Learning Communities Research and Practice

Volume 3 | Issue 1 Article 2

5-26-2015

Predicting First-Year Student Success in Learning Communities: The Power of Pre-College Variables

Rita A. Sperry
Texas A & M University - Corpus Christi, rita.sperry@tamucc.edu

Recommended Citation
Sperry, R. A. (2015). Predicting First-Year Student Success in Learning Communities: The Power of Pre-College Variables. *Learning Communities Research and Practice*, 3(1), Article 2.
Available at: https://washingtoncenter.evergreen.edu/lcrpjournal/vol3/iss1/2

Authors retain copyright of their material under a Creative Commons Non-Commercial Attribution 3.0 License.

Predicting First-Year Student Success in Learning Communities: The Power of Pre-College Variables

Abstract

The study used pre-college variables in the prediction of retention and probation status of first-year students in learning communities at a regional public university in South Texas. The correlational study employed multivariate analyses on data collected from the campus registrar about three consecutive cohorts (N = 4,215) of first-year students. Logistic regression models were developed to predict retention and probation status without respect to learning community membership, as well as for each learning community category.

The logistic regression model to predict retention regardless of learning community membership included five pre-college variables, while the model to predict probation status included eight pre-college variables, five of which overlapped with the retention model. The models for each learning community contained different sets of predictor variables; the most common pre-college predictors were high school percentile and the number of days since orientation. The results of the study provide practical implications for the learning communities program, as well as learning community scholars interested in targeting interventions to the students who need them most.

Keywords

first-year students, retention, probation, predictive models

Introduction

The Association for American Colleges and Universities (AAC&U) launched the Liberal Education and America's Promise (LEAP) initiative in 2005 to address ongoing issues in higher education, including a study of which educational practices have the greatest impact on the success of college students at all levels. As part of this work, Kuh (2008) outlined ten specific teaching and learning practices that the LEAP initiative found to be most effective at increasing student retention and engagement, each of which is linked to retention and graduation. One of the High-Impact Practices (HIPs) defined by the LEAP initiative to engage students from the onset of their academic careers was the implementation of learning communities.

Formed by the linking of two or more courses for a shared cohort of students, learning communities have demonstrated significant rewards for both students and faculty (Hill, 1985; Huerta, 2004; Lardner & Malnarich, 2008; Smith, MacGregor, Matthews, & Gabelnick, 2004; Tinto, 2000). Along with the Washington Center for Improving the Quality of Undergraduate Education, the National Resource Center for the First-Year Experience and Students in Transition supports the use of learning communities as an effective practice for integrating the entire first-year experience (Henscheid, 2004). Despite their identification as a HIP by the AAC&U and overwhelming support in the literature, it is still is an unfortunate reality that certain students continue to struggle to succeed in learning communities.

The issue of why some students struggle warrants a thorough analysis of recent learning communities. The goal is to develop models that identify which students are at most risk of not succeeding—in the study, of landing on probation or not being retained—which would allow faculty to intervene early and to target their interventions. In fact, learning communities might become more high impact with the use of data about entering students gathered before the first day.

Theoretical Framework

Tinto's (1975) Student Integration Model (SIM) postulated that the students who persist and succeed in college are those who are able to integrate successfully into an institution's social and academic environment. Alternatively, the students who are more likely to struggle and fail to persist are those who do not attempt or achieve social and academic integration. The SIM identified a variety of external or pre-college factors that play a role in college student integration, including past academic performance (prior qualifications), family background (family attributes), and personal goals (individual attributes), as well as experiences at the institution (inside and outside of the classroom). Tinto's model and these external factors provided the theoretical framework for the study.

Borrowing from Van Gennep's (1960) anthropological concept of rites of passage, Tinto (1988) defined three stages of student departure: separation, transition, and incorporation. At each of these three abstract stages, students decide whether or not to remain in college. Tinto argued that the first semester of college is particularly crucial to helping students make the successful social and academic transition that leads to persistence and ultimately graduation. Tinto (1997) later updated his SIM to include the significance of classroom experience and faculty interactions on student success and persistence and argued for the implementation of learning communities to assist in this critical academic period in students' lives.

Over the past several decades, Tinto's emphasis on the first semester of college has been supported and enhanced by numerous studies attempting to predict first-semester GPA and retention using a multitude of variables (Glynn, Sauer, & Miller, 2003; Herreid & Miller, 2009; Kahn & Nauta, 2001; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; Nora, Cabrera, Hagedorn, & Pascarella, 1996; Porter, 1999; Snyder, Hackett, Stewart, & Smith, 2002; Voorhees, 1987; Wetzel, O'Toole, & Peterson, 1999). Tinto's model has also helped to bring learning communities to the forefront of research in higher education as an intervention to assist students in making successful social and academic transitions to higher education. Most recently, this work has involved the creation of assignments in learning communities that require students to integrate content and skills across disciplines (Huerta & Sperry, 2010; Lardner & Malnarich, 2009); the analysis of social networks and peer relationships formed within learning communities (Chamberlain, 2011; Smith, 2010; Stuart, 2008); and the creation of learning communities to support developmental education (Hansen, Meshulam, & Parker, 2013; Heany & Fisher, 2011; Synder, Hakett, Stewart, & Smith, 2002).

Tinto's (1975; 1997) SIM is particularly salient to the prediction of outcomes such as retention and probation status because it provides a framework for identifying and categorizing the types of incoming variables that are related to student success. According to the SIM, student persistence is a function of various factors—past academic performance, personal and family background, personality and goals, and college experiences—each of which plays a role in explaining the end result. If these factors could each be measured, then it would be feasible to develop statistical models to predict whether or not a given student would be successful.

Literature Review

Despite the growth of learning communities as a movement in higher education, there is limited published research on their outcomes, especially in relation to first-year student success and persistence (Andrade, 2008; MacGregor & Smith, 2005). Taylor, Moore, MacGregor, and Lindbland (2003) identified 32 formal research studies and 119 institutional reports on learning communities programs. In a quasi-meta-analysis of first-year learning community programs, Andrade (2008) found 17 published articles on the impact of learning communities on first-year students, only 12 of which measured persistence. Fifteen of the studies addressed first-semester GPA. Andrade's (2008) results indicated that the research on learning communities appears to demonstrate their positive contribution to student outcomes—such as increased GPAs and persistence—but that it remains unclear as to what specific aspects of learning communities contribute the most to their success. The heterogeneity of programs across the country, as well as the self-selection effect common to most learning community programs, makes interpretation of the data difficult (Andrade, 2008; Habley & Bloom, 2012).

Roccini (2011) was also interested in the impact of learning communities on first-year students. His study explored the literature and identified more than 40 studies that supported the positive outcomes commonly attributed to learning communities (Habley &

Bloom, 2012). Roccini's (2011) study used path analysis to examine the responses to the College Student Experience Questionnaire, and the results indicated that, although learning communities do contribute to the success of first-year students, the impact is indirect via student engagement. In other words, learning community participation is related to increased student engagement, which is in turn related to educational gains. These findings echo the results of Zhao and Kuh's (2004) examination of the experiences of learning community participants who responded to the National Survey of Student Engagement.

A small number of published studies have attempted to aggregate findings about learning communities across multiple institutions. In 1993, Vincent Tinto directed a project for the National Center on Postsecondary Teaching, Learning and Assessment that examined three learning community programs for first-year students: The University of Washington's Freshman Interest Groups (FIGs), LaGuardia Community College's learning community clusters, and Seattle Central Community College's Coordinated Studies Program. Tinto, Love, and Russo (1993) found through quantitative and qualitative methods not only that students in learning communities report positive perceptions of classes, peers, faculty, and themselves at higher rates than non-learning community participants, but also that these students persisted at significantly higher rates.

The Manpower Demonstration Research Corporation (MDRC) recently partnered with the National Center for Postsecondary Research on a six-year grant to explore the variations of learning communities and their impact on student success for community college students. A study of six community colleges found that learning communities had small positive effects on overall academic progress but had no impact on persistence for developmental students (Visher, Weiss, Weissman, Rudd, & Wathington, 2012). These troubling MDRC findings spurred recent entreaties by the Washington Center for Improving the Quality of Undergraduate Education for learning community programs across the nation to conduct self-assessments in order to contest the results (E. Lardner, personal communication, April 25, 2014). Although the impact is often hard to isolate, learning communities are considered one of ten High-Impact Practices (HIPs) endorsed by the AAC&U's LEAP initiative because of their relationship to deep learning, effective educational practices, and self-reported personal and practical gains (Kuh, 2013).

Purpose of the Study

The first year of college is a critical period of transition for incoming college students. Learning communities have been identified as an approach to link students together in courses that are designed with first-year students' needs in mind. Yet learning community teaching teams may not be provided with data about their students prior to the start of the semester in order to strategically target interventions. One question then becomes, what variables known on or before the first day of classes are predictive of first-year student success, in terms of retention and probation status, for first-year college students in learning communities?

The present study sought to determine which pre-college variables—that is, independent variables that can be collected on or before the first day of classes—were predictors of retention or probation status for first-year students in a learning communities program. These variables informed our development of models to predict the probability of success as measured by retention or probation status for future students. The following questions informed the research study:

1. What pre-college variables are predictors of the retention of first-year students in learning communities?
2. What pre-college variables are predictors of the probation status of first-year students in learning communities?
3. What pre-college variables are predictors of retention and probation status for first-year students in *particular* learning communities?

Methods

A regional public four-year university in South Texas has required a learning community experience since it admitted its first cohort of first-year students in 1994. Several published studies have demonstrated the achievement of the program in helping students successfully make the transition from high school to college (Araiza, 2006; Huerta, 2004; Sterba-Boatwright, 2000). The program has also gained national recognition as a leader in the learning community movement (Kutil & Sperry, 2012; Smith, MacGregor, Matthews, & Gabelnick, 2004). However, the particular characteristics of the learning community program that contribute most to student success in terms of retention and probation status are relatively unknown. In addition, little information is available on the first day of class to faculty teaching teams about which students are most at risk of landing on probation or not returning for their sophomore year in order to target interventions.

Participants

The learning communities program for the study was located at a public university in South Texas that was designated as a Hispanic-Serving institution. At the onset of the study, the undergraduate student population of 9,152 students was composed of 46.02% Hispanic and 40.07% White students. All traditional incoming first-year students were required to enroll in a learning community during their first and second semesters. Over the three years studied, 1533 students enrolled in the learning communities in Fall 2010, 1503 in Fall 2011, and 1806 in Fall 2012.

Most of the learning communities in the program were triads, meaning that they contained three courses in which students co-enrolled in cohorts of 25. There were other learning communities ranging from two to five linked classes with varied cohort sizes. Every learning community contained a section of UCCP 1101 (First-Year Seminar) that supported the other courses in the learning community. The UCCP 1101 was a requirement for graduation from the institution. Most of the learning communities were also linked to ENGL 1301 (Composition I), a core curriculum first-year writing course. There were learning community options for students who had entered the program with credit for ENGL 1301.

Each learning community in the program was centered on one or two large core curriculum courses. For example, the Sociology learning community in Fall 2010 (Triad B) had 200 seats. All students in the learning community enrolled in the sociology course, Human Societies (SOCI 1301) on Tuesdays and Thursdays at 9:30am. The students were divided into eight groups of 25 for their UCCP 1101 course. Six of the sections were also linked to two sections of ENGL 1301, so the same 25 students who were in UCCP 1101 also attended their First-Year Composition course together. The instructors for the Sociology, First-Year Seminar, and First-Year Composition courses met weekly to plan assignments and activities. Students completed several assignments based around themes from the Sociology course, and grades were often shared in more than one of the linked classes.

Table 1 contains a summary of the learning communities offered in the Fall semesters of 2010, 2011, and 2012. For the purpose of the study, the learning communities were grouped by subject. The six learning community categories were as follows: Sociology (Triad B), History (Triads C, E, K, and M), Political Science (Triads F and L), Science (Triad S and Tetrads V and W), Developmental History (Tetrad N), and Other (Triads G and T).

Table 1
Learning Communities Offered in Fall 2010, Fall 2011, and Fall 2012

Learning Community	Fall 2010	Fall 2011	Fall 2012
Triad B – Sociology	✓	✓	✓
Triad C – History			✓
Triad E – History	✓	✓	✓
Triad F – Political Science	✓	✓	✓
Triad G – Geology			✓
Triad K – History	✓	✓	✓
Triad L – Political Science	✓	✓	✓
Triad M – History	✓	✓	
Tetrad N – Developmental History	✓	✓	✓
Triad S – Biology/Chemistry		✓	✓
Triad T – Chemistry		✓	✓
Tetrad V – Biology/Chemistry	✓	✓	✓
Tetrad W – Biology/Chemistry	✓	✓	✓

Procedure

The data used for the study were collected from the university's registrar's office. The researcher requested all the demographic information and student records from the department directly, a process that required both Institutional Review Board (IRB) and registrar approval. Data were obtained for the Fall 2010, Fall 2011, and Fall 2012 cohorts of students who were enrolled in learning communities. Table 2 lists the independent and dependent variables selected for the study.

Table 2
Dependent and Independent Variables

Variable	Explanation
Retention (into second fall semester)	1=retained, 0=not retained
Probation Status (after first semester)	1=on probation (below 2.0 GPA), 0=not on probation
Learning Community	Letter representing LC membership
First-Semester Hours	Number of hours attempted in first fall semester
High School Percentile	Percentile rank in high school class
Transferred Hours	Number of hours completed prior to college admission
SAT Score	SAT score or converted ACT score (ACT, 2008)
Age	Age in years on the first day of class
Days since Admission	Number of days elapsed between admission and fall
Days since Orientation	Number of days elapsed between orientation and fall
Developmental Status	1=not college-ready, 0=college-ready
Gender	1=female, 0=male
First-Generation Status	1=first-generation, 0=not first-generation
Ethnicity	1=Hispanic, 0=Non-Hispanic
Pell Grant Eligibility	1=Eligible, 0=Ineligible
Admission Status	1=Accepted, 0=Alternatively Admitted*

* Based on alternative qualification or committee review

Retention into the second fall semester and first-semester GPA have been previously identified at the institution as predictors of graduation, so both variables were selected as outcomes for the study. The independent variables were selected based on information that would be available in the students' records on the first day of classes, and were matched to one of the categories identified by Tinto's (1975; 1997) SIM classifications. Students who were not traditional first-time college students or who did not attend orientation were removed from the data set. The final data file consisted of 4,215 student records.

After a profile of participants was tabulated, multivariate analyses employed binary logistic regression using SPSS. Logistic regression is used to estimate the probability of an event occurring—in the study, either retention or probation—based on a set of predictor variables (Field, 2013). In binary logistic regression, a dichotomous outcome is transformed into a linear model by comparing each independent variable to the log odds of the event taking place. In an exploratory study, each independent variable is tested to determine its unique contribution to the prediction of the outcome, that is, its relationship to the log odds of the event to determine if it meets the inclusion criteria to be included in the final model. The model can then be used to estimate the probability of the event occurring as $p(\text{event}) = 1/(1 + e^{-z})$, where $z = \text{Constant} + B_1X_1 + B_2X_2 + \ldots + B_nX_n$. The Constant and Bs are coefficients obtained from the logistic regression. The Wald statistic tests whether the coefficient for each of the independent variables in the model is zero (0); it has a chi-square distribution (Field, 2013; Hosmer & Lemeshow, 2000). The Nagelkerke R^2 and classification table were examined to evaluate the practical significance and power of the logistic regression models. Similar to other coefficients of determination, the Nagelkerke R^2 represents the amount of variance in the outcome that is explained by the model's variables (Nagelkerke, 1991). The Hosmer and Lemeshow Test was used to examine the goodness-of-fit of logistic regression models (Hosmer & Lemeshow, 2000). Finally, odds ratios were calculated and examined to interpret the variables that defined the various models.

Results

Profile of Participants

The data for the study consisted of 4215 first-year student records for the Fall 2010, Fall 2011, and Fall 2012 semesters. The subjects were between the ages of 18 and 24, matriculated with less than 30 transferred hours, and enrolled in a First-Year Seminar (UCCP 1101) course in a learning community. The majority (62.20%) were enrolled in either History (35.40%) or Political Science (26.80%) learning communities, were college-ready in reading and mathematics (74.80%), were female (58.60%), were not first-generation (68.70%), and were alternatively admitted (54.10%). While no ethnicity was in the majority, 45.50% and 40.60% of the subjects had been identified as Hispanic and White, respectively.

The high school percentile data were treated as ordinal with the median of .68. The average SAT score was 966.22 (SD = 140.72). The distribution of age at the start of the fall semester was positively skewed; the median age was 18.59 years. A typical first-year student enrolled in 13.77 hours (SD = 1.25) during the first semester, was admitted 175.52 days (SD = 78.47) prior to the start of the semester, and attended orientation 37.06 days (SD = 24.21) before the first day of classes. The distribution of transferred hours brought in by first-year students was positively skewed with a median of 0.00 hours.

Descriptive statistics were obtained for each of the 13 predictor variables sorted by the learning community category to explore the role that learning community membership might have played in the retention and probation status of students (see Table 3). Group differences on the basis of means (for continuous variables) and frequencies (for categorical variables) were examined, using One-Way ANOVA and Chi-Square Test of Independence, respectively. Statistically significant group differences were found for all 13 predictor variables. For example, the average number of hours taken during the first semester by Science learning community students was 14.04 (SD = 1.46), which was significantly higher than the average number of first-semester hours for students in all other learning communities except those placed in the "Other" category, *Welch's* $F(5, 793.16) = 18.55$, $p < .01$. Students in the "Other" category of learning communities came in with the highest SAT scores, and group differences were statistically significant when compared to every other learning community category excluding the Science learning community, *Welch's* $F(5, 746.71) = 102.23$, $p < .01$.

Another example of notable group differences was developmental status, which ranged from 9.00% of the students in the Science learning community to 99.00% of the Developmental History learning community students; group differences were statistically significant, $\chi^2(5, N = 4215) = 607.32$, $p < .01$. Group difference based on admission status were also statistically significant, $\chi^2(5, N = 4215) = 223.42$, $p < .01$; while 65% of the students in the "Other" category of learning communities had been accepted based on standard admission criteria, 95% of the students in the Developmental History learning community had been alternatively admitted. Results are summarized in Table 3.

Table 3
Independent Variable Means by Learning Community (N = 4215)

	SOCI	HIST	POLS	SCI	DHIST	OTHER	ALL
First-semester Hours[a]	13.50	13.78	13.71	14.04	13.39	13.97	13.77
High School Percentile[a]	.63	.66	.64	.69	.54	.60	.68
Transferred Hours[a]	4.05	3.56	4.07	5.48	0.73	2.51	3.96
SAT Score[a]	953.46	965.89	947.84	1021.66	827.52	1027.10	966.22
Age[a]	18.66	18.67	18.73	18.60	18.71	18.87	18.68
Days since Admission[a]	174.21	174.70	167.05	197.13	166.60	150.71	175.52
Days since Orientation[a]	31.61	37.42	33.24	47.07	33.32	33.93	37.06
Developmental Status[b]	.29	.23	.27	.09	.99	.17	.25
Gender[b]	.56	.59	.59	.63	.70	.24	.59
First-Generation Status[b]	.34	.32	.33	.27	.41	.18	.31
Ethnicity[b]	.47	.46	.48	.42	.54	.29	.46
Pell Grant Eligibility[b]	.55	.50	.51	.44	.61	.31	.50
Admission Status[b]	.40	.46	.42	.62	.05	.65	.46

[a] One-Way ANOVA indicated significant differences between group means
[b] Chi-square Test of Independence indicated significant differences between group frequencies
Means for categorical data are reported for ease of interpretation.

Logistic Regression Models

Two logistic regression models were developed to identify the best predictors of retention and probation status regardless of learning community membership. The dependent variable for the first model was retention (1 = retained, 0 = not retained) into the second academic year. Probation status (1 = on probation, 0 = not on probation) following the first semester of college served as the outcome measure for the second model. Models 1 and 2 were then repeated for each learning community in order to develop prediction models for students in each of the six learning community (LC) categories.

Model 1: Predicting Retention Independent of Learning Community

Out of the 13 independent variables, five (high school percentile, SAT score, Pell Grant eligibility, days since admission, and days since orientation) met the criteria to be included in the model to predict retention. The model was statistically significant, $\chi^2(5) = 215.44$, $p < .01$, correctly classified 62.20% of the students, and accounted for 7.30% of the variance in retention. The goodness-of-fit test was not statistically significant, $\chi^2(8) = 9.69$, $p = .29$, indicating that the model fit the data. Inspection of the odds ratios revealed that retention was likely for students with higher high school percentiles, higher SAT scores, more days since admission, and more days since orientation. Students who were eligible for Pell Grants were less likely to be retained. Results are summarized in Table 4.

Model 2: Predicting Probation Status Independent of Learning Community

Eight variables were included in the final model to predict probation status. The model was statistically significant, $\chi^2(8) = 425.88$, $p < .01$, correctly classified 72.10% of the students, and accounted for 14.90% of the variance in retention. The goodness-of-fit test was not statistically significant, $\chi^2(8) = 13.83$, $p = .09$, indicating that the model fit the data. The odds ratios showed that probation was likely for students who were Hispanic and eligible for Pell Grants. Probation was less likely for students with higher high school percentiles and females, as well as for students with more transferred hours, higher SAT scores, more days since admission, and more days since orientation. Results are summarized in Table 5.

Table 4
Logistic Regression Model for Retention Independent of Learning Community (N = 3886)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	.96	.17	30.56*	2.62
SAT Score	.01	<.01	8.99*	1.00
Pell Grant Eligibility	-.34	.01	23.74*	0.71
Days since Admission	.01	<.01	28.08*	1.00
Days since Orientation	.01	<.01	15.95*	1.00
CONSTANT	-1.51	.26	33.13*	0.22

* $p < .01$

Table 5
Logistic Regression Model for Probation Independent of Learning Community (N = 3886)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	-2.06	.20	106.08*	0.13
Transferred Hours	-.03	.01	18.46*	0.97
SAT Score	-.01	<.01	12.38*	0.99
Gender	-.24	.08	9.47*	0.79
Ethnicity	.24	.08	9.02*	1.27
Pell Grant Eligibility	.34	.08	18.97*	1.41
Days since Admission	-.01	<.01	13.65*	0.99
Days since Orientation	-.01	<.01	19.01*	0.99
CONSTANT	2.01	.32	40.13*	7.48

* $p < .01$

Models 1a-1f: Predicting Retention within LC Categories

Six models were created to predict retention based on learning community category membership. Each model included different predictors, but high school percentile and the number of days since orientation were the most common variables, followed by Pell Grant eligibility and the number of days since orientation. All of the models were statistically significant, although none of the variables in the study met the criteria to be included in the model to predict retention in the Developmental History learning community. Results of the predictors identified for the various models can be found in Table 6.

Models 2a-2f: Predicting Probation Status within LC Categories

Similarly, six models were created to predict the probation status for students in each learning community category. More of the original 13 variables were included in these models, with the number of days since orientation serving as the most frequent predictor. High school percentile and SAT score also met the criteria to be included in half of the learning community predictive models. No variables were included in the model to predict probation status for students in the Developmental History learning community. Results are summarized in Table 7.

Table 6
Predictors of Retention in Various Logistic Regression Models

Predictor	M1	M1a	M1b	M1c	M1d	M1e	M1f
First-Semester Hours							
Developmental Status							
High School Percentile	✓		✓	✓	✓		
Transferred Hours			✓				
SAT Score	✓						✓
Age							
Gender							
First-Generation Status							
Ethnicity							
Pell Grant Eligibility	✓		✓	✓			
Days since Admission	✓		✓	✓			✓
Admission Status							
Days since Orientation	✓	✓			✓		

M1 = Model for predicting retention independent of learning community, M1a = Sociology Learning Community, M1b = History Learning Community, M1c = Political Science Learning Community, M1d = Science Learning Community, M1e = Developmental History Learning Community, M1f = Other Learning Communities

Table 7
Predictors of Probation Status in Various Logistic Regression Models

Predictor	M2	M2a	M2b	M2c	M2d	M2e	M2f
First-Semester Hours							
Developmental Status							
High School Percentile	✓		✓	✓	✓		
Transferred Hours	✓		✓		✓		
SAT Score	✓			✓	✓		✓
Age							
Gender	✓						
First-Generation Status							
Ethnicity	✓						
Pell Grant Eligibility	✓		✓				
Days since Admission	✓						
Admission Status							
Days since Orientation	✓	✓	✓	✓	✓		

M2 = Model for predicting probation status independent of learning community, M2a = Sociology Learning Community, M2b = History Learning Community, M2c = Political Science Learning Community, M2d = Science Learning Community, M2e = Developmental History Learning Community, M2f = Other Learning Communities

Discussion

The purpose of the study was to identify pre-college variables that could serve as predictors of retention and probation status of first-year students in learning communities. The results indicated that several of the 13 variables used in the study were useful in predicting the retention and probation status of first-year students, but also that the predictor variables changed, based on the learning community under scrutiny.

Additionally, although the students in each learning community were markedly different from one another when the groups were compared on each of the 13 pre-college variables (see Table 3), there was not a statistically significant difference in retention or probation status between any of the learning communities. This seems to indicate that some factor within the learning community experience or program—or some inexplicable outside factor—mitigated these incoming differences so students landed on probation and were retained at similar rates across the learning community categories. In fact, this result could be used to argue that the structure of learning communities themselves provide a significant inherent intervention beyond the specific subject areas of the courses.

The logistic regression models to predict retention and probation status independent of learning community category both indicated that high school percentile was a unique and significant predictor of student success. This finding is in agreement with the literature on first-year student success (Astin, 1971; Goldman & Widawski, 1976; Noble & Sawyer, 2004; Stiggins, Frisbie, & Griswald, 1989; Zheng, Saunders, Shelley, & Whalen, 2002) and indicates that students who did well in high school will continue to do well in learning communities for first-year college students. While this information is not surprising, it can be helpful in deciding which students are in most need of early intervention.

Comparing the prediction models for each learning community revealed some notable patterns within and among the learning community categories. Various predictors were included in the models to predict retention and probation status for students in each category, but they were different predictors for every category. At an individual institution, this type of information not only indicates the types of students that select particular community options, but also the impact of certain student traits—such as SAT score or orientation date—on student success in those learning communities. These differences invite future exploration by the learning community program.

Binary logistic regression models use existing data about past events to create equations to predict future outcomes. Thus, the study's models can be used to determine the probability and odds of landing on probation or being retained for new incoming students with similar characteristics based on their pre-college variables. For example, let us take a typical profile for a hypothetical incoming student named Jane Doe. Jane is college-ready in reading and mathematics, not first-generation, female, 18.68 years old, non-Hispanic, and ineligible for Pell Grants. She was ranked in the 68th percentile of her high school class, earned the SAT score of 960, is enrolled in 13 hours during her first semester, and has 3 transferred hours. She was alternatively admitted 176 days before the first day of class and attended the orientation 37 days before the start of the semester. Table 8 includes the calculations for determining the probability and odds of Jane being retained based on her pre-college variables. Table 9 details the probability and odds of Jane being on probation after her first semester using various models.

According to all of the retention models, Jane Doe would be likely to be retained and not be on probation. In fact, her odds for both measures, regardless of learning community membership, indicate a high probability that she would return for the second year of college and that she would have a 2.0 GPA or higher after the first semester. Within the learning communities, the odds of Jane being retained are the greatest in the "Other" category of learning communities, but Jane would need to be an engineering, geology, or environmental science major to register for those courses. The probabilities of Jane being retained after enrolling in the History or Political Science learning community are 85% and 87%, respectively. Jane's odds of returning after one year are smaller in the Science and Sociology learning communities, but are still in favor of a positive result. The odds of Jane landing on probation after her first semester are small to negligible in all of the learning communities. The highest probability that Jane would land on probation is in the Sociology learning community at 26%, but the odds are nearly three to one (3:1) against that outcome.

Implications

The results of the study have implications for the learning community program under review, as well as larger implications for learning community scholarship. Perhaps the most obvious implication is that pre-college variables can be useful as predictors of both retention and probation status, as suggested by Tinto's (1975; 1997) SIM, which classified student characteristics into four categories that work together to explain, or predict, outcomes such as retention and probation status. Although additional surveys and more student data could undoubtedly contribute to an increased understanding of learning community and student success in the first year, it is possible to formulate prediction models based on student data that are available on the first day of classes. This implication applies to any institution interested in first-year student success.

Table 8
Prediction of Retention using Binary Logistic Regression Models

Model 1 (Independent of Learning Community Membership)

Retention = -1.51 + .96(HS Percentile) + .01(SAT Score) —.34(Pell Grant Eligibility)
+ .01 (Days since Admission) + .01(Days since Orientation)
Retention = -1.51 + .96(.68) + .01(960) - .34(0) + .01(176) + .01(37) = 10.87

p(Retention) = 1 / (1 + e-10.87) = **.99**
odds(Retention) = .99/(1-.99) = **99.00 (in favor of retention)**

Model 1a: Sociology Learning Community

Retention = -.29 + .02(Days since Orientation)
Retention = -.29 + .02(37) = .45

p(Retention) = 1 / (1 + e-.45) = **.61**
odds(Retention) = .61/(1-.61) = **1.57 (in favor of retention)**

Model 1b: History Learning Community

Retention = -1.07 + 1.35(HS Percentile) + .03(Transferred Hours) - .30(Pell Grant Eligibility)
+ .01(Days since Admission)
Retention = -1.07 + 1.35(.68) + .03(3) - .30(0) + .01(176) = 1.70

p(Retention) = 1 / (1 + e-1.73) = **.85**
odds(Retention) = .85/(1-.85) = **5.46 (in favor of retention)**

Model 1c: Political Science Learning Community

Retention = -.46 + .88(HS Percentile) + .01(Days since Admission) - .56(Pell Grant Eligibility)
Retention = -.46 + .88(.68) + .01(176) - .56(0) = 1.90

p(Retention) = 1 / (1 + $e^{-1.90}$) = **.87**
odds(Retention) = .87/(1-.87) = **6.68 (in favor of retention)**

Model 1d: Science Learning Community

Retention = -1.44 + 1.57(HS Percentile) + .02(Days since Orientation)
Retention = -1.44 + 1.57(.68) + .02(37) = .37

p(Retention) = 1 / (1 + e-.37) = **.59**
odds(Retention) = .59/(1-.59) = **1.44 (in favor of retention)**

Model 1f: Other Learning Communities

Retention = -6.38 + .01(SAT Score) + .01(Days since Admission)
Retention = -6.38 + .01(960) + .01(176) = 4.98

p(Retention) = 1 / (1 + e-4.98) = **.99**
odds(Retention) = .99/(1-.99) = **99.00 (in favor of retention)**

Table 9
Prediction of Probation Status using Binary Logistic Regression Models

Model 2 (Independent of Learning Community Membership)

Probation = 2.01 – 2.06(HS Percentile) - .03(Transferred Hrs) - .01(SAT) - .24(Gender) + .24(Ethnicity) + .34(Pell) - .01(Days-Admission) - .01(Days-Orientation)
Probation = 2.01 - 2.06(.68) - .03(3) - .01(960) - .24(1) + .24(0) + .34(0) - .01(176) - .01(37)
Probation = -11.45

p(Probation) = 1 / (1 + $e^{11.45}$) = **1.06 * 10^{-5}**
odds(Probation) = 1.06 * 10^{-5}/ (1 - 1.06 * 10^{-5}) = **1.06 * 10^{-5}** (**not in favor of probation)**

Model 2a: Sociology Learning Community

Probation = -.31 - .02(Days since Orientation)
Probation = -.31 - .02(37) = -1.05

p(Probation) = 1 / (1 + $e^{1.05}$) = **.26**
odds(Probation) = .26/ (1 - .26) = **.35** (**not in favor of probation)**

Model 2b: History Learning Community

Probation = 1.13 - 2.67(HS Percentile) - .05(Transferred Hrs) + .42(Pell) - .01(Days-Orientation)
Probation = 1.13 – 2.67(.68) - .05(3) + .42(0) - .01(37) = -1.21

p(Probation) = 1 / (1 + $e^{1.21}$) = **.23**
odds(Probation) = .23/ (1 - .23) = **.30** (**not in favor of probation)**

Model 2c: Political Science Learning Community

Probation = 3.37 – 2.10(HS Percentile) - .01(SAT) - .01(Days-Orientation)
Probation = 3.37 – 2.10(.68) - .01(960) - .01(37) = -8.03

p(Probation) = 1 / (1 + $e^{8.03}$) = **3.26 x 10^{-4}**
odds(Probation) = 3.26 x 10^{-4}/ (1 - 3.26 x 10^{-4}) = **3.26 x 10^{-4}** (**not in favor of probation)**

Model 2d: Science Learning Community

Probation = 4.18 - 2.94(HS Percentile) - .04(Transferred Hrs) - .01(SAT) - .01(Days-Orientation)
Probation = 4.18 – 2.94(.68) - .04(3) - .01(960) - .01(37) = -7.91

p(Probation) = 1 / (1 + $e^{7.91}$) = **3.67 x 10^{-4}**
odds(Probation) = 3.67 x 10^{-4}/ (1 - 3.67 x 10^{-4}) = **3.67 x 10^{-4}** (**not in favor of probation)**

Model 2f: Other Learning Communities

Probation = 5.13 - .01(SAT Score)
Probation = 5.13 - .01(960) = -4.47

p(Probation) = 1 / (1 + $e^{4.47}$) = **.01**
odds(Probation) = .01/ (1 - .01) = **.01** (**not in favor of probation)**

On a program level, it would make sense to share information about the variables used in the study with the learning community teaching teams as soon as possible, preferably before the start of the semester. After students are registered, a profile of students in each learning community could be created with aggregated information about the students who would be in their linked classes to distribute to faculty. This profile could be integrated into planning meetings and would contain information such as the mean SAT score of the incoming students for the learning community, the number of first-generation students, and the number of students who were alternatively admitted. Learning community teaching teams could also request to have reports created for each student that include the pre-college variables, as well as the prediction models for retention and probation status (with and without respect to learning community membership). The learning community teams could elect to target specific interventions for students in a particular range of risk (for example, the students who attended the final orientation) for being on probation or not returning the next fall. These suggestions would require the learning community leadership team to run the models with incoming student data and create the reports to distribute to each of the teaching teams. Because interventions targeted at individual students are effective, the extra effort at the front end seems justified

Another implication of the findings for learning community leadership is that the differences among the learning community prediction models could be used for program assessment. Although there were no statistically significant differences in the ultimate retention and probation rates among the different learning communities, the odds of landing on probation or being retained went up or down depending on the learning community when the different models were used with pre-college data for a typical student. These differences seem to indicate inherent differences between the learning communities themselves that warrants further exploration by the program administrators. It seems that there might also be a way to use the prediction models to assess the effectiveness of individual learning communities, such as the Developmental History learning community, by comparing the predicted number of students on probation (or retained) at the start of the semester with the actual number of students on probation (or retained) at the end.

Beyond the institutional implications, this study contributes to a niche in the literature about first-year students in learning communities that had been lacking up to this point. The body of research on learning communities argues that they contribute to first-year student success and are a high-impact practice (Kuh, 2008), but little to no research had previously been published about the use of pre-college variables in logistic regression models to predict first-year student success in *various* learning community categories at an individual institution. The implication, therefore, for learning community scholars is the potential for regression analysis to be conducted in programs across the country as a way to explore the relationship between the information available about students on the first day of class and first-year student success rates. Prediction models can be employed on campuses in order to help target interventions led by learning community teaching teams. Learning community leaders could subsequently compile national data regarding the use of prediction models to share with the field. Administrators regularly require learning communities to justify their work with quantitative evidence of value added. The role that learning communities play in retention can be clarified with the use of predictive models that track targeted intervention.

Perhaps most importantly, the results of the study invite further analysis – of both qualitative and quantitative—to dive even deeper into the data available at individual institutions about the first-year students who engage in learning communities. Now that variables have been identified as unique predictors (such as orientation date) and non-predictors (such as admission or first-generation status) of retention or probation status, differences regarding their impact within the context of the particular learning communities program can be further investigated, either through additional quantitative analysis or by qualitative means to shed more light on the phenomenon. Because of the relative dearth in the literature regarding the impact of learning communities on first-year students (Andrade, 2008; MacGregor & Smith, 2005), it would seem that new avenues for research would be welcomed by anyone interested in the fate of the learning community movement and first-year college students as a whole. The MDRC's recent findings (Visher, Weiss, Weissman, Rudd, & Wathington, 2012) question the impact of learning communities on college student success and are a real threat to the future of learning communities; this study is only the first step in the formation of a complete, undeniable, and empirically-based repudiation.

Delimitations, Limitations, and Assumptions

The study was delimited to (a) first-year students at a single South Texas public university; (b) 13 pre-college variables which served as potential predictors; and (c) the outcome measures of retention and probation status. Due to the non-experimental nature of the study, no causal inferences were drawn. Some of the pre-college variables were self-reported; the underlying assumption was that students were truthful in reporting details such as high school rank and size, birthdate, and first-generation status on their applications for admission. It was also assumed that the data collected from the registrar were accurate and complete. Finally, this study assumed that the learning communities were relatively consistent from year to year and that there were no major pedagogical shifts within the three years under review.

Conclusions

According to the National Center for Higher Education Management Systems (2013), the national retention rate for first-time college freshmen at all four-year institutions in 2010 was 77.10%, while the Texas state retention rate came in a bit lower at 73.30%. Research has indicated that student performance in the first year is predictive of cumulative undergraduate GPA and subsequent graduation (Terenzini, Springer, Yaeger, Pascarella, & Nora, 1996). The aim of the study was to identify the pre-college student characteristics that are useful in prediction of the retention and probation status of first-year students in learning communities. Clearly, any empirical evidence to support a particular intervention (such as learning communities) to increase first-year student success would be worthy of note to any institution concerned about student persistence and graduation rates.

The results of the study shed light on particular characteristics of incoming students so that interventions can be targeted to the students who might profit from them the most. Learning community teaching team members can identify which students are most at risk of being on probation and not returning the following year. Because the results are based on former students who participated in the program, the learning communities that are shown to have been successful in the past in helping students stay off probation and remain at the institution can and should be further explored. Ideally, the traits of the successful learning communities could then be extended to the program as a whole in order to help the entire first-year student population. Although the particular predictors and models created in the study may not be generalizable to outside institutions because of differing student populations and situations, the process of analyzing pre-college traits of first-year students in learning communities using logistic regression models could easily be replicated in other settings. National supporters of the learning communities' movement, such as the Washington Center or the LEAP initiative, would undoubtedly be interested in results that contribute to the growing body of literature about how learning communities contribute to student success.

References

Andrade, M. S., (2008). Learning communities: Examining positive outcomes. *Journal of College Student Retention, 9*(1), 1-20.

Araiza, I. (2006). An assessment of the First-Year Learning Community Program: A report on three focus groups with Texas A&M University – Corpus Christi students enrolled in the First-Year Learning Communities Program. Unpublished manuscript, First-Year Experience Council, Texas A&M University – Corpus Christi, Corpus Christi, TX.

Astin, A. W. (1971). *Predicting academic performance in college: Selectivity data for 2300 American colleges.* New York, NY: The Free Press.

Chamberlain, T. A. (2011). *Branching out or breaking off? An exploration of first-year student social networks.* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3488197).

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage.

Glynn, J. G., Sauer, P. L., & Miller, T. E. (2003). Signaling student retention with prematriculation data. *NASPA Journal*, *41*(1), 41-67.

Goldman, R. D., & Widawski, M. H. (1976). A within-subjects technique for comparing college grading standards: Implications in the validity of the evaluation of college achievement. *Educational and Psychological Measurement*, *36*(2), 381-90.

Griffin, A. M & Romm, J. (Eds.). (2008). *Exploring the evidence: Reporting research on first-year seminars, Volume IV*. Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.

Habley, W. R., & Bloom, J. L. (2012). *Research-based strategies for college student success*. San Francisco, CA: Jossey-Bass.

Hansen, M., Meshulam, S., & Parker, B. (2013). Assessing the effectiveness of a learning community course design to improve the math performance of first-year students. *Learning Communities Research and Practice, 1*(1), 1-24.

Heaney, A., & Fisher, R. (2011). Supporting conditionally-admitted students: A case study of assessing persistence in a learning community. *Journal of the Scholarship of Teahcing and Learning, 11*(1), 62-78.

Henscheid, J. M. (Ed.). (2004). *Integrating the first-year experience: The role of first-year seminars in learning communities* (Monograph No. 39). Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.

Herreid, C. H., & Miller, T. E. (2009). Analysis of variables to predict first year persistence using logistic regression analysis at the University of South Florida: Model v2.0. *College and University, 84*(4), 12-19.

Hill, P. (1985, October). *The rationale for learning communities*. Speech presented at the Inaugural Conference on Learning Communities, Olympia, WA.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York, NY: Wiley.

Huerta, J. C. (2004). Do learning communities make a difference? *PS: Political Science and Politics, 37*(2), 291-296.

Huerta, J. C. & Sperry, R. A. (2010). Pulling it together: Using integrative assignments as empirical direct measures of student learning for learning community program assessment. *Journal of Learning Communities Research, 5*(1), 85-92.

Kahn, J. H., & Nauta, M. M. (2001). Social-cognitive predictors of first-year college persistence: The importance of proximal assessment. *Research in Higher Education, 42*(6), 633-652.

Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter.* Washington, DC: Association of American Colleges and Universities.

Kuh, G. D. (2013). Taking HIPs to the next level. In G. D. Kuh & K. O'Donnell (Eds.), *Ensuring quality and taking high-impact practices to scale*. Washington, DC: Association of American Colleges and Universities.

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. *The Journal of Higher Education, 79*(5), 540-563.

Kutil, B., & Sperry, R. (2012). The science of first-year learning communities. In M. Soven, D. Lehr, S. Naynaha, & W. Olsen (Eds.), *Linked courses for general education and integrative learning: A guide for faculty and administrators*. Sterling, VA: Stylus Publishing.

Lardner, E. & Malnarich, G. (2008). A new era in learning community work: Why the pedagogy of intentional integration matters. *Change, 40*(4), 30-37.

Lardner, E., & Malnarich, G. (2009). Assessing integrative learning: Insights from the Washington Center's National Project on Assessing Learning in Learning Communities. *Journal of Learning Communities Research, 3*(3), 1-20.

MacGregor, J., & Smith, B. L. (2005). Where are learning communities now? National leaders take stock. *About Campus, 10*(2), 2-8.

Nagelkerke, N. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692.

National Center for Higher Education Management Systems. (2013). *Retention rates – First-time college freshmen returning their second year.* Retrieved from http://www.higheredinfo.org/dbrowser/index.php?submeasure=223&year=2010

Noble, J. P., & Sawyer, R. L. (2004). Is high school GPA better than admission test scores for predicting academic success in college? *College and University, 79*(4), 17-22.

Nora, A., Cabrera, A., Hagedorn, L. S., & Pascarella, E. (1996). Differential impacts of academic and social experiences on college-related behavior outcomes across different ethnic and gender groups at four-year institutions. *Research in Higher Education*, *37*(4), 427-451.

Porter, S. R. (1999, June). *Viewing one-year retention as a continuum: The use of dichotomous logistic regression, ordered logit and multinomal logit.* Paper presented at the Annual Meeting of the Association of Institutional Research, Seattle, WA.

Roccini, L. M. (2011). The impact of learning communities on first year students' growth and development in college. *Research in Higher Education*, *52*, 178-193.

Smith, B., MacGregor, J., Matthews, R., & Gabelnick, F. (2004). *Learning communities: Reforming undergraduate education*. San Francisco, CA: Jossey-Bass.

Smith, R. A. (2010). *"Only connect": A mixed methods study of how first-year students create residential academic and social networks.* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3437586).

Snyder, V., Hackett, R. K., Stewart, M., & Smith, D. (2002, April). *Predicting academic performance and retention of private university freshmen in need of developmental education*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Sterba-Boatwright, B. (2000). The effects of mandatory freshman learning communities: A statistical report. *Assessment Update*, *12*(2), 4-5.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, *8*(2), 5-14.

Stuart, G. S. (2008). *Making connections: The role of social networks of first-year students enrolled in learning communities in predicting success in college*. (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3343351).

Taylor, K., Moore, J. S., MacGregor, J., & Lindbland, J. (2003). *Learning communities research and assessment: What we know now.* National Learning Communities Project Monograph Series, Olympia, WA: The Evergreen State College, Washington Center for Improving the Quality of Undergraduate Education and the American Association for Higher Education.

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences and cognitive development. *Research in Higher Education, 37*(1), 1-22.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89-125.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *The Journal of Higher Education, 59*(4), 438-455.

Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *The Journal of Higher Education, 68*(6), 599-623.

Tinto, V. (2000). What have we learned about the impact of learning communities on students? *Assessment Update, 12*(2), 1-2.

Tinto, V., Love, A. G., & Russo, P. (1993). *Building learning communities for new college students: A summary of research findings of the collaborative learning project*. University Park, PA: National Center on Postsecondary Teaching, Learning, and Assessment.

Van Gennep, A. (1960). *The rites of passage*. (M. B. Vizedom & G. L. Caffee, Trans.). Chicago, IL: University of Chicago Press (Original work published 1908).

Visher, M. G., Weiss, M. J., Weissman, E., Rudd, T., & Wathington, H. D. (2012). *The effects of learning communities for students in developmental education: A synthesis of findings from six community colleges* [Executive summary]. New York, NY: National Center for Postsecondary Research.

Voorhees, R. A. (1987). Toward building models of community college persistence: A logit analysis. *Research in Higher Education, 26*(2), 115-129.

Warburton, E., Bugarin, R., & Nunez, A. (2001). *Bridging the gap: Academic preparation and postsecondary success of first-generation students* (NCES 2001–153). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

Wetzel, J. N., O’Toole, D., & Peterson, S. (1999). Factors affecting student retention probabilities: A case study. *Journal of Economics and Finance, 23*(1), 45-55.

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education, 45*(2), 115-138.

Zheng, J. L., Saunders, K. P., Shelley, M. C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development, 3*(2), 267-83.