New Perspectives in Science Education
International Conference
NEW PERSPECTIVES in SCIENCE EDUCATION

The Role of College-Level Mathematics in STEM Major Persistence

Jaimi Paschal[1], Amanda Taggart[2]

California State University, Bakersfield, U.S.A.[1]
Utah State University, U.S.A.[2]

Abstract

The purpose of this study was to examine the association between a first-term college-level mathematics course and STEM major persistence utilizing data from students who declared STEM majors at matriculation at a 4-year university in the United States. Decreasing logistic regression was used to identify significant variables likely to increase student persistence in a STEM major through the sixth college term. Findings indicated that students who passed a first-term college-level mathematics course were at significantly greater odds of persisting.

Keywords: STEM, persistence

1. Introduction

As the focus on STEM major completion increases, there is a need to study the role of entry-level college-level mathematics courses that act as gatekeeper courses in most STEM disciplines. The purpose of this study was to examine the relationship between mathematics course-taking in the first term of college and college student persistence in STEM majors by asking the following research questions: (1) Are there differences between the characteristics of college students who do and do not persist in STEM majors to the sixth college term? (2) What is the contribution of demographic variables, pre-college variables, and college variables to persistence in STEM majors to the sixth college term?

2. Method

The site of the study was a 4-year public university in the western United States with an enrollment of approximately 9,500 undergraduate and graduate students. Participants consisted of all first-time freshmen who enrolled in the university between the Fall 2008 and Fall 2013 terms and had no missing data, (n = 625). At the time of matriculation, all the participants met college-ready mathematics requirements during their first college term. In addition, they declared and were enrolled in an undergraduate STEM major. Data were collected from archival records provided by the university under study. Three demographic variables were included in the model, including students' gender, race/ethnicity, and socioeconomic status (SES). Two pre-college variables, high school grade point average (GPA) and the determinants in high school that regulated student placement in a mathematics course in college, were also included in the model. The college variable included in the model was whether or not students passed a first-term college-level mathematics course. The outcome examined in this study was student retention in a STEM major at the sixth college term. The choice to study STEM major persistence through the sixth college term was made because most universities require students to declare a major by the end of the spring term prior to their junior year and because of its use in prior research [1], [2], [3].

Data analyses were conducted using SAS University Edition software. To answer the first research question, descriptive statistics were computed, and to answer the second research question, a regression model was used to examine the relationship between the dichotomous dependent variable and the set of predictor variables. Retention in a STEM major to the sixth college term was coded as the reference category. The form of the final model suggested with backwards model selection, the probability of a student being retained at the sixth college term, denoted as p, is

$$\ln[p(x)/1\text{-}p(x)] = \beta_0 + \beta_1 (Ethnic_i) + \beta_2 (Sex_i) + \beta_3 (Low\_SES_i) + \beta_4 (MCourse_i) + \beta_5 (GPA1_i) + \beta_6 (MthPlmt_i)$$

A second backwards model selection analysis was then run with all calculus courses removed from the data. The backwards model selection determined that logistic regression was an appropriate model. This was performed to determine whether the higher-level mathematics course data deletion yielded any difference in significance between the predictor and dependent variables.

The final logistic regression model was determined by utilizing backwards model selection. The model was fit with all the variables. The binary variable interactions were analyzed and the least significant

variable pair was removed. The model was refit with the remaining variables. These steps were repeated until only singular variables without significant interaction remained. The final model used the Wald chi-square statistic and *p*-values to determine model fit statistical significance. $P < .05$ was used to determine whether the predictor variables were significantly related to the dependent variable. The odds ratio variable was used to interpret the effect size of the predictor variables in the logistic regression model. A parameter estimate was run to determine whether the study sample means were indicative of the entire population.

3. Findings

The final sample included 47% female and 53% male students with a declared STEM major at matriculation. A large portion (65%) of the sample was classified as low-SES, and 73% of the participants were underrepresented minority (URM) students. Additionally, 63% of the sample entered college with a high school GPA of 3.0 or higher. Once enrolled, 63% of the students from the sample passed a college-level mathematics course. A descriptive comparison of students retained in a STEM major through their sixth college term (n = 231) versus students who were not retained in a STEM major through their sixth college term (n = 394) revealed that male students were retained at a higher rate than female students (75% versus 66%), non-URM students were retained at a higher rate than URM students (73% versus 68%), and students classified as low-SES were retained at a higher rate than students not classified as low-SES (73% versus 69%). Students entering college with a high school GPA of 3.0 or higher outnumbered students entering college with lower GPAs (61% versus 54%).

Students who passed a first-term college-level mathematics course were retained in a STEM major at a much higher rate than students who did not take or pass a first-term college-level mathematics course (75% versus 67%). In addition, students who placed into a first-term college level mathematics course by earning a high score on the SAT or ACT exam, earned a passing score of 3 or higher on the AP calculus exam, or passing two remedial mathematics courses during a summer bridge program showed higher rates of retention in a STEM major through the sixth college term when compared to the other means in which a student could place.

The variable parameters obtained from the dataset motivated the use of these same variables in answering research question two.

The first logistic regression model considered all variables (e.g., gender, SES, ethnicity, high school GPA, mathematics placement, first-term college-level mathematics course) and two-way interactions between the binary variables (gender, SES, ethnicity, high school GPA, first-term college-level mathematics course). The least significant variable was removed from the model, the model refit, and logistic regression rerun. This method was repeated, removing the least significant variable one by one, until only significant variables remained.

The final model using backwards model selection indicated that of the six predictor variables and 10 two-way interactions, only two had a statistically significant effect on the prediction of retention in a STEM major through the sixth college term (i.e., students who entered college with a high school GPA of 3.0 or higher, students who took and passed a college-level mathematics course during their first term in college; $p < .05$). The overall fit of the model was found to be significant with a Wald chi-square statistic of $p < .05$. The interactions between variables indicated no statistically significant effect on long-term STEM retention.

An interpretation of the significant predictor variables indicated that students who entered college with a high school GPA of 3.0 or higher increased their odds of STEM retention through the sixth college term by a factor of 1.3. Further, the confidence intervals suggest that there is 95% certainty that the odds of STEM retention through the sixth college term increased between a factor of 1.058 and 2.38. Additionally, students that took and passed a first-term college-level mathematics course increased their odds of long-term STEM retention by a factor of 1.3. The confidence intervals suggest that there is 95% certainty that the odds of STEM retention through the sixth college term for these students increased between a factor of 1.12 and 2.23. A two-way interaction was performed between the two statistically significant variables with results indicating a non-significant *p*-value.

Table 1 displays the estimated regression coefficients, standard errors, odds ratios, and 95% confidence intervals for the final logistic regression model.

Table 1. Logistic Regression Model: Analysis of Maximum Likelihood Estimates

Variable	β	Standard Error	Odds Ratio[1]	95% Confidence Intervals	
Pre-College Variables					
High school GPA					
GPA 3.00 or higher	0.235	0.112	1.265*	1.058	2.38
College Variables					
First-term college-level mathematics course Passed	0.271	0.1069	1.311*	1.12	2.23

*p < .05; [1]Only odds ratios with significant p-value are displayed.

To determine whether students who entered their first term of college and took a calculus course affected the probability of retention, logistic regression analyses were performed on a second dataset in which all calculus courses were removed, leaving only lower-level college-level mathematics courses while all other predictor variables remained the same. Backwards model selection was performed to determine the best model fit using the Chi-square test statistic and its corresponding *p*-value indicating significance at .05. The first logistic regression analysis considering all variables (e.g., gender, SES, ethnicity, high school GPA, mathematics placement, first-term college-level mathematics course) and two-way interactions between the binary variables (gender, SES, ethnicity, high school GPA, first-term college-level mathematics course). The least significant variable was removed from the model, the model refit, and logistic regression rerun. This method was repeated, removing the least significant variable pair one by one, until only significant variables remained.

As with the primary logistic regression model results, the final model using backwards selection indicated that of the six predictor variables and 10 two-way interactions, only two showed a statistically significant effect on the prediction of STEM retention through the sixth college term (i.e., students who entered college with a high school GPA 3.0 or higher, students who took and passed a college level mathematics course during their first term in college; $p < .05$). The overall fit of the model was found to be significant with a Wald chi-square statistic of $p < .05$. Analysis of the variable interactions indicated no statistically significant effect on long-term STEM retention. Moreover, a review of the results indicated that students with a high school GPA of 3.0 or higher who entered college eligible to take a college-level mathematics course below calculus increased their odds of retention in a STEM major through the sixth term by a factor of 1.21. Further, the confidence intervals suggest that there is 95% certainty that the odds of STEM retention through the sixth college term for these students increased between a factor of 1.042 and 2.365. Additionally, these students increased their odds of retention in a STEM major through the sixth term by a factor of 1.28 when they took and passed a first-term college-level mathematics course. The confidence intervals suggest that there is 95% certainty that the odds of STEM retention through the sixth college term for these students increased between a factor of 1.055 and 2.156. Table 2 displays the output including regression coefficients, standard errors, significance values, odds ratios, and model fit statistics.

Table 2. Logistic Regression Model: Analysis of Maximum Likelihood Estimates without Calculus

Variable	β	Standard Error	Odds Ratio[1]	95% Confidence Intervals	
Pre-College Variables					
High school GPA					
GPA 3.0 or higher	0.232	0.108	1.207*	1.042	2.365
GPA 2.99 or lower					

College Variables					
First-term college-level mathematics course					
Passed	0.248	0.101	1.281*	1.055	2.156
Did not pass or take					

*p < .05; [1] Only odds ratios with significant p-value are displayed.

4. References

[1] Griffith, A. L. “Persistence of women and minorities in STEM field majors: Is it the school that matters?”, Economics of Education Review, 2010, pages 911-922.

[2] Allen, J., & Robbins, S. B. “Prediction of college major persistence based on vocational interests, academic preparation, and first-year academic performance”, Research in Higher Education, 2008, pages 62-79

[3] Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. “Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation”, Journal of Educational Psychology, 2002, pages 562-575.