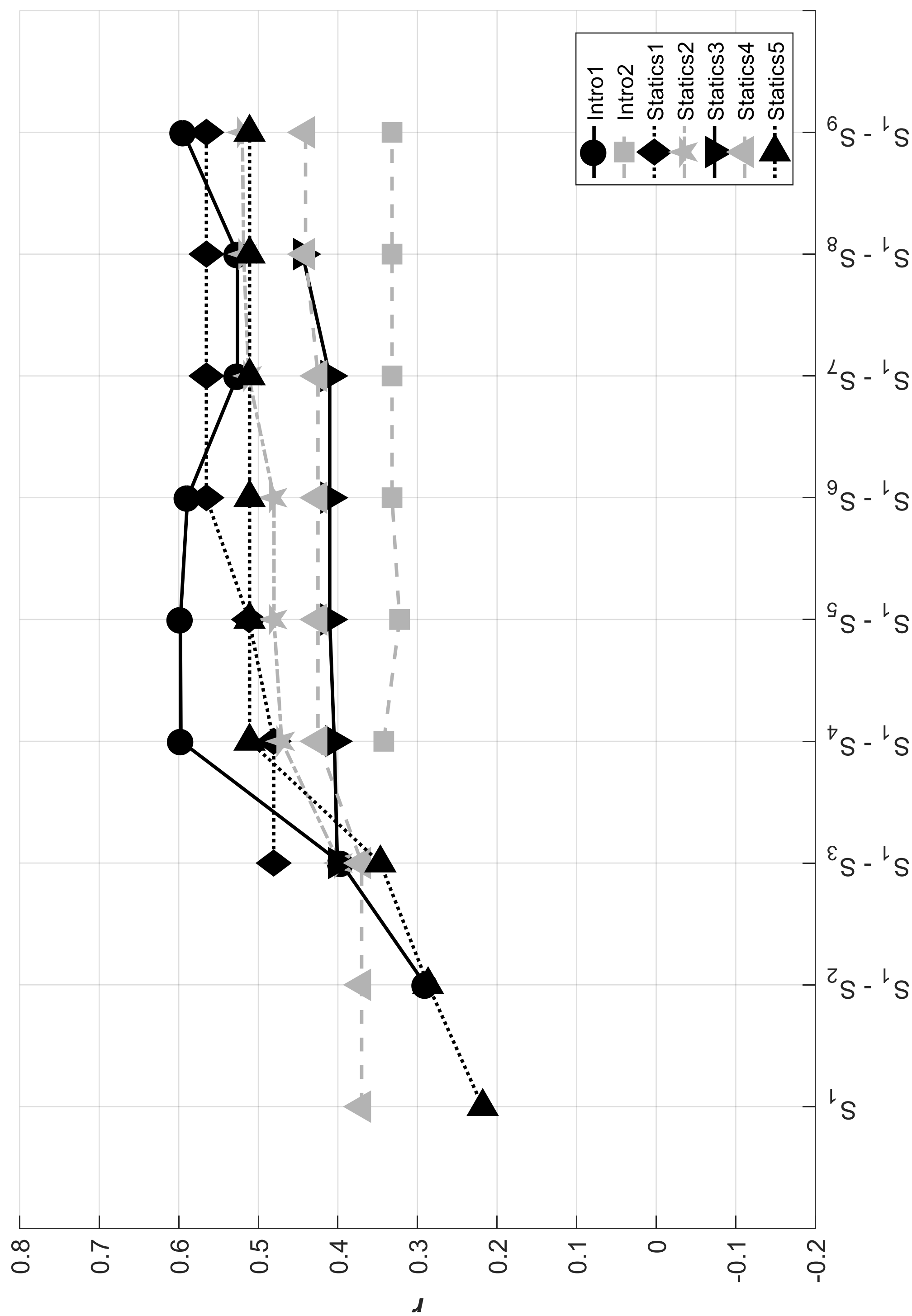


Figure 14. Correlation of stepwise regression models relating the sequence of *Stroke Count* (S) values to course grade.

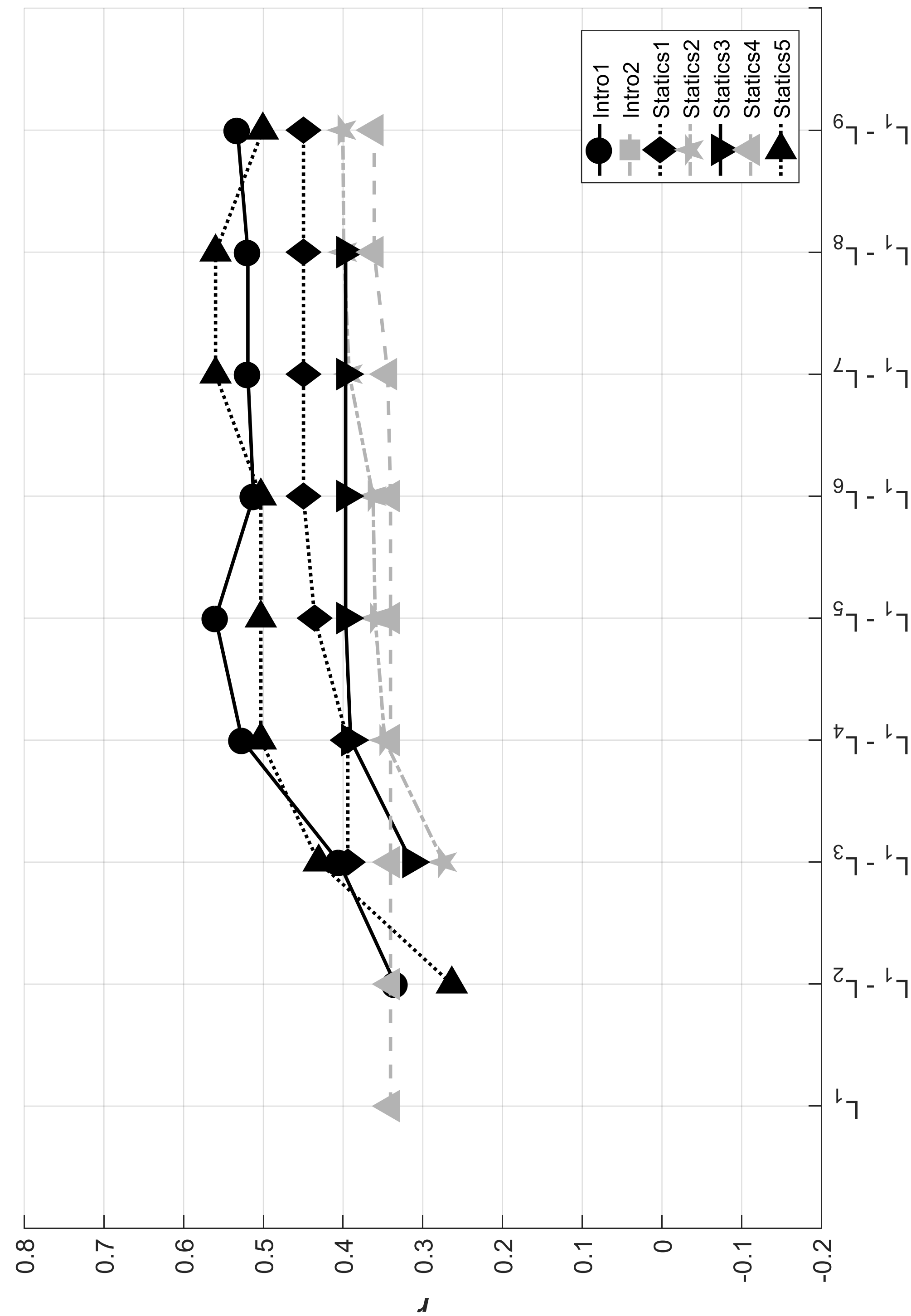


Figure 15. Correlation of stepwise regression models relating the sequence of *Length of Ink Written* (L) values to course grade.

Written values does generally result in a small increase in the predictive power compared to considering a single value of the measure. Intro2, however, was an outlier; the stepwise models produced no significant correlations. Otherwise, the correlation coefficients of the stepwise models increased on average by 9.3% over the single feature models.

Consistent with prediction 6 and based on stepwise regressions, we conclude that after a few weeks the correlations between homework habits and course grade are well established and homework habit data from subsequent weeks does not substantially improve the predictive power.

Discussion

Empirical Contribution

The primary finding is that final grade in an engineering course correlates with each of the four measures of conscientious habits in solving homework problems after the third week of the course (*Early Work Fraction* which measures completing work on time, *Problems Attempted* which measures being on task, and *Stroke Count* and *Length of Ink Written* which measure effort in working). Specifically, we observe after the third homework assignment significant and strong correlations for cumulative *Early Work Fraction* (except Statics1 and Statics4), *Problems Attempted, Stroke Count* (except Intro2), and *Length of Ink Written* (except Intro2 and Statics4) with course grade. Considering all four of these measures together, and using stepwise regression, we observed even stronger correlations between early homework habits and course grade. For example, using all four features in this way, the correlation coefficients achieved after the third homework were significant for all seven courses and were, on average, 88% as large as those obtained using the entire homework record. Likewise, the *Combined Measure of Conscientiousness*, which is a weighted combination of the four individual measures, produced significant correlations with course grade for all seven courses by the third homework assignment. Furthermore, the correlations at the third homework assignment were, on average, 86% as large as those obtained using the entire homework record. In short, we can predict the grade a student will eventually receive in a course based on how they solve their homework problems during the first few weeks of the course.

Theoretical Implications

Results support the early warning hypothesis and demonstrate the role of conscientiousness in academic success. Although no causal conclusions can be drawn, this study provides support for the theory that conscientiousness is an important predictor of academic success. The three core features of conscientiousness (Pellegrino & Hilton, 2012) are being on time, being on task, and producing high effort. The study demonstrates that *Early Work Fraction*, a measure of being on time, correlates positively and significantly with course grade. Similarly, the study demonstrates that *Problems Attempted*, a measure of being on task, correlates positively and significantly with course grade. Finally, the *Stroke Count* and the *Length of Ink Written*, which are measures of high effort, correlate positively and significantly with course grade. Thus, the results demonstrate that all three core features of conscientiousness are indicators of academic achievement. In addition, this study suggests an extension to this theory by demonstrating that conscientiousness early in a course is predictive of ultimate academic achievement in the course. However, future experiments involving interventions are needed to understand any causal implications of this extension to the theory.

Practical Implications

Our four measures of conscientious homework habits can be used as an early warning system to detect students who will need help in a course. Within the first few weeks of a course, these measures can help identify students who need support in an engineering course. Thus, these measures provide a practical and scalable mechanism for building an early warning system to detect students who are at risk of poor performance so that interventions can be applied early enough to produce an improved outcome. In particular, students who score low on the measures may benefit from training and guidance in how to manage their study time (Mayer, 2019).

Methodological Implications

This study highlights the value of educational data mining techniques using smartpen technology for educational research. In particular, smartpen technology allows for assessing student homework habits with a high level of detail in an unobtrusive manner and on a large scale. This study also speaks to the value of replication in education research (Shavelson & Towne, 2002) by demonstrating the same pattern of results across seven separate cohorts of students and two different courses.

Limitations and Future Directions

Future research is needed to determine whether similar results can be obtained in other STEM courses. Future research is needed to determine the effects on course grade of providing guidance and training to students who are detected (and not detected) by the early warning system. Finally, research is needed to determine whether conscientiousness in working on homework assignments is course specific or applies across courses.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60,* 383-398. doi:10.1007/s11423-012-9235-8

Anthony, L., Yang, J., & Koedinger, K. R. (2008). Toward next-generation intelligent tutors: Adding natural handwriting input. *IEEE MultiMedia, 15*(3), 64-68. doi:10.1109/MMUL.2008.73

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologis*t, *23*, 225-239. doi:10.1037/h0020791

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. doi:10.1007/s11251-005-1272-9

Beal, C.R., & Cohen, P.R. (2008). Temporal data mining for educational applications. In T.B. Ho & Z.H. Zhou (Eds.), *PRICAI 2008: Trends in Artificial Intelligence. PRICAI 2008. Lecture Notes in Computer Science, 5351* (pp. 66-77). Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-89197-0_10

Blanc, P. (1999). Solving a non-routine problem: What helps, what hinders. *Proceedings of the British Society for Research into Learning Mathematics, 19*(2), 1-6.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252. doi:10.1207/s1532690xci1302_3

Blumner, H. N., & Richards, H. C. (1997). Study habits and academic achievement of engineering students. *The Journal of Engineering Education, 86*, 125-132. doi:10.1002/j.2168-9830.1997.tb00275.x

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade point average from the revised and recentered SAT® I: Reasoning Test* (College Board Rep. No. 2000-1). New York: College Entrance Examination Board.

Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Cambridge, MA: Harvard University Press.

Chi, M.T., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use example in learning to solve problems. *Cognitive Science, 13*(2), 145-182. doi:10.1207/s15516709cog1302_1

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research, 76,* 1-62. doi:10.3102/00346543076001001

Crooks, T.J. (1980). *Grade prediction: The usefulness of context-specific predictors.* (Report No. TM 800676). Washington, D.C.: U.S. Department of Health, Education, and Welfare; National Institute of Education. (ERIC Document Reproduction Service no. ED194547)

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4-58. doi:10.1177/1529100612453266

Ebbinghaus, H. (1964). Memory. New York, NY: Dover. (Original work published 1885)

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding.* New York: Cambridge University Press.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4)*,* 717-741. doi:10.1007/s10648-015-9348-9

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. (Research and Occasional Paper Series No. CSHE.6.07). University of California, Berkeley, Center for Studies in Higher Education.

Gyllen, J., Stahovich, T., & Mayer, R. (2018). *How students read an e-textbook in an engineering course*. *Journal of Computer Assisted Learning.* doi:10.1111/jcal.12277

Gyllen, J. Stahovich, T., Mayer, R., Entezari, N., & Darvishzadeh, A. (2018). Priming productive study strategies with preparatory quizzes in an engineering course. Manuscript submitted for publication.

Hall, R. (2000). Video recording as theory. In D. Lesh & A. Kelley (Eds.), *Handbook of research design in mathematics and science education* (pp. 647-664). Mahwah, NJ: Lawrence Erlbaum.

Hall, S., & Vance, E.A. (2010). Improving self-efficacy in statistics: role of self-explanation & feedback. *Journal of Statistics Education, 18*(3). doi:10.1080/10691898.2010.11889583

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Herold, J., & Stahovich, T. (2012). Characterizing students handwritten self-explanations. In *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

Herold, J., Stahovich, T. F., & Rawson, K. (2013). Using educational data mining to identify correlations between homework effort and performance. In *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Hills, J.R., Klock, J.A., & Bush, M (1965). The Use of Academic Prediction Equations with Subsequent Classes. *American Educational Research Journal, 2*(4)*,* 203-206. doi:10.3102/00028312002004203

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: a perspective for enhancing meaningful learning. Educational *Psychology Review, 24*(3), 401-418. doi:10.1007/s10648-012-9202-2

Keith, T. Z. (1982). Time spent on homework and high school grades: A large-scale path analysis. *Journal of Educational Psychology, 74,* 248-253. doi: 10.1037/0022-0663.74.2.248

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science, 6,* 333-353. doi:10.1002/wcs.1350

Krüger, A., Merceron, A., & Wolf, B. (2010). A data model to ease analysis and mining of educational data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining*, (pp. 131-140). Pittsburgh, PA.

Lee, C.-K., Stahovich, T., & Calfee, R. C. (2011). A pen-based statics tutoring system. In *Proceedings of the American Society for Engineering Education 118th Annual Conference and Exposition*, Vancouver, BC, Canada.

Li, N., Cohen, W. W., Koedinger, K. R., & Matsuda, N. (2011). A machine learning approach for automatic student model discovery. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, C., & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 31-40). Eindhoven, the Netherlands.

Lounsbury, J.W., Sundstrom, E., Loveland, J.L., & Gibson, L.W. (2003). Broad versus narrow personality traits in predicting academic performance of adolscents. *Learning and Individual Differences, 14*(1), 65-75. doi:10.1016/j.lindif.2003.08.001

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*, 588-599. doi:10.1016/j.compedu.2009.09.008

Mayer, R. E. (2019). *How to be a successful student.* New York: Routledge.
Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. In *33rd Annual Frontiers in Education, 2003. FIE 2003* (pp. T2A 13-18). Westminster, CO. doi:10.1109/FIE.2003.1263284

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science, 13*, 390-407. doi:10.1177/1745691617710510

Mostow, J., González-Brenes, J. P., & Tan, B. H. L. (2011). Learning classifiers from a relational database of tutor logs. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 149-158). Eindhoven, Netherlands.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25*, 1159-1168. doi:10.1177/0956797614524581

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. In UIST '06 - *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (pp. 191-200). New York, NY: ACM Press.

Passons, W.R. (1967). Predictive validities of the ACT, SAT and high school grades for first semester GPA and freshman courses. *Educational and Psychological Measurement, 27*(4), 1132-1144. doi:10.1177/001316446702700469

Pellegrino, J. W., & Hilton, M. L. (Eds). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Washington, DC: National Academics Press.

Picciano, A.G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20. doi:10.24059/olj.v16i3.267

Rawson, K., & Stahovich, T. F. (2013). Predicting course performance from homework habits. In *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Rawson, K., Stahovich, T. F., & Mayer, R. E. (2017). Homework and achievement: Using smartpen technology to find the connection. *Journal of Educational Psychology, 109*(2), 208-219. doi:10.1037/edu0000130

Robyak, J.E., & Downey, R.G. (1979). The prediction of long-term academic performance after the completion of a study skills course. *Measurement and Evaluation in Guidance*, *12*(2), 108-111. doi:10.1080/00256307.1979.12022191

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

Romero, C., Romero, J., Luna, J., & Ventura, S. (2010). Mining rare association rules from e-learning data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 171-180). Pittsburgh, PA.
Romero, C., Ventura, S., Vasilyeva, E., & Pechenizkiy, M. (2010). Class association rules mining from students' test data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining*, (pp. 317-318). Pittsburgh, PA

Sawyer, R., & Maxey, J. (1979). The validity of college grade prediction equations over time. *Journal of Educational Measurement, 16*(4), 279-284. doi: 10.1111/j.1745-3984.1979.tb00109.x

Shanabrook, D. H., Cooper, D. G., Woolf, B. P., & Arroyo, I. (2010). Identifying high-level student behavior using sequence-based motif discovery. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining*, (pp. 191-200). Pittsburgh, PA.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Stahovich, T., & Pilegard, C. (2018) *Measuring Note-Taking Behaviors to Predict Academic Achievement in an Engineering Course.* Manuscript submitted for publication.

Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: Development and psychometric analysis. *The Journal of Engineering Education, 94,* 363-371. doi:10.1002/j.2168-9830.2005.tb00864.x

Stevens, R., Johnson, D. F., & Soller, A. (2005). Probabilities and predictions: Modeling the development of scientific problem-solving skills. *Cell Biology Education, 4,* 42-57. doi:10.1187/cbe.04-03-0036

Thorndike, E.L. (1910). The contribution of psychology to education. *Journal of Educational Psychology, 1*(1), 5-12. doi:10.1037/h0070113

Trivedi, S., Pardos, Z. A., Sárközy, G. N., & Heffernan, N. T. (2011). Spectral clustering in educational data mining. In M. Pechenizkiy, T.Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 129-138). Eindhoven, Netherlands.

Van Arsdale, T., & Stahovich, T. (2012). Does neatness count? What the organization of student work says about understanding. In *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

van Gog, T. (2013). Time on task. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 432-433). New York: Routledge.

Wilson, B.C., & Shrock, S. (2001). Contributing to success in an introductory computer science course: a study of twelve factors In *Proceedings of the Thirty-second SIGCSE Technical Symposium on Computer Science Education. ACM SIGCSE Bulletin, 33,* 184-188. doi:10.1145/366413.364581

Wood, D., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wollman, W., & Lawrenz, F. (1984). Identifying potential "dropouts" from college physics classes. *Journal of Research in Science Teaching, 21*(4), 385-390. doi: 10.1002/tea.3660210406

Xu, J. (2013). Homework and academic achievement. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 199-201). New York, NY: Routledge.

Zollanvari, A., Kizilirmak, R. C., Kho, Y. H., & Hernández-Torrano, D. (2017). Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors. IEEE Access, 5, 23792-23802. doi:10.1109/ACCESS.2017.2740980

Chapter 4

Effects of Weekly Reports of Study Behavior in a College Course: Evidence for a Two-Stage Model of Metacognition

Abstract

What can be done to help college students improve their study behaviors and grades in a college course? To address this issue, we provided students in an introductory-level engineering course with weekly reports summarizing their study behaviors for the week, such as time spent on homework, percentage of homework problems attempted, percentage of homework completed early, whether high-quality lecture notes were submitted, total time spent reading, and percentage of assigned pages read (experimental group). Compared to a control group that did not receive weekly reports, the experimental group generally did not show improvements in measures of study behaviors such as the percentage of homework completed early, the percentage of homework problems attempted, the total time spent on homework, the number of pen strokes written for homework problems, the length of ink written for homework problems, the percentage of lectures for which notes were submitted, the total reading time, and the percentage of assigned pages read. Nor did the experimental group achieve better course grades than the control group. We conclude that monitoring one's study behavior is insufficient to improve learning outcomes. Instead, improving learning

outcomes requires that monitoring be coupled with regulating one's study behavior, that is, adjusting how one studies based on feedback.

Objective and Rationale

Students in college courses sometimes engage in sub-optimal study behaviors, such as massing their study time into one long session, cramming their study time within 24 hours of the deadline, and misjudging the amount of study time that is needed (Gyllen, Stahovich, & Mayer, 2018; Miyatsu, Nguyen, & McDaniel, 2018; Rawson, Stahovich, & Mayer, 2017; Stahovich, Gyllen, & Mayer, in press). These study behaviors are inconsistent with a growing body of research that has applied the science of learning to education to produce guidelines for how to study effectively (Brown, Roediger, & McDaniel, 2014; Dunlowsky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2015, 2016; Mayer, 2019; Miyatsu, Nguyen, & McDaniel, 2018; Rhodes, Cleary, & DeLosh, 2020).

What can be done to help guide college students to study more effectively in their courses? We take a cue from the popularity of wearable fitness tracking systems, such as the FitBit and Apple Watch, which provide users with information about their progress towards fitness goals by quantifying fitness metrics such as physical activity, heartrate, and weight (Neff & Nafus, 2016). The present study examines the effects of applying technology-supporting self-tracking to the study behavior of college students taking an introductory engineering course. Analogous to how fitness trackers quantify physical activity, we use technology to provide students in a college course with weekly reports on their study activities, including their effort on homework and on reading assignments.

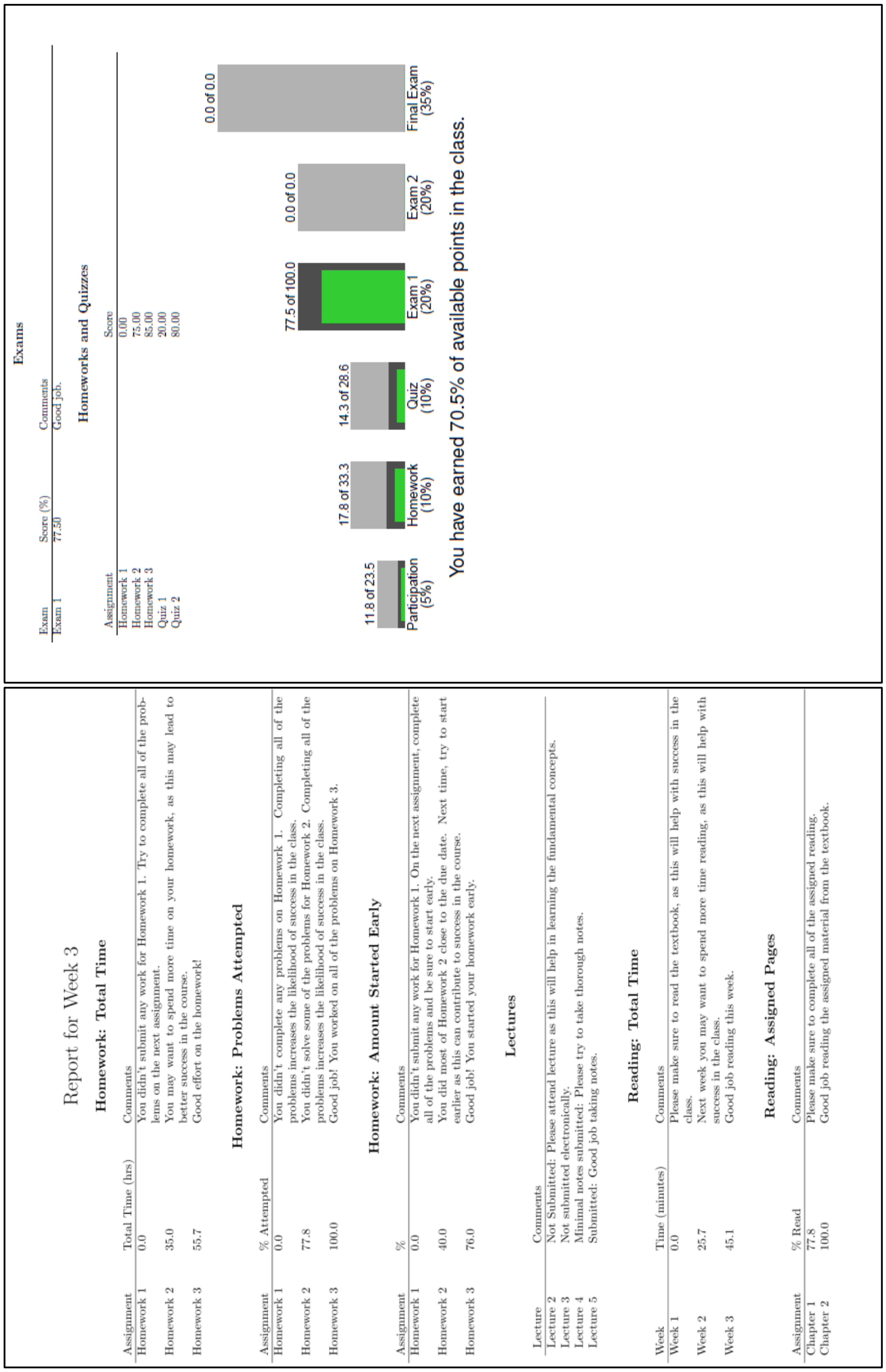
It is increasingly recognized that effective study skills are rarely taught, and that sub-optimal performance in college classes can be linked to sub-optimal students' study behaviors (Gyllen, Stahovich, & Mayer, 2018; Miyatsu, Nguyen, & McDaniel, 2018; Rawson, Stahovich, & Mayer, 2017; Stahovich, Gyllen, & Mayer, in press). In the present study, we examine the straightforward idea that regularly notifying students of deficiencies in their study activities during a course can help guide them to adjust their study activities and thereby improve their course grade.

Background on Self-Tracking

Wearable fitness tracking devices – a form of self-tracking – have become popular in an attempt to increase people's engagement in activities that promote health and fitness (Neff & Nafus, 2016). These devices objectively measure physical activity and provide the measurements to the user, often in relation to activity goals, such as a target number of footsteps to take each day. Research on the effectiveness of these devices has found some evidence that they can promote healthy lifestyles, but the results are mixed (Cadmus-Bertram, Marcus, Patterson, Parker, and Morey, 2015; Neff & Nafus, 2016; Schrager et al., 2017; Gordon and Bloxham, 2017). The present work examines an analogous approach to improving student learning. More specifically, we examine the idea that providing students with objective measurements of their study behavior, including homework and reading effort, will result in changes in their study behavior, and that that these changes will result in improved learning outcomes. We refer to this idea as the monitoring hypothesis.

This work builds upon recent research by Rawson, Stahovich, and Mayer (2017) that found that objective measures of students' homework effort in an introductory-level engineering course are positively and significantly correlated with achievement in the course. Students were given smartpens to complete their homework assignments. The devices recorded temporal-spatial information about each pen stroke the students produced while solving homework problems. The recorded pen strokes were then used to objectively measure student homework effort, including the total time spent on assignments, the amount of writing (including the number of strokes and the total amount of ink written), the number of problems attempted, and the fraction of work done within 24 hours of the due date. There were strong correlations across multiple course offerings between course grade and total time spent on homework (r = .44), the number of strokes written (r = .49), the amount of ink written (r = .42), the number of problems attempted (r = .45), and the fraction of work done near the due date (r = -.32).

In the present study, we provided students in a sophomore-level engineering course with weekly progress reports (Figure 16) describing their efforts in the class to date. The reports included objective measures of their homework effort, note-taking effort, and reading effort; qualitative feedback about their effort and suggestions for improving performance; and grades for individual deliverables and the course overall.



Report for Week 3

Homework: Total Time

Assignment	Total Time (hrs)	Comments
Homework 1	0.0	You didn't submit any work for Homework 1. Try to complete all of the problems on the next assignment.
Homework 2	35.0	You may want to spend more time on your homework, as this may lead to better success in the course.
Homework 3	55.7	Good effort on the homework!

Homework: Problems Attempted

Assignment	% Attempted	Comments
Homework 1	0.0	You didn't complete any problems on Homework 1. Completing all of the problems increases the likelihood of success in the class.
Homework 2	77.8	You didn't solve some of the problems for Homework 2. Completing all of the problems increases the likelihood of success in the class.
Homework 3	100.0	Good job! You worked on all of the problems on Homework 3.

Homework: Amount Started Early

Assignment	%	Comments
Homework 1	0.0	You didn't submit any work for Homework 1. On the next assignment, complete all of the problems and be sure to start early.
Homework 2	40.0	You did most of Homework 2 close to the due date. Next time, try to start earlier as this can contribute to success in the course.
Homework 3	76.0	Good job! You started your homework early.

Lectures

Lecture	Comments
Lecture 2	Not Submitted: Please attend lecture as this will help in learning the fundamental concepts.
Lecture 3	Not submitted electronically.
Lecture 4	Minimal notes submitted: Please try to take thorough notes.
Lecture 5	Submitted: Good job taking notes.

Reading: Total Time

Week	Time (minutes)	Comments
Week 1	0.0	Please make sure to read the textbook, as this will help with success in the class.
Week 2	25.7	Next week you may want to spend more time reading, as this will help with success in the class.
Week 3	45.1	Good job reading this week.

Reading: Assigned Pages

Assignment	% Read	Comments
Chapter 1	77.8	Please make sure to complete all of the assigned reading.
Chapter 2	100.0	Good job reading the assigned material from the textbook.

Exams

Exam	Score (%)	Comments
Exam 1	77.50	Good job.

Homeworks and Quizzes

Assignment	Score
Homework 1	0.00
Homework 2	75.00
Homework 3	85.00
Quiz 1	20.00
Quiz 2	80.00

Figure 16. An example of the weekly progress report.

We used smartpens to measure students' effort on homework and lecture notes, and used an instrumented document viewing program to measure reading effort. To determine the effects of the weekly progress reports on student learning behavior, we compared the students' learning behavior to that of students in a control group who took a different offering of the course. We also used smartpens and the instrumented document viewing program to measure the learning behavior of the students in the control group, but those students received no weekly progress reports.

Students who exhibit certain homework behaviors, such as spending sufficient time on assigned problems, attempting to complete all the assigned problems, and working on the assignment days before the due date instead of waiting until the proverbial last minute have a higher likelihood of doing well in a course than those who do not. Suppose we had an automated system that could inform students of their behaviors on homework assignments, and suggest changes to their behavior on subsequent assignments to improve their outcome in the course. In our present study, we developed such a system which provides auto-generated weekly reports that informed students of their behavior. We use this system to examine two hypotheses: 1) informing students of how their current study behaviors will motivate the students to change their behaviors on subsequent homework assignments and study tasks, and 2) influencing student study behavior will result in better course outcomes (as reflected in final course grade).

Homework problems are a ubiquitous part of college courses in science, technology, engineering, and mathematics (STEM). Rawson, Stahovich, and Mayer (2017) showed objective measures of students' total homework effort in an introductory-level engineering course are positively and significantly correlated with achievement in the course, and the objective measures from early in the course are nearly as predictive as those taken over the entire course (Rawson, Stahovich, & Mayer, 2018). This present study utilizes the same technology of Rawson et al. (2017) to digitally record student behavior during homework problem solving. We use smartpens to unobtrusively record the student's handwritten homework solutions as timestamped, individual digitized pen strokes. From the digitized homework, we report to the students objective measures of their behavior, and based on their effort, suggestions for improvement on subsequent homework assignments, and observe whether their behavior changes on the subsequent homework assignments.

Literature Review

Personality traits have been explored as factors in student performance. For example, Lounsbury, Sunstrom, Loveland, and Gibson (2002) examined work drive, general intelligence (measured via the Otis-Lennon test), and the Big Five personality traits, and found that they could explain up to 26.8% of the variance in a student's course grade. However, the Big Five personality traits did not add significantly to the prediction of course grade above and beyond work drive.

Incentives can serve as a powerful force for changing student behavior. Cullen, Cullen, Hayhow, and Plouffe (1975) found that high school students given additional grade points for completing homework assignments resulted in a higher completion of assignments. Tuckman (1998) analyzed the use of tests on textbook material as an incentive to motivate procrastinators. The students who were given this treatment scored significantly higher on a final achievement test compared to other students who were asked to outline the reading material for homework. Radhakrishnan, Lam, and Ho (2009) observed that university students with a higher incentive to complete homework performed better in the class than students with less incentive. Our work is aimed at influencing student's homework behavior without an explicit incentive.

Robyak and Downey (1979) examined the effectiveness of a study skills course and found a significant increase in academic performance for students one year after they took this course. Blumner and Richards (1997) examined the study habits of engineering students by measuring distractibility, inquisitiveness, and compulsiveness, and determined that more successful students tended to be less distractible and more inquisitive. Dunlosky, Rawson, Marsh, Nathan, & Willingham (2013) performed a detailed literature review of 10 learning techniques, and found distributed practice and practice testing having the highest relative utility across learning conditions, student characteristics, course material, and criterion tasks.

Research in which data recorded from students' learning activities is used to model student learning has a long and rich history. Such work dates back to large-scale studies of computer assisted instruction (CAI) in schools in the 1960s (Atkinson, 1968). More recently, researchers have made significant progress in educational data mining, a research methodology in which data mining techniques are used to extract models of student learning from large datasets (Koedinger, D'Mello, McLaughlin, Pardos, & Rosé, 2015; Romero, Romero, Luna, & Ventura, 2010). Much of the data used in this work is extracted from log files of intelligent tutoring systems (Beal & Cohen, 2008; Li, Cohen, Koedinger, & Matsuda, 2011; Mostow, González-Brenes, & Tan, 2011; Shanabrook, Cooper, Woolf, & Arroyo, 2010; Stevens, Johnson, & Soller, 2005; Trivedi, Pardos, Sráközy, & Heffernan, 2011) and learning management systems such as Moodle and Blackboard (Minaei-Bidgoli, Kashy, Kortenmeyer, & Punch, 2003; Krüger, Merceron, & Wolf, 2010; Picciano, 2012; Romero, Ventura, Vasilyeva, & Pechenizkiy, 2010).

In our work, we use data recorded from students' learning activities written on paper, rather than activities involving typing on a computer keyboard. Research by Oviatt, Arthur, and Cohen (2006) shows that natural work environments are critical to student performance. In their examinations of computer interfaces for completing geometry problems, they found that "as the interfaces departed more from familiar work practice…, students would experience greater cognitive load such that performance would deteriorate in speed, attentional focus, metacognitive control, correctness of problem solutions, and memory." Similarly, Anthony, Yang, and Koedinger (2008) found that handwriting interfaces were more beneficial than keyboard interfaces for math tutoring systems. Mueller and Oppenheimer (2014) made a similar finding in relation to note taking. They examined student note taking using both longhand and laptops, and found that the latter can lead to shallower processing. Lectures were shown on a screen, with students taking notes, followed by distractor tasks. Using a model including both word count and verbatim overlap (three-word chunks from student notes matching the lecture transcript), they were able to predict performance on a test of the lecture material with a correlation coefficient of $r = .41$.

Researchers have only recently begun using smartpens for assessment. For example, Herold and Stahovich (2012) used smartpens to examine the homework of students who were asked to provide self-explanations for their solutions to statics problems. The study found that students who generated self-explanations were more likely to complete homework problems in the order assigned (i.e., complete one problem before beginning the next) than were students who did not generate self-explanations.

Van Arsdale and Stahovich (2012) demonstrated that the temporospatial organization of a student's solution to an engineering problem is indicative of the correctness of that solution. They recorded students' work on exam problems using smartpens and characterized the problem-solving activity in terms of the sequence of problem-solving steps and the arrangement of work on the page. While they focused on a microscale analysis of problem-solving behavior on individual exam problems (taking minutes), we consider a macroscale analysis of homework habits (from days to over a week) when providing suggestions to the student in weekly progress reports.

Herold, Stahovich, and Rawson (2013) used smartpens to examine the correlation between effort on a homework assignment and grade on that particular assignment. They measured effort on a per-problem basis and its relationship with performance on subsequent, related homework, quiz, and exam problems. They characterized effort by the amount of time the pen was in contact with the paper, which is only a small fraction of the time spent on a problem. They found that this “writing time” was correlated with performance on subsequent problems. Our work is similar in that we calculate measures of time spent on homework, but our focus is on providing suggestions to the student to modify their behaviors towards more positive outcomes, rather than course grade prediction.

Rawson and Stahovich (2013) used smartpens as part of a method for making early predictions of student success or failure in a statics course. The method examined digital ink data captured with smartpens from a homework assignment and a corresponding quiz given early in the course. To make predictions, they constructed a set of features that included measures extracted from the digital ink, scores on the homework and quiz, a force concept inventory score (Hestenes, Wells, & Swackhamer, 1992), and self-reported measures of how much written work was produced directly versus being copied from scratch work. They then used these features to train regression models to predict course grade. These models achieved a correlation coefficient of r = .66.

Table 7

Two Metacognitive Processes Related to Study Behavior

Process	Description	Example
Monitoring	Awareness and evaluation of study behavior	I see that I am not studying until the day the assignment is due
Regulating	Adjustment of study behavior	I will study earlier in the week

Two-stage Model of Metacognition for Study Behaviors

Building on contemporary conceptions of metacognition (Azevedo & Aleven, 2013; Mayer, 2011), Figure 17 presents a two-stage model of metacognition for study behaviors. On the objective side, students engage in study behaviors and perform on learning outcome tests. On the internal side, students may monitor their study behaviors (i.e., become aware of what they are doing and how useful it is) and regulate their study behaviors (i.e., plan to change how they study). Students may not improve their learning outcome because they are unaware of the impact of their study behaviors (i.e., they have problems with monitoring) or because they do not change their study behaviors even when they are aware (i.e., they have problems with regulating).

Table 7 summarizes the two metacognitive processes of monitoring and regulating. Monitoring involves awareness and evaluation of one's study behavior, whereas regulating involves adjusting one's study behavior based on that information.

Providing weekly reports of study activities is intended to foster the process of monitoring by making the learner's study behavior more visible to the learner. Figure 18 shows what happens when we add weekly reports of study activities, which are

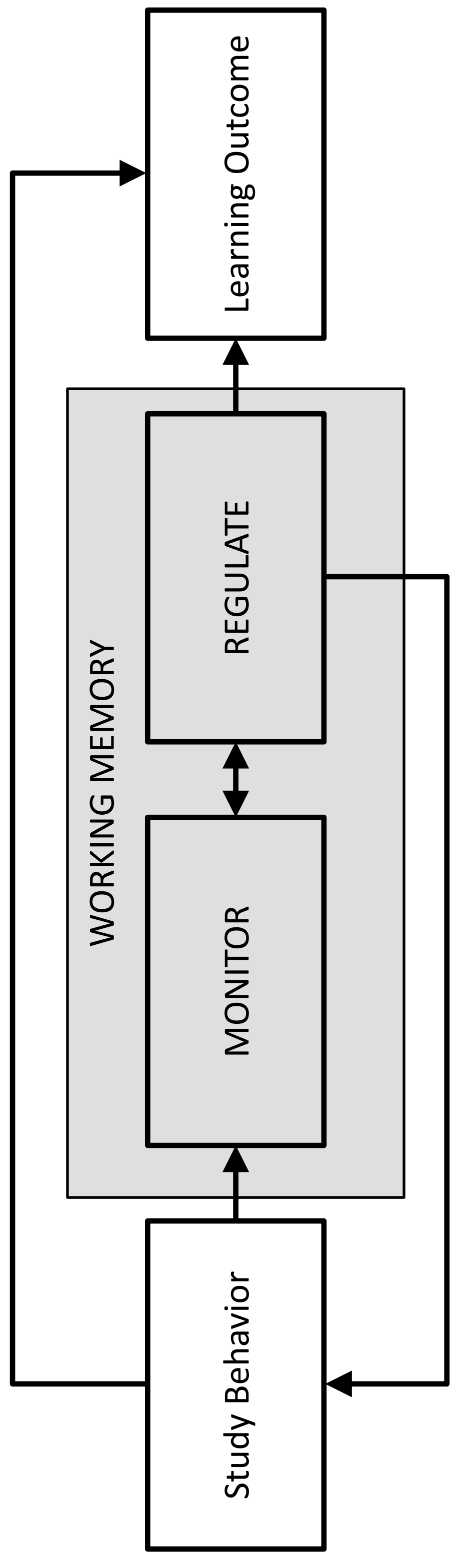


Figure 17. A two-stage model of metacognition for study behaviors.

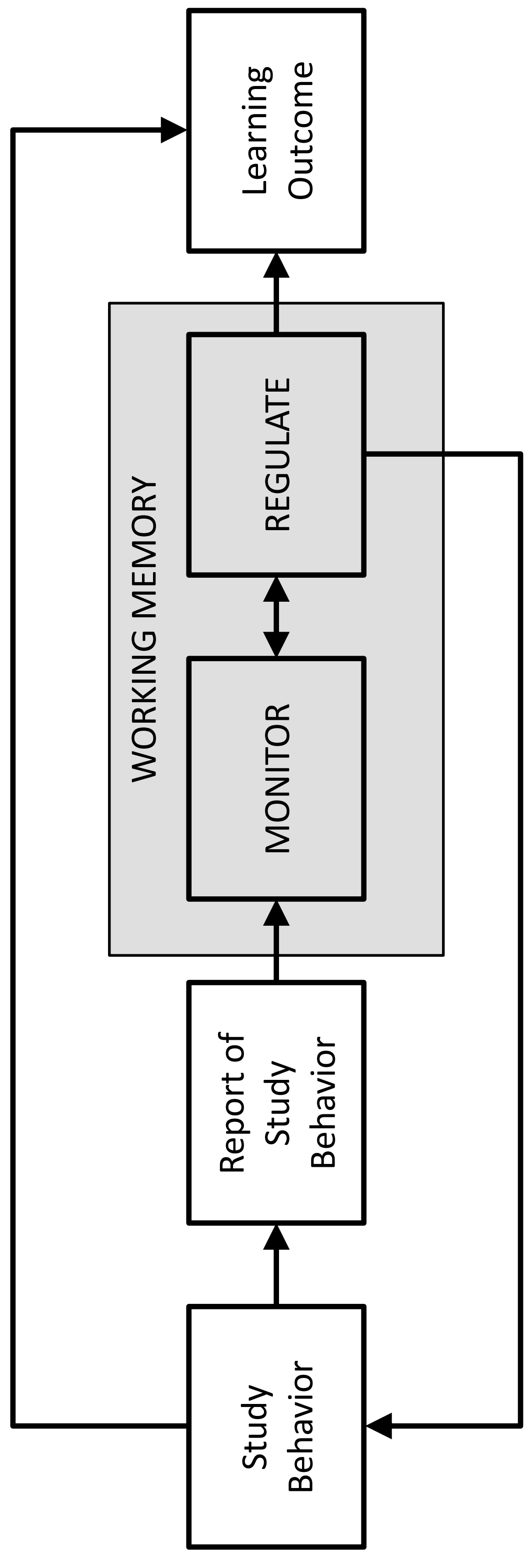


Figure 18. A model of how weekly reports affect metacognition for study behaviors.

intended to mainly affect the metacognitive process of monitoring. If monitoring is necessary and sufficient for improving learning, then adding weekly reports should improve study behaviors and learning outcomes, as described below as the monitoring hypothesis.

Monitoring hypothesis. Students may not study appropriately because they are unaware of their study behaviors. If students are made aware, they will improve their study behaviors and thereby improve course grade. According to the monitoring hypothesis, providing weekly reports should make students aware of their study behaviors, and thereby result in improvements in study behavior and course grades.

In contrast, if monitoring is necessary but not sufficient for improving learning, then adding weekly reports should not improve study behaviors and learning outcomes, as described below as the regulation hypothesis.

Regulation hypothesis. Students may not study appropriately because they do not know when and how to regulate their study behaviors. Even if students are aware of their study behavior, they do not know how to interpret and use information to adjust their behavior. If students are made aware, they will still not improve their study behaviors and course grade. According to the regulation hypothesis, providing weekly reports may make students aware of their study behavior, but will not necessarily cause a change in study behavior or course grades. If students do not adjust their study behaviors based on the weekly reports, they will not improve on learning outcomes.

Providing feedback is a fundamental instructional tactic in psychology and education (Johnson & Priest, 2014; Mayer, 2011; Shute, 2008), and is consistent with Hattie's (2009) admonition to make learning visible to the learner. Our weekly reports provide feedback concerning what kinds of study behaviors the learner is engaging in and thereby makes learning efforts visible to the learner. The key issue addressed in this study is whether making study behaviors more visible to the learner - mainly supporting the metacognitive process of monitoring - is sufficient to affect the learner's study behavior and learning outcomes.

We examined these competing predictions by comparing the study behaviors and course grades of students in an engineering course who received weekly reports of their study behavior against students who did not receive weekly reports.

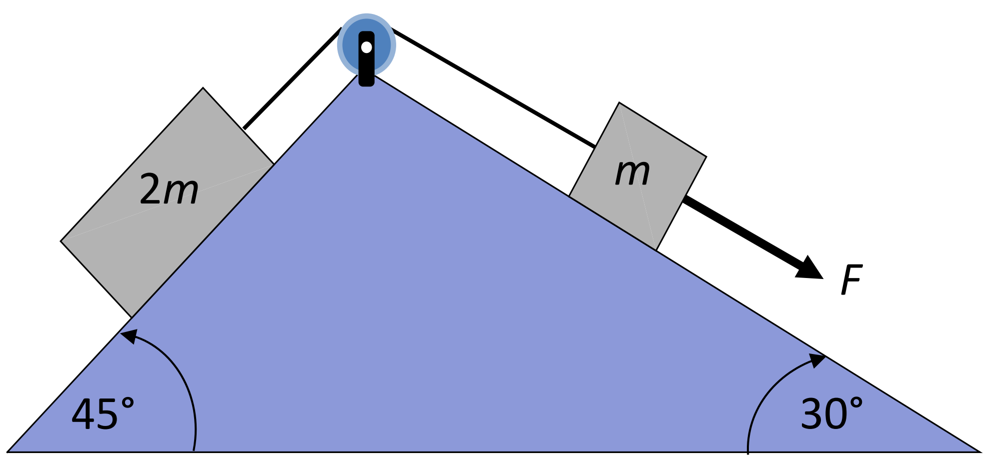


Figure 19. A typical statics problem. In this example, the student must solve for the minimum and maximum force *F* that will keep the two blocks with mass *m* and 2*m* from moving, given the blocks are attached together by a rope and the coefficient of static friction between the blocks and the inclined plane is μ_s.

Method

Subjects and Design

The participants in this study were 135 students who completed a lower division statics course at a university in southern California during the winter quarter of 2015 (control group) and 65 students who completed the same course during the spring quarter of 2016 (experimental group). This study used a quasi-experimental design in which the experimental group received weekly reports of their study behavior and the control group did not. The courses were equivalent in content and exams, and had the same instructor.

The Course Context

Statics is the part of engineering mechanics focused on the equilibrium of objects subject to forces and moments. Topics include force systems, equilibrium of two- and three-dimensional structures, and equilibrium of frames and machines. Figure 19 shows a problem typical of those solved on homework assignments, quizzes, and exams. To solve such problems, a student must draw diagrams (i.e., free body diagrams), write force and moment equilibrium equations, and solve the equations for the unknowns. Statics is a required course for some engineering majors, especially Mechanical Engineering. The particular statics course used in the present study employed a traditional lecture format, meeting on a Tuesday-Thursday schedule, with two 80-min lectures per week and an additional 50-min weekly discussion section.

Students used smartpens to complete the homework assignments, quizzes, and exams. Students also used them to take lecture notes, which were included in the course grade as class participation. Smartpens are ink pens that digitize the writing as it occurs. The smartpens are used with a special dot-patterned paper and create two records, ink on paper and timestamped digitized pen strokes. Students used a computer program called InkViewer to submit the digitized ink for their coursework so that it could be graded. We used the digitized ink data to analyze student homework activities.

Students used an instrumented document viewing program called DocViewer (Gyllen, Stahovich, & Mayer, 2018) to read encrypted course documents including the syllabus, textbook, handouts, the instructor's lecture notes, assigned homework problems, graded work, and weekly progress reports (e.g., Figure 16). DocViewer allows measurement of student reading by recording time-stamped events comprising activities such as opening a document, changing the currently displayed page, zooming on the page, searching, scrolling, or dimming of the screen as a result of lack of interaction with the program. The screen dims after a preset duration of two minutes of inactivity so as to help determine when the student is actively engaged with the reading. Once the screen dims, the student can "wake" the screen by interacting with the program.

Both the experimental and control group offerings of the statics course were taught by the same instructor and covered the same material with the same lecture content. Both offerings employed the same weights for computing final course grade:

5% for taking lecture notes, 10% for the homework, 10% for quizzes, 20% for the first midterm exam, 20% for the second midterm exam, and 35% for the final exam. Students in the control offering recorded all of their coursework with smartpens and used DocViewer for reading course materials just as the students in the experimental offering did.

The homework assignments in the two offerings were similar to enable meaningful comparison. Homework assignments 1, 3, 6, 7, and 8 were identical between the two offerings. For the control offering, the second assignment had nine problems, but only eight of these were assigned for the experimental offering. Conversely, for the experimental offering, the fourth assignment included the five problems from the control offering plus two additional problems. For both offerings, the fifth assignment had five problems, but these differed between the two offerings, although they covered similar material. The ninth assignment had completely different problems and subject matter between the offerings. The experimental offering had an additional tenth assignment, but students were required only to complete a course survey and a statics concept inventory (Steif & Dantzler, 2005) to receive credit, and did not submit any statics problems utilizing the smartpens.

Weekly Progress Reports

Starting in week three, students in the experimental offering received weekly progress reports describing their efforts in the class to date. The reports for a given week were prepared and distributed to the students as soon as the prior week's homework was graded. For example, homework assignment three was assigned in the third week of the course and submitted in the fourth week. Thus, the progress report for the fourth week included feedback on the first three assignments, as well as feedback on all other coursework that had been completed by that point. As shown in Figure 16, the reports provided students with measures of their homework effort, lecture note-taking effort, reading effort, and grades.

Table 8
Feedback Related to Total Time Spent on Homework

Total Time (hrs)	Feedback
TT = 0	You didn't submit any work for Homework *X*. Try to complete all of the problems on the next assignment.
TT < HT	You may want to spend more time on your homework, as this may lead to better success in the course.
TT ≥ HT	Good effort on the homework!

Homework X is a specific assignment number (e.g., "Homework 3"), TT is the student's Total Time spent on Homework X, and HT is the class mean Total Time for Homework X.

Homework effort. Students were provided with three measures of homework effort: *Total Time*, *Problems Attempted*, and *Proportion Started Early*. *Total Time*, as the name suggests, is the total time spent by a student to complete a homework assignment. This is calculated as the time from the first pen stroke written until the last, excluding any intervals of 10 or more minutes during which no writing occurred. The weekly reports listed the *Total Time* for each assignment in units of hours. The weekly reports also provided qualitative feedback based on the *Total Time*. As there is no a priori method to determine how long it should take to complete an assignment, we used the class mean for an assignment as a reference to determine the appropriate feedback. Students were not informed that the class mean was used as the reference. As shown in Table 8, if a student's *Total Time* for an assignment was at least as large as the class mean for that assignment, the student received praise indicating that he or she had made "Good effort on the homework!" Conversely, students who did some work, but spent less time than the class mean were instructed to spend more time on homework "as this may lead to better success in the course." Likewise, students who submitted no work were instructed to "Try to complete all of the problems for the next assignment." Each report was cumulative, detailing the total time spent for each homework assignment that had been submitted thus far.

Table 9
Feedback Related to Problems Attempted on Homework

Problems Attempted	Feedback
PA = 0%	You didn't complete any problems on Homework *X*. Completing all of the problems increases the likelihood of success in the class.
0% ≤ PA < 100%	You didn't solve some of the problems for Homework *X*. Completing all of the problems increases the likelihood of success in the class.
PA = 100%	Good job! You worked on all of the problems on Homework *X*.

Homework X is a specific assignment number (e.g., "Homework 3"). PA is the percentage of assigned problems that the student attempted on Homework X.

We define *Problems Attempted* as the number of problems on an assignment for which the student made at least some minimum level of effort as indicate by writing at least 50 pen strokes. As simply writing “Problem 1” takes at least eight pen strokes, a threshold of 50 strokes corresponds to only a small amount of work. The weekly reports listed the *Problems Attempted* for each assignment as a percentage. Furthermore, as shown in Table 9, students received praise for working on all problems, stating “Good job! You worked on all of the problems on Homework *X*.” Conversely, if a student did not attempt all of the problems on an assignment, or attempted none of them, he or she was encouraged to do so on the next assignment and was informed that “Completing all of the problems increases the likelihood of success in the class.”

Table 10
Feedback Related to Early Work on Homework

Early Work	Feedback
EW = Ø	You didn't submit any work for Homework *X*. On the next assignment, complete all of the problems and be sure to start early.
EW < 50%	You did most of Homework *X* close to the due date. Next time, try to start earlier as this can contribute to success in the course.
50% ≤ EW ≤ 100%	Good job! You started your homework early.

Homework X is a specific assignment number (e.g., “Homework 3”). EW is the percentage of work done for Homework X at least 24 hours before the due date.

The final measure of homework effort is the *Proportion Started Early*. (In the weekly reports, this was called "*Amount Started Early.*" However, for clarity in the present analysis we use the term "*Proportion Started Early.*") This is calculated as the percentage of the work on a homework assignment – measured in terms of the number of pen strokes – that was completed at least 24 hours before the due date. For example, if 35% of the pen strokes for an assignment were written 24 or more hours before the due date, the *Proportion Started Early* would be 35%. The reports list the *Proportion Started Early* for each assignment as a percentage. The reports also contained feedback encouraging students to complete each assignment early. As shown in Table 10, students who completed at least 50% of an assignment early received praise stating "Good job! You started your homework early." Conversely, students that submitted the homework but completed less than 50% early were informed that they did most of the assignment close to the due date, were encouraged to start the next assignment early, and were informed that "this can contribute to success in the course." Finally, students who submitted no work were encouraged to complete all of the problems on the next assignment and to begin working on it early.

Effort on lecture notes. Students received class participation credit by submitting their lecture notes. They were instructed to write their notes with their smartpens and to submit their notes using InkViewer in the same way they submitted other course work. To receive class participation credit for a particular lecture, we

Table 11
Feedback Related to Lecture Notes

Notes Submitted	Feedback
None	Not Submitted: Please attend lecture as this will help in learning the fundamental concepts.
Paper copy	Not submitted electronically.
Below Threshold	Minimal notes submitted: Please try to take thorough notes.
Meet Threshold	Submitted: Good job taking notes.

required students to submit 200 or more pen strokes with at least 80% of the writing occurring within the lecture period.

As shown in Table 11, each weekly report provided feedback for all prior lectures. If the student submitted notes meeting the requirements for credit, the report indicated that the notes had been submitted and that the student did a “Good job taking notes.” If the submitted notes did not meet the requirements for credit, the report indicated that minimal notes were submitted and encouraged the student to “Please try to take thorough notes.” If a student did not submit notes for a particular lecture, the report indicated that none were submitted and the student was encouraged to “Please attend lecture as this will help in learning the fundamental concepts.” On rare occasions, some students forgot to turn on their smartpens, or had hardware issues during lecture. In these situations, students could submit the paper copy of their lecture notes to the teaching assistant. The student then received credit for taking those notes and the report indicated that they were "Not submitted electronically."

Students were encouraged to submit their lecture notes each week with the weekly homework assignments. However, students could submit their notes any time before the end of the quarter. (Because the smartpens record the pen strokes with timestamps, it was possible to determine when the notes were written even if they were submitted at the end of the quarter.) If notes for a past lecture were not yet submitted, the weekly reports listed them as such. Once the notes were submitted, the status was updated in subsequent reports. Notes were not required for the first class period of the quarter and for the two periods with midterm exams.

Reading effort. Students were provided with two measures of effort on reading the textbook: the *Total Time* spent reading each week was listed on all progress reports and, starting in the fifth week, the reports also listed the fraction of *Assigned Pages* read. We computed both measures from the reading data collected by the DocViewer program that students used to read the course materials. In computing these measures, we define a reading episode as an interval of time during which a particular textbook page is continuously displayed with DocViewer for at least 15 seconds but no more than an hour. If the interval was less than 15 seconds, we considered that to be page navigation rather than reading and discarded that time. Ordinarily, DocViewer's timeout function dimmed the screen after a two minute period of inactivity, indicating that the student was no longer interacting with the program. Thus, if a student left the program unattended, the reading episode would be automatically terminated. However, during the experiment we discovered that if the student were to put his or her computer in standby mode with the DocViewer program running (or terminate the program at the operating system level), the inactivity timeout might not be recorded. As a remedy, in processing the data, if an episode exceeded an hour, we assumed that the computer was in standby mode and discarded that time. We computed the total reading time as the sum of all of the valid reading episodes (i.e., episodes between 15 seconds and one hour in duration). Likewise, we considered a page to have been read if there was at least one valid reading episode for that page.

The progress reports listed the *Total Time* spent reading for each week in units of minutes. The reports also provided qualitative feedback about the reading time. To determine the appropriate feedback, we had to determine a suitable minimum reading time for each week. We started by computing the class mean reading time for the week and then adjusted the value if necessary. We believe that, on average, students should spend at least two minutes reading each assigned page. Thus, if the weekly class mean was less than this, we took two minutes per assigned page as the minimum acceptable reading time for the week. Conversely, we believe that spending an average of four minutes per page is acceptable. Thus, if the class mean for the week exceeded this, we took four minutes per assigned page as the minimum acceptable reading time. Otherwise, we used the class mean as the minimum acceptable reading time. We computed the minimum acceptable reading time in this way for all weeks except for the

Table 12
Feedback Related to Total Time Spent Reading

Total Time (minutes)	Feedback
T = 0	Please make sure to read the textbook, as this will help with success in the class.
T < RT	Next week you may want to spend more time reading, as this will help with success in the class.
T ≥ RT	Good job reading this week.

T is the student's reading time for the week and RT is the minimum acceptable reading time for the week.

two weeks with midterms. For the week of the first midterm, we set the minimum acceptable reading time to be 100 minutes as we expected students to reread the textbook to prepare for the exam. For the week of the second midterm, we took the minimum acceptable reading time to be the mean for the week.

As shown in Table 12, students whose *Total Time* reading exceeded the minimum acceptable reading time (RT) for the week received praise stating "Good job reading this week." Students who did some reading but less than the minimum acceptable amount were told "Next week you may want to spend more time reading, as this will help with success in the class." Finally, students who did not read at all during the week were told "Please make sure to read the textbook, as this will help with success in the class."

Starting in the fifth week, the progress reports listed the percentage of *Assigned Pages* read each week. As shown in Table 13, students who read all of the assigned pages received praise stating “Good job reading the assigned material from the textbook.” Students who did not read all of the assigned pages were encouraged to “complete all of the assigned reading.”

Table 13
Feedback Related to Assigned Pages of Reading

% Read	Feedback
RA < 100	Please make sure to complete all of the assigned reading.
RA = 100	Good job reading the assigned material from the textbook.

RA is the percentage of assigned pages read by the student.

Table 14
Feedback Related to Midterm Exam Scores

Score (%)	Comments
S < 66	It is recommended that you gain substantial additional practice in problem solving to better prepare for the next exam. You may also want to meet with Professor <X> to discuss things you can do to improve your performance in the class.
66 ≤ S < 74	It is recommended that you gain additional practice in problem solving to better prepare for the next exam.
74 ≤ S < 86	Good job.
S ≥ 86	Excellent job!

S is the exam score.

Feedback about grades. The progress reports included the grades on each individual homework assignment, quiz, and exam. As shown in Table 14, the reports also provided qualitative feedback about students' performance on the exams. Students who received an A+, A, or A- (grade ≥ 86%) were informed that they had done an "Excellent Job." Students who received a B+, B, or B- (i.e., 74% ≤ grade < 86%) were informed that that had done a "Good job" on the exam. Students who received a C+, C, or C- (i.e., 66% ≤ grade < 74%) were encouraged to "gain additional practice in problem solving to better prepare for the next exam." Finally, students who received a D+ or lower (i.e., grade < 66%) were given strong feedback stating "It is recommended that you gain substantial additional practice in problem solving to better prepare for the next exam. You may also want to meet with Professor <X> to discuss things you can do to improve your performance in the class."

After receiving feedback about exam scores on the Week 6 report, students also requested that the weekly reports include their grades on homework and quizzes. This was added to the weekly report starting with the Week 8 report.

The final element of the weekly progress reports was a graphical representation of the student's grade to date. As shown in Figure 16, each component of the grade, including class participation, homework, quizzes, the first midterm exam (i.e., Exam 1), the second midterm exam (i.e., Exam 2), and the final exam is represented by a shaded bar. The relative heights of the bars indicate the relative weights of the various components. For example, the bar for homework, which comprises 10% of the final course grade, is twice as tall as the bar for class participation, which comprises only 5% of the final grade. The relative weights of the various grade components are also given numerically below each bar.

The bars are shaded to indicate how much of the various grade components have been completed to date and how much of the possible credit has been achieved. Portions of a bar shown in light gray represent work that had not yet been assigned or graded. In Figure 16, for example, because the final exam had not yet been completed, the entire bar was shaded in light gray. Likewise, 66.7% of the homework bar is shaded light gray because at that point in the course, six of the nine homework assignments remained to be graded. The green portion of a bar represents the fraction of the possible points the student received. For the homework bar in Figure 16, for example, the area of the green region is 53.3% of the non-light-gray region, indciating that the student received 53.3% of the possible homework points thus far in the course. By comparing the relative sizes of the shaded regions the student can quickly determine how much of the coursework has been completed and how much of the possible credit was received. The numbers above each column also report the precise numerical values. For example, the numbers above the homework bar in Figure 16 indicate that 33.3 percentage points were available and the student received 17.8 percentage points, which is 53.3% of the available points. The caption below the chart informs the student of the overall percentage of the available points achieved to date. For example, in the sample report in Figure 16, the student has received 70.5% of the available points which is computed by summing the product of the points achieved and the weights for each grade component.

Procedure

The students received smartpens and dot-patterned notebooks at the beginning of the term and used them for all of the homework assignments, exams, quizzes, and lecture notes. They used the InkViewer software to submit the digital copies of this work. The homework, quizzes, and exams were graded and returned electronically. Students then used the DocViewer software to view their graded work.

Starting just before they submitted their third homework assignment, students in the experimental group received a weekly progress report. This first report included information on work up through and including the second homework assignment. As this report was distributed just prior to the due date for the third homework assignment, any impact on student behavior would not be expected until the fourth homework assignment.

Similarly, for the remainder of the course, a progress report was provided each week with the exception of the ninth week. The second midterm exam was given during that week and the teaching staff was unable to finish grading all of the submitted work in time to produce a progress report. Students received the last progress report during finals week. This report included all course work up through and including the work of the tenth week, which was the final week of instruction. Students in the control did not receive weekly reports.

Table 15
List of Weekly Reports Provided to Students in the Experimental Group

Week	Work Included	Percentage of Students Reading Report within a Week of Distribution
1	No Report	N/A
2	No Report	N/A
3	H1-2, L2-4	49.2
4	H1-3, L2-6,	66.2
5	H1-4, L2-8	64.6
6	H1-5, L2-9, E1	73.8
7	H1-6, L2-9,11-12, E1	58.5
8	H1-7, L2-9,11-14, E1, Q1-5	53.8
9	No Report	
10	H1-8, L2-9,11-16,18, E1-2, Q1-5	75.4
11 (Finals)	H1-9, L2-9,11-16,18-20, E1-2, Q1-7	7.7

H corresponds to homework assignments, L corresponds to lectures, E corresponds to exams, and Q corresponds to quizzes (e.g., H1-4 is the first four homework assignments). Students did not take notes during Lectures 10 and 17, as the midterm exams were given during these lecture periods.

Results

Did Students Read the Progress Reports?

A preliminary issue concerns whether students in the experimental group actually consulted the weekly progress reports. Table 15 describes the set of weekly reports that were provided to the students in the experimental group. As shown in the table, with the exception of the last weekly report, roughly 50% to 75% of students read each report within a week of distribution. Most students did not read the final progress report, as it was provided during the week of the final exam. Only 7.7% of the students read it, with one student reading it after the final exam.

Very few students read all of the reports within a week of distribution. As shown in Table 16, the majority (84.6%) of students read at least one of the reports within a week of distribution, while 15.4% of students read no reports within a week. Over half of the students (58.5%) read at least five reports within a week of distribution and about a quarter (23.1%) read at least seven of the possible eight within a week. Overall, we conclude that students in the experimental group did not examine all of the reports on a consistent basis.

Table 16
Weekly Reports Read and Course Statistics

Number of Weekly Reports Read Within a Week	Percentage of Students	Course Grade M	Course Grade S.D.
0	15.4	.53	.24
1 or more	84.6	.68	.16
5 or more	58.5	.67	.16
7 or more	23.1	.73	.11

Table 16 also shows the average course grade disaggregated by the number of reports read within a week of distribution. Students who read more reports tended to have higher grades than those who did not, although no causality can be inferred.

Do the weekly reports change homework behavior?

To examine if the weekly reports changed student behavior, we compare measures of homework effort between the experimental and control groups of the course. We measure the effort using the three measures included in the weekly reports: *Proportion Started Early*, *Problems Attempted*, and *Total Time*. We also consider two other measures not included in the reports: *Stroke Count* and *Length of Ink Written*. The former is the total number of pen strokes written for an assignment, while the latter is the distance travelled by the pen when in contact with the paper as measured in inches. To enable meaningful comparison between the two groups, we normalize *Total Time*, *Stroke Count*, and *Length of Ink Written* by the number of problems on the assignments, thus accommodating the variation between groups in the number of assigned problems on some assignments. (*Proportion Started Early* and *Problems Attempted* are inherently normalized.) In makings these comparisons, we use a two-tailed *t*-test assuming unequal variances and use 5% as the significance threshold.

Proportion started early. Table 17 shows the mean and standard deviation of the *Proportion Started Early* for both the experimental and control groups for the eight homework assignments we consider in our analysis. (Recall that we exclude the ninth assignment as the content differed between the two courses.) The differences between

Table 17
Comparison of Proportion Started Early Between the Experimental and Control Group

Assignment	Experimental Group		Control Group		
	M	SD	M	SD	*p*
1	.248	.344	1.000	.001	< .001
2	.300	.385	.819	.285	< .001
3	.289	.387	.318	.341	.617
4	.329	.400	.266	.346	.303
5	.279	.415	.285	.402	.924
6	.189	.252	.465	.353	< .001
7	.267	.363	.398	.398	.029
8	.331	.367	.187	.297	.011
1-8	.279	.246	.469	.194	< .001
4-8	.267	.252	.337	.254	.071

the two courses are significant for only five of the eight assignments, but the results cast doubt on the hypothesis that the weekly progress reports influence student behavior. For assignments 1 and 2, the means for the experimental offering are significantly smaller than for the control offering ($p < .001$). However, we believe that this was due to an error in the homework prompts for the control offering which erroneously indicated a due date that was earlier than the actual due date, thus causing many of the students to begin working earlier than they otherwise would have. The mean *Proportion Started Early* for assignments 6 ($p < .001$) and 7 ($p = .029$) are also significantly smaller for the experimental offering than for the control offering. Assignment 8 is the only one for which the mean is significantly (p = .011) greater for the experimental offering than for the control offering.

The penultimate row of Table 17 includes the mean and standard deviation of the *Proportion Started Early* for the eight assignments combined for each group. The mean for the experimental offering is significantly smaller ($p < .001$) than for the control offering. However, this difference is difficult to interpret. As described above, the first two homework assignments are somewhat anomalous for the control offering. Furthermore, the first three assignments for the experimental offering were not influenced by the progress reports; the fourth was the first that could have been influenced. To provider a better comparison, the last row of Table 17 includes the mean and standard deviation of the *Proportion Started Early* for assignments 4 - 8 combined for each offering. These five assignments are comparable (similar content) and had the potential to be influenced by the reports (they were assigned after students in the experimental offering had received prior reports). Here too, the mean is smaller for the experimental offering than for the control offering, although the difference is non-significant (p = .071)

Problems attempted. Table 18 shows the mean and standard deviation of *Problems Attempted* for both the experimental and control groups for the eight homework assignments individually, for all eight assignments combined, and for assignments 4 – 8 combined. For all cases, the differences between the means are non-significant.

Table 18
Comparison of Problems Attempted (Normalized) Between the Experimental and Control Group

Assignment	Experimental Group		Control Group		
	M	S.D.	M	S.D.	*p*
1	.859	.259	.933	.184	.060
2	.960	.112	.946	.139	.439
3	.894	.229	.950	.130	.079
4	.962	.134	.973	.098	.609
5	.963	.122	.956	.165	.743
6	.839	.243	.865	.209	.471
7	.864	.266	.842	.263	.600
8	.860	.208	.835	.272	.512
1-8	.825	.238	.840	.180	.668
4-8	.650	.189	.617	.226	.288

Total time. Table 19 shows the mean and standard deviation of normalized *Total Time* for both the experimental and control groups for the eight homework assignments individually, for all eight assignments combined, and for assignments 4 – 8 combined. The normalized *Total Time* is significantly different only for homework assignment 1 (p = .038). The students in the experimental group spent, on average, less time on this assignment than did the students in the control group. However, we believe this is a result the due date issue described above. If students in the control group started early because of an erroneous due date, it is not surprising that they had more time to spend on the assignment. For all eight assignments taken together and for the combined set of assignments 4 – 8, the differences are also non-significant.

Table 19
Comparison of Total Time (Normalized) Spent on Homework Between the Experimental and Control Group

Assignment	Experimental Group		Control Group		
	M	S.D.	M	S.D.	*p*
1	.214	.130	.261	.162	.038
2	.290	.137	.319	.140	.175
3	.293	.163	.336	.160	.092
4	.362	.168	.396	.190	.221
5	.264	.135	.294	.151	.167
6	.513	.304	.531	.366	.721
7	.419	.260	.418	.295	.978
8	.490	.239	.461	.329	.525
1-8	.323	.158	.337	.173	.572
4-8	.317	.162	.306	.213	.700

Stroke count. The progress reports did not present the number of pen strokes written on each assignment and thus students received no feedback about this measure of homework effort. Here we examine if this measure was affected by the progress reports. More specifically, we compare the number of pen strokes written, normalized by the number of assigned problems, between the experimental and control groups. Table 20 shows the mean and standard deviation of the normalized number of strokes for the eight homework assignments individually, for all eight combined, and for assignments 4 – 8 combined. For four of the first five assignments (assignments 1, 3, 4, and 5), the normalized *Stroke Count* is significantly less for the experimental offering. For the other assignments, the differences are non-significant. Likewise, for all eight

Table 20
Comparison of Stroke Count (Normalized) Between the Experimental and Control Group

Assignment	Experimental Group		Control Group		
	M	S.D.	M	S.D.	*p*
1	225.405	121.846	281.498	134.027	.007
2	371.404	144.922	394.943	164.781	.322
3	343.253	152.284	413.064	177.31	.006
4	395.013	157.409	454.241	185.895	.027
5	360.000	136.668	422.641	162.483	.006
6	643.481	362.996	649.367	401.895	.921
7	511.742	265.935	481.825	290.284	.491
8	626.726	252.646	554.532	314.678	.112
1-8	392.896	177.147	405.759	187.252	.638
4-8	400.949	188.002	376.173	225.909	.416

assignments combined and for assignments 4 – 8 combined the differences between the means are non-significant.

Table 21
Comparison of Length of Ink Written (Normalized) Between the Experimental and Control Group

Assignment	Experimental Group		Control Group		
	M	S.D.	M	S.D.	*p*
1	67.724	39.807	89.047	53.77	.003
2	97.31	49.263	112.112	60.947	.077
3	89.068	44.476	119.174	67.508	< .001
4	120.285	54.332	145.117	75.727	.012
5	100.95	48.837	122.004	65.666	.014
6	173.159	99.611	190.958	120.214	.290
7	151.489	93.716	157.516	100.247	.692
8	169.627	85.029	172.972	115.522	.834
1-8	109.31	55.039	122.628	70.274	.146
4-8	111.604	59.163	115.121	78.624	.725

Length of ink written. Just as with the *Stroke Count*, the weekly reports did not present the length of ink written on each assignment and thus students received no feedback about this measure of homework effort. To determine if this measure was affected by the weekly reports, we compare the length of ink written, normalized by the number of assigned problems, between the experimental and control offerings. Table

21 shows the mean and standard deviation of the normalized *Length of Ink Written* for the eight homework assignments individually, for all eight assignments combined, and for assignments 4 – 8 combined. For four of the first five assignments (assignments 1, 3, 4, and 5), the normalized *Length of Ink Written* is significantly smaller for the experimental group. This is consistent with the fact that the experimental group had a smaller average *Stroke Count* for these assignments than did the control group. For the other assignments, the differences are non-significant. Likewise, for all eight assignments combined and for assignments 4 – 8 combined the differences between the means are non-significant.

Table 22
Comparison of Percentage of Lectures Submitted Between the Experimental and Control Group

	Experimental Group		Control Group		
	M.	S.D.	M.	S.D.	*p*
	76.1	28.6	80.4	28.9	.319

Lecture notes submitted. As students could submit their lectures notes for full credit at any time during the course, it is not meaningful to compare lecture note submission on a weekly basis. Instead we consider only the number of lecture notes submitted over the entire quarter. As shown in Table 22, students in the experimental group submitted an average of 76% out of a possible 16 lectures while those in control group submitted an average of 80% out of 15 lectures. This difference between the means is non-significant (p = .319).

Reading time. The assigned reading materials for the experimental and control offerings were nearly identical. The only difference was that the experimental offering was assigned 1.25 additional pages of material (which was assigned during the last week of the quarter). While the assigned material was essentially the same, there were differences in the schedules of reading assignments. More specifically, there were three sections of the third chapter of the textbook that were assigned to the control offering with the second homework assignment, but which were assigned to the experimental offering with the fourth homework assignment. Likewise, the entirety of chapter three was assigned to the control offering with the fifth homework assignment, whereas for the experimental offering it was assigned with the sixth homework assignment.

During the experiment we used a simple algorithm to parse the log files from DocViewer and compute the reading time and page views. We subsequently developed an improved parsing algorithm that produces a much more accurate measure of reading time, which we call the *True Total Time*. The *Total Time*, which was reported in the progress reports, was smaller than the *True Total Time*. For example, for the entire 10-week quarter, the average *Total Time* was 45% less than the average *True Total Time*.

Table 23 shows the *Total Time* for the experimental group and the *True Total Time* for both the experimental and control groups. Our analysis focuses only on the *True Total Time*. For eight of the 10 weeks, the mean *True Total Time* for the control group was greater than for the experimental group, but this was significant for only for three of the weeks (1, 5, and 10). For two weeks (2 and 6), the mean for the

Table 23
Comparison of Total Time Reading (in Minutes) Between the Experimental and Control Group

	Experimental Group				Control Group		
	Total Time		True Total Time		True Total Time		
Week	M.	S.D.	M.	S.D.	M.	S.D.	*p*
1	5.152	18.939	4.365	11.277	16.990	44.883	.002
2	66.503	100.169	123.783	146.860	39.776	77.137	< .001
3	156.723	250.469	199.200	199.191	257.135	185.684	.052
4	119.185	202.016	198.294	183.095	213.292	130.727	.559
5	151.882	180.277	197.981	143.173	253.010	175.469	.019
6	48.185	86.402	119.045	170.267	56.519	120.111	.010
7	101.698	179.898	225.421	209.240	258.176	222.367	.312
8	123.734	165.204	299.018	299.316	357.313	300.986	.201
9	122.222	145.785	174.339	140.245	213.092	223.463	.134
10	92.431	171.134	261.705	316.720	444.463	381.632	< .001
1-10	987.714	1082.38	1803.149	1232.032	2109.767	1291.497	.108
4-10	759.335	852.885	1475.802	1000.123	1795.866	1140.361	.045

experimental group was significantly greater than for the control group. However, because the schedules of reading assignments varied between the two groups, weekly comparisons are difficult to interpret. The penultimate row of the table compares the means of *True Total Time* for all 10 weeks combined. The difference is non-significant.

Additionally, the last row of the table compares the means of *True Total Time* for weeks 4 – 10 combined. These are the weeks in which the students in the experimental group could have been influence by the progress reports. Here the difference is significant (p = .045), with the experimental group (M = 1475.8 min) reading less than the control group (M = 1795.9 min).

Table 24
Comparison of Explanatory Text Reading Time (in Minutes) Between the Experimental and Control Group

Week	Experimental Group		Control Group		
	M	S.D.	M	S.D.	*p*
1	1.669	5.341	10.087	33.596	.005
2	26.45	53.061	16.1	46.578	.207
3	45.46	115.72	34.023	47.376	.477
4	21.865	46.688	15.277	23.782	.319
5	24.153	34.196	25.399	57.473	.853
6	22.507	72.511	7.164	21.805	.125
7	10.739	24.259	23.783	35.577	.004
8	20.377	50.685	16.786	41.794	.641
9	10.616	19.829	13.351	46.355	.569
10	23.963	51.437	28.255	49.561	.597
1-10	207.798	332.717	190.225	258.18	.725
4-10	134.219	197.554	130.015	190.365	.893

Both the *Total Time* and the *True Total Time* measures include all documents read with the DocViewer program, including explanatory text (i.e., the main text of the chapters), the assigned homework questions at the ends of the chapters, the homework assignments, and graded work (homework, quizzes, and exams), for example. Only the explanatory text comprises reading for understanding. Here, we compare reading of explanatory text between the two groups. (We used the more accurate log file parser for this analysis.) Table 24 shows the *Explanatory Text Reading Time* for both groups for all weeks individually, for all weeks combined, and for weeks 4 – 10 combined. One surprising result is that the students in both groups did very little reading for understanding. Over the entire 10 weeks of the quarter, students in the experimental group spent only 207.8 minutes (3.5 hours) reading explanatory text, while the students in the control group spent only 190.2 minutes (3.2 hours). For both groups, this is an average of about 20 minutes per week. Most of the other time recorded by DocViewer was related to viewing the homework questions at the ends of the chapters.

There was essentially no difference in how the two groups read the explanatory text. With the exception of the first week, the differences in the mean *Explanatory Text Reading Time* were non-significant. For the first week, the experimental group averaged 1.7 minutes of reading while the control group averaged 10.1 minute (p = .005). The difference in the means for the 10 weeks combined and weeks 4-10 combined were non-significant.

Table 25

Comparison of Cumulative Percentage of Assigned Pages Read Between the Experimental and Control Group

	Experimental Group		Control Group		
Week	M.	S.D.	M.	S.D.	*p*
1	2.2	6.9	6.1	15.8	.014
2	11.3	17.7	8.1	11.6	.202
3	17.3	23.6	16.3	15.4	.765
4	18.2	23.7	19.2	16.6	.749
5	24.8	25.3	23.0	18.1	.617
6	24.5	24.9	23.7	18.6	.828
7	24.5	24.5	25.5	19.0	.769
8	25.2	24.4	27.7	19.5	.482
9	26.4	25.0	26.8	18.6	.892
10	24.4	23.0	24.2	16.8	.929

Reading assigned pages. Table 25 shows the percentage of assigned pages read (i.e., "*Assigned Pages*") for both groups. As some sections of the textbook were assigned multiple times, we report the reading of the cumulatively assigned pages through the end of each week. For example, the results listed for the fifth week represent the percentage of all assigned pages read from weeks 1 – 5. If a page of the textbook was assigned multiple times, reading it once is sufficient for it to be considered read. In the first week, the experimental group read fewer (2.2%) of the assigned pages than did the control group (6.1%). This difference is significant (p = .014). However, there were no differences between the groups for the other nine weeks. For example, students in the experimental group read on average only 24.4% of the assigned pages by the end of the quarter while students in the control group read on average only 24.2%. This difference is non-significant (p = .929).

Is Reading the Weekly Report Related to Course Success?

Our results suggest that the weekly progress reports did not lead to changes in students' learning behaviors. More specifically, there were no significant differences

Table 26

Correlation Between the Number of Read Weekly Reports and Measures of Success in the Course

Measure	*r*	*p*
Amount Started Early	.26	.037
Problems Attempted	.39	.001
Total Time	.27	.029
Length of Ink Written	.31	.013
Stroke Count	.35	.004
Lectures	.12	.336
Reading: Total Time	-.03	.818
Reading: Assigned Pages	-.11	.431
Course Grade	.23	.067

between the experimental and control groups in effort related to homework, note taking, or reading. Since there were no changes in behavior, the question of whether or not the changes in behavior led to changes in learning outcomes is moot. Instead, we consider the question of whether or not the act of reading the weekly progress reports is itself related to behaviors associated with academic achievement. Table 26 lists the Pearson correlations between the number of weekly progress reports a student read within a week of their being provided and our five measures of homework effort. Prior work by Rawson et al. (2017) has demonstrated that these measures are in fact related to academic achievement as measured by course grade. Here we find that all five measures correlate positively and significantly ($p < .05$) with the number of reports read within a week of their distribution. These two facts – that behavior did not change as a result of reading the reports and that reading the reports correlated with behaviors that are related to academic achievement – suggest that the students who read the weekly progress reports are the ones who already exhibited successful learning behaviors. The converse is perhaps more interesting: those students who most need help achieving success in the course are the ones that do not read the progress reports. This is a major finding of this study.

Table 26 also includes the Pearson correlation between the number reports read within a week of distribution and the number of lectures for which lecture notes were submitted. The correlation is positive, but non-significant. Similarly, Table 26 lists the Pearson correlations between the number reports read and our two measures of reading effort. Both correlations are negative, but non-significant. Apparently, reading the progress reports is unrelated to taking lecture notes and reading the textbook.

Finally, Table 26 lists the Pearson correlation between the number of reports read within a week distribution and overall grade in the course. Consistent with the results in Table 16, this correlation is positive. However, it is non-significant ($p = .067$). Thus, while students who read the progress reports are also the ones who tended to exhibit high effort on homework, which is associated with academic success, the number of reports read is not itself an effective predictor of academic success.

Did the Groups Differ on Course Grade?

Our results also suggest that the weekly progress reports did not lead to changes in final course grades. More specifically, we compared final course grades of the experimental group (M = .659, S.D. = .182) to those of the control group (M = .670, S.D. = .146) using a two-tailed t-test with unequal variances and found the difference in the means to be non-significant ($p = .666$). In short, we conclude that providing students with objective measures of study behavior was ineffective at improving final course grade.

Discussion

Empirical Contribution

Our primary finding is that providing objective measures of the students' learning effort – including effort on homework, taking lecture notes, and reading –

results in no significant changes to students' learning behavior. Over the weeks that could have been influenced by prior progress reports (i.e., after the third week), there were on average no significant differences between the experimental and control groups in the fraction of homework started early (i.e., *Proportion Started Early*), the *Total Time* spent on homework, and the percentage of *Problems Attempted*. Additionally, there was no significant difference over these weeks in the *Length of Ink Written* and the *Stroke Count*, two additional measures of homework that were not reported to the students. Furthermore, there was no significant difference between the two groups in the *Percentage of Lectures Submitted* and, on average over the entire quarter, there was no significant difference in the *True Total Time* spent reading, the *Explanatory Text Reading Time*, and the *Percentage of Assigned Pages Read*. Finally, the weekly reports did not affect course grade. We conclude that the opportunity to monitor diagnostic information about one's study behavior is not sufficient to cause a change in study behavior or learning outcome.

Theoretical Implications

This research provides evidence for the two-stage model of metacognition as summarized in Figure 18. First, the results do not support the monitoring hypothesis, which asserts that providing feedback about one's study behavior will cause improvements in their study behavior and learning outcome. Apparently, monitoring one's study behavior is insufficient to produce useful changes in student studying behavior.

Second, the results do support the regulation hypothesis, which asserts that students must be able to monitor how they are doing and plan for how to adjust their behavior. This study suggests that the two metacognitive processes of monitoring and regulating must both be engaged. Simply presenting feedback barely primes the monitoring process - given that students do not always access the weekly reports - and apparently does not sufficiently prime the regulating process.

Practical Implications

Having the equivalent of a wearable fitness tracking device for education is appealing, but our results suggests that this concept does not work in practice. While providing objective measures of physical activity may influence some to improve their fitness, the analogy appears not to hold for undergraduate engineering education. Students may need explicit guidance and training in how to study in addition to simply being given information about how they are doing.

While providing objective measures via weekly reports to students does not result in improved outcomes, the measures can be used to detect students with low motivation (Bandura, 1977), as these students will avoid tasks (such as reading weekly reports).

Methodological Implications

This study highlights the value of educational data mining techniques using smartpens. Smartpen technology allows for assessing student activity objectively with high levels of detail in an unobtrusive manner. In addition, as the student activity can be measured in an automated fashion without need of manual intervention, the student study size is scalable as well.

Limitations and Future Directions

Our analysis for this study was with a single engineering statics course, and conclusions drawn for this course might not be applicable to other STEM courses. The tone of the feedback may have been ineffective, so future research is needed on how best to convey diagnostic information. While weekly reports in statics did not influence student behavior in this study, we are planning on using a gamification framework with a concrete grade benefit serving as motivation to observe if giving tasks which require a student to meet specified behavior thresholds will result in the desired behavior outcomes.

References

Anthony, L., Yang, J., & Koedinger, K. R. (2008). Toward next-generation intelligent tutors: Adding natural handwriting input. *IEEE MultiMedia, 15*(3), 64-68. doi: 10.1109/MMUL.2008.73

Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*(6, Pt. 1), 359-372. doi:10.1037/h0043445

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologis*t, *23*, 225-239. doi:10.1037/h0020791

Azevedo, R., & Aleven, V. (Eds.). (2013). *International handbook of metacognition and learning technologies.* New York: Springer. doi:10.1007/978-1-4419-5546-3

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. doi:10.1037/0033-295X.84.2.191

Beal, C.R., & Cohen, P.R. (2008). Temporal data mining for educational applications. In T.B. Ho & Z.H. Zhou (Eds.), *PRICAI 2008: Trends in Artificial Intelligence. PRICAI 2008. Lecture Notes in Computer Science, 5351* (pp. 66-77). Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-89197-0_10

Brown, P. C., Roediger, III., H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Cambridge, MA: Harvard University Press.

Blumner, H. N., & Richards, H. C. (1997). Study habits and academic achievement of engineering students. *The Journal of Engineering Education, 86*, 125-132. doi:10.1002/j.2168-9830.1997.tb00275.x

Cadmus-Bertram, L.A., Marcus, B.H., Patterson, R.E., Parker, B.A., & Morey, B.L. (2015). Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. *American Journal of Preventive Medicine*, *49*(3), 414-418. doi:10.1016/j.amepre.2015.01.020

Chi, M.T., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use example in learning to solve problems. *Cognitive Science, 13*(2), 145-182. doi:10.1207/s15516709cog1302_1

Cullen, F.T., Cullen, J.B., Hayhow, V.L., & Plouffe, J.T. (1975). The effects of the use of grades as an incentive. *The Journal of Educational Research*, *68*(7), 277-279. doi: 10.1080/00220671.1975.10884770

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4-58. doi:10.1177/1529100612453266

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding.* New York: Cambridge University Press.

Fiorella, L. & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28,* 717-741. doi:10.1007/s10648-015-9348-9

Gordon, R., & Bloxham, S. (2017) Influence of the Fitbit Charge HR on physical activity, aerobic fitness and disability in non-specific back pain participants. *The Journal of Sports Medicine and Physical Fitness, 57* (12). 1669-1675. doi: 10.23736/S0022-4707.17.06688-9

Gyllen, J., Stahovich, T., & Mayer, R. (2018). How students read an e-textbook in an engineering course. *Journal of Computer Assisted Learning.* doi:10.1111/jcal.12277

Hattie, J. (2009). *Visible learning.* New York: Routledge.

Herold, J., & Stahovich, T. (2012). Characterizing students handwritten self-explanations. In *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

Herold, J., Stahovich, T. F., & Rawson, K. (2013). Using educational data mining to identify correlations between homework effort and performance. In *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Hestnes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, *30* (3), 141-158. doi:10.1119/1.2343497

Johnson, C. I., & Priest, H. A. (2014). The feedback principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed), 449–463. New York: Cambridge University Press.
Keith, T. Z. (1982). Time spent on homework and high school grades: A large-scale path analysis. *Journal of Educational Psychology, 74,* 248-253. doi: 10.1037/0022-0663.74.2.248

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science, 6,* 333-353. doi:10.1002/wcs.1350

Krüger, A., Merceron, A., & Wolf, B. (2010). A data model to ease analysis and mining of educational data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining*, (pp. 131-140). Pittsburgh, PA.

Li, N., Cohen, W. W., Koedinger, K. R., & Matsuda, N. (2011). A machine learning approach for automatic student model discovery. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, C., & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 31-40). Eindhoven, the Netherlands.

Lounsbury, J.W., Sundstrom, E., Loveland, J.L., & Gibson, L.W. (2003). Broad versus narrow personality traits in predicting academic performance of adolscents. *Learning and Individual Differences, 14*(1), 65-75. doi:10.1016/j.lindif.2003.08.001

Mayer, R. E. (2011). *Applying the science of learning.* Boston: Pearson.

Mayer, R. E. (2019). *How to be a successful student.* New York: Routledge.

McDermott, P. A., Mordell, M., & Stoltzfus, J. C. (2001). The organization of student performance in American schools: Discipline, motivation, verbal learning, nonverbal learning. *Journal of Educational Psychology, 93*(1), 65-76. doi: 10.1037/0022-0663.93.1.65

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology, 106*(1), 121-131. doi:10.1037/a0033546

Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. In *33rd Annual Frontiers in Education, 2003. FIE 2003 (pp. T2A 13-18).* Westminster, CO. doi:10.1109/FIE.2003.1263284

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science, 13*, 390-407. doi:10.1177/1745691617710510

Mostow, J., González-Brenes, J. P., & Tan, B. H. L. (2011). Learning classifiers from a relational database of tutor logs. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 149-158). Eindhoven, Netherlands.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25*, 1159-1168. doi:10.1177/0956797614524581

Neff, G. & Nafus, D. (2016). *Self-tracking.* Cambridge, MA: MIT Press.

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. In UIST '06 - *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (pp. 191-200). New York, NY: ACM Press.

Picciano, A.G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20. doi:10.24059/olj.v16i3.267

Radhakrishnan, P., Lam, D., & Ho, G. (2009). Giving University Students Incentives to Do Homework Improves Their Performance. *Journal of Instructional Psychology*, *36*(3), 219-225.

Rawson, K., & Stahovich, T. F. (2013). Predicting course performance from homework habits. In *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Rawson, K., Stahovich, T. F., & Mayer, R. E. (2017). Homework and achievement: Using smartpen technology to find the connection*. Journal of Educational Psychology, 109*(2)*,* 208-219. doi:10.1037/edu0000130

Rhodes, M. G., Cleary, A. M., & DeLosh, E. L. (2019). *A guide to effective studying and learning: Practical strategies from the Science of Learning.* New York: Oxford University Press.

Robyak, J.E., & Downey, R.G. (1979). The prediction of long-term academic performance after the completion of a study skills course. *Measurement and Evaluation in Guidance*, *12*(2), 108-111. doi:10.1080/00256307.1979.12022191

Romero, C., Romero, J., Luna, J., & Ventura, S. (2010). Mining rare association rules from e-learning data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 171-180). Pittsburgh, PA.

Romero, C., Ventura, S., Vasilyeva, E., & Pechenizkiy, M. (2010). Class association rules mining from students' test data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining*, (pp. 317-318). Pittsburgh, PA

Shanabrook, D. H., Cooper, D. G., Woolf, B. P., & Arroyo, I. (2010). Identifying high-level student behavior using sequence-based motif discovery. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining*, (pp. 191-200). Pittsburgh, PA.

Schunk, D.H. (1991). Self-Efficacy and Academic Motivation. *Educational Psychologist*, *26*(3-4), 207-231. doi:10.1080/00461520.1991.9653133

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189. doi:10.3102/0034654307313795

Schrager, J. D., Shayne, P., Wolf, S., Das, S., Patzer, R. E., White M., & Heron, S. (2017). Assessing the Influence of a Fitbit Physical Activity Monitor on the Exercise Practices of Emergency Medicine Residents: A Pilot Study. *JMIR Mhealth Uhealth,* 5(1):e2. doi: 10.2196/mhealth.6239

Stahovich, T., Gyllen, J., & Mayer, R. E. (in press). Accuracy in judgments of study time predicts academic success in an engineering course. *Metacognition and Learning*.

Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: Development and psychometric analysis. *The Journal of Engineering Education, 94,* 363-371. doi:10.1002/j.2168-9830.2005.tb00864.x

Stevens, R., Johnson, D. F., & Soller, A. (2005). Probabilities and predictions: Modeling the development of scientific problem-solving skills. *Cell Biology Education, 4,* 42-57. doi:10.1187/cbe.04-03-0036

Trivedi, S., Pardos, Z. A., Sárközy, G. N., & Heffernan, N. T. (2011). Spectral clustering in educational data mining. In M. Pechenizkiy, T.Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 129-138). Eindhoven, Netherlands.

Tuckman, B.W. (1998). Using Tests as an Incentive to Motivate Procrastinators to Study, *The Journal of Experimental Education*, *66*(2), 141-147. doi: 10.1080/00220979809601400

Van Arsdale, T., & Stahovich, T. (2012). Does neatness count? What the organization of student work says about understanding. In *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

Chapter 5

Conclusions

The work presented in this dissertation details three major studies on student homework behaviors and the relationship to course achievement. In the first study, we looked at the relationship between homework behaviors and course achievement. Additionally, work relying upon self-reports was scrutinized by comparing self-reporting of time spent on homework versus time spent on homework recorded objectively with smartpens.

For our second study, we extended our analysis to answer whether the homework behaviors measured over an entire course offering (which showed strong positive correlations to academic achievement) are also present earlier in the course.

And in our last study, we examined whether we could influence students to change their homework behaviors. We informed students with a weekly report about their current homework behaviors, with tailored feedback on ways to improve. We then examined whether the students changed their behavior on subsequent homework assignments.

Homework and Achievement

For our first study, we studied the connection between student behaviors and academic achievement across three different cohorts of an undergraduate engineering statics course. We found consistent and positive correlations between student behaviors and course grade, including the time spent working on the homework, excluding breaks greater than 10 minutes (r = .44), how many strokes were written (r = .49), how many problems were attempted (r = .45), and average time per problem (r = 0.33). We also found a strong negative correlation between the amount of homework done within 24 hr of the due date and course grade (r = -.32).

We also examined time spent on homework when utilizing students' self-reported estimates. In contrast to our objective measure of time spent on homework having strong, positive correlations, the relationship between self-reported time spent on homework and course grade for the three cohorts had a negative correlation (r = -.16), and only one of the three cohorts had a correlation that was statistically significant. Furthermore, all three cohorts greatly overestimated how much time they spent on the homework when compared to the time recorded with the smartpens. 88.5%, 85.5%, and 85.5% of students in each cohort overestimated the time spent working on homework, confirming the unreliable nature of self-reports and study effort that other researchers have proposed (Blumner & Richards, 1997; Schuman, Walsh, Olson, & Etheridge , 1985).

Homework and Predicting Success

In our second study, we analyzed conscientious homework habits (Pellegrino & Hilton, 2012) on a weekly basis, rather than only over the entire course (as we had for the first study). We examined four measures and combinations of conscientious homework habits on assignments over seven offerings of two introductory engineering courses. By the third week of the course we observed significant and strong correlations between the measures and course grade. Working on time (as measured by the *Early Work Fraction*, which is the fraction of work done at least 24 hr before the homework due date) was statistically significant for five of the seven offerings. Being on task (as measured by the *Number of Problems Attempted*) was statistically significant for all seven offerings. Working with high effort as measured by *Stroke Count* was statistically significant for 6 offerings, and *Length of Ink Written* was statistically significant for 5 offerings.

When these four measures of conscientious habits were considered together, and using stepwise regression, we observed stronger correlations with course grade. The correlations with these four features together and course grade were statistically significant for all seven offerings by the third week of the course and were, on average, 88% as large as correlations utilizing the entire homework record.

When using the *Combined Measure of Conscientiousness*, a weighted combination of the four individual measures, correlations by the third homework assignment were statistically significant for all seven offerings, and on average, 86% as large as correlations utilizing the entire homework record.

This study showed the three core features of conscientiousness are indicators of academic achievement, and further suggests that conscientiousness early in the course is indicative of ultimate academic achievement in the course.

Effect of Weekly Reports on Study Behavior

For our third study, we gave an experimental group a weekly report detailing their current homework behaviors and suggestions of changes to their behavior that lead to increased likelihood of success in the course. We then examined whether student behavior changed on subsequent assignments, and compared behaviors to that of a control group that did not receive a weekly report.

Most students did not read all of the provided weekly reports, with only 7.7% of students reading all eight weekly reports, and only 23.1% reading seven of eight reports within a week of it being initially provided.

We observed that providing the objective measures of learning effort – effort on homework, taking lecture notes, reading –results in no significant change to students' learning behavior. We also performed a two-tailed t-test with unequal variances and observed no significant difference of final course grade (p = .666) between the weekly report experimental group (n = 65, M = .659, S.D. = .182) and the control group (n = 135, M = .670, S.D. = .146).

Limitations and Future Directions

Several concerns arise from relying upon just the homework data collected from smartpens. The first concern is that not all of the student's work is captured with the smartpen, but rather some is done using scratch paper. In this scenario, the student's complete homework behavior would not be measured. As detailed in Chapter 2, a postexperimental questionnaire asked students how much on a scale of 1 (“doing all homework elsewhere”) to 7 (“using the pen to do everything”), the mean rating was 5.1 (S.D. = 1.7). Another concern is that not all of the time spent working on homework is measured by starting measurement at the beginning at the first pen stroke, as this can neglect time spent reading and thinking about the problem before starting to answer. Future work pairing the student’s written work (recorded with the smartpen) with that of their reading (Gyllen, Stahovich, & Mayer, 2018) would provide a more accurate measurement of total time spent on the homework.

Recording of student behaviors can also be affected by “collaboration” with fellow classmates. If students are copying work from one another, the recorded work is not truly indicative of their normal homework behaviors.

Weekly reports were found to be an insufficient motivator to change student behavior. Future work could examine using a gamification framework (Dichev & Dicheva, 2016) with a concrete grade benefit serving as motivation to observe if giving tasks which require a student to meet specified behavior thresholds will be a sufficient mechanism for “self-regulation” of behavior.

One of the potential uses of these objective measures of homework behaviors is detection of at-risk students, allowing for intervention early enough in the course to produce more desirable course outcomes for the student. These same techniques applied to a small class could also be applied to a massive online open course (MOOC) environment, as the measures of student behavior in this paper do not require manual intervention, and are scalable as well.

Contributions

This work makes both applied and methodological contributions to educational research. The findings illustrate the relationship between students' homework behaviors and academic achievement. Additionally, this work shows students' homework behaviors are established early in a course, and remain consistent throughout the entire course. Finally, this work details that monitoring of one's behaviors is insufficient motivation to improve learning outcomes, and most likely must be coupled with regulating the study behavior to affect change.

This work highlights the potential of educational data mining and smartpen technology for educational research. Our results confirm that unreliability of studies relying upon self-reports. Our studies also speak to the value of replication in education research.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4,* 167–207. doi:10.1207/s15327809jls0402_2

Anthony, L., Yang, J., & Koedinger, K. R. (2008). Toward next-generation intelligent tutors: Adding natural handwriting input. *IEEE MultiMedia, 15,* 64–68. doi:10.1109/MMUL.2008.73

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60,* 383–398. doi:10.1007/s11423-012-9235-8

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist, 23,* 225–239. doi:10.1037/h0020791

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. doi:10.1007/s11251-005-1272-9

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. doi:10.1037/0033-295X.84.2.191

Beal, C. R., & Cohen, P. R. (2008). Temporal data mining for educational applications. *Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence, 5351,* 66–77.

Berliner, D. (1984). The half-full glass: A review of research on teaching. In P. L. Hosford (Ed.), *Using what we know about teaching* (pp. 51–84). Alexandria, VA: Association for Supervision and Curriculum Development.

Berliner, D. (1991). What's all this fuss about instructional time? In M. Ben-Peretz & R. Bromme (Eds.), *The nature of time in schools*. New York, NY: Teachers College Press.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252. doi:10.1207/s1532690xci1302_3

Blanc, P. (1999). Solving a non-routine problem: What helps, what hinders. *Proceedings of the British Society for Research into Learning Mathematics, 19,* 1–6.

Blumner, H. N., & Richards, H. C. (1997). Study habits and academic achievement of engineering students. *The Journal of Engineering Education, 86,* 125–132. doi:10.1002/j.2168-9830.1997.tb00275.x

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade point average from the revised and recentered SAT® I: Reasoning Test* (College Board Rep. No. 2000-1). New York: College Entrance Examination Board.

Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Cambridge, MA: Harvard University Press.

Cadmus-Bertram, L.A., Marcus, B.H., Patterson, R.E., Parker, B.A., & Morey, B.L. (2015). Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. *American Journal of Preventive Medicine*, *49*(3), 414-418. doi:10.1016/j.amepre.2015.01.020

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723–733.
Clark, R. C., & Mayer, R. E. (2011). *e-Learning and the science of instruction* (3rd ed.). San Francisco, CA: Pfeiffer.

Chi, M.T., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use example in learning to solve problems. *Cognitive Science, 13*(2), 145-182. doi:10.1207/s15516709cog1302_1

Cooper, H. (1989). *Homework*. White Plains, NY: Longman. doi:10.1037/11578-000

Cooper, H. (2001). *The battle over homework* (2nd ed.). Thousand Oaks, CA: Corwin.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research 1987–2003. *Review of Educational Research, 76,* 1–62. doi:10.3102/00346543076001001

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3,* 425–453. doi:10.1111/j.1745-6924.2008.00089.x

Crooks, T.J. (1980). *Grade prediction: The usefulness of context-specific predictors.* (Report No. TM 800676). Washington, D.C.: U.S. Department of Health, Education, and Welfare; National Institute of Education. (ERIC Document Reproduction Service no. ED194547)

Cullen, F.T., Cullen, J.B., Hayhow, V.L., & Plouffe, J.T. (1975). The effects of the use of grades as an incentive. *The Journal of Educational Research*, *68*(7), 277-279. doi: 10.1080/00220671.1975.10884770

Dichev, C., & Dicheva, D. (2016). Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Education Technology in Higher Education*, *14*:9. doi:10.1186/s41239-017-0042-5

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. doi:10.1177/1529100612453266

Ebbinghaus, H. (1964). *Memory*. New York, NY: Dover. (Original work published 1885)

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding.* New York: Cambridge University Press.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717-741. doi:10.1007/s10648-015-9348-9

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. (Research and Occasional Paper Series No. CSHE.6.07). University of California, Berkeley, Center for Studies in Higher Education.

Gordon, R., & Bloxham, S. (2017) Influence of the Fitbit Charge HR on physical activity, aerobic fitness and disability in non-specific back pain participants. *The Journal of Sports Medicine and Physical Fitness, 57* (12). 1669-1675. doi: 10.23736/S0022-4707.17.06688-9

Graham, S., & Williams, C. (2009). An attributional approach to motivation in school. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 11–34). New York, NY: Routledge.

Gyllen, J., Stahovich, T., & Mayer, R. (2018). *How students read an e-textbook in an engineering course*. *Journal of Computer Assisted Learning.* doi:10.1111/jcal.12277

Gyllen, J. Stahovich, T., Mayer, R., Entezari, N., & Darvishzadeh, A. (2018). Priming productive study strategies with preparatory quizzes in an engineering course. Manuscript submitted for publication.

Hall, R. (2000). Video recording as theory. In D. Lesh & A. Kelley (Eds.), *Handbook of research design in mathematics and science education* (pp. 647–664). Mahwah, NJ: Lawrence Erlbaum.

Hall, S., & Vance, E.A. (2010). Improving self-efficacy in statistics: role of self-explanation & feedback. *Journal of Statistics Education, 18*(3). doi:10.1080/10691898.2010.11889583

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 metaanalyses relating to achievement*. New York, NY: Routledge.

Herold, J., & Stahovich, T. (2012). Characterizing students handwritten self-explanations. *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

Herold, J., Stahovich, T., Lin, H.-L., & Calfee, R. C. (2011). The effectiveness of “pencasts” as an instructional medium. *Proceedings of the American Society for Engineering Education 118th Annual Conference and Exposition*, Vancouver, British Columbia, Canada.

Herold, J., Stahovich, T. F., & Rawson, K. (2013). Using educational data mining to identify correlations between homework effort and performance. *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Hestnes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, *30* (3), 141-158. doi:10.1119/1.2343497

Hills, J.R., Klock, J.A., & Bush, M (1965). The Use of Academic Prediction Equations with Subsequent Classes. *American Educational Research Journal, 2*(4)*,* 203-206. doi:10.3102/00028312002004203

Johnson, C. I., & Priest, H. A. (2014). The feedback principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed), 449–463. New York: Cambridge University Press.

Jones, L., & Ruch, G. M. (1928). The twenty-seventh yearbook of the national society for the study of education: Nature and nurture, part ii-their influence upon achievement. In G. M. Whipple (Ed.), *Achievement as affected by amount of time spent in study* (pp. 131–134). Bloomington, IL: Public School Publishing Co.

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: a perspective for enhancing meaningful learning. Educational *Psychology Review, 24*(3), 401-418. doi:10.1007/s10648-012-9202-2

Karweit, N. (1984). Timed-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership, 41,* 32–35.

Keith, T. Z. (1982). Time spent on homework and high school grades: A large-scale path analysis. *Journal of Educational Psychology, 74,* 248–253. doi: 10.1037/0022-0663.74.2.248

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science, 6,* 333–353. doi:10.1002/wcs.1350

Krüger, A., Merceron, A., & Wolf, B. (2010). A data model to ease analysis and mining of educational data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 131–140). Pittsburgh, PA.

Lee, C.-K., Stahovich, T., & Calfee, R. C. (2011). A pen-based statics tutoring system. *Proceedings of the American Society for Engineering Education 118th Annual Conference and Exposition*, Vancouver, British Columbia, Canada.

Li, N., Cohen, W. W., Koedinger, K. R., & Matsuda, N. (2011). A machine learning approach for automatic student model discovery. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 31–40). Eindhoven, the Netherlands.

Lin, H., Stahovich, T. F., & Herold, J. (2012). Automatic handwritten statics solution classification and its applications in predicting student performance. *Proceedings of the 119th American Society for Engineering Education Annual Conference and Exposition*, San Antonio, TX.

Lounsbury, J.W., Sundstrom, E., Loveland, J.L., & Gibson, L.W. (2003). Broad versus narrow personality traits in predicting academic performance of adolscents. *Learning and Individual Differences, 14*(1), 65-75. doi:10.1016/j.lindif.2003.08.001

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54,* 588–599. doi:10.1016/j.compedu.2009.09.008

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511811678

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139547369.005

Mayer, R. E. (2019). *How to be a successful student: 20 study habits based on the science of learning.* New York: Routledge.

McDermott, P. A., Mordell, M., & Stoltzfus, J. C. (2001). The organization of student performance in American schools: Discipline, motivation, verbal learning, nonverbal learning. *Journal of Educational Psychology, 93*(1), 65-76. doi:10.1037/0022-0663.93.1.65

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology, 106*(1), 121-131. doi:10.1037/a0033546

Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. In *33rd Annual Frontiers in Education, 2003. FIE 2003* (pp. T2A 13-18). Westminster, CO. doi:10.1109/FIE.2003.1263284

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science, 13*, 390-407. doi:10.1177/1745691617710510

Mostow, J., González-Brenes, J. P., & Tan, B. H. L. (2011). Learning classifiers from a relational database of tutor logs. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C.

Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 149–158). Eindhoven, Netherlands.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25,* 1159 –1168. doi:10.1177/0956797614524581

Neff, G. & Nafus, D. (2016). *Self-tracking.* Cambridge, MA: MIT Press.

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology - UIST '06* (pp. 191–200). New York, NY: ACM Press.

Passons, W.R. (1967). Predictive validities of the ACT, SAT and high school grades for first semester GPA and freshman courses. *Educational and Psychological Measurement, 27*(4), 1132-1144. doi:10.1177/001316446702700469

Pellegrino, J. W., & Hilton, M. L. (Eds). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Washington, DC: National Academics Press.

Picciano, A.G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20. doi:10.24059/olj.v16i3.267

Radhakrishnan, P., Lam, D., & Ho, G. (2009). Giving University Students Incentives to Do Homework Improves Their Performance. *Journal of Instructional Psychology*, *36*(3), 219-225.

Rawson, K., & Stahovich, T. F. (2013). Predicting course performance from homework habits. *Proceedings of the American Society for Engineering Education 120th Annual Conference and Exposition*, Atlanta, GA.

Rawson, K., Stahovich, T. F., & Mayer, R. E. (2017). Homework and achievement: Using smartpen technology to find the connection. *Journal of Educational Psychology, 109*(2), 208-219. doi:10.1037/edu0000130

Rhodes, M. G., Cleary, A. M., & DeLosh, E. L. (2019). *A guide to effective studying and learning: Practical strategies from the Science of Learning.* New York: Oxford University Press.

Robyak, J.E., & Downey, R.G. (1979). The prediction of long-term academic performance after the completion of a study skills course. *Measurement and Evaluation in Guidance*, *12*(2), 108-111. doi:10.1080/00256307.1979.12022191

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

Romero, C., Romero, J., Luna, J., & Ventura, S. (2010). Mining rare association rules from e-learning data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 171–180). Pittsburgh, PA.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384. doi:10.1016/j.compedu.2007.05.016

Romero, C., Ventura, S., Vasilyeva, E., & Pechenizkiy, M. (2010). Class association rules mining from students' test data. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 131–140). Pittsburgh, PA.

Sawyer, R., & Maxey, J. (1979). The validity of college grade prediction equations over time. *Journal of Educational Measurement, 16*(4), 279-284. doi: 10.1111/j.1745-3984.1979.tb00109.x

Schuman, H., Walsh, E., Olson, C., & Etheridge, B. (1985). Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces, 63,* 945-966. doi:10.1093/sf/63.4.945

Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35–54). New York, NY: Routledge.

Schrager, J. D., Shayne, P., Wolf, S., Das, S., Patzer, R. E., White M., & Heron, S. (2017). Assessing the Influence of a Fitbit Physical Activity Monitor on the Exercise Practices of Emergency Medicine Residents: A Pilot Study. *JMIR Mhealth Uhealth,* 5(1):e2. doi: 10.2196/mhealth.6239

Shanabrook, D. H., Cooper, D. G., Woolf, B. P., & Arroyo, I. (2010). Identifying high-level student behavior using sequence-based motif discovery. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 191–200). Pittsburgh, PA.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189. doi:10.3102/0034654307313795

Stahovich, T., Gyllen, J., & Mayer, R. E. (in press). Accuracy in judgments of study time predicts academic success in an engineering course. *Metacognition and Learning*.

Steif, P. S., & Dantzler, J. A. (2005). A Statics Concept Inventory: Development and psychometric analysis. *The Journal of Engineering Education, 94,* 363–371. doi:10.1002/j.2168-9830.2005.tb00864.x

Steif, P. S., & Dollár, A. (2009). Study of usage patterns and learning gains in a web-based interactive static course. *The Journal of Engineering Education, 98,* 321–333. doi:10.1002/j.2168-9830.2009.tb01030.x

Steif, P. S., Lobue, J. M., Kara, L. B., & Fay, A. L. (2010). Improving problem solving performance by inducing talk about salient problem features. *The Journal of Engineering Education, 99,* 135–142. doi:10.1002/j.2168-9830.2010.tb01050.x

Stevens, R., Johnson, D. F., & Soller, A. (2005). Probabilities and predictions: Modeling the development of scientific problem-solving skills. *Cell Biology Education, 4,* 42–57. doi:10.1187/cbe.04-03-0036

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. doi:10.1007/978-1-4419-8126-4

Thorndike, E.L. (1910). The contribution of psychology to education. *Journal of Educational Psychology, 1*(1), 5-12. doi:10.1037/h0070113

Trivedi, S., Pardos, Z. A., Sárközy, G. N., & Heffernan, N. T. (2011). Spectral clustering in educational data mining. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 129–138). Eindhoven, Netherlands.

Tuckman, B.W. (1998). Using Tests as an Incentive to Motivate Procrastinators to Study, *The Journal of Experimental Education*, *66*(2), 141-147. doi:10.1080/00220979809601400

Van Arsdale, T., & Stahovich, T. (2012). Does neatness count? What the organization of student work says about understanding. *Proceedings of the American Society for Engineering Education 119th Annual Conference and Exposition*, San Antonio, TX.

van Gog, T. (2013). Time on task. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 432–433). New York, NY: Routledge.

van Gog, T., Ericsson, K. A., Rikers, R. M. J. P., & Paas, F. (2005). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educational Technology Research and Development, 53,* 73–81. doi:10.1007/BF02504799

Wilson, B.C., & Shrock, S. (2001). Contributing to success in an introductory computer science course: a study of twelve factors In *Proceedings of the Thirty-second SIGCSE Technical Symposium on Computer Science Education. ACM SIGCSE Bulletin, 33,* 184-188. doi:10.1145/366413.364581

Wood, D., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wollman, W., & Lawrenz, F. (1984). Identifying potential "dropouts" from college physics classes. *Journal of Research in Science Teaching, 21*(4), 385-390. doi:10.1002/tea.3660210406

Xu, J. (2013). Homework and academic achievement. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 199–201). New York, NY: Routledge.

Zollanvari, A., Kizilirmak, R. C., Kho, Y. H., & Hernández-Torrano, D. (2017). Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors. IEEE Access, 5, 23792-23802. doi:10.1109/ACCESS.2017.2740980

Appendix

Table 27
Correlation Between Early Work Fraction (E) *and Course Grade*

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9
Intro1	.54*	.51*	.51*	.47*	.48*	.50*	.51*	.50*	.53*
Intro2	.34*	.38*	.40*	.41*	.41*	.42*	.42*	.40*	.40*
Statics1			.14	.19	.29*	.31*	.35*	.37*	.38*
Statics2			.43*	.43*	.41*	.48*	.49*	.47*	.47*
Statics3			.22*	.29*	.32*	.25*	.20*	.19*	
Statics4	.19	.15	.24	.23	.29*	.28*	.33*	.35*	.34*
Statics5	.45*	.42*	.47*	.46*	.46*	.45*	.46*	.47*	.48*

*$p < .05$.

Table 28

Correlation Between Problems Attempted (P) *and Course Grade*

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
Intro1	0.52*	0.47*	0.50*	0.59*	0.61*	0.61*	0.60*	0.61*	0.63*
Intro2	0.04	0.04	0.19*	0.26*	0.31*	0.32*	0.28*	0.28*	0.32*
Statics1			0.27*	0.24*	0.25*	0.33*	0.37*	0.33*	0.33*
Statics2			0.31*	0.52*	0.53*	0.54*	0.60*	0.63*	0.65*
Statics3			0.27*	0.22*	0.35*	0.28*	0.24*	0.22*	
Statics4	0.28*	0.34*	0.34*	0.37*	0.49*	0.50*	0.51*	0.54*	0.54*
Statics5	0.31*	0.44*	0.39*	0.50*	0.49*	0.50*	0.50*	0.51*	0.54*

*$p < .05$.

Table 29

Correlation Between Stroke Count (S) *and Course Grade*

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
Intro1	.14	.31*	.40*	.48*	.51*	.52*	.53*	.52*	.54*
Intro2	-.08	.02	.12	.22*	.24*	.25*	.23*	.21*	.24*
Statics1			.48*	.47*	.51*	.57*	.51*	.52*	.53*
Statics2			.40*	.51*	.51*	.51*	.55*	.56*	.57*
Statics3			.40*	.41*	.46*	.40*	.37*	.35*	
Statics4	.29*	.26*	.27*	.35*	.42*	.42*	.42*	.44*	.44*
Statics5	.22*	.35*	.40*	.49*	.48*	.49*	.50*	.50*	.53*

*$p < .05$.

Table 30

Correlation Between Length of Ink Written (L) *and Course Grade*

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9
Intro1	.19	.36*	.41*	.48*	.51*	.51*	.52*	.52*	.53*
Intro2	-.12	-.06	.04	.13	.14	.15	.14	.14	.16
Statics1			.39*	.38*	.44*	.45*	.44*	.44*	.43*
Statics2			.27*	.40*	.40*	.41*	.44*	.45*	.46*
Statics3			.31*	.40*	.44*	.39*	.36*	.34*	
Statics4	.25	.18	.19	.26*	.33*	.34*	.34*	.36*	.35*
Statics5	.20	.33*	.39*	.46*	.45*	.46*	.48*	.48*	.51*

*$p < .05$.

Table 31

Correlation of Stepwise Regression Models Relating All Four Measures of Conscientiousness Homework Habits to Course Grade

	Through HW1	Through HW2	Through HW3	Through HW4	Through HW5	Through HW6	Through HW7	Through HW8	Through HW9
Intro1	.72*	.66*	.65*	.67*	.68*	.68*	.68*	.67*	.70*
Intro2	.34*	.38*	.40*	.41*	.46*	.47*	.42*	.40*	.44*
Statics1			.48*	.47*	.51*	.57*	.51*	.52*	.56*
Statics2			.55*	.61*	.60*	.62*	.67*	.67*	.68*
Statics3			.40*	.41*	.46*	.40*	.37*	.35*	
Statics4	.29*	.34*	.34*	.37*	.49*	.50*	.51*	.54*	.54*
Statics5	.50*	.57*	.57*	.64*	.63*	.63*	.63*	.64*	.66*

$^*p < .05$.

Table 32

Correlation Between Combined Measure of Conscientiousness (C) *in Homework Habits and Course Grade*

	Through HW1	Through HW2	Through HW3	Through HW4	Through HW5	Through HW6	Through HW7	Through HW8	Through HW9
Intro1	0.64*	0.61*	0.62*	0.66*	0.67*	0.67*	0.66*	0.66*	0.69*
Intro2	0.14	0.18	0.30*	0.38*	0.41*	0.42*	0.39*	0.38*	0.40*
Statics1			0.45*	0.44*	0.50*	0.54*	0.55*	0.54*	0.54*
Statics2			0.49*	0.62*	0.60*	0.62*	0.66*	0.66*	0.67*
Statics3			0.39*	0.40*	0.47*	0.40*	0.34*	0.31*	
Statics4	0.31*	0.30*	0.33*	0.37*	0.48*	0.47*	0.50*	0.51*	0.51*
Statics5	0.46*	0.56*	0.57*	0.64*	0.63*	0.63*	0.63*	0.64*	0.66*

*p < .05.

Table 33

Correlation of Stepwise Regression Models Relating the Sequence of Early Work Fraction (E) *Values to Course Grade*

	E_1	E_1 - E_{1-2}	E_1 - E_{1-3}	E_1 - E_{1-4}	E_1 - E_{1-5}	E_1 - E_{1-6}	E_1 - E_{1-7}	E_1 - E_{1-8}	E_1 - E_{1-9}
Intro1	.54*	.55*	.56*	.56*	.56*	.56*	.56*	.56*	.56*
Intro2	.34*	.40*	.42*	.43*	.43*	.44*	.44*	.44*	.44*
Statics1					.29*	.31*	.35*	.37*	.38*
Statics2			.43*	.46*	.49*	.54*	.54*	.54*	.54*
Statics3			.22*	.28*	.28*	.28*	.28*	.28*	
Statics4					.29*	.29*	.32*	.33*	.35*
Statics5	.45*	.45*	.45*	.45*	.45*	.45*	.45*	.53*	.50*

*$p < .05$.

Table 34

Correlation of Stepwise Regression Models Relating the Sequence of Problems Attempted (P) *Values to Course Grade*

	P_1	P_1 - $P_{1\text{-}2}$	P_1 - $P_{1\text{-}3}$	P_1 - $P_{1\text{-}4}$	P_1 - $P_{1\text{-}5}$	P_1 - $P_{1\text{-}6}$	P_1 - $P_{1\text{-}7}$	P_1 - $P_{1\text{-}8}$	P_1 - $P_{1\text{-}9}$
Intro1	.27*	.38*	.56*	.66*	.66*	.61*	.61*	.61*	.67*
Intro2			.27*	.34*	.31*	.32*	.32*	.32*	.32*
Statics1			.27*	.27*	.27*	.33*	.37*	.37*	.37*
Statics2			.31*	.48*	.53*	.53*	.57*	.60*	.61*
Statics3			.27*	.27*	.27*	.27*	.27*	.27*	
Statics4	.38*	.38*	.43*	.46*	.62*	.50*	.51*	.54*	.54*
Statics5	.31*	.31*	.33*	.41*	.41*	.41*	.41*	.41*	.43*

*p < .05.

Table 35

Correlation of Stepwise Regression Models Relating the Sequence of Stroke Count (S) *Values to Course Grade*

	S_1	S_1 - S_{1-2}	S_1 - S_{1-3}	S_1 - S_{1-4}	S_1 - S_{1-5}	S_1 - S_{1-6}	S_1 - S_{1-7}	S_1 - S_{1-8}	S_1 - S_{1-9}
Intro1		.29*	.40*	.60*	.60*	.59*	.53*	.53*	.59*
Intro2				.34*	.32*	.33*	.33*	.33*	.33*
Statics1			.48*	.48*	.51*	.57*	.57*	.57*	.57*
Statics2			.40*	.47*	.48*	.48*	.51*	.52*	.52*
Statics3			.40*	.40*	.41*	.41*	.41*	.44*	
Statics4	.37*	.37*	.37*	.43*	.43*	.43*	.43*	.44*	.44*
Statics5	.22*	.29*	.35*	.51*	.51*	.51*	.51*	.51*	.51*

*$p < .05$.

Table 36

Correlation of Stepwise Regression Models Relating the Sequence of Length of Ink Written (L) *Values to Course Grade*

	L_1	L_1 - $L_{1\text{-}2}$	L_1 - $L_{1\text{-}3}$	L_1 - $L_{1\text{-}4}$	L_1 - $L_{1\text{-}5}$	L_1 - $L_{1\text{-}6}$	L_1 - $L_{1\text{-}7}$	L_1 - $L_{1\text{-}8}$	L_1 - $L_{1\text{-}9}$
Intro1		.33*	.41*	.53*	.56*	.51*	.52*	.52*	.53*
Intro2									
Statics1			.39*	.39*	.44*	.45*	.45*	.45*	.45*
Statics2			.27*	.35*	.36*	.36*	.39*	.40*	.40*
Statics3			.31*	.39*	.40*	.40*	.40*	.40*	
Statics4	.34*	.34*	.34*	.34*	.34*	.34*	.34*	.36*	.36*
Statics5		.26*	.43*	.50*	.50*	.50*	.56*	.56*	.50*

**p* < .05.