Learning Strategies for First-Year Biology: Toward Moving the "Murky Middle"

Angelique Kritzinger,†* Juan-Claude Lemmens,‡ and Marietjie Potgieter§

†Department of Plant and Soil Sciences, ‡Department of Education Innovation, and §Department of Chemistry, University of Pretoria, Pretoria 0002, South Africa

ABSTRACT

Higher education faces the challenge of high student attrition, which is especially disconcerting if associated with low participation rates, as is the case in South Africa. Recently, the use of learning analytics has increased, enabling institutions to make data-informed decisions to improve teaching, learning, and student success. Most of the literature thus far has focused on "at-risk" students. The aim of this paper is twofold: to use learning analytics to define a different group of students, termed the "murky middle" (MM), early enough in the academic year to provide scope for targeted interventions; and to describe the learning strategies of successful students to guide the design of interventions aimed at improving the prospects of success for all students, especially those of the MM. We found that it was possible to identify the MM using demographic data that are available at the start of the academic year. The students in the subgroup were cleanly defined by their grade 12 results for physical sciences. We were also able to describe the learning strategies that are associated with success in first-year biology. This information is useful for curricular design, classroom practice, and student advising and should be incorporated in professional development programs for lecturers and student advisors.

INTRODUCTION

Researchers in the United States who have studied retention over a number of decades have aptly noted that student success is a complex puzzle (Baird, 2000; Bean and Eaton, 2000; Braxton, 2000; Kuh *et al.*, 2007). Despite many research studies, models, frameworks, and interventions based on the findings from this research, U.S. institutions have experienced slow growth in graduation rates. One notable way in which the college system in the United States has approached the student retention and success puzzle is the use of analytics to transform data into actionable information that could be used to move the needle on student success indicators.

It is widely acknowledged that, across the higher education sector in South Africa, the level of attrition at first-year level is high, overall completion rates are low, and the majority of students do not complete their degrees in regulation time (Scott *et al.*, 2007). In South Africa, the higher education system comprises universities and universities of technology where students can study for a bachelor's degree and continue to postgraduate level. According to the Council on Higher Education in South Africa, only 25% of the 2008 student intake in the sciences completed their degrees in the minimum regulation time of 3 years; after 6 years, only 58% of the students had completed their degrees (Council on Higher Education, 2014). Developed countries have similar attrition rates but are able to compensate to some extent for the losses with high participation rates in higher education. However, in developing countries where participation in higher education is low, a poor success rate leading to high attrition is problematic.

With the objective of identifying students at risk of failing, a fair amount of institutional research over the past few years has been done to describe the student population in South Africa in terms of demographic characteristics, including prior learning,

student readiness, the impact of financial aid, race, and gender, (Van der Merwe and Pina, 2008; Lemmens, 2010; Van Zyl *et al.*, 2012; Van Zyl, 2013). The purpose of these descriptive studies was to identify factors that affect student retention and success by describing the characteristics of the students who are successful compared with those who are not.

In South Africa, as in other countries, the focus of many of these studies is to provide an early alert of high-risk student groups through the prediction of poor academic performance and attrition, mainly in the first year. Lacking from these research studies is evidence of the impact of interventions on student success, and from a learning analytics perspective, the impact on making learning more effective and efficient (Vuorikari *et al.*, 2016). Vuorikari *et al.* (2016) point out that a reason for limited evidence on impact is the short time frame of many of the studies, specifically in Europe, and in the case of North America, the possible inefficient use of data by focusing on annual and retrospective reporting practices of student data, which are more associated with academic analytics (DeBerard *et al.*, 2004; Siemens and Long, 2011). Another possible reason, albeit not as obvious, is the perpetual focus on identifying and supporting the "at-risk" students in many of the early-alert models currently used, in order to improve the retention and success rates of institutions (Siemens and Long, 2011; Sclater *et al.*, 2016). The question thus emerges whether the primary support of the at-risk student group should not be expanded to include students whose chances of success are also dubious but, due to unknown reasons, are not flagged as being at risk through conventional early warning models. A relatively small investment of resources for the support of these students could potentially make a big difference in their prospects of success.

In 2014, the Education Advisory Board (EAB) of the United States defined the "murky middle" (MM) as students who are at risk of dropping out of university later than their first year (Student Success Collaborative, 2014). These students are not included in the category of at-risk students who are flagged for possible drop out in the first year; they are part of a student population with a grade point average (GPA) of between 2.0 and 2.99 who progressed to the second year of study; yet a significant proportion of them drop out later in their studies. The EAB collected data from a large number of diverse American higher education institutions (740,000 unique student records from 73 institutions) and analyzed those data for trends and patterns that can inform the design of effective intervention strategies. Three findings are important for the MM; first, these students do not conform to the characteristics that are used to flag at-risk students. Second, the leading indicator before some of the students drop out is not a general decrease in GPA but an increase in the number of courses failed. The MM dropouts earn roughly the same number of "A's" per term, fewer "B's" and "C's", but the number of "F" increases shortly before they terminate their studies. The distribution of grades in successive semesters thus shows an increase in the number of courses failed while still maintaining a GPA that will not point to the student being at risk of dropping out. The third important finding is that outcomes improve dramatically when the downward trends in grades are reversed (Student Success Collaborative, 2014). Very little is known about this group of students other than the characteristics just described. However, the possibility that interventions targeted

at these students will have a high return on investment is an attractive prospect.

While the concept of the MM was defined in the context of a very large student group at a large number of institutions, we propose that it can also be applied within a single course to facilitate analysis of student characteristics and the relation of this analysis to student performance. In principle, students enrolled for a specific course can also be divided into three groups: those who are at risk of failing, those who are likely to pass with relative ease, and those students for whom the outcomes are uncertain (whom we also label the MM). In learning analytics and prediction studies, characteristics such as prior learning, demographics, gender, race, and financial status are usually used to predict performance; however, very few studies try to characterize students in terms of other factors, such how well students can regulate their learning and which learning strategies good students use to be successful.

Self-Regulated Learning

Self-regulated learning (SRL) is a learning theory that describes the self-directed processes learners use to "transform their mental abilities into task-related academic skills" (Zimmerman and Schunk, 2001). Although there are many theoretical perspectives concerning SRL (Boekaerts, 1999; Boekaerts *et al.*, 2000; Butler and Winne, 1995; Winne, 1995; Pintrich, 1995; Greene and Azevedo, 2007), most come to the same conclusion—that the most effective learners are skillful self-regulators. According to Pintrich and de Groot (1990), there are a variety of definitions of self-regulated learning. In all of them, three components seem to be important in classroom performance. The first, metacognition, includes students' strategies for planning, modifying, and monitoring their cognitions. The second is students' management and control of their effort in classroom academic tasks (behavior); and the third is the actual cognitive strategies that students use to learn, remember, and understand material. Goals have been shown to play an important role in SRL, and Zimmerman (2002) states that the skills inherent to SRL include setting specific goals, adopting strategies for attaining goals, using time management skills, monitoring performance, and managing physical and social contexts. In a recent study, Sebesta and Bray Speth (2017) show that various self-regulation strategies such as self-evaluation, goal setting and planning, and information seeking were associated with improved grades in a large biology class. Very few studies in biology education focus on SRL as a construct in itself, they rather focus on aspects of SRL such as metacognition and metacognitive regulation.

With the SRL theory as a basis, Pintrich *et al.* (1991) developed the Motivated Strategies for Learning Questionnaire (MSLQ) to assess students' motivational orientations and their use of different learning strategies. The final version of the MSLQ was completed in 1990 and was presented formally for the first time in the *Journal of Educational and Psychological Measurement* in 1991 (Pintrich *et al.*, 1991). The MSLQ has two sections: the Motivation section consists of 31 items that assess a student's goals and value beliefs for a course; and the Learning Strategies section includes 50 items related to a students' use of different cognitive and metacognitive strategies and resource management. The different subscales will be discussed in more detail later in the paper. The MSLQ has a modular

design, enabling researchers to use each section separately. The MSLQ was developed to be course specific; however, according to Conley (2007), cognitive and metacognitive strategies are transferable. This means that successful learning strategies documented for one course in biology could also be effective in other biology courses, but there might be some course-specific requirements.

The MSLQ was chosen as the self-report instrument to measure learning strategies in our study. The MSLQ has been widely used in research, and a meta-analysis of these studies was recently published by Credé and Phillips (2011). Credé and Phillips (2011) indicate that total scores on the MSLQ were better correlated with single-course grades than with GPA. Research by Hilpert *et al.* (2013), Credé and Phillips (2011), and Dunn *et al.* (2012), suggests that the effort regulation and metacognitive regulation subscales should be combined, as they operate at the same level of analysis (executive function). The researchers also point out that these two subscales mediate motivation and academic performance. However, the other subscales that measure "strategy use" are not good predictors of academic outcomes. The relationships between MSLQ subscale scores and academic performance were generally weak to moderate, with "effort regulation," "self-efficacy," and "time and study environment" for individual classes showing the highest validity. They also showed that specific learning approaches such as rehearsal, organization, and peer learning were largely unrelated to academic performance, while the less contextual abilities, such as metacognitive self-regulation and effort regulation were most strongly related to academic performance. In addition the relationship between peer learning and GPA was not in the expected direction. That being said, the MSLQ in this study was not specifically used for prediction of academic performance but used as a way to describe students in terms of their enacted learning strategies and to pinpoint possible opportunities for interventions. The MSLQ has also been used in first-year biology classes to determine the effect of the flipped classroom approach as reported by van Vliet *et al.* (2015). They found that the flipped classroom approach changed the learning strategies of students and enhanced learning strategies such as critical thinking, task value, and peer instruction (as measured by the MSLQ) but that the effect of change in learning strategies were not long-lasting. According to Zeegers (2004), researchers found that there was a consistent correlation ($r = 0.30$) between academic achievement and measures of self-regulation.

In line with the SRL theory, Conley (2007) defines academic preparedness as the single most important determining factor for success in higher education. Conley (2007) defined college readiness operationally as the level of preparation a student needs to enroll and succeed—without remediation—in a credit-bearing general education course at a postsecondary institution. The Conley framework foregrounds and repackages most of the elements of SRL as operationalized by the MSLQ. Conley identifies four key components that form the foundation of academic readiness. These components are multifaceted and include factors both internal and external to the university environment. In the current study, *key cognitive strategies*, *key content,* and *academic behaviors* are the facets of interest. *Contextual skills and awareness*, the fourth facet of the model, refers to the systemic understanding on how higher education institutions operate as a system and as a culture (Conley, 2007). This facet was beyond the scope of the study.

The most central of the facets of readiness, key cognitive strategies, is important in the broader context of higher education. It focuses on learning skills that are not content specific, such as critical thinking and problem solving. These skills enable students to learn content from a range of disciplines (Conley, 2007).

The second element of the framework is key content, which refers to general academic knowledge and overarching academic skills such as writing and research. Although some discipline-specific knowledge is essential, overarching academic skills that are not content specific are needed for students to be ready for the demands of university. An example of these academic skills is areas in which science studies instill the knowledge of how to use all the steps in the scientific method, how to communicate science, and to appreciate that scientists think in terms of models and systems to interpret complex phenomena.

Academic behaviors refer to a range of behaviors that reflect self-awareness, self-monitoring, and self-control and are independent of specific content areas. Key academic behaviors consist largely of self-monitoring skills and study skills. Self-monitoring skills are a form of metacognition that requires students to be aware of their understanding of a subject, to reflect on task effectiveness, and to adjust learning strategies by transferring skills from familiar settings to new settings.

Present Study

To enhance student success and throughput, we need to take a different approach to the analysis of student performance by going beyond prediction and description of at-risk students. In this study, we chose to focus instead on a group of students who are typically not flagged as at risk for possible drop out in the first year of study. However, many of these students drop out later during their undergraduate studies without triggering any alerts within normal systems. We have also labeled them the "murky middle," because their graduation outcomes are difficult to predict. Thus, the aim of this paper is to investigate whether it is possible to categorize students in a single course (microlevel) similar to what was done by the Student Success Collaborative on the macrolevel and derive meaningful information about student success from the difference in the resultant subsets of student characteristics in which categorization is based only on demographics and prior learning achievements. In essence, we designed this study to explore the potential of learning analytics applied to a single course as the unit of analysis, in combination with strong learning theories such as SRL and college readiness, to generate useful pointers about effective learning strategies that can inform course design and advisory practices. Specifically, we set out to answer the following research questions:

RQ1. Which pre-entry characteristics differentiate effectively between students who are likely to pass (LTP), the murky middle (MM), and students at risk of failing?

RQ2. Which learning strategies in first-year biology are associated more strongly with good performance than with marginal or poor performance?

Our approach was to compare the self-reported learning strategies of the strong and weak groups to identify those strategies that are significantly different to inform the design of

activities that could move the MM toward behaviors associated with success rather than failure.

METHODOLOGY: PARTICIPANTS AND CONTEXT

Molecular and Cell Biology (MLB 111) is a first-year course in the Faculty of Natural and Agricultural Sciences at the University of Pretoria (UP), South Africa. It is a service course for students in the biological sciences program and other programs such as medicine and veterinary science. Only 27% of the students enrolled in the Faculty of Natural and Agricultural Sciences, the faculty where this course is housed, chose the program that they are registered for as a first choice of study. The fact that the majority of the MLB 111 students are not enrolled in their preferred program influences the level of motivation for this course as well as the "expectancy value" (Ambrose *et al.*, 2010) for the course. A further complicating factor is the wide range in academic preparedness of the students upon entry, with the majority of them being underprepared for the demands of higher education (Scott *et al.*, 2007). Generally, a small number of those students enrolled for MLB 111 succeed in their objective to transfer to their program of choice; the rest either drop out or continue to complete their studies in biological sciences.

The course has an average enrollment of 1500 students and is presented during the first semester of each academic year. A team of three lecturers is involved; each teaches a specific component of the course and repeats the same lecture for three groups of ~500 students. In the interest of consistency, the instructors use the same assessment approaches throughout the semester (e.g., formative assessment by means of clickers (Miller and Tanner, 2015) during class and in tutorial classes, online tests via the learning management systems, and two high-stakes summative assessments during the course of the semester).

DESCRIPTION OF THE DATA SET

Data for this study were obtained from a convenience sample of students enrolled during the first semester of 2015 for MLB 111. The data set comprised biographical data, prior achievement data, longitudinal data, and MSLQ survey data.

Biographical Data

Biographical data were obtained from the university student information system. Only students entering higher education for the first time were included. A total of 1084 student records were used for the study. This number represented 68.3% of the total population of students registered for this course. The remainder (31.7%) were either students repeating the course or students who transferred from another university. The decision was made to use only first-time entering first-year students in this study. The rationale for delimiting the sample was to allow the research team to focus on students who are new to the academic environment, as repeating or transfer students had exposure to the course content and the demands of the higher education environment. The experiences that repeating and transferring students bring with them are confounding factors and would lead to spurious associations. The students excluded from the study might also benefit from the actionable interventions emanating from the findings of this study. The sample comprised 730 females (67%) and 354 males (33%), and the ages of the students ranged between 17 and 31, with an average age of 18.3 years. Field of study was determined by the faculty in which the students were enrolled; 801 students majored in science (74%), 174 in medical sciences (16%), 77 in veterinary sciences (7%), and 32 in a non–science major (3%). The racial division of the students was 58% white, 29% Black, and 13% racial minorities.

Prior Achievement Data

High School Achievement Data. High school achievement data such as GPA have been used in nearly all retention and prediction studies and appear to be consistently significant predictors of student success internationally (Kokaua *et al.*, 2014) and in South Africa (Lourens and Smit, 2003; Baard *et al.*, 2010; Van Zyl *et al.*, 2012; Kirby and Dempster, 2015a). Performance scores achieved in the grade 12 school leaving examination (National Senior Certificate, NSC) that were deemed relevant to this study are those for mathematics, physical sciences (chemistry and physics), biology, and English as first language or as first additional language. In the South African secondary school system, the subject of physical sciences consists of roughly equal amounts of chemistry and physics content.

National Benchmark Tests (NBTs). The National Benchmark Tests (NBTs) are designed to support decision making in South African universities about the academic readiness of first-year students, university placement, course development, and program planning. In a recent study, Rankin *et al.* (2012) found that both the school leaving examination (NSC) and the NBTs were useful predictors of academic performance in higher education in South Africa. However, research at the institution where the study of this article was completed found that high school results were better predictors of university success than proficiency tests such as the NBT (unpublished technical reports by Lemmens [2011, 2014] and Lemmens and Schaap [2012]). According to Conley (2007), admission tests have been reasonably effective methods of identifying students who are potentially college ready in the United States. NBT test scores were obtained from the university's enrollment system.

Computer Literacy. The MLB 111 course has an online component in which students are required to complete weekly online assignments. Van Zyl *et al.* (2012) found that self-reported computer skills were a significant predictor of student success in the first semester of higher education studies at another South African University. Students at the UP take a computer literacy placement test upon entry, and the performance results for this test are included in the analysis.

Longitudinal Data

Student academic standing and GPAs were obtained at the end of 2015 and 2016 with the aim of validating the categorization of students.

Data on Learning Strategies

The MSLQ has two sections: Motivation and Learning Strategies. Learning strategies have been shown to be strongly predictive of grades (Credé and Kuncel, 2008), but motivation is more difficult to influence than learning strategies, and its association

with performance is complex. As previously mentioned, research by Hilpert *et al.* (2013), Credé and Phillips (2011), and Dunn *et al.* (2012) points out that effort regulation and metacognitive regulation, traits represented in the Learning Strategies section of the MSLQ, mediate motivation and academic performance. We have therefore decided to use only the Learning Strategies section of the MSLQ for data collection. The Learning Strategies section has nine subscales, broadly divided into cognitive and metacognitive strategies and resource management strategies. Rehearsal, organization, metacognitive self-regulation, elaboration, and critical thinking are all part of the cognitive and metacognitive self-regulation strategies, whereas time and study environment, effort regulation, help seeking, and peer learning are part of the resource management strategies. The wording of the MSLQ items was slightly adapted to fit South African colloquial use (Appendix A in the Supplemental Material), without making changes to the psychometric properties of the survey. The seven-point Likert scale was changed to a four-point scale: very true of me (scored 3), mostly true of me (2), seldom true of me (1), and not at all true of me (0). We reduced the number of response categories to lower the cognitive load (Revilla *et al.*, 2014) and opted for an even number of categories to avoid distortions often associated with the middle category of Likert scales (Nadler *et al.*, 2015). The questionnaire was administered electronically in the last 2 weeks of the semester, directly before the examination, via the university's learning management system. Negatively worded items were reverse scored. Reverse-scored items are indicated as such in Appendix A in the Supplemental Material and Table 2 later in this article. Of the sample of 1084 students, 715 students completed the questionnaire, representing a response rate of 66%. Only 528 (49%) of the 715 responses were usable after correcting for incomplete response sets. This subgroup consisted of 380 females (72%) and 148 males registered for the following programs: 80% science, 10% medical sciences, 8% veterinary sciences, and 2% from non–science majors. The racial division of the students was 59% white, 29% Black, and 12% racial minorities. Thus, the subgroup of students who completed the questionnaire was comparable to the bigger sample of students used for classification.

The first objective of this study was to define and describe the MM group of students within the first-year biology course. The first year is the time with the highest attrition in higher education. The second objective of this study was to identify the particular learning strategies that seem to be most effective for the disciplinary context. Effective learning strategies are essential for being successful in college (Tuckman and Kennedy, 2011), and the use of inappropriate learning strategies is one of many reasons for students dropping out, but one that can be addressed with well-designed interventions.

The research commenced with an inductive analytical approach to explore the relationship between an outcome variable and possible predictor variables. The possible predictor variables were limited to pre-entry data that are readily available at the start of the academic year. These include demographic and prior achievement data. The outcome variables were summative assessments conducted during the semester, namely, two semester tests and the exam marks for the module.

The research questions dictated the use of a statistical method that allows for the investigation of the patterns of relationship as well as the interaction effects among variables. Many of the variables do not have a linear relationship and, arguably, many of the academic-related variables are not at an interval level, hence the need for nonparametric statistical techniques. CHAID (chi-squared automatic interaction detection) analysis was chosen as the preferred method of analysis to answer the first research question, largely due to the nonparametric nature of the data, the need to segment students according to their academic outcomes, and the need to understand the interaction effects of the predictor variable and for its ability to generate a simple, yet powerful display that supports the interpretation of the results. The CHAID technique, which is similar to a classification and regression tree analysis, was developed by Kass (1980), and is a model often used in data mining, such as prediction modeling and segmentation (Nisbet *et al.*, 2009). The technique has been used recently by various authors for performance analysis of student populations (Kirby and Dempster, 2014, 2015b; Baran and Kilic, 2015). The CHAID model creates a tree diagram by identifying the most important predictor variables associated with the outcome variable and then splits the first predictor variable into groups that are significantly different from one another. After the first split is made, based on the chi-square test to determine the best split at each step, the model proceeds to the remainder of the predictor variables (see Figure 1). The model repeats this process until there are no significant contributions left (Nisbet *et al.*, 2009). When using educational data, we frequently use two states: pass versus fail. The outcome variables (i.e., summative assessments) were transformed into these two states for the CHAID analysis. However, school performance data were left as continuous data and not transformed. The IBM SPSS Statistical package, v. 23, was used for the CHAID analysis, and the Bonferroni adjustment was applied to limit the type I error.

The term "MM," as described by the EAB, can only be claimed if the group identified as the MM had similar medium-term success prospects as defined in the larger-scale study (Student Success Collaborative, 2014). Thus, the results of the CHAID division have to be validated using different strategies; first, a cross-tabulation was constructed, and the associated Goodman and Kruskal's coefficient was calculated to describe the correlation between the strongest predictor variable and the first summative assessment of the course. Second, the three groups of students had to be tracked in terms of their GPAs and academic standings at the end of successive years to verify performance patterns relative to their classification (Table 1). In South Africa, GPA is a number that expresses student performance as a percentage point (%). The students can achieve a maximum of a 100%, with 50% representing a pass and 75% a distinction.

The second objective of this study was to formulate a rich description of effective learning strategies to guide the design of suitable interventions aimed at improving the prospects of success of all students. To answer the second research question, we divided the MSLQ data into three subsets for the at-risk, MM, and LTP students. Of the 528 complete questionnaires, 199 students were of the at-risk group, 166 from the MM, and 163 from the LTP. The results were analyzed for significant differences between the composite scores for each subscale in the Learning Strategies section of the instrument. To test the validity of the instrument, which was changed from a seven-point
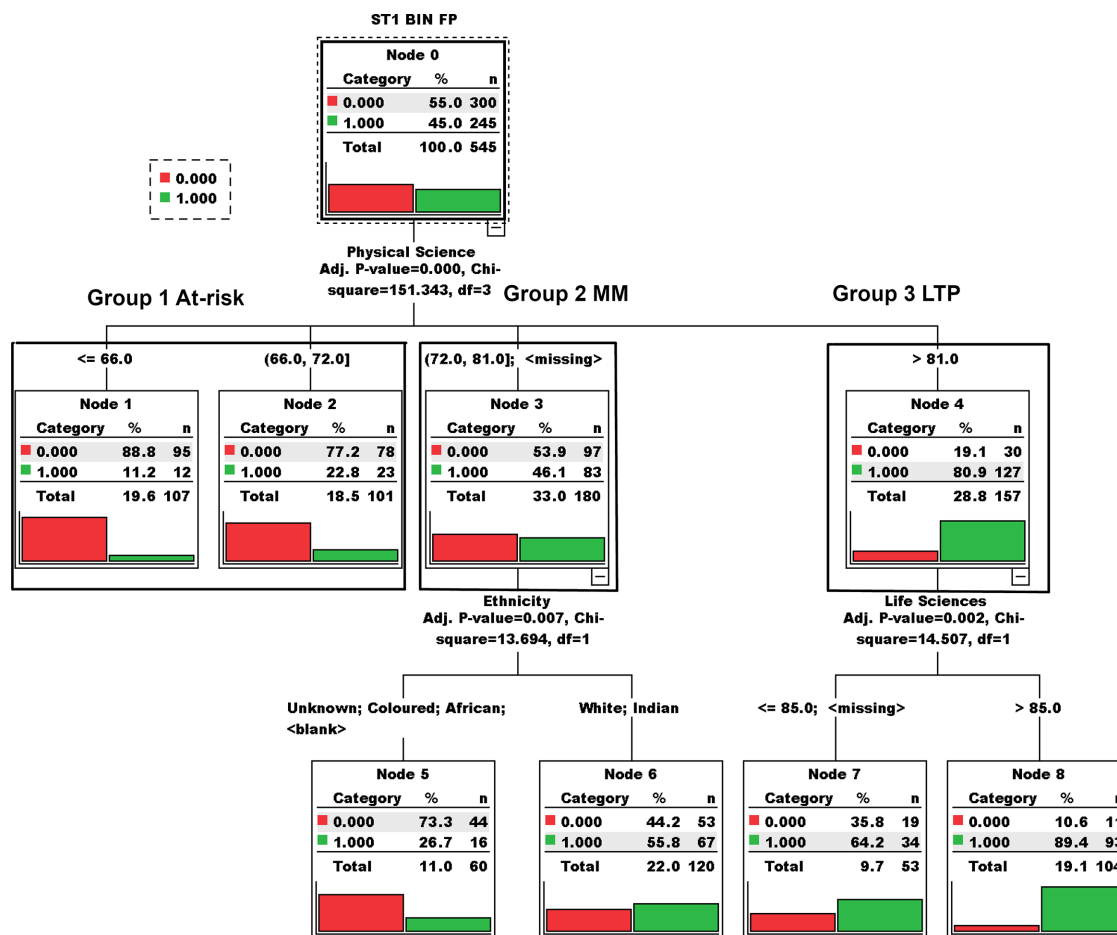
FIGURE 1. CHAID analysis with semester test 1 as the outcome variable.

Likert to a four-point Likert, we performed a Cronbach alpha analysis (α = 0.92), which compares favorably with the Cronbach alpha values reported in the MSLQ manual for the Learning Strategies subscales (α = 0.52–0.79). One-way analysis of variance (ANOVA) tests were performed on these composite scores to detect any statistically significantly differences between groups. Significant differences on the ANOVA were further evaluated with Tukey post hoc tests to see which groups were significantly different. (A summary of the means of the subscale scores is presented in Appendix B in the Supplemental Material.) This was then followed by a Kruskal-Wallis H analysis

of item responses to identify items with significantly different response patterns. The "between-group" differences for individual items were determined by means of a Mann-Whitney *U* test. These results are reported in Table 2 and are discussed in the next section.

RESULTS

Categorization and Validation
To divide the students into at-risk, MM, and likely-to-pass students, we performed a CHAID analysis to determine the prior learning factor that best predicted the outcome of the first

TABLE 1. Overall performance of students at the end of the first and second year

		Group 1: at risk	Group 2: MM	Group 3: LTP
2015	*N* at start of 2015	426	315	343
	% of group that passed MLB 111	49.1	67.9	93
	Average % of final mark per group	49.9	55.7	67.4
	Dismissed due to poor academic performance	27	6	1
	Mean GPA (2015)	54	60	71
	Mean credit pass ratio	0.78 ± 0.26	0.88 ± 0.20	0.98 ± 0.67
2016	Active students at end of 2016	332	254	301
	Mean GPA (2016)	54	60	71
	Mean credit pass ratio (2016)	0.83 ± 0.20	0.90 ± 0.16	0.98 ± 0.06

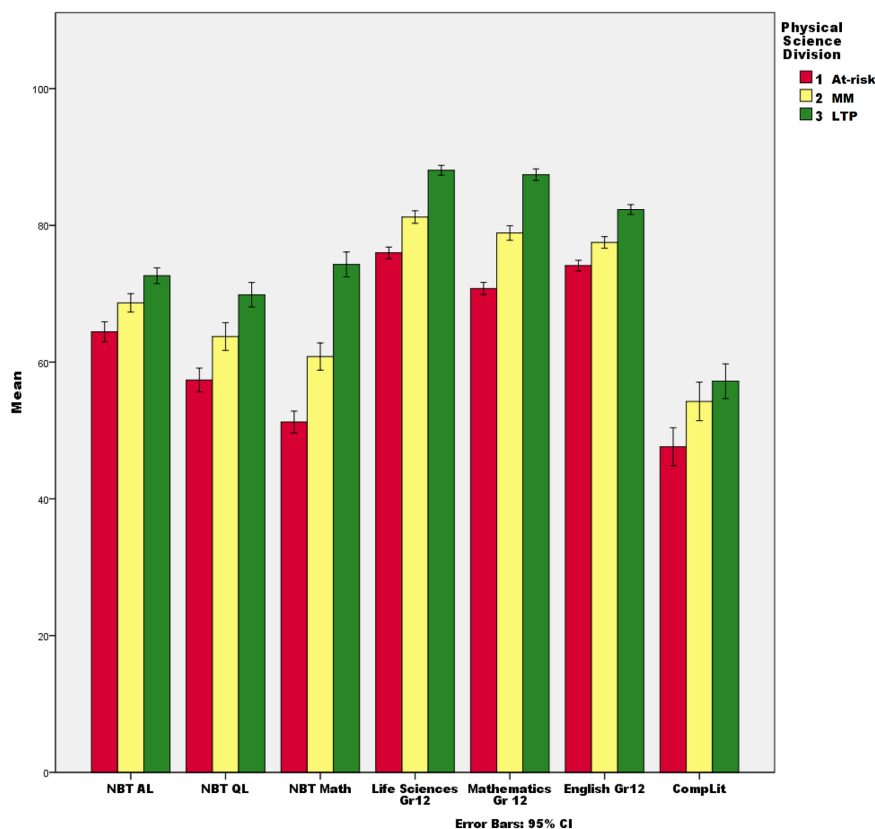**TABLE 2. Subscales of the MSLQ and items showing significant difference between groups**

Scale name	Number of items	Significance of ANOVA[a]	Items with significance (5% level)[b]	Kruskal-Wallis asymptotic significance
			Cognitive and metacognitive strategies	
Rehearsal	4	**0.005** (at risk and LTP)	46. (Groups 1 and 2 and groups 2 and 3) When studying for this course, I work through my class notes and the course materials a number of times	0.000 0.026
		At risk and MM $p = 0.037$	59. (Groups 1 and 2) I memorize key words to remind myself of important concepts in this course.	
Organization	4	0.053	49. I make simple charts, diagrams, or tables to help me organize course material.	0.036
Metacognitive self-regulation	12	**0.002**	41. When I become confused about something I'm reading for this course, I go back and try to figure it out	0.023
			61. (Groups 2 and 3) I determine what I am supposed to learn from the material before I start studying	0.009
			78. (Groups 1 and 2) When I study for this course, I set goals for myself in order to direct my activities in each study session.	0.001
			79. (Groups 1 and 2) If I get confused taking notes in this course, I make sure I sort it out afterwards.	0.001
Elaboration	6	0.013	62. (Groups 2 and 3) I try to relate ideas in this subject to those in other courses whenever possible.	0.032
			81. (Groups 2 and 3) I try to apply ideas from course material in other course activities such as lectures and discussions	0.000
Critical thinking	5	0.021		
			Resource management strategies	
Time and study environment	8	0.013	35. (Groups 1 and 2) I study in a place where I can concentrate on my course work.	0.000
Effort Regulation	4	**0.000** (at risk and LTP)	37. (Groups 1 and 2 and groups 2 and 3) I feel so lazy or bored when I study for this course that I give up before I finish what I planned to do. (reverse coded)	0.000
		MM and LTP **$p = 0.010$**	48. (Group 1 and 2 and 2 and 3) I work hard to do well in this course even if I don't like what we are doing.	0.000
			60. When course work is difficult, I either give up or only study the easy parts. (reverse coded)	0.005
Peer learning	3	0.015	34. (Groups 2 and 3) When studying for this course, I try to explain the material to a classmate or friend.	0.003
			45. (Groups 2 and 3) I try to work with other students from this course to complete the course assignments.	0.032

[a]Scales indicated in bold are scales that show "convincing" evidence ($p < 0.01$) of differences between the two extreme groups.
[b]All significant differences indicated are between the at-risk and LTP group, unless otherwise indicated.

summative assessment. The CHAID analysis showed that grade 12 performance in physical sciences was the best predictor of success in the first semester test of MLB 111. The results for the test sample are represented in the tree diagram shown in Figure 1. The green bars indicate the students who passed the first semester test and the red bars the students who failed, with the relative height of the bars representing proportions of the sub-set. In each box, the total number of cases is reported as well as its proportion of the test sample. Figure 1 indicates that the student group was divided into four subgroups or nodes based on performance in grade 12 physical sciences (chi-square 151.34, df 3, $p < 0.001$). However, upon inspection, it was clear that both nodes 1 and 2 represent subgroups of students with a poor chance of success: 11.2% and 22.8%, respectively. These two nodes can therefore be combined to generate three distinct subgroups of comparable sizes, namely, 208, 180, and 157 students. Thus, students with a grade 12 physical sciences mark below 72% were designated as at risk (group 1). Students with a mark between 72 and 81% were designated the MM (group

2), and students with a mark higher than 81% were designated as LTP (group 3). Those students for whom no physical sciences marks (6%) were captured in the system (missing) were grouped with the MM (group 2). All subsequent analyses were performed in SPSS with the student sample split into these three groups (426 at risk, 315 MM, and 343 LTP). Although the CHAID indicated that ethnicity and life science marks were contributing factors that were also predictive within the model, a decision was made not to include the second level of analysis to determine the categorization of students. Ethnicity was shown to be a predictor for success in higher education in other studies (Kuh *et al.*, 2008; Van Zyl *et al.*, 2012; Tejada *et al.*, 2016), but in our case, physical sciences marks proved to be the stronger predictor, thereby allowing us to categorize on neutral grounds. The results indicate that the model has good overall precision; it is able to correctly classify 75% of the cases, and detection of false positives is also 75%. The exclusion of the second level of the CHAID from the classification reduced the model accuracy by only 2% (72.8%). Research by Lemmens (Lemmens, 2015;

FIGURE 2. Summary of all prior achievement scores showing statistically significant differences in the means of the scores for the at-risk, MM, and LTP groups.

learning and demographic data is the highest for summative assessments done early in the semester; and early identification of students at risk and, in this case, the MM, creates opportunities for student support to prevent failure or drop out.

The validity of the categorization of the sample into three subsets based on performance in grade 12 physical sciences was checked in terms of its correlation with the other prior learning variables. An ANOVA was performed to determine individual correlation between the physical sciences performance categories and performance on the other variables. A significant difference at the 5% level was found between the three groups for all other predictor variables. All the variables, with the exception of computer literacy (CompLit), had significance levels of $p < 0.001$. Where between-group values were significant, Tukey post hoc tests were performed to see which groups showed significant differences. The results are shown in Figure 2 with red, yellow, and green bars for the at-risk, MM, and LTP groups, respectively.

The final test for the categorization of students into groups based on grade 12 physical sciences marks was whether a significant difference between the prospects of success of the three groups could be found in the short and medium term of higher education studies. The pass rate for MLB 111 and the credit pass ratios for 2015 and 2016 were compared for the three subgroups, where the mean credit pass ratio reflects the percentage of credits passed compared with the total number of credits enrolled for. The results are reported in Table 1.

There was a significant difference between the "mean GPA" for the different groups ($F = 40.42$, $df = 2$, $p < 0.001$). At the end of 2015, the at-risk group passed, on average, 78% of credits enrolled for, the MM 88%, and the LTP group 98%. The mean GPA for the at-risk group was 54, the MM 60, and the LTP group 71 (expressed as a percentage). This means that, in 2015, the at-risk group failed on average two 16-credit courses, while the MM group failed one 16-credit course out of the first-year total of 144 credits. In 2016, the picture stayed mostly the same. While the prospects of success for the MM are significantly higher than those of the at-risk group (88 vs. 78%), it still means that the MM group will have to repeat courses that they failed, which will extend their study duration and increase their risk of dropping out during later years.

Difference in Learning Strategies between the Three Defined Groups

The analysis of the MSLQ Learning Strategies subscales indicates distinctly different response patterns for the three subgroups of students for eight of the nine subscales at the 5% level of significance. The means for all the groups were always with the at-risk group the lowest mean and the LTP the highest mean (Appendix B in the Supplemental Material). The results of the

Lemmens and Kebalepile, 2017) on the grade 12 high school results, notably physical sciences and Mathematics, tend to have weak Spearman's rank-order correlations with the first-year academic performance when the unit of analysis is at the department or faculty level. This correlation improves significantly when the unit of analysis is at the course level. This finding could be due to less spurious relationships at the course level. Thus, the ability of the CHAID model to classify students with a 75% accuracy with only physical sciences marks is regarded as very promising in this instance. It is important to note that the 25% level of classification inaccuracy does not present a threat to the overarching aim of the project, because the purpose of classification was to enable a more nuanced understanding of student behavior. No student would be excluded from the opportunity to benefit from the interventions that would result from it. The ability of the CHAID classification model to classify students correctly with 75% accuracy is regarded as acceptable to classify students into the three performance categories. Using the categories created by the CHAID, a Spearman's rank-order correlation was calculated between physical sciences marks and the first summative assessment, namely, semester test 1. There was a strong, positive correlation that was statistically significant ($G = 0.621$, $p < 0.001$). By comparison, the precision of the models wherein the second semester test or the final marks for the course were used as the outcome variable were 52% and 56%, respectively. The first summative assessment was chosen as the outcome variable of choice for the following reasons: the predictive power of prior

statistical analyses (for the different subscales) are reported in Table 2. In all but one of these subscales (help seeking), there is a difference between the at-risk and LTP groups; in addition, one subscale differentiated between the at-risk and MM groups (rehearsal) and one subscale differentiated between the MM and LTP students (effort regulation). Three of the scales (rehearsal, metacognitive self-regulation, and effort regulation) provided strong evidence of differences between groups (see Appendix B in the Supplemental Material for p values showing statistical differences between groups).

The results of item analyses presented in Table 2 will be discussed next with reference to the findings of the meta-analysis of other studies by Credé and Phillips (2011). They found that some of the constructs in the MSLQ exhibit meaningful relationships with academic performance. Their study showed the effort regulation subscale to have the highest correlation with academic performance ($\rho = 0.41$), which is on par with more traditional predictors of academic performance such as admission tests, prior learning, and study skills (Credé and Phillips, 2011). In the present study, effort regulation distinguished strongly between two pairs of groups, with two items (37 and 48) differentiating between all three groups. The effort regulation subscale probes students' ability to apply sustained effort and persist even when the work is difficult, there are distractions, or tasks are boring.

Credé and Phillips (2011) have shown that the meta-analysis is broadly supportive of the basic assumptions that underpin the theory of self-regulated learning: those students who can engage metacognitively, regulate their effort, and have appropriate learning strategies have higher average grades than students who cannot do so. This is in agreement with our results, in which the subscale of metacognitive self-regulation strongly differentiated between the at-risk and LTP groups ($p = 0.002$), with two items (78 and 79) also differentiating between the at-risk and MM and one item (61) between the MM and LTP groups. Item 61 provided strong evidence that the LTP group members had better planning capabilities and were more skilled in the selection of important learning material. The results for item 78 suggest that the at-risk students were less adept at setting goals before each study session and monitoring their progress. Items 79 and 41 point to the importance of students taking control of their own learning.

Credé and Phillips (2011) found that many of the specific learning strategies, such as rehearsal, elaboration, organization, critical thinking, peer learning, and help seeking, appeared to be unrelated to academic performance as operationalized in the MSLQ, but that less-contextual abilities, such as metacognitive self-regulation and effort regulation, were most strongly related to academic performance. In this study, scales such as rehearsal, peer learning, organization, and elaboration distinguished, at least, between the two extreme groups. Help seeking did not distinguish between any of the groups, and the reason might be related to the explanation provided by Credé and Phillips (2011). Good students might not need help or may not report on their help-seeking behavior, and low-performing students might not realize they need help and would therefore also not report any help-seeking behavior. Rehearsal strategies differentiated between the at-risk and MM students ($p = 0.037$) and between the at-risk and LTP students ($p = 0.005$). Of specific importance is item 46, which distinguished significantly between all groups, thereby providing evidence for the importance of repetition for mastering biology.

The fact that peer learning showed a significant difference ($p = 0.015$) between groups in our study is noteworthy. Two items differed significantly between the MM and LTP groups as well as between the two extreme groups, which indicates that the need to work with fellow students was recognized more strongly by the LTP group than the others. Conley (2007) states that academic behaviors such as the ability to participate successfully in study groups are critical for success in certain disciplines. It would therefore be advisable to intentionally build peer-learning activities into course design for first-year biology.

Elaboration strategies help students store information in long-term memory by building connections to related concepts. Elaboration strategies are considered deep processing strategies and are important for higher education (Weinstein and Mayer, 1986). The difference in the response patterns of the three groups was significant ($p = 0.013$). The two items (items 62 and 81) that differentiated did so between the at-risk and LTP but also between the MM and the LTP groups. These items probed whether students made an effort to connect concepts within the course and also relate it to other courses. The MLB 111 course that was the focus of this study includes material from other disciplines such as chemistry, biochemistry, physics, and plant science, and the awareness of alignment with other disciplines is thus important for success.

The previous discussion focused on the differences between the three groups of students that manifested most prominently between the at-risk and likely to pass students. The at-risk students need substantial academic and psychosocial support to be successful, while the LTP students will probably be successful without a lot of intervention. However, as this paper aims to describe the MM and generate actionable data regarding this group, it is important to look more closely at this group of students. By definition, MM students displayed characteristics of both the at-risk and LTP students. This is evident in the distribution of questionnaire responses; the average scores for the MM falls roughly in the middle of the subscale (Appendix B in the Supplemental Material). The subscales for which this is not the case are therefore of interest, particularly if their responses resemble those of the at-risk subgroup rather than those of the LTP. One scale for which the MM *clearly* tended more toward the at-risk group was effort regulation (see Appendix B in the Supplemental Material). Duckworth (2016) eloquently explains in her recent book that "effort counts twice," because effort improves skills, and skills combined with effort equals achievement. Admittedly, many other factors might influence the success of the MM students, but this study highlights the importance of the students putting effort into their academic work, even when it is a boring or difficult, a skill that in this study seems to be not as strongly developed in the MM. Building a culture in class that promotes perseverance (or grit) by demonstrating the value and relevance of the content and advocating for increased effort. This is a low-cost intervention that, for the MM, may have a large impact and high return on investment.

DISCUSSION

We have shown in this study that it was possible to identify indicator(s) that would differentiate between three groups of

students: those who are likely to pass (LTP), the murky middle (MM), and those who are at risk of failing. Our analysis has identified grade 12 performance in physical sciences as the most powerful predictor for the first 2 years of study, and it defined the MM as students within the performance band of 72–81% for this subject. The findings suggest that students seem to be confined to a specific performance band and will find it hard to break free from that group without interventions and additional effort. The categorization of students was confirmed by various outcomes such as the mean GPA and mean credit pass ratio. At the UP, all the prior learning variables analyzed in this study showed significant differences between the three groups ($p < 0.05$), indicating that these students enter the university system with distinctly different academic competencies and prospects of success. This finding is noteworthy and could be useful on an administrative level with regard to admission of students and subsequent support of students admitted into the first year of a science degree. Students who are in the at-risk category should not be admitted unless extensive support is provided and embedded in their programs. The UP offers an academic development program (ADP) with a lower entrance requirement, which is specifically designed for this purpose. However, our results suggest that careful consideration should be given to raising the entrance requirements for MLB 111—70% for physical sciences, which would channel at-risk students into the ADP rather than the mainstream.

After categorization of the students, the self-reported learning strategies of the three groups were compared to identify strategies that are associated with success, information that will guide the design of future teaching and learning interventions. The results of the MSLQ analysis indicate distinct differences between learning strategies of the at-risk and LTP students. Metacognitive self-regulation, rehearsal, and effort regulation were the subscales that differentiated most convincingly between the groups. In addition, the ability to relate course material to other courses and the ability to manage the resources well were linked to good performance. Thus, a profile of learning strategies associated with good academic outcomes can be derived from the items that differentiated most convincingly between students who were LTP (group 3) or at risk (group 1).

The productive strategies of the LTP students can be summarized as follows:

1. Work with other students to complete assignments and clarify concepts (items 34 and 45)
2. Apply deep learning by relating ideas to other courses and connecting concepts within a course (items 62 and 81)
3. Sort out any confusion in a timely manner (items 41 and 79)
4. Persist even when work is difficult or not of interest (items 37, 48, and 60)
5. Choose suitable spaces to study (item 35)
6. Apply good time management and thus have time for revision and rehearsal (item 46)
7. Employ appropriate study methods that include memorization and organization (items 49 and 59)
8. Plan study activities and set goals to direct these study activities (items 78 and 61)

The list of effective learning strategies for MLB 111 resonates well with the literature on student success in science.

Multiple studies have been reported of teaching practices that incorporate these pointers and deliver improved student outcomes. This confirms the credibility and usefulness of these results. For example, active-learning strategies such as peer instruction (Crouch and Mazur, 2001; Lasry *et al.*, 2008) and think–pair–share (Miller and Tanner, 2015) encourage students to work with peers to solve problems (items 34 and 45). Clickers (Miller and Tanner, 2015) are widely used during lectures as a form of formative assessment to provide immediate feedback and encourage students to sort out confusing work in a timely manner (items 41 and 79). Research has also shown that group work in and out of class can transform course experiences for the better and help students to become actively engaged with the work (Wood, 2009). These strategies are suitable for large lecture halls and in smaller tutorial classes *and* can be taught by student advisors. Making connections (items 62 and 81) between new information and what is already known or understood is an essential part of the learning process and something that can be explicitly taught in class. Knowing in advance what the big ideas are and how they relate to one another helps learners make sense of information (Hammond *et al.*, 2001). Items 37, 48, and 60 clearly distinguish between the three different groups and point to effort regulation and the willingness to persevere even when the work is not of interest or difficult. This was also the subscale for which the MM were statistically much closer to the at-risk than to the LTP, indicating the need to focus efforts to improve these skills for the MM specifically. Grit (Duckworth, 2016) has been shown to be predictive of academic success and is a teachable skill (Duckworth *et al.*, 2007).

The challenge is to ensure that first-year biology students, especially MM and at-risk students, develop these strategies in time to help them to be successful. In our view, this is the joint responsibility of lecturers and student advisors. Lecturers should be guided to adjust the design of their courses to purposefully incorporate SRL skills during lectures or in the online environment. Recent research has shown that changing course design to be more student centered increases learning gains in biology classes (Connell *et al.*, 2016). This, coupled with the research by Owens *et al.* (2018), who showed that professional development for faculty helped them make significant changes in their course design, makes the case for developing professional development courses for lecturers and student advisors to facilitate the explicit teaching and demonstration of SRL skills inside and outside the classroom.

Student advisors, on the other hand, should take responsibility for assisting students in the learning process by teaching "soft skills" that students can adapt as needed. According to Simpson *et al.* (1997), *learning to learn* programs that are focused on assisting students to become self-regulated learners and that are based in conceptual work in psychology are the most successful interventions for teaching students learning strategies. These courses focus on developing a repertoire of learning strategies such as time management (item 46) and study methods (items 49 and 59) that students can adapt as necessary. These courses have been shown to increase GPA, retention, and graduation rates (Weinstein *et al.*, 2000). On the basis of these results, we propose that large introductory courses should be supported by a learning to learn intervention, which is designed in conjunction with student advisors and incorporates these findings.

CONCLUSION

The first aim of this study was to determine whether the concept of the MM, as defined by the EAB (Student Success Collaborative, 2014), was applicable in our context of a single course within a program (microscale) to generate useful insights similar to its application to programs within institutions (macroscale). Our results indicate that it is possible to define the MM students in a specific course using only prior achievement data. Furthermore, the validity of the categorization was confirmed for subsequent performance and longitudinal data for academic standing. The MM students were consistently unsuccessful in passing all courses that they registered for as indicated by the mean credit pass ratio for the group in 2015 and 2016. This means that they were falling behind because of having to repeat courses, which would hamper progression and prolong their study time. This situation makes them more vulnerable to drop out in later years due to financial constraints and discouragement. The middle group can therefore be labeled the MM, in line with the findings of the Student Success Collaborative, which first coined the term. In our context, more than half of attrition of students happens after the first year. However, most efforts are currently focused on the first year, with limited resources allocated to subsequent years. The development of focused interventions for students who persist beyond the first year but continue to fall behind is thus an important area for investigation.

Related to our second research question, the study identified productive learning strategies that were used by the LTP group more than the other two groups. In addition, the study identified that effort regulation was the learning strategy for which intervention could make a meaningful contribution toward the prospect of success for the MM. These findings can be used to inform classroom practice and student advising. We argue that identifying the MM early and providing interventions with a focus on this group of students can significantly increase throughput rates. Well-designed interventions aimed at the development of effective learning strategies at the start of an academic career will have an impact beyond the first year. The MM falls just short of success by failing on average one course per year, but this shortfall accumulates over time with possible disastrous effects. By "moving the middle" toward success, these effects can be mitigated. Finally, within an environment of resource constraints, well-designed interventions aimed at the MM are expected to have a higher return on investment than those aimed at the at-risk group, which requires much more comprehensive support to achieve success.

Lecturers may have limited control over the admission of students into their programs, but they do have an obligation to teach the students admitted into the system. Typically, lecturers do not take responsibility for the development of soft skills, presumably because they are unaware of the specific needs of students, or because they see that as the task of student advisors. Learning strategies such as effort regulation, setting study goals, and having good time management skills are all skills that can be developed within a classroom setting, and it would be of great benefit to the students if lecturers were aware of the contribution that they can make in this regard. Course design that incorporates the principles of deep learning (making connections), peer learning, and metacognitive monitoring will benefit all students, especially the MM and at-risk students. On the basis of our results, we advocate for the inclusion of research-based material on effective learning strategies in staff development programs to empower faculty to actively promote or incentivize these learning strategies. In conclusion, we propose that our definition of the murky middle at course level and our study of the learning strategies of groups with different prospects of success have delivered actionable information that will be of interest for university management, curriculum designers, lecturers, and student advisors in the quest to improve retention and student success.

REFERENCES

Ambrose, S. A., Bridges, M. W., & DiPietro, M. (2010). *How learning works: Seven research-based principles for smart teaching*. Hoboken, NJ: Jossey-Bass. Retrieved February 5, 2016, from https://site.ebrary.com/lib/pretoria/docDetail.action?docID=10385881

Baard, R. S., Steenkamp, L. P., Frick, B. L., & Kidd, M. (2010). Factors influencing success in first-year accounting at a South African university: The profile of a successful first-year accounting student. *South African Journal of Accounting Research*, 24(1), 129–147. https://doi.org/10.1080/10291954.2010.11435150

Baird, L. (2000). College climate and the Tinto model. In Braxton, J. M. (Ed.), *Reworking the student departure puzzle* (pp. 62–80). Nashville, TN: Vanderbilt University Press.

Baran, B., & Kilic, E. (2015). Applying the CHAID algorithm to analyze how achievement is influenced by university students' demographics, study habits, and technology familiarity. *Journal of Educational Technology & Society*, 18(2), 323–336. Retrieved October 27, 2015, from https://go.galegroup.com/ps/i.do?id=GALE%7CA419268912&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=14364522&p=AONE&sw=w

Bean, J. P., & Eaton, S. B. (2000). A psychological model of college student perception. In Braxton, J. M. (Ed.), *Reworking the student departure puzzle* (pp. 48–61). Nashville, TN: Vanderbilt University Press.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, 31(6), 445–457. https://doi.org/https://doi.org/10.1016/S0883-0355(99)00014-2

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (2000). Self-regulation: An introductory overview. In Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Eds.), *Handbook of self-regulation* (pp. 1–9). San Diego: Academic. https://doi.org/https://doi.org/10.1016/B978-012109890-2/50030-5

Braxton, J. M. (2000). Reinvigorating theory and research on the departure puzzle. In Braxton, J. M. (Ed.), *Reworking the student departure puzzle* (pp. 257–274). Nashville, TN: Vanderbilt University Press.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. https://doi.org/10.2307/1170684

Conley, D. T. (2007). *Redefining college readiness* (Vol. 3). Eugene: Center for Educational Policy Research, University of Oregon.

Connell, G. L., Donovan, D. A., & Chambers, T. G. (2016). Increasing the use of student-centered pedagogies from moderate to high improves student learning and attitudes about biology. *CBE—Life Sciences Education*, 15(1), ar3. https://doi.org/10.1187/cbe.15-03-0062

Council on Higher Education. (2014). *VitalStats public higher education 2014*. Pretoria, South Africa.

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, 3(6), 425–453. https://doi.org/10.1111/j.1745-6924.2008.00089.x

Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences*, *21*(4), 337–346. https://doi.org/10.1016/j.lindif.2011.03.002

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977. https://doi.org/10.1119/1.1374249

DeBerard, M., Spielmans, G., & Julka, D. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal*, *38*(1), 66–80.

Duckworth, A. (2016). *Grit: The power of passion and perseverance*. New York: Scribner.

Duckworth, A., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Dunn, K. E., Lo, W.-J., Mulvenon, S. W., & Sutcliffe, R. (2012). Revisiting the Motivated Strategies for Learning Questionnaire: A theoretical and statistical reevaluation of the Metacognitive Self-Regulation and Effort Regulation subscales. *Educational and Psychological Measurement*, *72*(2), 312–331. https://doi.org/10.1177/0013164411413461

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, *77*(3), 334–372. https://doi.org/10.2307/4624902

Hammond, L.-D., Austin, K., Orcutt, S., & Rosso, J. (2001). Episode #1. Introduction chapter. How people learn: Introduction to learning theories. In *The learning classroom: Theory into practice a telecourse for teacher education and professional development* (pp. 1–22). Stanford, CA: Stanford University. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.3636&rep=rep1&type=pdf

Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: A new conceptualization of an established questionnaire. *SAGE Open*, *3*(4), 1–10. https://doi.org/10.1177/2158244013510305

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29*(2), 119–127. Retrieved February 17, 2017, from https://links.jstor.org/sici?sici=0035-9254%281980%2929%3A2%3C119%3AAETFIL%3E2.0.CO%3B2-N

Kirby, N. F., & Dempster, E. R. (2014). Using decision tree analysis to understand foundation science student performance. Insight gained at one South African university. *International Journal of Science Education*, *36*(17), 2825–2847. https://doi.org/10.1080/09500693.2014.936921

Kirby, N. F., & Dempster, E. R. (2015a). Not the norm: The potential of tree analysis of performance data from students in a foundation mathematics module. *African Journal of Research in Mathematics, Science and Technology Education*, *19*(2), 131–142. Retrieved February 11, 2015, from www.tandfonline.com/doi/abs/10.1080/10288457.2015.1028716

Kirby, N. F., & Dempster, E. R. (2015b). A picture paints a thousand numbers—and tells as story: A post-positivist approach to building substantive grounded theory using decision tree analysis. *South African Journal of Higher Education*, *29*(6), 124–144.

Kokaua, J., Sopoaga, F., Zaharic, T., & Van der Meer, J. (2014). The development of a pre-enrolment screening tool to inform targeted support services in the first year in health sciences. *International Journal of the First Year in Higher Education*, *5*(1), 55–66. https://doi.org/10.5204/intjfyhe.v5i1.205

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *Journal of Higher Education*, *79*(5), 540–563.

Kuh, G. D., Kinzie, J. L., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2007). *Piecing together the student success puzzle: Research, propositions, and recommendations* (ASHE Higher Education Report, Vol. 32, No. 5). Hoboken, NJ: Wiley. Retrieved September 9, 2016, from https://eric.ed.gov/?id=EJ791634

Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, *76*(11), 1066–1069. https://doi.org/10.1119/1.2978182

Lemmens, J.-C. (2010). *Student's readiness for university education* (Unpublished doctoral thesis). University of Pretoria, Pretoria, South Africa.

Lemmens, J.-C. (2015). *Evaluation of the NSC and NBT results: 2014 and 2015*. Pretoria, South Africa: University of Pretoria.

Lemmens, J.-C., & Kebalepile, M. M. (2017). *Evaluation of the NSC*. Pretoria, South Africa: University of Pretoria.

Lourens, A., & Smit, I. P. J. (2003). Retention: Predicting first-year success: Research in higher education. *South African Journal of Higher Education*, *17*(2), 169–176.

Miller, S., & Tanner, K. D. (2015). A portal into biology education: An annotated list of commonly encountered terms. *CBE—Life Sciences Education*, *14*(2), fe2. https://doi.org/10.1187/cbe.15-03-0065

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology*, *142*(2), 71–89. https://doi.org/10.1080/00221309.2014.994590

Nisbet, R., Elder, J., & Gary, M. (2009). *Handbook of statistical analysis and data mining applications*. Amsterdam: Elsevier.

Owens, M. T., Trujillo, G., Seidel, S. B., Harrison, C. D., Farrar, K. M., Benton, H. P., … & Tanner, K. D. (2018). Collectively improving our teaching: Attempting biology department–wide professional development in scientific teaching. *CBE—Life Sciences Education*, *17*(1), ar2. https://doi.org/10.1187/cbe.17-06-0106

Pintrich, P. R. (1995). Understanding self-regulated learning. *New Directions for Teaching and Learning*, *1995*(63), 3–12. https://doi.org/10.1002/tl.37219956304

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Rankin, N., Schöer, V., Sebastiao, C., & Walbeek, C. (2012). Predictors of academic performance: National Senior Certificate versus National Benchmark Test. *South African Journal of Higher Education*, *26*(3), 564–585.

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research*, *43*(1), 73–97. https://doi.org/10.1177/0049124113509605

Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education: A review of UK and international practice*. Bristol, UK: JISC. Retrieved April 20, 2016, from www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v3.pdf

Scott, I., Yeld, N., & Hendry, J. (2007). *Higher Education Monitor: A case for improving teaching and learning in South African higher education* (HE Monitor No. 6). Pretoria, South Africa: Council on Higher Education. Retrieved September 5, 2016, from www.che.ac.za

Sebesta, A. J., & Bray Speth, E. (2017). How should I study for the exam? Self-regulated learning strategies and achievement in introductory biology. *CBE—Life Sciences Education*, *16*(2), ar30. https://doi.org/10.1187/cbe.16-09-0269

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(4), 30–32. https://doi.org/10.1145/2330601.2330605

Simpson, M. L., Hynd, D. R., Nist, S. L., & Burrell, K. I. (1997). College academic assistance programs and practices. *Educational Psychology Review*, *9*(1), 39–87.

Student Success Collaborative. (2014). *The Murky Middle Project*. www.eab.com/technology/student-success-collaborative/members/white-papers/the-murky-middle-project

Tejada, F. R., Parmar, J. R., Purnell, M., & Lang, L. A. (2016). Admissions criteria as predictors of academic performance in a three-year pharmacy program at a historically Black institution. *American Journal of Pharmaceutical Education*, *80*(1), 6. https://doi.org/10.5688/ajpe8016

Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *Journal of Experimental Education*, *79*, 478–504.

Van der Merwe, A., & Pina, R. (2008). Use of portals, CRM, assessment: Stellenbosch University Case Study. *Workshop presented at the 1st Southern African Conference on the First-Year Experience: Using ICT to Track, Profile and Individually Help Students*, Stellenbosch.

van Vliet, E. A., Winnips, J. C., & Brouwer, N. (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. *CBE—Life Sciences Education*, *14*(3), ar26. https://doi.org/10.1187/cbe.14-09-0141

Van Zyl, A. (2013). Teaching the students we have: Two perspectives on first year students at the University of Johannesburg and the UJ first year experience initiative. In *Higher Education Learning and Teaching Association of South Africa (HELTASA)*. Pretoria: Higher Education Learning and Teaching Association of South Africa. Retrieved January 12, 2017, from http://heltasa.org.za/wp-content/uploads/2014/09/HELTASA-2013-Teaching-the-Students.pdf

Van Zyl, A., Gravett, S., & De Bruin, G. P. (2012). To what extent do pre-entry attributes predict first year student academic performance in the South African context? *South African Journal of Higher Education*, *26*(5), 1095–1111.

Vuorikari, R., Muñoz, J. C., Ferguson, R., Brasher, A., Clow, D., Cooper, A., … Ullmann, T. (2016). *Research evidence on the use of learning analytics: Implications for education policy (Joint Research Centre Science for Policy Report)*. https://doi.org/10.2791/955210

Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-regulation interventions with a focus on learning strategies. In *Handbook of self-regulation* (pp. 727–747). https://doi.org/10.1016/B978-012109890-2/50051-2

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In Wittrick, M. (Ed.), *Handbook of research and teaching* (3rd ed.) (pp. 315–327). New York: Macmillan.

Winne, P. H. (1995). Self-regulation is ubiquitous but its forms vary with knowledge. *Educational Psychologist*, *30*(4), 223–228.

Wood, W. (2009). Innovations in teaching undergraduate biology and why we need them. *Annual Review of Cell and Developmental Biology*, *25*, 93–112.

Zeegers, P. (2004). Student learning in higher education: A path analysis of academic achievement in science. *Higher Education Research & Development*, *23*(1), 35–56. https://doi.org/10.1080/0729436032000168487

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70.

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