Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lssp20

The effect of cognitive and behavioral factors on student success in a bottleneck business statistics course via deeper analytics

Sinjini Mitra & Kenny Le

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2022, VOL. 51, NO. 6, 2779–2808
https://doi.org/10.1080/03610918.2019.1700279

The effect of cognitive and behavioral factors on student success in a bottleneck business statistics course via deeper analytics

Sinjini Mitra and Kenny Le

Information Systems and Decision Sciences Department, California State University, Fullerton, California, USA

ABSTRACT

In this article, we study a set of factors underlying student success in a bottleneck business course using statistical and data mining techniques. Factors included learning styles, motivational and other cognitive factors, personality traits, learning analytics, along with background demographic and academic ones. Our analysis yielded interesting insights that show some of these factors play significant roles in predicting both student performance and their propensity to utilize resources that help improve their performance, such as additional support services. The predictive accuracy of both of our models were over 95% (error rate <5%). Moreover, quantile regression models were used to determine factors that specifically affect the performance of low-performing students so that targeted intervention and support services can be developed specifically for them. In conclusion, deeper analytics via statistical models are crucial for forming an in-depth understanding of how to improve student performance in a bottleneck course and this has far-reaching implications for both educators and administrators in higher education.

ARTICLE HISTORY
Received 23 March 2019
Accepted 28 November 2019

KEYWORDS
Bottleneck course; Student success; Cognitive factors; Learning analytics; Personality traits; Prediction

1. Introduction

Studies of student success in higher education are widely available today as educational institutions strive to reduce the time it takes students to complete their degrees (Crisp, Doran, and Reyes 2018). One of the primary factors that contribute to longer graduation times are "bottleneck" courses, which refer to those courses that have a high level of difficulty and hence historically high failure rates. To mitigate this, colleges and universities have conducted research to identify which factors affect performance in such courses, the knowledge of which can empower them to introduce appropriate help and support programs to enhance academic success in those. These factors range from demographic factors (such as, gender, ethnicity) to academic factors (like prior academic history, grade point average or GPA), although there are several cognitive (like learning styles, motivational factors) and behavioral (like personality) factors that play a role in determining student success but are often overlooked (Glesne 1998). More recently,

CONTACT Sinjini Mitra smitra@fullerton.edu California State University, Information Systems and Decision Sciences, 800 N State College Blvd, Fullerton, California 92831, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

institutions of higher education are increasingly collecting data on student learning behavior to improve educational practices via data-driven decision-making (Elias 2011). Commonly referred to as "learning analytics", they use tools to track students' interaction and engagement with their learning management system (or, LMS) yielding data such as the number of times a student accesses the course website, how long they spend on it, how they navigate through different course activities like lecture slides, assignments, projects, etc. Thus, studying such factors in conjunction with other background factors has the potential to offer useful insights into academic performance that can be used not only to devise intervention techniques but also to inform instructors about how to prepare course materials, design class delivery methods, choose educational technologies, and develop sensitivity to differing student learning preferences in ways to enhance student success in bottleneck courses, thereby reducing the time to graduation.

An important set of resources that play a critical role in student success efforts are academic support services and intervention programs that are particularly designed to help students who have difficulty in certain courses. Some such programs are remedial courses, tutoring services, and Supplemental Instruction (SI). Although available for free, one concern is that a majority of students who need to use them do not do so. The success of one program, SI, is well-known in terms of improving student performance and increasing passing rates in several disciplines such as STEM (science, technology, engineering and mathematics) fields, humanities and business (Martin and Arendale 1992; Mitra and Goldstein 2018). However, attendance at SI sessions is often low in some fields, of which one is Business. This necessitates that research studies be undertaken in order to understand factors that drive consistent participation of students in them. This is the key to increasing participation in such sessions, which in turn can increase student success rates in difficult bottleneck courses and facilitate on-time graduation.

In this article we present a comprehensive research study involving a set of diverse factors to investigate their role in determining student success in a bottleneck Business course at a large public university in the western United States. This study builds upon an earlier study (Mitra and Goldstein 2015) which discovered certain background academic and demographic factors that contributed to the risk of failing such a course and used the insights to design a two-stage intervention method for at-risk students. The current study considers several additional variables representing different aspects of students' cognitive abilities and behavior as well as learning analytics that were not included in the prior study. Specifically, we also isolate factors that affect performance of low-performing students who need the most assistance to succeed and compare with those for high-performing ones. Further, we explore which subset of factors help predict a student's willingness in seeking additional help and support meant to help improve student performance and success in such a course. The findings from our analyses conducted using advanced statistical and data mining methods will thus significantly contribute toward improving performance and success in this difficult course for a diverse group of students.

The rest of the article is organized as follows. We start with an overview of existing literature on factors contributing to student success and participation in academic support programs like SI, followed by our research goals, data, methods and results. We conclude with a discussion of the practical implications, limitations and opportunities for future research.

2. Background and literature review

The level of success students achieve in their years at college has far-reaching implications for students' personal and professional lives, persistence in elected majors, perseverance in higher education, and their success in career. However, disaffection with low performance in introductory college classes is a serious problem at colleges and universities nationwide (Horn and Premo 1995; Horn, Peter, and Rooney 2002). For more than two decades, research has shown that student success in STEM disciplines is most negatively affected by students' lack of success in gateway courses that develop essential skills and introduce students to disciplinary studies (Tobias 1990; Seymour and Hewitt 1997) but often have high failure rates, thus causing a "bottleneck" in the graduation timeline. The problem is especially evident in introductory business, mathematics, and science courses that are often required and are integral components of an undergraduate curriculum. Yet, many students who enroll in these courses achieve moderate or low levels of success in them, resulting in delayed graduation and attrition of students (Gainen 1995).

Much research has been conducted on identifying factors that affect student performance and success in bottleneck courses, from demographic ones such as, gender, ethnicity (Herndon and Moore 2002; Brower and Ketterhagen 2004) to academic factors like GPA, prior academic history and class attendance (Romer 1993; Devadoss and Foltz 1996). Similarly, noncognitive factors or qualitative factors have been used successfully to predict grades in many courses in several areas of study. One such factor is *academic self-concept*, defined as a person's general composite view of themselves toward their own academic skills, knowledge, and capabilities based on past experiences and self-evaluations (Snow, Corno, and Jackson 1996; Byrnes 2003). Although such factors are often overlooked (Glesne 1999), some studies have shown that these variables are more useful than cognitive variables in predicting academic success of nontraditional students (e.g., Sedlacek 2004). Meece, Anderman, and Anderman (2006) found a relationship between student motivation and academic self-concept in introductory Mathematics courses. Academic self-concept was shown to be a strong predictor of persistence in undergraduate math programs and in fact, a stronger predictor of final grades for females than for males (House 1995).

Although research on student success has been conducted for a long time, the widespread availability of data that are collected by educational institutions have recently paved the path for application of more rigorous data mining techniques in order to gather deeper insights about student performance and success in order to facilitate decision-making by students, teachers and administrators (Dutt, Ismail, and Herawan 2017). This field, commonly known as Educational Data Mining (EDM), is growing. Fatima, Fatima, and Krishna Prasad (2015) conducted a survey of EDM tools to develop models for improving student experiences. Manjula (2018) implemented k-means clustering algorithm for the purpose of analyzing an educational dataset. Kaur and Bathla (2018) applied Naïve Bayes and Support Vector Machines (SVM) to predict student performance based on data obtained from Kaggle. Bakhshinategh et al. (2018) presented a survey of various EDM techniques and applications that have existed over the last 10 years.

2.1. Personality traits

Personality has been defined in several ways but according to the American Psychology Association (APA), personality refers to individual differences in characteristic patterns of thinking, feeling, and behaving (American Psychological Association (APA) 2018). Various personality assessment tools have been studied in the literature, but the most important one is the Big Five Inventory (BFI) model (Goldberg 1992; Saucier and Goldberg 1996), which we employ in this study. The Big Five personality dimensions include the following:

- *openness:* this trait is associated with imagination and curiosity, and makes people more tolerant, creative, and interested in variety;
- *conscientiousness:* this trait is associated with efficiency, preparedness, and dependability and makes people work hard, be disciplined, and be achievement-oriented;
- *extraversion:* this trait is linked with sociability and positive affect, and makes people fun-loving, assertive, and optimistic;
- *agreeableness:* this trait is linked with kindness and sensitivity, and makes people sympathetic, trusting, supportive, and nurturing, and
- *neuroticism (or emotional instability):* this trait is associated with the ability to control one's emotions and people who are highly neurotic tend to be anxious, temperamental, and prone to emotional distress.

There is evidence from earlier research studies that personality traits can predict academic achievement (Costa and McCrae 1992). According to O'Connor and Paunonen (2007), students' behavioral tendencies are often reflected in their personality traits that can in turn influence certain behaviors related to academic performance such as perseverance and conscientiousness. A number of works has shown that conscientiousness affects academic motivation and success (Wagerman and Funder 2007; Clark and Schroth 2010). Often academic motivation is seen to mediate the relationship between personality traits and academic achievement (Hazrati-Viari, Rad, and Torabi 2012). Payne, Youngcourt, and Beaubien (2007) found that students who have high levels of conscientiousness, extraversion and openness have greater motivation to learn and hence perform well whereas Komarraju, Karau, and Schmeck (2009) observed that students with low levels of extraversion and high levels of neuroticism have low motivation in studies and do not perform well. In fact, many education researchers believe that neuroticism has a negative relationship with academic motivation and performance and the other three traits have a positive relationship with these factors (Komarraju and Karau 2005).

2.2. Learning styles

The idea that students have different learning styles and preferences is now widely accepted (Wratcher et al. 1997). It is important that educators today are aware of these diverse learning styles in the student population which have a direct impact on their performance in college. Yet not all faculty take these into account while teaching. Hence, teaching methods adopted by instructors are not always effective, particularly across different class formats like face-to-face and online. For instance, research suggests that online classes are best suited for students who are self-motivated, self-disciplined and have effective time management skills (McEwan 2001; Huber and Lowry 2003). This suggests that students who lack these characteristics yet enroll in online courses, may struggle and subsequently drop these courses, delaying their degree completion and graduation. It is thus essential that instructors have knowledge of students' learning styles and other cognitive factors and their impact on performance in order to design and deliver effective instruction for a diverse group of students (Lo and Shu 2005).

Individuals learn in different ways, like watching, listening, questioning, doing, and helping others to learn (Rogers and Freiberg 1994). Different learners process information differently because of different cognitive and learning styles, and exposure to different learning environments and experiences (Scarr 1992; Cassidy 2004). An individual's learning style will affect how information is processed during learning and thinking, thereby impacting learning effectiveness and efficiency (Bencheva 2010). One of the most broadly known learning style inventories used in education research today is *Kolb's Learning Style Inventory* (*LSI)* (Kolb 1981; 1986), which we focus on in our study. The LSI posits learning is a four-stage circular process, and classifies learners as convergers, divergers, assimilators and accommodators. Each of the four stages correspond to four broad types of learning styles: (1) Concrete experience – CE (feeling), (2) Abstract conceptualization – AC (thinking), (3) Active experimentation – AE (doing), and (4) Reflective observation – RO (watching). CE and RO constitute the "diverging" style of learning which enables a person to look at things from different perspectives and learn by watching and viewing concrete situations (rather than doing). People with this style typically have broad interests and like to gather information. The "assimilating" learning style is comprised of AC and RO, which utilizes a concise and logical approach via clear explanations to understand ideas and concepts rather than practical experiences. AC and AE characterize people with a "converging" learning style prefer to solve problems and find practical uses for the concepts and theories learned. Lastly, the "accommodating" learning style encompasses CE and AE and is hands-on, relying on intuition rather than logic. People with this style often use other people's analysis and adopt a practical and experiential approach.

A substantial literature in business education studies the effect of learning styles on student performance. Kolb (1981) found that learning styles of over 800 managers and business graduate students varied with their undergraduate major. Other research found finance and accounting majors prefer a convergent learning style while marketing and management students were mostly experience-oriented (accommodators and divergers) (Bergevin 1993). Students in accounting, finance, marketing, and management information systems preferred an assimilator style using the Kolb LSI (Loo 2002), while visual and kinesthetic styles were preferred for introductory accounting students (Shoemaker and Kelly 2015). On the other hand, a study of over 400 undergraduate students in a financial management class using Kolb's LSI found that learning styles did not impact student performance (Fox and Bartholomae 1999).

2.3. Goal orientations, motivational factors, and other cognitive factors

Goal orientations are cognitive frameworks that influence motivation in achievement settings (Elliot and Dweck 1988). Two types of goal orientations have implications for the tasks students choose, how they approach these tasks, how they react to outcomes, and ultimately, what they learn: (1) *learning (mastery) goals*, which focus students' attention and effort on developing and improving their competencies, and (2) *performance goals*, which focus students' attention and effort on demonstrating their competencies and outperforming their peers. Learning and performance goal orientations can be crossed with approach and avoid goals to form the 2 x 2 model of achievement goal orientations which include: (i) learning approach goals (LGO), (ii) learning avoid goals (LAGO), (iii) performance approach goals (PGO), and (iv) performance avoid goals (PAGO) (e.g., Elliot and McGregor 2001; Kozlowski and Bell 2006). For instance, learning avoidance goals focus attention and effort on avoiding skill losses. Performance avoidance goals focus on avoiding the appearance of incompetence or avoiding doing worse than one's peers. Research with college students has shown learning approach goals predicted intrinsic interest, while performance approach goals predicted grades (Elliot and Church 1997; Elliot and McGregor 2001).

For students to succeed in a course, motivational factors play a significant role. Self-determination theory posits motivation can be assessed along an intrinsic-extrinsic continuum (Deci et al. 1991; Ryan and Deci 2000). Intrinsically motivated behaviors include task engagement the individual finds genuinely interesting and inherently satisfying. An autonomous motivational style toward the intrinsic end of the continuum and is thus characterized by self-regulated and self-determined motivation. On the other hand, a controlled style lies toward the extrinsic end of the continuum and is experienced as motivation driven by external factors rather than inherent interest. In educational settings, students with an autonomous motivational style are more likely to engage in activities that enhance their learning while individuals with a controlled motivational style are more inclined to focus their effort and attention on activities that can improve their grades (Black and Deci 2000; Ryan and Deci 2000).

Among other meta-cognitive factors that have a potential impact on academic achievement of students, the most important ones are self-regulated learning such as intrinsic value (Pintrich and De Groot 1990). Self-regulated learning may include students' meta-cognitive strategies for planning, monitoring, and modifying their cognition (Zimmerman and Pons 1988), their management of effort on various tasks and activities pertaining to the course, and the cognitive processes they use for learning, understanding and mastering course content. Intrinsic value is also an important component of students' own choice or interest in becoming cognitively engaged in their academic work. A study of science learners showed that both self-regulated learning and intrinsic value were positively related to students' engagement and academic performance (Pintrich and De Groot 1990).

Finally, procrastination, defined as a voluntary, irrational delay of behavior, has been found to impair individual academic and work group performance (Jiao et al. 2011; Lakshminarayan, Potdar, and Reddy 2012). Groups with the lowest levels of achievement tended to include students who procrastinated most frequently on administrative tasks and academic work. Procrastination is widely prevalent among college students across the world, and hence an important factor to consider in a comprehensive study of student success. Interventions such as stimulus control, stimulus cues, and cognitive restructuring have been successful in remediating procrastination (Rozental and Carlbring 2014).

2.4. Learning analytics

With the advance of technology in the last decade, we have the emergence of a field called *learning analytics* in which sophisticated analysis are performed to improve learning and education (Elias 2011). Learning analytics is defined by the Society for Learning Analytics Research (SOLAR) as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (SOLAR website). It uses tools and methods from business intelligence, web analytics, statistics and data mining to analyze large repositories of data available at colleges and universities, mostly through integration with their Learning Management Systems (LMS). Such vast amounts of data on student behavior may include user visits, attempts on assignments, number of downloads, LMS tools accessed, and number of contributions to a discussion forum, that can provide the instructor with real-time information to improve student engagement (Educause 2010; Beer, Clark, and Jones 2010). This helps the instructor to track and determine which students may need help and when so that he or she may reach out to those students via email as a way of intervention. This also helps the development of personalized learning environments to refine course offerings for users.

An Educause survey conducted recently also showed that institutions are more commonly using learning analytics to monitor and study student progress and success although there are concerns regarding data quality, difficulties in system integration, and ethical issues (Arroway et al. 2016) while Borden and Coates (2017) discussed how learning analytics can be used effectively to provide insights into collegiate experience via the development of a conceptual framework and advancement of institutional systems. Many studies have also focused on the correlation between students' LMS usage and their performance (Filippidi, Tselios, and Komis 2010; Jo, Kim, and Yoon 2014) as well as with their satisfaction with courses offered via LMS (Naveh, Tubin, and Pliskin 2012). Kesahavamurthy and Guruprasad (2014) presented a survey of research conducted on learning analytics along with several applications to predict student performance and detect at-risk students. According to MacFayden and Dawson (2010) predictive models for monitoring student achievement should be built at the course level, and they were able to identify struggling students in need to academic support based on LMS data. Whitmer (2012) observed that learning analytics explained over four times the variation in final grades compared to traditional student characteristics. Yu and Jo (2014) predicted students' learning achievement through analyzing data from Moodle LMS for 84 students at a university in South Korea. Mwalumbwe and Mtebe (2017) discovered that time spent in the LMS and number of logins and downloads did not have any significant impact on students' learning performance in blended courses, while discussion posts and peer interaction through the LMS did. West, Searle, and

Price (2018) explored the use of learning analytics to enhance and improve teaching practices based on information regarding student engagement and motivation.

2.5. Supplemental Instruction (SI) and help-seeking behavior

Research studies have shown that Supplemental Instruction (SI) is effective at improving academic performance of students who participate (Ogden et al. 2003; Malm, Bryngfors, and Morner 2010; 2011), with the greatest impacts seen on students who are at risk (Shaya, Petty, and Petty 1993) and ethnic minority students (Wilcox and Koehler 2007).

Moreover, SI has been found to significantly enhance reasoning and problem-solving skills (Shaya, Petty, and Petty 1993) and increase retention (Lyle and Robinson 2003), GPA (Fayowski and MacMillan 2008), and graduation rates (Congos 2003; Bowles, McCoy, and Bates 2008), specifically in STEM-related fields (Peters et al. 2007). In addition, SI provides students the training necessary to learn and assimilate study skills that they take with them to future semesters (Peters et al. 2007; Fayowski and MacMillan 2008). In the context of Business courses, Kenney (1989) and Kenney and Kallison (1994) conducted several studies that demonstrated the effectiveness of SI in first-semester calculus classes for business and economics majors at a university. Mitra and Goldstein (2018) recently discovered a positive impact of SI on the performance and success of students in Business courses, particularly for at-risk students. In fact, the success of this program led the United States Department of Education (USDOE) to designate it as an "*Exemplary Educational Program*" in 1981 (Burmeister 1996; Widmar 1994), which resulted in widespread SI implementation throughout the country. SI is one of the academic success programs that is expected to benefit students in bottleneck courses, which we study in this article.

Despite the extensive study of the impact of SI, participation in SI sessions, particularly regarding consistent participation, is understudied in the current literature. Yet, this is important to understand because an inherent problem for any academic assistance program in colleges (such as tutoring or SI), is that these are voluntary and require initiative on the part of the students. Hence it is important for academic advisors to have some insights and understanding of how students seek help voluntarily and also how intervention impacts the likelihood of participation. This will ensure that students are utilizing all available resources that will help them achieve academic success. *Help-seeking* has been identified as one of the processes of self-regulated learning (Zimmerman 1998; Newman 2002) and is typically exhibited by students with high levels of self-efficacy and social skills. Schunk and Ertmer (2000) pointed out that help-seeking is a complex process that involves several dimensions–cognitive and motivational. Karabenick and Knapp (1991) looked at help-seeking in higher education settings and found that help-seeking could be properly viewed as an achievement-related learning strategy rather than as a manifestation of dependency. Magnusson and Perry (1992) found that help-seeking in college students was influenced by both classroom factors and teacher characteristics. Ryan, Gheen, and Midgley (1998) mentioned that help-seeking behavior was positively correlated with high self-efficacy. Often students who are at risk of failing are unable to seek proper help; according to Hodges and White (2001), such students may not be as capable of understanding the need to ask for help when needed as those students who are not detected to be at risk. The literature also contains evidence that sometimes interventions can be effective at improving help-seeking behavior in such students (Dweck 1986). This is particularly pertinent and important for at-risk students and has not been studied comprehensively for academic support programs like SI.

3. Research goals and methods

The goal of this article is to form a deep understanding of which factors affect the performance of students in a bottleneck Business course via statistical models that are able to provide very useful insights which can in turn be used to improve performance for students to a considerable extent. In particular, we consider a diverse set of factors, from students' background demographic and academic factors to cognitive factors like learning and motivational styles, help-seeking behavior to personality traits and learning analytics. Such a comprehensive and rigorous study, to the best of our knowledge, has not been undertaken before for such a course. Our specific study objectives as follows:

1. identify which of these cognitive, behavioral, and engagement factors affect student performance;
2. identify which of the above-mentioned factors particularly have an influence on the performance and success of low-performing students, and how this differs from factors affecting the performance of high-performing students;
3. identify which of these factors are associated with students' propensity to utilize extra resources available to aid at-risk students (such as, Supplemental Instruction or SI);
4. use the above outcomes to build statistical predictive models for student success and participation in support programs.

The results from the above analyses will thus provide the huge benefit of informing instructors and advisors in the university beforehand which subgroups of students are likely to succeed in this difficult course and which subgroups are likely to regularly take advantage of additional resources (like SI) to get help during the course, so that suitable steps may be taken to address them. Such outcomes will go a long way in diminishing the bottleneck in such courses and leading to shorter overall graduation time for students enrolled in a degree program.

3.1. Data collection and study sample

167 undergraduate students enrolled in a business statistics course over one semester participated in this study. Students belonged to 4 different sections, taught by 2 different instructors. All course sections used the same learning management system (LMS) and had the same course structure in terms of content, schedule, assignments, and exams. Of these four sections, two were offered in an online format. There were no significant differences among the four sections with respect to the basic demographic composition. The traditional course sections used about 150 minutes of class time each week for

lecture and in-class problem-solving activities, whereas the online classes required students to master the course content on their own, with on-campus meetings focused exclusively on conducting exams. These exams were held on campus and proctored by the instructor just like the face-to-face sections, thus removing any bias in that respect. Out of the total of 167 students, 86 (51.5%) belonged to face-to-face sections and 81 (48.5%) to online sections, an almost equal representation from the two different class formats.

Data on students' background demographic and academic background were obtained from the university's office of institutional research, whereas data on the cognitive and behavioral variables were obtained via a survey questionnaire that was posted as a link on the course website (with IRB approval). Out of the total of 167 students, 128 voluntarily completed this survey yielding a response rate of 76.7%. The sample size for our analyses was thus $n = 128$.

3.2. Measures

We assessed personality traits, help-seeking behavior, learning styles, motivational styles, goal orientations, meta-cognitive strategies, procrastination, academic self-concept, and class performance as the main predictor variables in this study. Data on students' personality traits were collected using the 44-item Big Five Inventory (BFI) model (Goldberg 1992; Saucier and Goldberg 1996). The responses to the item questions followed a 7-point Likert scale with the following categories: 1 = strongly disagree, 7 = strongly agree. Extraversion was measured using 8 item questions from the inventory (3 with reversed scales), one representative question being "*I see myself as someone who is talkative*". Agreeableness was measured using 9 questions (4 with reversed scales), one typical question being "*I see myself as someone who is helpful and unselfish with others*". Conscientiousness was measured using 9 questions (4 with reversed scales), a typical question being "*I see myself as someone who does a thorough job*". Neuroticism was measured using 8 questions (with reversed scales for 3 questions), a representative question was "*I see myself as someone who can be moody*." Finally, openness was measured with responses to 10 questions (2 with reversed scales) and a typical question for this trait was "*I see myself as someone who is original and comes up with new ideas*". The final value for each of the five personality traits for each student was obtained by averaging the responses to the individual questions (after considering the reversed scales for the relevant questions), thus yielding a quantitative or continuous variable for our models.

Following Ames and Lau (1982), we devised a 6-item Likert scale with responses on a 5-point scale (1 = strongly disagree, 5 = strongly agree) in order to assess individual differences in help-seeking behavior. The items, which are listed below, were designed to measure students' perceptions of threat in seeking help from both personal and public sources (others, professors): (a) Even if I have trouble learning the material in this class, I try to do the work on my own, without help from anyone; (b) I ask the instructor to clarify concepts I don't understand well; (c) When I can't understand the material in this course, I ask another student in this class for help; (d) I try to identify students in this class whom I can ask for help if necessary; (e) I try to take advantage of resources available on campus (like SI, advisor, tutoring) if I need help; and (f) If I seek help in class, my teachers and peers will think negatively about me.

Learning styles of students were measured with a short form of Kolb's learning style inventory (Kolb 1981; 1986) – 7 items for reflective observation and active experimentation (ROAE), and 5 items each for abstract conceptualization (AC) and concrete experience (CE). An example from ROAE scale is "when I learn, I like to think about ideas." An example from the AC scale is "I learn best when I trust my hunches and feelings", and one from the CE scale is "when I learn, I like to watch and listen." All scales used 5 points (1 = strongly disagree, 5 = strongly agree). We measured motivational style with 8 items from the self-regulated learning scale (Black and Deci 2000; Ryan and Deci 2000), four of which focus on autonomous motivation (AUTREG) and four on controlled motivation (CONREG). An example from the autonomous scale is "I would like to participate actively in class because a solid understanding of statistics is important to my intellectual growth", while an example from the controlled scale is "I would like to actively participate in class because I would feel proud of myself if I did well in the course" (1 = strongly disagree, 5 = strongly agree).

We measured goal orientations using a seven-point scale (1 = not at all true of me, 7 = very true of me) consisting of 12 items and four subscales with three items each to assess learning approach goals (LGO), learning avoid goals (LAGO), performance approach goals (PGO), and performance avoid goals (PAGO) (Elliot and McGregor 2001). An example learning approach goal item is "I want to learn as much as possible from this class". For the learning avoid goal, an example question is "I worry that I may not learn all that I possibly could in this class." A performance approach goal example question is "it is important for me to do better than other students" while a performance avoid goal example is "I just want to avoid doing poorly in this class".

Intrinsic value (INTRVAL) consisted of a nine-item scale, an example question being "I like what I'm learning in this class." We measured procrastination (PROC) using 6 items from the 16-item Tuckman procrastination scale (Tuckman 1991) with a sample question being "when I have a deadline, I wait till the last minute". For the latter two variables, the same scale as before was used (1 = strongly disagree, 5 = strongly agree).

Academic self-concept was measured using three different questions to understand differences among student's self-perceptions about their ability to attain academic success – (i) "what is your level of interest and motivation in this course" which was measured using a 5-point Likert scale (1 = not interested, 5 = very interested), (ii) "what is your level of self-confidence in your overall intellectual ability in mathematics and quantitative courses in general", measured using a 5 point Likert scale (1 = very low, 5 = very high), and (iii) "how academically prepared do you think you were at the beginning of the course" measured using a 3-point scale (1 = not prepared, 2 = somewhat prepared but lacking important skills and knowledge, 3 = well-prepared).

Lastly, we describe the response variables. Class performance was measured as a continuous variable based on the final course grade for each student, measured on a standard percentage grading scale (0 to 100). Students' success was measured by a binary variable which assumed the value 1 for a student who obtained a letter grade of C or above in the course, and the value 0 for a student who obtained a grade lower than C or dropped/withdrew from the course. We call this variable PASS. Students' propensity to take advantage of campus resources to improve performance in difficult courses is measured by participation in Supplemental Instruction (SI) sessions. Since the benefits of SI is not fully realized if a student only attends them a couple of times (typically just before exams), we measure consistent participation as attending at least 3 sessions during a semester. We call this variable "SI attend". So "SI attend" takes the value 1 for a student who attends 3 or more SI sessions during a semester, and the value 0 otherwise. All measurement scales demonstrated adequate reliability ($\alpha > .80$).

Control variables included gender (1 = female, 2 = male), ethnicity (categories: White, Hispanic, Asian/Pacific Islander, International, Others), first generational college status (1 = Yes, 0 = No), transfer status (1 = first time freshman, 2 = transfer student), class format (0 = traditional face-to-face; 1 = online), self-reported GPA, and age (the last two included as continuous variables).

3.3. Analysis and methods

Descriptive statistics are computed first in order to summarize the distribution of the main study and control variables, frequency distributions for the categorical variables and means and standard deviations for continuous variables. A correlation analysis was also performed to assess the level of association among the different study variables to determine potential multicollinearity issues. Some graphical representations are also used to visualize the pattern of distribution of certain variables, particular those exhibiting trends over time.

In order to identify the variables that affect students' chances of success in this bottleneck course and in order to determine variables that have an effect on students' consistent participation in SI sessions during a semester, we used *logistic regression* (Hosmer and Lemeshow 2000). For the success model, the binary variable PASS is used as the dependent or response variable, while for the SI participation model we use "SI attend" as the dependent variable. The individual test statistic used to test for the significance of each coefficient in logistic regression models is called the "Wald" statistic and has a chi-square distribution with one degree of freedom.

Quantile regression (QR) is a statistical technique that is used to estimate and conduct inference about quantile function as opposed to means that are modeled in multiple linear regression (Koenker and Hallock 2001). This type of regression is commonly used in econometrics applications. The main advantage of quantile regression over linear regression is that it provides the ability to model the entire family of quantile functions from in the distribution of the response variable (instead of a single mean), thus yielding a more comprehensive statistical analysis of the relationships among different variables. In this study we use quantile regression to determine the specific groups of factors that affect student performance in this bottleneck course for different groups of students such as, low-performing, high-performing, and so on. These student groups are represented by the quantiles of the distribution of the final grades in this course. Thus, the insights obtained from these models will yield an understanding of which factors particularly are the most relevant in determining success for low-performing students, which is important for the purpose of devising intervention and support services to provide additional to those students to help them be successful.

Table 1. Descriptive statistics summary of the background (control) variables used in the study.

Variables	Frequencies	Relative Frequencies
Gender (n = 128)		
Male	69	53.9%
Female	59	46.1%
Transfer Status (n = 128)		
First time freshmen (FTF)	31	24.2%
Transfer	97	75.8%
Ethnicity (n = 128)		
White	34	26.6%
Hispanic	33	25.8%
Asian/Pacific Islander	37	28.9%
International	8	6.3%
Others	16	12.5%
First generational status (n = 128)		
Yes	31	24.2%
No	97	75.8%
Class format (n = 128)		
Face-to-face (traditional)	63	49.2%
Online	65	50.8%
GPA	Mean = 3.15, SD = 0.37	
Age	22.5 years, SD = 3.67 years	

Finally, classification models based on logistic regression and Random Forests (Breiman 2001) are employed to predict which students are likely to succeed in this course and which ones are most likely to take advantage of SI as an academic support service. *Random Forests* (RF) are an ensemble learning model for performing classification tasks that operates by constructing multiple decision trees during training and using majority voting to combine the decisions from the individual trees on the test to output a final overall outcome. We further compare our results with other common classifiers used in data mining (Shmueli, Bruce, and Patel 2016) such as Naïve Bayes, Decision Trees and Support Vector Machines (SVM). All of these models are built based on the factors identified earlier via the logistic regression and quantile regression models, and the best predictive model in each case is selected as the one that has the lowest Akaike's Information Criterion or AIC value (Anderson 2008). The accuracies of the classification models are evaluated using two types of cross-validation methods. First, we divide the entire dataset into a training set (65%) and a test or validation set (35%). The random partition is repeated 20 times in order to remove selection bias, and the final model accuracies are computed by averaging over the results from the various repetitions. Second, we use the leave-one-out (LOO) method whereby the classifier is trained on all the data except for one record which is used for testing. This is repeated for all data records and the final model accuracies are averaged over all the repetitions. The classification results are presented as the overall misclassification error rates (%) calculated on the test sets in both cases.

However, for both of our models, the datasets are somewhat unbalanced since the classification categories are not approximately equally represented. For the student success model, the number of students who failed is significantly lower than those who passed, and for the SI attendance model, the number of students attending SI sessions consistently during a semester is also considerably low. We thus decided to employ an over-sampling technique called "Synthetic Minority Over-sampling Technique

Table 2. Means and standard deviations of the study variables – cognitive and behavioral variables. The sample size $n = 128$.

Variable	Mean	Std. dev.
Learning Styles		
ROAE	4.27	0.51
AC	4.31	0.67
CE	4.17	0.71
Motivational factors		
AUTREG	3.57	0.71
CONREG	3.28	0.67
Goal orientations		
LGO	4.99	1.36
LAGO	4.53	1.49
PGO	4.00	1.72
PAGO	5.33	1.28
Self-regulated learning		
INTRVAL	3.71	0.62
PROC	2.65	1.05
Personality traits		
Openness	4.71	0.75
Neuroticism	3.83	1.10
Conscientiousness	4.99	0.91
Agreeableness	5.32	0.88
Extraversion	4.37	1.09
Help-seeking	3.15	0.73
Academic self-concept	2.78	0.62

(SMOTE)" which was introduced by Chawla et al. (2002) to address this issue of imbalance in datasets. This is because the classifier is often not able to learn the characteristics of the class which has relatively fewer observations (called "rare" or "minority" class) as well as it does for the other class with a large number of observations (called "majority" class), thus leading to poor classification outcomes for new test data belonging to the rare class. This results in low F_1 scores for the classifiers also. Our goal was thus to employ SMOTE to obtain a relatively "balanced" sample for each of our predictive models using the same classifiers. Accuracies of the models are computed in each case via both the LOO and training/test set methods.

4. Results

A summary of the demographic distributions of the study sample is included in Table 1. Nearly 54% of the students enrolled in these classes are men and 76% are transfer students. Regarding ethnic composition, we saw almost equal proportion of Asian, White and Hispanic students. Moreover, about 24% of students are first generational college students. Of the 128 students, 51% were enrolled in the two online sections and 49% in the two traditional face-to-face sections. The average (self-reported) GPA of these students who participated was 3.15. The average age of the student participants was 22.5 years.

The means and standard deviations of the different cognitive and motivational variables, personality traits, and academic self-concept are included in Table 2, and those of the outcome variables (grades, SI attendance) appear in Table 3. Of the 128 students, 89% passed the course whereas 11% did not. Further, 16.4% attended at least one SI session during the semester.

Table 3. Descriptive statistics for the study variables (outcomes). The sample size $n = 128$.

Variable	Mean	Std. dev.
Final grades	80.54%	9.46
SI session visit	1.05	3.89
SI attendance	**Frequencies**	**Relative Frequencies**
Yes	21	16.4%
No	107	83.6%
PASS		
1	114	89.0%
0	14	11.0%

Figure 1 shows the trend in the pattern of students' interaction with the LMS during the entire course of the semester (16 weeks). It is clearly seen that the weeks around examination times have prominent spikes (Weeks 5–6, Week 12, Week 16). The overall average mean number of LMS visits for all students is 25.51, while that for online classes alone is 26.83 and face-to-face classes is 24.26. The online class engagement is slightly higher than face-to-face as expected, however the difference is not statistically significant at the 5% level ($p = 0.153$).

Initial exploratory analyses revealed that a few of the background variables like ethnicity, transfer status and age had very weak association with either of our outcome variables (student success and SI participation), hence were not included in our models. This helped somewhat reduce the complexity of our models that already contained a large number of study variables.

4.1. Quantile regression models

The dependent or outcome variable used for the quantile regression models is the final course grades (continuous variable). We used the quantiles 0.2 to 0.8 to represent the overall spectrum of performance, from low-performing (lower quantiles) to high-performing (higher quantiles) students, and thereby study the effects of different factors on these different sub-groups.

Bivariate correlations among some of the cognitive and behavioral variables were found to be significant, which were: (i) CE and ROAE ($r = 0.68$), (ii) AC and ROAE ($r = 0.83$), (iii) PROC and conscientiousness ($r = -0.61$), (iv) AUTREG and INTRVAL ($r = 0.72$), (v) LGO and AUTREG ($r = 0.62$), and (vi) LGO and INTRVAL ($r = 0.61$). Hence Variance Inflation Factors (or, VIFs) were used to examine multicollinearity issues, the correlation among the predictor variables, in the models (Kutner, Nachtsheim, and Neter 2004; Sheather 2009) that is known to adversely affect regression results. VIF is an index that measures the severity of multicollinearity, and the square-root of VIF for a certain independent variable can be interpreted as the extent to which the standard error of the model is inflated due to correlations with other independent variables in the model. None of the VIFs for our study variables exceeded 3.5 (typical threshold used is 5), hence none of them was eliminated on account of multicollinearity.

The model results are shown in Table 4. They show the p-values of the different predictor variables for the seven different quantile regression models. Due to space constraints, we do not include the detailed model output for the seven models. The p-values are the most important quantities that demonstrate the strength of the effects

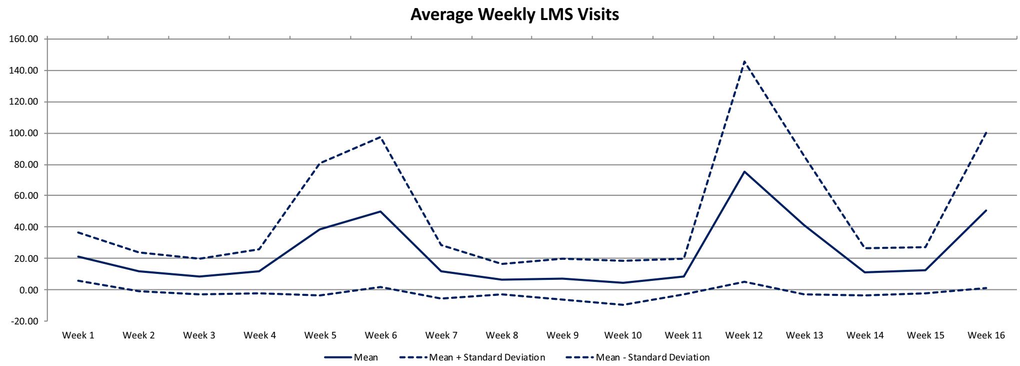

Figure 1. Line graph showing the trend of average LMS visits for each of the 16 weeks during the semester. The solid line represents the mean and the dotted lines represent the confidence bands. The overall mean number of LMS visits was 25.51 with a standard deviation on 10.11.

Table 4. *p*-values from the 7 quantile regression models (Q = 0.2 to Q = 0.8).

	Q = .2	Q = .3	Q = .4	Q = .5	Q = .6	Q = .7	Q = .8
Gender	0.72	0.28	0.37	0.56	0.48	0.47	0.67
First Generation	0.75	0.84	0.96	0.48	0.83	0.52	0.54
Class format	0.03**	0.02**	0.01**	0.01**	0.01**	<0.0001***	<0.0001***
GPA	<0.0001***	<0.0001***	<0.0001***	<0.0001***	<0.0001***	<0.0001***	<0.0001***
LMS Visits	0.69	0.91	0.81	0.79	0.83	0.74	0.55
SI Attendance	0.60	0.23	0.19	0.11	0.17	0.15	0.05*
Help-Seeking	0.54	0.57	0.75	0.86	0.97	0.73	0.72
Openness	0.19	0.11	0.35	0.37	0.58	0.38	0.39
Neuroticism	0.36	0.50	0.77	0.42	0.31	0.54	0.68
Conscientiousness	0.22	0.28	0.61	0.76	0.88	0.91	0.94
Agreeableness	0.85	0.91	0.78	0.61	0.28	0.40	0.51
Extraversion	0.96	0.92	0.58	0.79	0.88	0.84	0.98
CE	0.39	0.18	0.37	0.13	0.08*	0.55	0.69
AC	0.43	0.40	0.48	0.45	0.53	0.41	0.46
ROAE	0.33	0.51	0.51	0.54	0.74	0.54	0.66
PROC	0.72	0.76	0.98	0.48	0.54	0.78	0.75
INTRVAL	0.04**	0.01**	0.05*	0.16	0.40	0.70	0.79
AUTREG	<0.0001***	<0.0001***	0.04**	0.07	0.22	0.39	0.46
CONREG	0.90	0.31	0.33	0.51	0.68	0.51	0.81
PGO	0.03**	0.01**	0.30	0.78	0.70	0.40	0.78
LGO	0.27	0.37	0.85	0.76	0.84	0.94	0.90
PAGO	0.24	0.28	0.65	0.61	0.38	0.81	0.46
LAGO	0.19	0.31	0.23	0.67	0.75	0.32	0.38
Academic Self-Concept	0.79	0.99	0.69	0.41	0.33	0.44	0.58

*$p < 0.10$,
**$p < 0.05$,
***$p < 0.0001$.

of the different predictor variables on the outcome that can be compared at the different quantiles, and hence are shown. The distribution of the model coefficients, along with the respective 95% confidence bands, for all the models appear in Figure 2. Class status and GPA have a significant association with final grades for all students (*p*-values <0.05 in all cases). The model coefficients show that students in face-to-face classes performed significantly better than those in the online classes whereas students with higher GPA had better performance. Looking carefully at the results, we see that the effect of class status on performance is stronger for high-performing students than low-performing students (*p*-values get lower and lower as the quantile value increases)

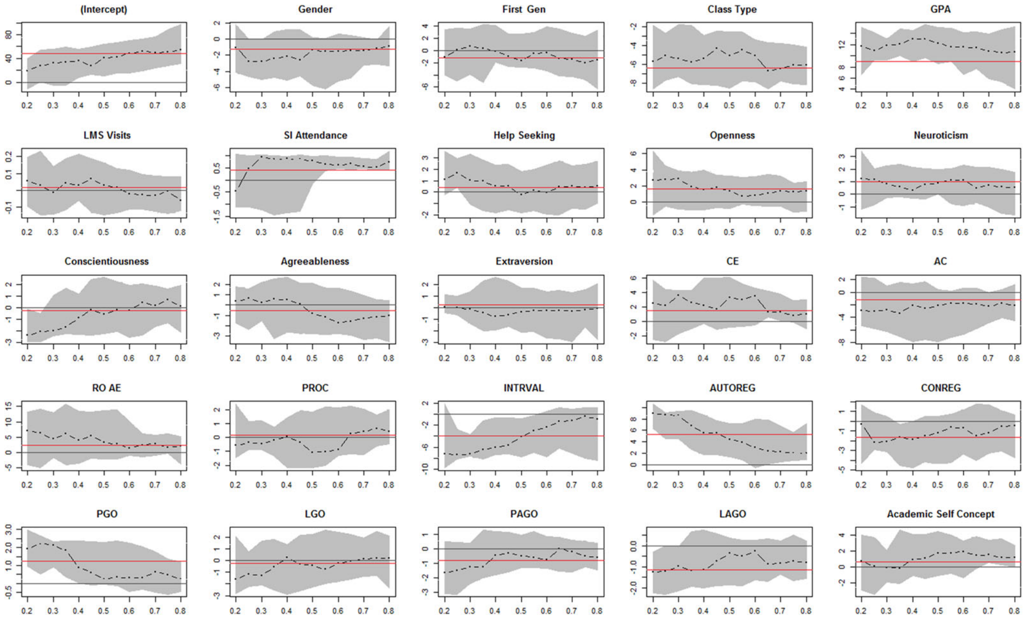

Figure 2. Estimated coefficients and confidence bands from the different quantile regression models. The red solid line is the coefficient of the variables for the OLS (Ordinary Least Squares) regression model and the dotted black line represents the coefficients of the variables at each quantile. The shaded area represents the 95% confidence band for the QR model coefficients.

whereas that of the GPA is uniform across all student groups. Among the cognitive variables, intrinsic value (INTRVL), autonomous regulation (AUTREG) and controlled regulation (CONREG) had significant effect on student performance for low-performing students, but not for high-performing students. The model coefficients show that among the low-performing students, those with higher levels of both autonomous and controlled regulation and intrinsic value performed better. None of the five personality traits had any significant effect on the performance of any of the student groups and neither did the number of LMS visits (learning analytics). SI attendance had a moderately significant effect (p-value $= 0.05$) on the performance of good students only so that the number of SI sessions attended was positively correlated with their performance (more sessions attended implied better performance).

4.2. Logistic regression model for student success

We use PASS as the binary outcome variable for the logistic regression model of student success in this course. We considered several models with different subsets of the predictor variables along with the control variables and used AIC to determine the final model. The model with the highest AIC value was chosen as the best predictive model, shown in Table 5. The strongest predictors of student success among the host of cognitive and motivational variables appear to be concrete experience (CE), autonomous regulation (AUTREG), performance approach goal orientation (PGO) and academic self-concept, all of which have p-values less than 0.05. Among the five personality traits considered in the study, only extraversion has a statistically significant effect on student

Table 5. Logistic regression model output for determining student success. Odds ratio is calculated as exp(β).

Variables	Coefficient (β)	SE (β)	p-value	Odds Ratio
Class Type	−5.82	2.26	0.01*	0.0029
SI Attend	4.38	3.07	0.15	79.7
GPA	6.11	2.60	0.02*	451
CE	3.78	1.65	0.02*	43.9
Extraversion	1.81	0.84	0.03*	6.11
INTRVAL	−5.78	2.94	0.05	0.0031
AUTREG	7.62	3.64	0.04*	2050
CONREG	−4.29	2.35	0.07	0.0137
PGO	2.25	0.97	0.02*	9.45
LAGO	−1.10	0.60	0.06	0.331
Academic self-concept	−4.20	1.81	0.02*	0.015

*$p < 0.05$,
**$p < 0.0001$.

success (p-value $<$ 0.05). Among the control variables, class type and GPA both have significant effects on a student's probability of passing this course. Finally, the variable representing learning analytics of students did not appear to be a significant predictor of student success for this course as it was not even included in the final predictive model.

The odds ratio, furthermore, helps explain the nature of association of the outcome variable with each of the significant predictor variables in the final model. For instance, students with higher levels of autonomous regulation, those whose goal in class is geared more toward performance (that is, performing well in the course is the main goal as opposed to learning) and those who possess a learning style which is based on hands-on and concrete experience are expected to have a significantly higher probability of passing the course. Similarly, students who possess a higher degree of extraversion (that is, have a personality which is more outgoing and social) have a higher chance of being successful in this course. On the other hand, students with higher level of academic self-concept have a significantly lower probability of passing the course than those who have lower levels of self-concept. The latter finding implies that students who perceive to have more interest in the course, more self-confidence in their own intellectual ability and believe they are more well-prepared for the course with the requisite skills are more likely to fail the course. This shows that sometimes there is a disconnect between students' own perception of their own abilities and expectation of performance in a course and their actual performance.

4.3. Logistic regression model for utilizing help and support (SI participation)

For studying students' propensity to utilize additional resources as a means of support in their course, we used "SI attend" as our outcome or dependent variable in the logistic regression model that was introduced earlier. Just as with the model for student success, several models were tested in this case also with different subsets of variables and the best predictive model was selected based on the AIC value. Table 6 shows the model output. We see that the cognitive and behavioral variables that have a statistically significant effect on consistent SI attendance are help-seeking, controlled regulation (CONREG), and learning approach goal (LGO), all p-values being less than 0.05. Among the personality traits, openness has a significant effect on determining whether a student will attend SI sessions.

Table 6. Logistic regression model output for determining students' willingness to seek support by attending SI sessions. Odds ratio is calculated as exp(β).

Variables	Coefficient (β)	SE (β)	p-value	Odds Ratio
Class Type	−1.79	1.15	0.12	0.167
Openness	−1.78	0.84	0.03**	0.169
Neuroticism	−0.83	0.54	0.12	0.434
Extraversion	−1.13	0.63	0.07	0.322
Help-seeking	2.87	1.02	0.01**	17.6
AUTREG	−1.91	1.02	0.06	0.148
CONREG	1.74	0.72	0.02**	5.7
LGO	1.74	0.75	0.02**	5.71
Academic self-concept	−1.07	0.78	0.17	0.345

*$p < 0.05$,
**$p < 0.0001$.

Extraversion and neuroticism are included in the final model, yet their effects are not significant ($p > 0.05$). The only control variable included in the final predictive model is class type whose effect is not significant. Just as with the model for student success, engagement with the LMS (learning analytics) did not come out to be a significant predictor of students' SI participation in the presence of other variables in the study.

The odds ratios from the model indicate that students who have higher levels of controlled regulation, are more likely to ask for help, and have a learning-oriented goal (that is, focused on learning more rather than performance) have higher probabilities of attending at least 3 SI sessions as a means of obtaining help on the difficult course. On the other hand, students that are less open seem to be more likely to seek help in SI sessions that those who have a more open personality. The latter finding is surprising because individuals who are more open-minded are expected to be more willing to try new things (like SI, for example) than those who are less open. However, some of these students may have experienced SI in previous courses and hence they may have been already familiar with it and hence more inclined toward attending them for this course.

4.4. Classification results from the two predictive models

Next, we assessed the predictive accuracy of the two predictive models via the two cross-validation methods described earlier. The results using a cutoff probability of 0.50 for classification on the test set are shown in Table 7. The logistic regression model for student success yielded error rates of 4.95% for the training/test method (45 records included in the test set) and 10.2% for the LOO method. We compared these results with other classification techniques such as Naïve Bayes, SVM, Decision Trees, and RF (built using an ensemble of 1000 decision trees) and these results are also included in Table 7. So, although LOO is expected to be more efficient for smaller datasets, the training/test method yielded higher accuracies for logistic regression and Naïve Bayes ($p < 0.0001$). For the other three methods (SVM, Decision Trees and RF), there was no statistically significant difference between the two sets of results at the 5% level ($p = 0.10$, 0.11, 0.08 respectively). Further, the results also showed clearly that logistic regression using the training/test method outperforms the other methods significantly in this case ($p = 0.01$ for SVM, $p < 0.0001$ for Naïve Bayes, RF and Decision Trees).

On the other hand, the results from the two validation methods were not significantly different for the consistent SI participation model for all the classifiers ($p = 0.19$, 0.07, 0.59,

Table 7. Classification results from the different classifiers for the student success model.

	Error rates	
Classification technique	Training/test set	LOO
Logistic Regression	4.95%	10.20%
Naive Bayes	6.67%	11.02%
SVM	8.00%	10.24%
Decision Trees	6.67%	8.67%
RF	11.1%	8.67%

0.34 and 0.93 for logistic regression, Naïve Bayes, SVM, Decision Trees and RF respectively). These results appear in Table 8. In this case, the minimum error rates were produced by Naïve Bayes and Random Forests (4.72% for both) for the LOO method although they were not significantly different from that of the logistic regression model using the training/test method which was 4.85% ($p = 0.97$ in both cases). Moreover, the low standard deviations in each case (1 – 1.3%) showed that the results were fairly consistent across different repetitions for each cross-validation method and are hence reliable.

Thus, although both models proved to demonstrate considerable accuracy in predicting future outcomes (> 95%) and can be valuable in practical applications, the imbalance in our datasets implied that a new test record was more likely to get classified as belonging to the "majority" class than the "rare" class. Hence, we calculated F_1 scores to accompany the error rates reported earlier. These scores provide a means to assess the accuracy of a test by considering both precision and recall (Powers 2011). The F_1 scores for the logistic regression models (shown in Table 9) based on the training/test validation method proved to be optimal in each case; that is, they were the highest among all the classifiers (0.40 and 0.67 for the student success and SI attendance models respectively). However, these values were quite low.

Table 10 shows the results of applying the technique of SMOTE by over-sampling at 500% to partially obviate the issue of data imbalance in our study for both models. Logistic regression yielded the best results, which we report here. For the student success model, the error rates were 8.5% for both the LOO and the training/test set methods. Although this value was significantly higher than those obtained earlier ($p < 0.0001$) in a statistical sense, the F_1 scores were significantly higher – 0.93 and 0.98 for the two methods. For the SI attendance model, the error rates obtained were 1.3% and 3.4% respectively for LOO and training/test set method using logistic regression. These values are significantly lower than our previous results ($p < 0.0001$ in both cases). The F_1 scores were also over 0.90 for all cases. Thus, the SMOTE technique not only helped us improve the F_1 scores but also enhance the prediction accuracy for one of the models significantly. We henceforth choose the logistic regression model with the training/test set validation method as our optimal student success model and the logistic regression model with SMOTE using the LOO method as the optimal SI attendance model.

5. Discussions and conclusions

This study provided the first comprehensive research into a set of diverse factors and their effect on student performance and success in a bottleneck Business course and on students' likelihood of seeking additional help for their learning by attending support

Table 8. Classification results from the different classifiers for the SI Attendance model.

	Error rates	
Classification technique	Training/test set	LOO
Logistic Regression	4.85%	6.30%
Naive Bayes	6.67%	4.72%
SVM	7.82%	7.09%
Decision Trees	6.67%	7.88%
RF	4.90%	4.72%

Table 9. F_1 scores of the logistic regression-based predictive models for student success and SI attendance.

Predictive Model (Logistic regression)	F_1 scores
Student success	0.40
SI Attendance	0.67

Table 10. Classification results from the logistic regression model using the technique of SMOTE. Error rates and F_1 scores are reported for both LOO and training/test validation methods.

Logistic Regression (using SMOTE)	Error rates (F_1 scores)	
	Training/test set	LOO
Student success model	8.50% (0.98)	8.50% (0.93)
SI attendance model	3.4% (0.98)	1.3% (0.99)

programs offered at their institution. Factors included cognitive and motivational variables like controlled and autonomous regulation, learning styles, intrinsic value and procrastination, personality and behavioral factors, and academic self-concept. Moreover, a factor representing students' engagement with the course materials via the LMS (learning analytics) was also integrated into the models. Finally, demographic and academic background factors including class type (online or face-to-face) were used as control variables in our models.

Our analyses yielded interesting insights as to which factors among these are the key in predicting student success as well as in predicting which students are likely to utilize available resources to improve their learning and performance in the course. To the best of our knowledge, cognitive and behavioral factors have never been used for the latter purpose, so this contribution helps expand the field of education research significantly in that direction and significantly extends the work that was initially undertaken in Mitra and Goldstein (2015). It is not always the case that students that need help actually seek it, hence it was important to understand factors underlying such behavior because institutions spend a lot of funds on developing intervention and services to support student learning but many of these resources are not utilized sufficiently by students. As an example, we observed how students who are less open-minded, who have learning-oriented goals toward their courses, have more controlled regulation, and are likely to seek help have higher chances of attending more SI sessions during a semester for this bottleneck course. Mitra and Goldstein (2018) previously observed that students who had attended SI in previous courses are more likely to attend SI consistently in the future. This, along with our results that less open-minded students are more likely to attend SI consistently, shows that students who are not familiar with SI are not likely to try it out, rather those who have some level of familiarity with it through earlier courses tend to keep attending these sessions. This, in turn, seems to indicate that generally students are satisfied with their SI experience and realize its benefits in terms of improved learning and performance and hence are more likely to put in the extra effort to take advantage of it whenever they are available for a course. On the other hand, it is expected that students who are focused on learning well in the course (and not just on passing the course) will attend SI session more frequently because SI provides them with the opportunity to reinforce the materials learned in their classes, thus resulting in a better overall learning experience for the course. Finally, students who are more likely to seek help also naturally were found to be more inclined toward attending SI sessions, a result which is consistent with expectations.

Regarding factors that affect student success in this course most significantly in our predictive model, students who have higher levels of autonomous regulation, who are have performance-oriented goals and those who prefer a learning style that is based on concrete hands-on experience, are expected to have a higher chance of success in the course. Students who have greater autonomous regulation are typically more organized and have greater motivation (Black and Deci 2000) and hence are expected to put more effort in their courses and thus perform well. Similarly, students who rely on more hands-on experiences are likely to engage in more problem-solving which form the basis of a quantitative course like this one, and hence will probably perform better and have higher chances of successfully completing the course. In terms of goal orientations, students who have a more performance-oriented goal also are expected to put in more hard work to achieve their aim of performing well in the course. This is very interesting, because our earlier finding showed that students who have a more learning-focused goal (instead of a performance-focused one) are more likely to utilize additional resources to support their learning. Among the five primary personality traits studied here, students who are more social and have a positive outlook (that is, possess a higher degree of "extraversion") are likely to succeed in this course. This result aligns well with findings in earlier studies (Komarraju and Karau 2005) that have found a positive relationship between extraversion and academic achievement. Lastly, students with higher levels of academic self-concept were found to have a lower chance of succeeding in the course. Such students usually believe that they have higher confidence in their own intellectual abilities and level of preparedness; however, this does not reflect in their actual performance in the course. This is a phenomenon that has been visible before in several aspects, most prominently in technology usage, and shows that individuals often do not have an accurate perception of their true understanding and abilities. In fact, in psychology, the Dunning-Kruger effect describes a cognitive bias in which people of low ability often mistakenly assess their cognitive ability as greater than it actually is (Kruger and Dunning 1999). It is indeed very interesting to note that how certain behavior and attitudes of people remain consistent across different applications and in different aspects of life.

An important contribution of this work is identifying the subset of factors that specifically impact the performance of low-performing students who are at a greater risk of failing the course. Given the diverse nature of student populations, it is reasonable to assume that different factors may play a role in determining success of high and low-performing students. Indeed, we found that several cognitive factors such as intrinsic value and self-regulation play a greater role in low-performing student performance, and such insights can be extremely beneficial in developing targeted resources for such students.

Learning analytics or students' engagement behavior and pattern with their LMS did not appear to be a strong predictor of either student success or of help-seeking behavior in terms of SI attendance. Although some existing research (Yu and Jo 2014) suggest that students who are more engaged in the course are likely to perform well, we did not observe this as one of the most significant factors in the presence of stronger factors representing a student's cognitive, behavioral traits as well as personality characteristics. Even the correlation coefficient between students' average LMS visits and their final course grades was only 0.10 (not statistically significant), this leading to the conclusion that this variable did not have any association with student success in the course.

5.1. Practical implications

Knowledge about underlying factors that significantly contribute toward different aspects of student success in a bottleneck course is extremely valuable to educators and administrators of higher education where the current focus is increasing on-time graduation rates, mainly for undergraduate students. This is because such findings and insights provide in-depth understanding of the type of targeted intervention and support that can be designed and implemented for such a diverse student population as served by our institution. For instance, efforts may be undertaken to encourage and motivate students who are not likely to utilize additional support services to do so via strategies like email reminders, incentives from instructors, and so on. Moreover, instructors can take into account the different motivational styles and learning styles that characterize their students to improve upon the design and delivery of courses in a way that facilitates learning in a wider student subpopulation.

Additionally, knowledge about factors specific to the success of low-performing students can be used by education administrators to develop resources and programs to discuss skills (such as, setting specific learning goals, developing certain learning styles that boost motivation, self-regulation, and so on) before they begin their degree program that are likely to help them achieve successful outcomes and eventually graduate on time. Many students entering college do not possess these skills that often are associated with poor performance, hence they will benefit significantly.

Finally, our study can be replicated for any course in any discipline, bottleneck or not, and hence has far-reaching impacts for the general higher educational system in terms of boosting student graduation rates.

5.2. Limitations

The primary limitation of the study is the sample size. Although sufficiently large for the purpose of fitting logistic regression models, the sample size played a role in the classification experiments because of the partitions done for the purpose of cross-validation. Moreover, the imbalance created in the datasets owing to the wider prevalence of one of the outcomes used in classifications also affected their validity as is evident by the low F_1 score for the student success model. This issue was addressed in the case of the SI attendance model by employing the SMOTE technique; however, we hope to collect additional data and explore other statistical modeling techniques and methods in the near future to further improve our student success model. Another potential limitation is the scope for generalizability of our results. Although the demographic composition of our sample closely represents that of the entire student population at our university, it is probably not reasonable to make general inferences about all students due to the restricted sample size. However, our study design is quite general and can be used as a model to adapt for student success studies in other disciplines, especially for bottleneck courses in such disciplines, and also at other educational institutions in the region and nationwide.

5.3. Conclusions and future work

Our study yielded interesting and useful insights about the academic performance of students in a bottleneck Business course at a large public university beyond just background demographic and academic factors that are traditionally studied. The study included a diverse set of cognitive, behavioral and motivational factors, including learning analytics, and showed their importance in determining student success as well as in understanding which of the students are likely to seek additional support services in order to improve their performance in the course. In fact, it was very insightful to observe how different sets of factors proved to be significant predictors for these two phenomena studied here, both that are extremely crucial to overall academic achievement. A large-scale study like this thus has far-reaching impacts and the findings therefrom can be used broadly for developing appropriate intervention and support services for students as well as targeted advising and mentoring guidelines so that students receive the help they need to be successful in the course, and as a result, in complete their degree requirements in a timely manner.

Our current and future direction of study include scaling the study to include more students and extending to other bottleneck courses. Further, we will explore the opportunities for applying more sophisticated statistical models like hierarchical models and mixed effects models to this data, especially those that are efficient for unbalanced datasets.

References

American Psychological Association (APA). 2018. American Psychological Association (APA) Website. https://www.apa.org/.

Ames, R., and S. Lau. 1982. An attributional analysis of student help-seeking in academic settings. *Journal of Educational Psychology* 74 (3):414–23. doi:10.1037//0022-0663.74.3.414.

Anderson, D. R. 2008. *Model based inference in the life sciences*. New York: Springer.

Arroway, P., G. Morgan, M. O'Keefe, and R. Yanosky. 2016. Learning analytics in higher education. Research Report. Educause, Louisville, CO.

Bakhshinategh, B., O. R. Zaiane, S. Elatia, and D. Ipperciel. 2018. Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies* 23 (1):537–53. doi:10.1007/s10639-017-9616-z.

Bencheva, N. 2010. Learning styles and e-learning face-to-face to the traditional learning. https://pdfs.semanticscholar.org/0927/8f8fa4f4243dbdb316cfa07b8f6d6fa466a7.pdf

Beer, C., K. Clark, and D. Jones. 2010. Indicators of engagement. In *Proceeding ASCILITE: Curriculum, technology & transformation for an unknown future*, ed. C. Steel, M. J. Keppell, and P. Gerbic. Sydney: ASCILITE.

Bergevin, P. M. 1993. The relationship between language and learning style. *The Journal of Language for International Business* 4:1–6.

Black, A. E., and E. L. Deci. 2000. The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education* 84 (6):740–56. doi:10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3.

Borden, V. M. H., and H. Coates. 2017. Learning Analytics as a counterpart to surveys of student experience. *New Directions for Higher Education* 2017 (179):89–102. doi:10.1002/he.20246.

Bowles, T., A. McCoy, and S. Bates. 2008. The effect of Supplemental Instruction on timely graduation. *College Student Journal* 42:853–9.

Breiman, L. 2001. Random Forests. *Machine Learning* 45 (1):5–32. doi:10.1023/A:1010933404324.

Brower, A. M., and A. Ketterhagen. 2004. Is there an inherent mismatch between how black and white students expect to succeed in college and what their colleges expect from them? *Journal of Social Issues* 60 (1):95–116. doi:10.1111/j.0022-4537.2004.00101.x.

Burmeister, S. 1996. Supplemental Instruction: An interview with Deanna Martin. *Journal of Developmental Education* 20:22–6.

Byrnes, J. P. 2003. Factors predictive of mathematics achievement in white, black, and Hispanic 12th graders. *Journal of Educational Psychology* 95 (2):316–26. doi:10.1037/0022-0663.95.2.316.

Cassidy, S. 2004. Learning styles: An overview of theories, models, and measures. *Educational Psychology* 24 (4):419–44.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16:321–57. doi:10.1613/jair.953.

Clark, M. H., and C. Schroth. 2010. Examining relationships between academic motivation and personality among college students. *Learning and Individual Differences* 20 (1):19–24. doi:10.1016/j.lindif.2009.10.002.

Congos, D. H. 2003. Health checklist for Supplemental Instruction (SI) programs. *The Learning Assistance Review* 8 (2):29–45.

Costa, P. T., and R. R. McCrae. 1992. *Revised NEO personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Crisp, G., E. Doran, and N. A. S. Reyes. 2018. Predicting graduation rates at 4-year broad access institutions using a Bayesian modeling approach. *Research in Higher Education* 59 (2):133–55. doi:10.1007/s11162-017-9459-x.

Deci, E. L., R. J. Vallerand, L. G. Pelletier, and R. M. Ryan. 1991. Motivation and education: The self-determination perspective. *Educational Psychologist* 26 (3):325–46.

Devadoss, S., and J. Foltz. 1996. Evaluation of factors influencing student attendance and performance. *American Journal of Agricultural Economics* 78 (3):499–507. doi:10.2307/1243268.

Dutt, A., M. A. Ismail, and T. Herawan. 2017. A systematic review on educational data mining. *IEEE Access* 5:15991–6005. doi:10.1109/ACCESS.2017.2654247.

Dweck, C. S. 1986. Motivational processes affecting learning. *American Psychologist* 41 (10):1040–8. doi:10.1037/0003-066X.41.10.1040.

Educause. 2010. 7 Things you should know about analytics, EDUCAUSE 7 things you should know series. http://www.educause.edu/ir/library/pdf/ELI7059.pdf.

Elias, T. 2011. Learning analytics: Definitions, processes and potential. Athabasca University. http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

Elliot, A. J., and M. A. Church. 1997. A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology* 72 (1):218. doi:10.1037//0022-3514.72.1.218.

Elliot, E. S., and C. S. Dweck. 1988. Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology* 54 (1):5–12. doi:10.1037//0022-3514.54.1.5.

Elliot, A. J., and H. A. McGregor. 2001. A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology* 80 (3):501. doi:10.1037//0022-3514.80.3.501.

Fatima, D., S. Fatima, and A. V. Krishna Prasad. 2015. A survey on research work in educational data mining. *IOSR Journal of Computer Engineering* 17 (2):43–9.

Fayowski, V., and P. MacMillan. 2008. An evaluation of the Supplemental Instruction programme in a first-year calculus course. *International Journal of Mathematical Education in Science and Technology* 39 (7):843–55. doi:10.1080/00207390802054433.

Filippidi, A., N. Tselios, and V. Komis. 2010. *Impact of Moodle usage practices on students' performance in the context of a blended learning environment*. Patras, Greece: Social Applications for Lifelong Learning, 1–6.

Fox, J., and S. Bartholomae. 1999. Student learning style and educational outcomes: Evidence from a family financial management course. *Financial Services Review* 8 (4):235–51. doi:10.1016/S1057-0810(00)00042-1.

Gainen, J. 1995. Barriers to success in quantitative gatekeeper courses. In J. Gainen and E. W. Williamson (Eds.) *Fostering student success in quantitative gateway courses*, 5–14. San Francisco: Jossey-Bass.

Glesne, C. 1998. *Becoming qualitative researchers: An introduction*. New York: Longman.

Glesne, C. 1999. *Becoming qualitative researchers: An introduction*, 2nd ed. Don Mills, ON: Longman.

Goldberg, L. R. 1992. The development of markers for the Big-Five factor structure. *Psychological Assessment* 4 (1):26–42. doi:10.1037/1040-3590.4.1.26.

Hazrati-Viari, A., A. T. Rad, and S. S. Torabi. 2012. The effect of personality traits on academic performance: The mediating role of academic motivation. *Procedia - Social and Behavioral Sciences* 32:367–71. doi:10.1016/j.sbspro.2012.01.055.

Herndon, M. K., and J. L. Moore. 2002. African American factors for student success: Implications for families and counselors. *The Family Journal* 10 (3):322–7. doi:10.1177/10680702010003009.

Hodges, R., and W. G. White. 2001. Encouraging high-risk student participation in tutoring and Supplemental Instruction. *Journal of Developmental Education* 24 (3):2–10.

Horn, L., K. Peter, and K. Rooney. 2002. Profile of undergraduates in U.S. postsecondary institutions: 1999–2000. NCES 2002–168. U.S. Department of Education, National Center for Education Statistics, Washington, DC.

Horn, L. J., and M. D. Premo. 1995. Profile of undergraduates in U.S. postsecondary education institutions: 1992–93, with an essay on undergraduates at risk. NCES 96–237. U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied logistic regression*, 2nd ed. New York: Wiley-Interscience Publication.

House, J. D. 1995. Non-cognitive predictors of achievement in introductory college Mathematics. *Journal of College Student Development* 36 (2):171–81.

Huber, H., and J. C. Lowry. 2003. Meeting the needs of consumers: Lessons for post-secondary environments. *New Directions for Adult and Continuing Education* 100:79–88. doi:10.1002/ace.121.

Jiao, Q. G., D. A. DaRos-Voseles, K. M. T. Collins, and A. J. Onwuegbuzie. 2011. Academic procrastination and the performance of graduate-level cooperative groups in research methods courses. *Journal of the Scholarship of Teaching and Learning* 11 (1):119–38.

Jo, I.-H., D. Kim, and M. Yoon. 2014. Analyzing the log patterns of adult learners in LMS using learning analytics. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14)*, 183–7. Indianapolis, IN: Association for Computing Machinery.

Karabenick, S. A., and J. R. Knapp. 1991. Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology* 83 (2):221–30. doi:10.1037/0022-0663.83.2.221.

Kaur, H., and E. G. Bathla. 2018. Student performance prediction using educational data mining techniques. *International Journal on Future Revolution in Computer Science & Communication Engineering* 4 (12):93–7.

Kenney, P. 1989. Effects of Supplemental Instruction on student performance in a college level mathematics course. PhD diss., University of Texas-Austin. Dissertation Abstracts for Social Sciences, 50.

Kenney, P. A., and J. M. Kallison. 1994. Research studies of the effectiveness of Supplemental Instruction in mathematics. In *Supplemental Instruction: Increasing achievement and retention*, ed. D. C. Martin and D. Arendale, 75–82. San Francisco, CA: Jossey-Bass. doi:10.1002/tl.37219946010.

Kesahavamurthy, U., and H. S. Guruprasad. 2014. Learning analytics: A survey. *International Journal of Computer Trends and Technology* 18 (6):260–4. doi:10.14445/22312803/IJCTT-V18P155.

Koenker, R., and K. F. Hallock. 2001. Quantile regression. *Journal of Economic Perspectives* 15 (4):143–56. doi:10.1257/jep.15.4.143.

Kolb, D. A. 1981. Learning styles and disciplinary differences. In *The modern American college*, ed. A. Clickering. San Francisco, CA: Jossey-Bass.

Kolb, D. A. 1986. *Learning style inventory: Technical manual*. Boston, MA: Hay Group, Hay Resources Direct.

Komarraju, M., S. J. Karau, and R. R. Schmeck. 2009. Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences* 19 (1):47–52. doi:10.1016/j.lindif.2008.07.001.

Komarraju, M., and S. J. Karau. 2005. The relationship between the Big Five personality traits and academic motivation. *Personality and Individual Differences* 39 (3):557–67. doi:10.1016/j.paid.2005.02.013.

Kozlowski, S. W. J., and B. S. Bell. 2006. Disentangling achievement orientation and goal setting: Effects on self-regulatory processes. *Journal of Applied Psychology* 91 (4):900–16. doi:10.1037/0021-9010.91.4.900.

Kruger, J., and D. Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77 (6):1121–34. doi:10.1037/0022-3514.77.6.1121.

Kutner, M. H., C. J. Nachtsheim, and J. Neter. 2004. *Applied linear regression models*. 4th ed. New York: McGraw-Hill Irwin.

Lakshminarayan, N., S. Potdar, and S. G. Reddy. 2012. Relationship between procrastination and academic performance among a group of undergraduate dental students in India. *Journal of Dental Education* 77 (4):524–8.

Lo, J., and P. Shu. 2005. Identification of leaning styles online by observing learners' browsing behavior through a neural network. *British Journal of Educational Technology* 36 (1):43–55. doi:10.1111/j.1467-8535.2005.00437.x.

Loo, R. 2002. A meta-analytic examination of Kolb's learning style preferences among Business majors. *Journal of Education for Business* 77 (5):252–6. doi:10.1080/08832320209599673.

Lyle, K., and W. Robinson. 2003. A statistical evaluation: Peer-led team learning in an organic chemistry class. *Journal of Chemical Education* 80 (2):132–4. doi:10.1021/ed080p132.

MacFayden, L. P., and S. Dawson. 2010. Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education* 54 (2):588–99. doi:10.1016/j.compedu.2009.09.008.

Magnusson, J.-L., and R. P. Perry. 1992. Academic help-seeking in the university setting: The effects of motivational set, attributional style, and help source characteristics. *Research in Higher Education* 33 (2):227–45. doi:10.1007/BF00973580.

Malm, J., L. Bryngfors, and L. Morner. 2010. Supplemental Instruction (SI) at the Faculty of Engineering (LTH), Lund University, Sweden: An evaluation of the SI program at five LTH engineering programs, Autumn 2008. *Australian Journal of Peer Learning* 3 (1):38–50.
Malm, J., L. Bryngfors, and L. Morner. 2011. Supplemental Instruction: Whom Does it Serve? *International Journal of Teaching and Learning in Higher Education* 23 (3):282–91.
Manjula, M. 2018. A systematic review on educational data mining. *International Journal of Scientific Research in Science, Engineering and Technology* 4 (4):164–70.
Martin, D. C., and D. Arendale. 1992. *Supplemental Instruction: Improving first-year student success in high-risk courses*. Columbia, SC: National Resource Center for the Freshman Year Experience.
McEwan, B. C. 2001. Web-assisted and online learning. *Business Communication Quarterly* 64 (2):98–103.
Meece, J. L., E. M. Anderman, and L. H. Anderman. 2006. Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology* 57 (1):487–503. doi:10.1146/annurev.psych.56.091103.070258.
Mitra, S., and Z. Goldstein. 2015. Designing early detection and intervention techniques via predictive statistical models: A case study on improving student performance in a business statistics course. *Communications in Statistics: Case Studies, Data Analysis and Applications* 1 (1):9–21. doi:10.1080/23737484.2015.1063409.
Mitra, S., and Z. Goldstein. 2018. Impact of Supplemental Instruction on business courses: A statistical study. *INFORMS Transactions on Education* 18 (2):89–101. doi:10.1287/ited.2017.0178.
Mwalumbwe, I., and J. S. Mtebe. 2017. Using learning analytics to predict students' performance in Moodle learning management system: A case of Mbeya University of science and technology. *The Electronic Journal of Information Systems in Developing Countries* 79 (1):1–13. doi:10.1002/j.1681-4835.2017.tb00577.x.
Naveh, G., D. Tubin, and N. Pliskin. 2012. Student satisfaction with learning management systems: A lens of critical success factors. *Technology, Pedagogy and Education* 21 (3):337–50. doi:10.1080/1475939X.2012.720413.
Newman, R. S. 2002. How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory into Practice* 41 (2):132–8. doi:10.1207/s15430421tip4102_10.
O'Connor, M. C., and S. V. Paunonen. 2007. Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences* 43 (5):971–90. doi:10.1016/j.paid.2007.03.017.
Ogden, P., D. Thompson, A. Russell, and C. Simons. 2003. Supplemental Instruction: Short and long-term impact. *Journal of Developmental Education* 26 (3):2–8.
Payne, S. C., S. S. Youngcourt, and J. M. Beaubien. 2007. A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology* 92 (1):128–50. doi:10.1037/0021-9010.92.1.128.
Peters, A., D. Mani, S. Rasathurai, and M. Greene. 2007. *The effectiveness of Supplemental Instruction and technology in increasing student performance in mathematics (Supplemental Instruction update)*. Kansas City: The International Center for SI, University of Missouri–Kansas City.
Pintrich, P. R., and E. V. De Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* 82 (1):33–40. doi:10.1037//0022-0663.82.1.33.
Powers, D. M. W. 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2 (1):37–63.
Rogers, C., and J. Freiberg. 1994. *Freedom to learn*, 3rd ed. Upper Saddle River, NJ: Merrill Publishing.
Romer, D. 1993. Do students go to class? Should they? *Journal of Economic Perspectives* 7 (3):167–74. doi:10.1257/jep.7.3.167.
Rozental, A., and P. Carlbring. 2014. Understanding and treating procrastination: A review of a common self-regulatory failure. *Psychology* 5 (13):1488–502. doi:10.4236/psych.2014.513160.

Ryan, R. M., and E. L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55 (1):68–78. doi:10.1037//0003-066X.55.1.68.

Ryan, A. M., M. H. Gheen, and C. Midgley. 1998. Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social-emotional role, and the classroom goal structure. *Journal of Educational Psychology* 90 (3):528–35. doi:10.1037/0022-0663.90.3.528.

Saucier, G., and L. R. Goldberg. 1996. Evidence of the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality* 10 (1):61–77. doi:10.1002/(SICI)1099-0984(199603)10:1<61::AID-PER246>3.0.CO;2-D.

Scarr, S. 1992. Developmental theories of the 1990s: Development and individual differences. *Child Development* 63 (1):1–19. doi:10.2307/1130897.

Schunk, D. H., and P. A. Ertmer. 2000. Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation*, ed. M. Boekaerts, P. R. Pintrich, and M. Zeidner, 631–49. San Diego: Academic Press.

Sedlacek, W. E. 2004. Why we should use noncognitive variables with graduate and professional students. *The Advisor: The Journal of the National Association of Advisors for the Health Professions* 24 (2):32–9.

Seymour, E., and N. Hewitt. 1997. *Talking about leaving: Why undergraduates leave the sciences*. Boulder: Westview.

Shaya, S. B., H. R. Petty, and L. I. Petty. 1993. Education: A case study of Supplemental Instruction in biology focuses on at-risk students. *Bioscience* 43 (10):709–11. doi:10.2307/1312343.

Sheather, S. 2009. *A modern approach to regression with R*. New York: Springer NY.

Shmueli, G., P. C. Bruce, and N. R. Patel. 2016. *Data mining for business analytics: Concepts, techniques, and applications with XLMiner*. 3rd ed. New York: Wiley.

Shoemaker, N., and M. Kelly. 2015. How college business students learn with emphasis on differences between majors. *Journal of College Teaching & Learning (TLC)* 12 (3):223–30. doi:10.19030/tlc.v12i3.9314.

Snow, R. E., L. Corno, and D. Jackson. 1996. Individual differences in a affective and conative functions. In *Handbook of educational psychology*, ed. D. C. Berliner, and R. C. Calfee. New York: Simon & Schuster Macmillan

Society for Learning Analytics Research (SOLAR). 2019. Society for Learning Analytics Research (SOLAR) website. https://solaresearch.org/.

Tobias, S. 1990. *They're not dumb, they're different: Stalking the second tier*. Tucson, AZ: Research Corporation.

Tuckman, B. W. 1991. The development and concurrent validity of the procrastination scale. *Educational and Psychological Measurement* 51 (2):473–80. doi:10.1177/0013164491512022.

Wagerman, S. A., and D. C. Funder. 2007. Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality* 41 (1):221–9. doi:10.1016/j.jrp.2006.03.001.

West, L., T. Searle, and R. Price. 2018. The use of learning analytics to support improvements in teaching practice. Innovative Research Universities, Melbourne, Australia. https://www.iru.edu.au/wp-content/uploads/2018/04/MRUN-Learning-Analytics-report.pdf

Whitmer, J. 2012. Logging on to improve achievement: Evaluating the relationship between use of the learning management system, student characteristics, and academic achievement in a hybrid large enrollment undergraduate course. University of California. http://johnwhitmerdotnet.files.wordpress.com/2013/01/jwhitmer_dissertation_complete_1-21-2013.pdf.

Widmar, G. E. 1994. Supplemental Instruction: From small beginnings to a national program. *New Directions for Teaching and Learning* 60:3–10. doi:10.1002/tl.37219946003.

Wilcox, F. K., and C. Koehler. 2007. Supplemental Instruction: Critical thinking and academic assistance. *Metropolitan Universities. An International Forum* 6 (4):87–99.

Wratcher, M. A., E. E. Morrison, V. L. Riley, and L. S. Scheirton. 1997. *Curriculum and program planning: A study guide for the core seminar*. Fort Lauderdale, FL: Nova Southeastern University Programs for Higher Education.

Yu, T., and I. Jo. 2014. Educational technology approach toward learning analytics relationship between student online behavior and learning performance in higher education. *In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, 269–70. Indianapolis, IN: Association for Computing Machinery.

Zimmerman, B. J. 1998. Academic studying and the development of personal skill: A self-regulatory perspective. *Educational Psychologist* 33 (2–3):73–86. doi:10.1080/00461520.1998.9653292.

Zimmerman, B. J., and M. M. Pons. 1988. Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology* 80 (3):284–90. doi:10.1037//0022-0663.80.3.284.