# The Relationship between High-Risk Courses and Retention in University

Chen Zong and Suzann M. Koller

## About the Authors

Chen Zong is Institutional Research Analyst in the Office of Institutional Analysis at the University of Wyoming. Suzann M. Koller is Manager of Institutional Analysis in the Office of Institutional Analysis at the University of Wyoming.

## Abstract

Understanding the relationship between high-risk courses and Fall-to-Fall retention is essential to enhance student persistence and academic achievement in higher education institutions. The purpose of this study is to examine the relationship between high-risk courses and Fall-to-Fall retention of first-time, full-time students. The course data of 8,220 students between 2016 and 2020 at a large public research university were analyzed using descriptive statistics, correlation, and logistic regression methods: First, the characteristics of high-risk courses and the students who took the most high-risk courses were identified. Second, the findings of correlation analysis indicate that there was a statistically significant correlation between Fall-to-Fall retention and the number of high-risk courses students take in their first year. Third, the significant predictors of retention include the following: first-semester GPA, high school GPA, tuition residency, total number of courses taken in their first year, whether the student is first-generation, whether the student has an undeclared major, and the number of high-risk courses the student takes in their first year. The results of model likelihood ratio test indicate that the final model provides a significantly better fit to the data than the null model ($\chi^2 = 2393.9$, $df = 7$, $p < .001$, $R^2 = 39.9\%$). The findings of this study will provide useful information that institutions can use to identify the high-risk courses and to increase retention rate.

**Keywords:** high-risk courses, retention, higher education, logistic regression

**The AIR Professional File, Summer 2025**
**Article 178**

https://doi.org/10.34315/apf1782025
Copyright © 2025, Association for Institutional Research

## INTRODUCTION

In higher education institutions, high-risk courses can be a challenge for students, faculty, and administrators. These courses, distinguished by their demanding curriculums and elevated difficulty levels, can impact academic performance, retention rates, and overall student success (Martin & Arendale, 1992). In this study, the high-risk course is defined as a course with a low percentage of students who pass (i.e., including courses where students earn letter grades A, B, C, or D), which is slightly different from high DFW rate courses (i.e., including courses where students earn grades D, F, or Withdrawal) or high-failure rate courses (i.e., including courses where students earn grade F only).

Fall-to-Fall retention rates (i.e., the percentage of students who persist from one academic year to the next) serve as a vital indicator of student success and institutional effectiveness in higher education. Because retention rates in higher education institutions are usually focused on the first-year to second-year performances when there are no data about first-year students' previous college-level course work, it is difficult to identify and provide supports to high-risk students (Martin & Arendale, 1992). Therefore, understanding the factors contributing to the risk associated with the first-year, high-risk courses is essential for educators and administrators to develop effective strategies to support first-time, first-year students; and understanding the relationship between high-risk courses and Fall-to-Fall retention is essential to enhance student persistence and academic achievement.

The purpose of this study is to examine the relationship between high-risk courses and Fall-to-Fall retention of the first-time, full-time students at a large public research university. In this study, a first-time, full-time student is defined as a student who has no prior postsecondary education experience attending any institution for the first time at the undergraduate level, who is enrolled for 12 more semester credits (Integrated Postsecondary Education Data System [IPEDS], 2025). A high-risk course is defined as a first-year, for-credit course with 50 or more first-time, full-time students enrolled during the 5 years between Fall 2016 and Fall 2020, with fewer than 80% of the enrolled students passing the course. The findings of this study will provide useful information to identify the high-risk courses, improve the student success in these courses, and increase the Fall-to-Fall retention rate. There were three research questions:

- RQ1: What are the average number of high-risk courses that students take in their first year, by student characteristics and retention status?
- RQ2: Is there a statistically significant association between high-risk courses and Fall-to-Fall retention?
- RQ3: How well does a combination of student demographics, high school academic background, university academic experience, and first-year course enrollments predict Fall-to-Fall retention?

## LITERATURE REVIEWS

To answer the research questions, this literature review explores the characteristics of high-risk courses for college students, the types of students who are most likely to take high-risk courses, the association between high-risk courses and college student retention, and predictive models for college student retention, with a focus on the inclusion of high-risk courses.

## Characteristics of High-Risk Courses

The academic success of first-time, full-time students is a critical focus for higher education institutions. High-risk courses, often characterized by high failure rates and significant academic challenges, can impact students' academic success and retention at the institution. Identifying these courses and understanding their characteristics can help institutions implement strategies to improve student retention and success. High-risk courses "include those that have the following characteristics: large amounts of weekly readings from both difficult textbooks and secondary library reference works, infrequent examinations that focus on higher cognitive levels of Bloom's taxonomy, voluntary and unrecorded class attendance, and large classes in which each student has little opportunity for interaction with the professor or the other students" (Martin & Arendale, 1992, p. 14).

STEM (science, technology, engineering, and mathematics) courses, online courses, remedial courses, and gateway courses have often been considered to be high-risk courses in previous studies. STEM courses such as computer science, biology, and chemistry are frequently identified as high risk due to the difficulty or rigor of the course; however, non-STEM courses, including public speaking, critical reading, and writing, can also be high risk (Daniel, 2022). Bambara et al.'s (2009) study found that students who were enrolled in online high-risk courses, with over 30% withdrawal and failure rates, often had the academic experience of isolation, challenge, ownership, and acquiescence as the structural themes; the researchers suggested that there was a need for future research examining how other factors affect student retention and positive completion in high-risk courses. Remedial coursework was designed to help students who are not adequately prepared to succeed in college-level courses (Sanabria et al., 2020). Sanabria et al. (2020) found that students who took and passed remedial coursework were more likely to graduate, compared to peers who did not take remedial coursework, while students who did not pass remedial coursework were less likely to obtain a bachelor's degree or took longer than their peers to graduate. Although gateway courses are often considered to be high risk, Sargent et al.'s (2022) study indicated that receiving a DFW grade (i.e., a grade of D, F, or Withdrawal) in a gateway course did not significantly impact graduation rates over a 36-semester study period involving 3,667 students.

## Students in High-Risk Courses

High-risk courses present significant challenges to student success in higher education. Understanding the characteristics of college students who are most likely to take high-risk courses is crucial for developing targeted interventions to support at-risk students and to improve their academic success. Daniel (2022) found that students who met developmental course requirement criteria due to poorer academic performance were less likely to retain or persist; they emphasized the importance of enrolling high-risk students in skill-appropriate courses during their first semester to improve long-term retention. Salazar-Fernandez et al. (2021) analyzed how educational trajectories of undergraduate students in high-failure rate courses can lead to late dropouts. Their study suggested that institutions should monitor high-failure rate courses that students enroll in after a stopout, because students who took a stopout while having high-risk courses they must retake were more likely to have a late dropout. Haynes Stewart et al.'s (2011) study indicated that age, gender, high school performance, registration status (full time or part time), and participation in a first-year orientation program significantly predicted course outcome (pass or fail).

### Predictors for Student Retention

Predictive models for college student retention are useful tools for higher education institutions aiming to improve student success. These models typically use a combination of academic, demographic, and behavioral data to predict student outcomes. According to Paterson and Guerrero (2023), logistic regression is one of the commonly used techniques in these models; it allows institutions to identify significant factors that influence student success such as demographics, GPA, and course performance.

Predictive models for college student retention that include high-risk courses can provide valuable insights for higher education institutions. Daniel (2022) emphasized that early success in high-risk courses was a key factor in student resilience and retention, and that identifying and supporting students in these courses could significantly improve their chances of persistence. Higgs et al. (2021) highlighted that course-specific data (e.g., types of activities required in the courses, hours spent studying, teaching methods) could be important indicators for students' performance, retention, and passing rates.

Although it is a common problem in higher education, course failure or high-risk courses receive relatively little research attention (Haynes Stewart et al., 2011). Many previous research studies focused on high-risk students rather than on high-risk courses (Martin & Arendale, 1992). When searching for the keywords "high risk courses" and "retention" on Google Scholar, almost all top results are related to at-risk/high-risk students instead of high-risk courses (e.g., Daniel, 2022; Laskey & Hetzel, 2011; Valentine et al., 2011), and only a few studies explored college students enrolled in high-risk courses (e.g., Bambara et al., 2009; Martin & Arendale, 1992). Some studies analyzed the relationship between course failure and graduation or degree completion instead of between course failure and retention rates (e.g., Sanabria et al., 2020; Sargent et al., 2022). Some studies focused only on instructional approaches to improve course performance but did not analyze the relation between high-risk courses and retention (e.g., Martin & Arendale, 1992; Stone & Jacobs, 2008). Most studies focused on only one type of high-risk course; for example, some studies focused on online or distance learning courses (Baker et al., 2015; Bambara et al., 2009; Nash, 2005; Simpson, 2013), gateway courses (Bloemer et al., 2017; Sargent et al. 2022), remedial courses (Gajewski & Mather, 2015; Sanabria et al., 2020), or only one specific course such as calculus (Norton et al., 2018), geology (Roberts et al. 2018), or psychology (Haynes Stewart et al., 2011). Therefore, the research findings of this study will fill the gap in current literature with useful information about the relationship between high-risk courses and retention, with the goal of improving student success in higher education institutions.

## METHODS

This study used 5 years of data of first-time, full-time students enrolled in a public research university located in a town in the United States. Overall, the average total enrollment of the institution was around 10,000 to 12,000, including both undergraduate and graduate academic programs, with a student-to-faculty ratio of 13:1. In addition, first-time, full-time student enrollments were between 1,400 and 1,800 for each Fall cohort. The overall Fall-to-Fall retention rate of first-time, full-time students has ranged from 75% to 79% for the past 5 years. Using descriptive statistics, correlation, and logistic regression methods, the students' course and retention data were analyzed to explore the research questions.

## Data Source and Sample

The data of first-time, full-time students (8,220 students) enrolled in Fall semesters between 2016 and 2020 were used in this study. The students' demographic information is presented in Table 1. About 77% of the first-time, full-time students between Fall 2016 and Fall 2020 were White (n = 6,355); the balance (23%) represented the other Race/Ethnicity groups, including Race and Ethnicity unknown (*n* = 564), Hispanics of Any Race (*n* = 557), Two or More Races (*n* = 372), Nonresident Alien (*n* = 114), Black or African American (*n* = 111), Asian (*n* = 95), American Indian or Alaska Native (*n* = 42), and Native Hawaiian or Other Pacific Islander (*n* = 10). About the same numbers of Female (*n* = 4,128) and Male (*n* = 4,092) students were represented in this sample. Most of the students in this sample were aged 19 and younger (*n* = 7,976); there were 217 students were aged 20–24, and only 27 students were aged 25 and older. About 24% of the students in this sample were first-generation college students (*n* = 1,941); the balance (76%) were not first-generation college students (*n* = 6,279). Finally, the number of in-state students (*n* = 4,283) was slightly higher than out-of-state students (*n* = 3,937) in this study.

**Table 1. Demographic Information of First-time, Full-time Students between Fall 2016 and Fall 2020**

| Student Characteristics | Total | |
|---|---|---|
| **Race/Ethnicity** | **#** | **%** |
| American Indian or Alaska Native | 42 | 1% |
| Asian | 95 | 1% |
| Black or African American | 111 | 1% |
| Hispanics of Any Race | 557 | 7% |
| Native Hawaiian or Other Pacific Islander | 10 | 0% |
| Nonresident Alien | 114 | 1% |
| Race and Ethnicity unknown | 564 | 7% |
| Two or More Races | 372 | 5% |
| White | 6,355 | 77% |
| | | |
| **Gender** | | |
| Female | 4,128 | 50% |
| Male | 4,092 | 50% |
| | | |
| **Age at Entry** | | |
| 19 and under | 7,976 | 97% |
| 20–24 | 217 | 3% |
| 25+ | 27 | 0% |
| | | |
| **First-Generation Status*** | | |
| First-Generation | 1,941 | 24% |
| Not First-Generation | 6,279 | 76% |
| | | |
| **Tuition Residency** | | |
| In-State | 4,283 | 52% |
| Out-of-State | 3,937 | 48% |
| | | |
| **Total** | **8,220** | **100%** |

Note: *First-generation status is based on answers to the question: "Do either of your parents have a 4-year baccalaureate degree?"

## High-Risk Courses

There were 77,455 undergraduate-level, for-credit course records that the 8,220 first-time, full-time students took in their first year at the university. There were 21 courses identified as high risk for first-time, full-time students using these criteria (Table 2):

- Undergraduate level courses only
- First-year courses only
- Credit courses only (attempted credit > 0)
- 5 years' total enrollment > = 50
- Pass rate < 80%

**Table 2. High-Risk First-Year Courses for First-Time, Full-Time Students**

| Course Name | Pass | | Fail | | Incomplete | | Withdraw | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | # | % | # | % | # | % | # | % | # |
| General Biology | 2,145 | 79.6% | 310 | 11.5% | 2 | 0.1% | 239 | 8.9% | 2,696 |
| College Algebra | 1,191 | 77.1% | 176 | 11.4% | 2 | 0.1% | 175 | 11.3% | 1,544 |
| Trigonometry | 618 | 78.5% | 65 | 8.3% | 4 | 0.5% | 100 | 12.7% | 787 |
| Pre-college: Algebra II | 378 | 66.8% | 125 | 22.1% | 1 | 0.2% | 62 | 11.0% | 566 |
| Business Calculus | 420 | 79.7% | 48 | 9.1% | 0 | 0.0% | 59 | 11.2% | 527 |
| Problem Solving | 281 | 73.2% | 39 | 10.2% | 0 | 0.0% | 64 | 16.7% | 384 |
| Intro Computer Science I | 293 | 77.5% | 44 | 11.6% | 0 | 0.0% | 41 | 10.8% | 378 |
| Pre-college: Algebra I | 277 | 79.6% | 49 | 14.1% | 0 | 0.0% | 22 | 6.3% | 348 |
| Academic Success Skills* | 170 | 60.5% | 73 | 26.0% | 1 | 0.4% | 37 | 13.2% | 281 |
| U.S. From 1865 | 204 | 78.8% | 31 | 12.0% | 0 | 0.0% | 24 | 9.3% | 259 |
| Introduction to Philosophy | 190 | 74.8% | 34 | 13.4% | 1 | 0.4% | 29 | 11.4% | 254 |
| Sports Economics | 101 | 78.9% | 15 | 11.7% | 0 | 0.0% | 12 | 9.4% | 128 |
| Intro to American Studies | 80 | 75.5% | 10 | 9.4% | 5 | 4.7% | 11 | 10.4% | 106 |
| Insect Biology | 73 | 77.7% | 8 | 8.5% | 0 | 0.0% | 13 | 13.8% | 94 |
| Pre-college: Pre-Algebra | 73 | 77.7% | 10 | 10.6% | 0 | 0.0% | 11 | 11.7% | 94 |
| World Archaeology | 62 | 77.5% | 11 | 13.8% | 1 | 1.3% | 6 | 7.5% | 80 |
| Problems in: Electrical and Computer Engineering | 57 | 77.0% | 11 | 14.9% | 0 | 0.0% | 6 | 8.1% | 74 |
| Spec. Topics: Build the Future | 54 | 75.0% | 0 | 0.0% | 10 | 13.9% | 8 | 11.1% | 72 |
| 1st Yr German I | 54 | 77.1% | 6 | 8.6% | 1 | 1.4% | 9 | 12.9% | 70 |
| Class Piano II | 49 | 76.6% | 6 | 9.4% | 0 | 0.0% | 9 | 14.1% | 64 |
| 1st Year Japanese I | 39 | 69.6% | 10 | 17.9% | 0 | 0.0% | 7 | 12.5% | 56 |

Note: * Probation course. Pass includes A, B, C, D, and Satisfied; Fail includes F and Unsatisfied.

The characteristics of 77,455 course records were analyzed using descriptive statistics. Table 3 compares the high-risk courses (pass rate < 80%) and other courses (pass rate > = 80%) based on the course characteristics. Overall, in the 77,455 course records, there were 8,862 records of high-risk courses (11.4%), and 68,593 records of other courses (88.6%).

**Table 3. Characteristics of High-Risk Courses**

| | High-Risk Course | | Other Courses | | Total |
|---|---|---|---|---|---|
| **Course Characteristics** | # | % | # | # | # |
| **Delivery Method** | | | | | |
| Face-to-Face | 7,711 | 12.1% | 56,078 | 87.9% | 63,789 |
| Hybrid | 72 | 67.3% | 35 | 32.7% | 107 |
| Online | 1,036 | 7.7% | 12,373 | 92.3% | 13,409 |
| Unknown | 43 | 28.7% | 107 | 71.3% | 150 |
| | | | | | |
| **Core Course*** | | | | | |
| Yes | 6,985 | 12.6% | 48,567 | 87.4% | 55,552 |
| Communication 1 | 0 | 0.0% | 4,731 | 100.0% | 4,731 |
| Communication 2 | 0 | 0.0% | 3,894 | 100.0% | 3,894 |
| Communication 3 | 0 | 0.0% | 18 | 100.0% | 18 |
| First Year Seminar | 0 | 0.0% | 8,105 | 100.0% | 8,105 |
| Human Culture | 694 | 4.7% | 14,014 | 95.3% | 14,708 |
| Physical and Natural World | 2,790 | 24.6% | 8,546 | 75.4% | 11,336 |
| Quantitative Reasoning | 3,242 | 40.4% | 4,783 | 59.6% | 8,025 |
| U.S. and State Constitutions | 259 | 5.5% | 4,476 | 94.5% | 4,735 |
| No | 1,877 | 8.6% | 20,026 | 91.4% | 21,903 |
| | | | | | |
| **Math or English Gateway**** | | | | | |
| English Gateway | 0 | 0.0% | 4,676 | 100.0% | 4,676 |
| Math Gateway | 3,242 | 38.9% | 5,099 | 61.1% | 8,341 |
| Not Gateway | 5,620 | 8.7% | 58,818 | 91.3% | 64,438 |
| | | | | | |
| **Grand Total** | **8,862** | **11.4%** | **68,593** | **88.6%** | **77,455** |

Note: *Core courses refer to the general education (University Studies Program) courses in this study. ** Math or English Gateway courses are the first course for any program to fulfill the single-course college-level math or English requirement.

Regarding course delivery methods, 67.3% of the 107 hybrid course records and 12.1% of the 63,789 face-to-face course records were identified as a high-risk course. Only 7.7% of the 13,409 online course records were identified as high-risk courses.

Core courses are more likely to be high risk: 12.6% of the 55,552 core course records were identified as high risk, compared to 8.6% of the 21,903 non-core course records. Core courses were identified as courses that meet the general education requirements. Among the eight types of core courses, 40.4% of the 8,025 Quantitative Reasoning core course records and 24.6% of the 11,336 Physical and Natural World core course records were high risk, numbers that are much higher than any other types of core courses including Communication 1–3 (0%), First-Year Seminar (0%), Human Culture (4.7%), and U.S. and State Constitutions (5.5%).

Comparing the high-risk percentages of math and English gateway courses, 38.9% of the 8,341 math gateway course records were identified as a high-risk course, and none of the 4,676 English gateway course records was identified as a high-risk course. Of the 64,438 other non-gateway course records, 8.7% were identified as a high-risk course.

## Data Analysis and Variables

To answer the first research question, the course data and retention data of the 8,220 first-time, full-time students were analyzed using descriptive statistics (e.g., mean, percentage). To answer the second research question, bivariate correlation analyses were conducted using R to investigate if there was a statistically significant association between Fall-to-Fall retention and the selected student characteristics, including how many high-risk courses students take. To answer the third research question, binary logistic regression was conducted using R to investigate the best predictive model of Fall-to-Fall retention. The method of model selection is used to simplify the logistic regression model by removing variables (Dey et al., 2025; Starbuck, 2023). This approach can enhance the model's interpretability and performance by eliminating irrelevant or redundant predictors. Backward elimination is a common method that starts with the full model and iteratively removes the least significant variables based on a chosen criterion, such as the *p*-value (Starbuck, 2023). This method helps in identifying the most impactful variables while discarding those that do not contribute significantly to the model's predictive power (Dey et al., 2025).

The dependent (outcome) variable was Fall-to-Fall retention, and 13 independent (predictor) variables were selected based on literature reviews for the base model (Bass & Ballard, 2012; DeNicco et al., 2015; Djulovic & Li, 2013; Johnson et al., 2022; Ram et al., 2015):

- **Fall-to-Fall retention:** whether a first-time, full-time student retained after 1 year (retained = 1, not retained = 0)
- **Student demographics:** gender (M = 1, F = 0), age at entry, race/ethnicity (White = 1, minority = 0), first-generation (first-gen = 1, non-first-gen = 0), tuition residency (resident = 1, non-resident = 0)
- **High school academic background:** high school GPA, test score (ACT and SAT converted to ACT scale)
- **University academic experience:** student classification (freshman = 1, sophomore = 2, junior = 3, senior = 4), undeclared major (undeclared = 1, major declared = 0), first-semester GPA, on-campus or distance education (on-campus = 1, distance = 0)
- **First-year course enrollment:** total number of courses taken in their first year, number of high-risk courses taken in their first year.

Data issues were checked before the statistical analyses. The outliers due to data entry errors were removed: (a) a student with an age of 0, (b) a student with a high school GPA of 0, and (c) a student with a high school GPA of 4.15, which exceeds the maximum of 4.0 based on the university policy. The intercorrelations of all the independent variables were tested and no multicollinearity issue ($r > 0.8$) was found between any of them.

# FINDINGS

This section will discuss the findings for each of the three research questions. The characteristics of students who took the most high-risk courses were identified. The correlation between students' Fall-to-Fall retention and the number of high-risk courses they took in their first year was investigated, and the other significant predictors associated with retention were explored.

### RQ1: What are the average number of high-risk courses that students take in their first year, by student characteristics and retention status?

To answer RQ1, the total number of high-risk courses that each student had taken in their first year was computed, then the means of all students' first-year high-risk courses were computed and compared based on race/ethnicity, gender, age at entry, first-generation status, tuition residency, and Fall-to-Fall retention (Table 4). Overall, the average number of high-risk courses that all 8,220 students took in their first year was 1.08 courses; students who did not retain after 1 year (1.17, $n$ = 1,853) took more high-risk courses than those who retained (1.05, $n$ = 6,367).

**Table 4. Average Number of High-Risk Courses in First Year by Student Characteristics and Retention Status**

| | Retained after 1 Year | | | Not Retained after 1 Year | | | Total | |
|---|---|---|---|---|---|---|---|---|
| **Student Characteristics** | Average # high-risk courses | Headcount | % of total | Average # high-risk courses | Headcount | % of total | Average # high-risk courses | Headcount |
| **Race/Ethnicity** | | | | | | | | |
| American Indian or Alaska Native | 1.36 | 22 | 52.4% | 1.40 | 20 | 47.6% | 1.38 | 42 |
| Asian | 1.16 | 69 | 72.6% | 1.04 | 26 | 27.4% | 1.13 | 95 |
| Black or African American | 1.25 | 85 | 76.6% | 1.08 | 26 | 23.4% | 1.21 | 111 |
| Hispanics of Any Race | 1.18 | 418 | 75.0% | 1.42 | 139 | 25.0% | 1.24 | 557 |
| Native Hawaiian or Other Pacific Islander | 1.14 | 7 | 70.0% | 1.67 | 3 | 30.0% | 1.30 | 10 |
| Nonresident Alien | 0.97 | 92 | 80.7% | 0.95 | 22 | 19.3% | 0.96 | 114 |
| Race and Ethnicity unknown | 1.18 | 392 | 69.5% | 1.23 | 172 | 30.5% | 1.19 | 564 |
| Two or More Races | 1.01 | 268 | 72.0% | 0.99 | 104 | 28.0% | 1.00 | 372 |
| White | 1.03 | 5,014 | 78.9% | 1.16 | 1,341 | 21.1% | 1.06 | 6,355 |
| | | | | | | | | |
| **Gender** | | | | | | | | |
| Female | 1.05 | 3,352 | 81.2% | 1.20 | 776 | 18.8% | 1.08 | 4,128 |
| Male | 1.05 | 3,015 | 73.7% | 1.15 | 1,077 | 26.3% | 1.08 | 4,092 |
| | | | | | | | | |
| **Age at Entry** | | | | | | | | |
| 19 and under | 1.04 | 6,198 | 77.7% | 1.18 | 1,778 | 22.3% | 1.07 | 7,976 |
| 20–24 | 1.27 | 152 | 70.0% | 0.97 | 65 | 30.0% | 1.18 | 217 |
| 25+ | 1.76 | 17 | 63.0% | 1.50 | 10 | 37.0% | 1.67 | 27 |
| | | | | | | | | |
| **First-Generation Status** | | | | | | | | |
| First-Generation | 1.12 | 1,351 | 69.6% | 1.24 | 590 | 30.4% | 1.16 | 1,941 |
| Not First-Generation | 1.03 | 5,016 | 79.9% | 1.14 | 1,263 | 20.1% | 1.06 | 6,279 |
| | | | | | | | | |
| **Tuition Residency** | | | | | | | | |
| In-State | 1.04 | 3,405 | 79.5% | 1.18 | 878 | 20.5% | 1.07 | 4,283 |
| Out-of-State | 1.07 | 2,962 | 75.2% | 1.17 | 975 | 24.8% | 1.09 | 3,937 |
| | | | | | | | | |
| **Grand Total** | **1.05** | **6,367** | **77.5%** | **1.17** | **1,853** | **22.5%** | **1.08** | **8,220** |

### RACE/ETHNICITY

The results show that American Indian or Alaska Native, Hispanics of Any Race, Native Hawaiian or Other Pacific Islander, White, and students with Race and Ethnicity unknown who did not retain took more high-risk courses than those who retained. American Indian or Alaska Native students had the highest average number of high-risk courses in their first year (1.38, $n$ = 42) compared to any other race/ ethnicity group. American Indian or Alaska Native students who did not retain after 1 year (1.40, $n$ = 20) had a higher average number of high-risk courses than their peers who retained (1.36, $n$ = 22). The Fall-to-Fall retention rate of American Indian or Alaska Native students (52.4%, $n$ = 22) was also the lowest among all race/ethnicity groups. However, students who identified as Asian, Black or African American, Nonresident Alien, and Two or More Races who did not retain took fewer risk courses than those who retained.

### GENDER

Male and female students had the same average number of high-risk courses in their first year (1.08, female $n$ = 4,128, male $n$ = 4,092). For the students who retained after 1 year, male and female students also had the same average number of high-risk courses (1.05, female $n$ = 3,352, male $n$ = 3,015). For the students who did not retain after 1 year, both male (1.15, $n$ = 1,077) and female students (1.20, $n$ = 776) had higher average numbers of high-risk courses than those who retained.

### AGE AT ENTRY

Students aged 25 and older had the highest average number of high-risk courses (1.67, $n$ = 27) compared to all other age groups, but the sample size of this group was small. Surprisingly, for students aged 25 and older, those who retained after 1 year had a higher average number of high-risk courses (1.76, $n$ = 17) than those who did not retain (1.50, $n$ = 10). The Fall-to-Fall retention rate of age 25 and older group (63.0%, $n$ = 17) was the lowest among all age groups. Students aged 19 and under had the lowest average number of high-risk courses (1.07, $n$ = 7,976) and highest retention rate (77.7%, $n$ = 6,198).

### FIRST-GENERATION STATUS

First-generation students (1.16, $n$ = 1,941) took more high-risk courses in their first year than the other students (1.06, $n$ = 6,279) on average. Furthermore, first-generation students had a lower retention rate (69.6%, $n$ = 1,351) than the other students (79.9%, $n$ = 5,016). First-generation students who did not retain after 1 year (1.24, $n$ = 590) also had a higher average number of high-risk courses than those who retained (1.12, $n$ = 1,351).

### TUITION RESIDENCY

Out-of-state students (1.09, $n$ = 3,937) had a slightly higher average number of high-risk courses than in-state students (1.07, $n$ = 4,283), and out-of-state students (75.2%, $n$ = 2,962) also had lower retention rate than in-state students (79.5%, $n$ = 3,405). In-state students who did not retain (1.18, $n$ = 878) had a higher average number of high-risk courses than those who retained (1.04, $n$ = 3,405). Again, out-of-state students who did not retain (1.17, $n$ = 975) had a higher average number of high-risk courses than their peers who retained (1.07, $n$ = 2,962).

### RQ2: Is there a statistically significant association between high-risk courses and Fall-to-Fall retention?

A descriptive analysis was conducted to compare the retention rates by the total number of high-risk courses taken in their first year. Table 5 shows that the students who took three or more high-risk courses had the lowest Fall-to-Fall retention rate (72.2%). The students who took one or two high-risk courses had higher retention rates. The students who did not take any high-risk courses had the highest retention rate (79.4%).

**Table 5. Comparison of Retention Rates by Total Number of High-Risk Courses Taken in Their First Year**

| | Retained after 1 Year | | Not Retained after 1 Year | | Total |
|---|---|---|---|---|---|
| # of high-risk courses | # | % | # | % | # |
| 0 | 2,087 | 79.4% | 543 | 20.6% | 2,630 |
| 1 | 2,503 | 77.5% | 728 | 22.5% | 3,231 |
| 2 | 1,258 | 76.7% | 382 | 23.3% | 1,640 |
| 3+ | 519 | 72.2% | 200 | 27.8% | 719 |
| **Grand Total** | **6,367** | **77.5%** | **1,853** | **22.5%** | **8,220** |

To address both RQ2 and RQ3, a correlation matrix was computed to examine the intercorrelations (i.e., bivariate/one-to-one correlation) of Fall-to-Fall retention and all the 13 selected independent variables of student characteristics, including the total number of high-risk courses taken in their first year. Table 6 shows that all 13 selected independent variables were significantly correlated with the dependent variable of Fall-to-Fall retention.

**Table 6. Intercorrelations for Student Characteristics Variables and Fall-to-Fall Retention**

| | Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Fall-to-Fall Retention | Gender | Age | Race/ Ethnicity | First-Generation Status | Tuition Residency | HS GPA | ACT & SAT | Student Classification | Undeclared Major | 1st Sem GPA | On-Campus | Total Courses Y1 | Risk Total Y1 |
| 1 | Fall-to-Fall Retention | 1 | | | | | | | | | | | | | |
| 2 | Gender | –0.09*** | 1 | | | | | | | | | | | | |
| 3 | Age | –0.04*** | 0.12*** | 1 | | | | | | | | | | | |
| 4 | Race/Ethnicity | 0.05*** | –0.03* | 0.02 | 1 | | | | | | | | | | |
| 5 | First-Generation Status | –0.11*** | –0.02* | 0.03* | –0.12*** | 1 | | | | | | | | | |
| 6 | Tuition Residency | 0.05*** | –0.02* | 0.06*** | 0.03* | 0.11*** | 1 | | | | | | | | |
| 7 | HS GPA | 0.30*** | –0.22*** | –0.09*** | 0.16*** | –0.11*** | 0.18*** | 1 | | | | | | | |
| 8 | ACT & SAT | 0.18*** | 0.01 | –0.05*** | 0.15*** | –0.18*** | 0.05*** | 0.50*** | 1 | | | | | | |
| 9 | Student Classification | 0.07*** | –0.06*** | –0.01 | 0.03* | –0.03** | 0.06*** | 0.21*** | 0.23*** | 1 | | | | | |
| 10 | Undeclared Major | –0.03** | –0.06*** | 0.01 | –0.02 | –0.02 | 0.05*** | –0.08*** | –0.09*** | –0.03** | 1 | | | | |
| 11 | 1st Sem GPA | 0.48*** | –0.18*** | 0.01 | 0.10*** | –0.14*** | 0.01 | 0.56*** | 0.39*** | 0.15*** | –0.04*** | 1 | | | |
| 12 | On-Campus | 0.02* | –0.01 | 0.01 | 0.00 | 0.01 | –0.00 | –0.02 | 0.00 | 0.01 | –0.03** | –0.01 | 1 | | |
| 13 | Total Courses Y1 | 0.39*** | –0.08*** | –0.06*** | –0.02 | –0.04** | –0.02 | 0.13*** | 0.09*** | 0.04*** | –0.02 | 0.30*** | 0.04** | 1 | |
| 14 | Risk Total Y1 | –0.05*** | 0.00 | 0.04*** | –0.04** | 0.04*** | –0.01 | –0.21*** | –0.27*** | –0.16*** | 0.05*** | –0.25*** | 0.02 | 0.02 | 1 |

Note: Variable Definitions: **Fall-to-Fall Retention** (whether a student retained after 1 year, retained = 1, not retained = 0); **Gender** (male = 1, female = 0); **Age** (age at entry); **Race/Ethnicity** (White = 1, minority = 0); **First-Generation Status** (first-gen = 1, non-first-gen = 0); **Tuition Residency** (resident = 1, non-resident = 0); **HS GPA** (high school GPA); **ACT & SAT** (test score, ACT and SAT converted to ACT scale); **Student Classification** (freshman = 1, sophomore = 2, junior = 3, senior = 4); **Undeclared Major** (undeclared = 1, major declared = 0); **1st Sem GPA** (first-semester GPA); **On-Campus** (on-campus = 1, distance = 0); **Total Courses Y1** (total number of courses taken in their first year); **Risk Total Y1** (the number of high-risk courses taken in their first year).

Pearson's $r$ was used in the correlation analysis because (a) variables are normally distributed, (b) relationship between the variables is linear, and (c) there are no outliers in the data.

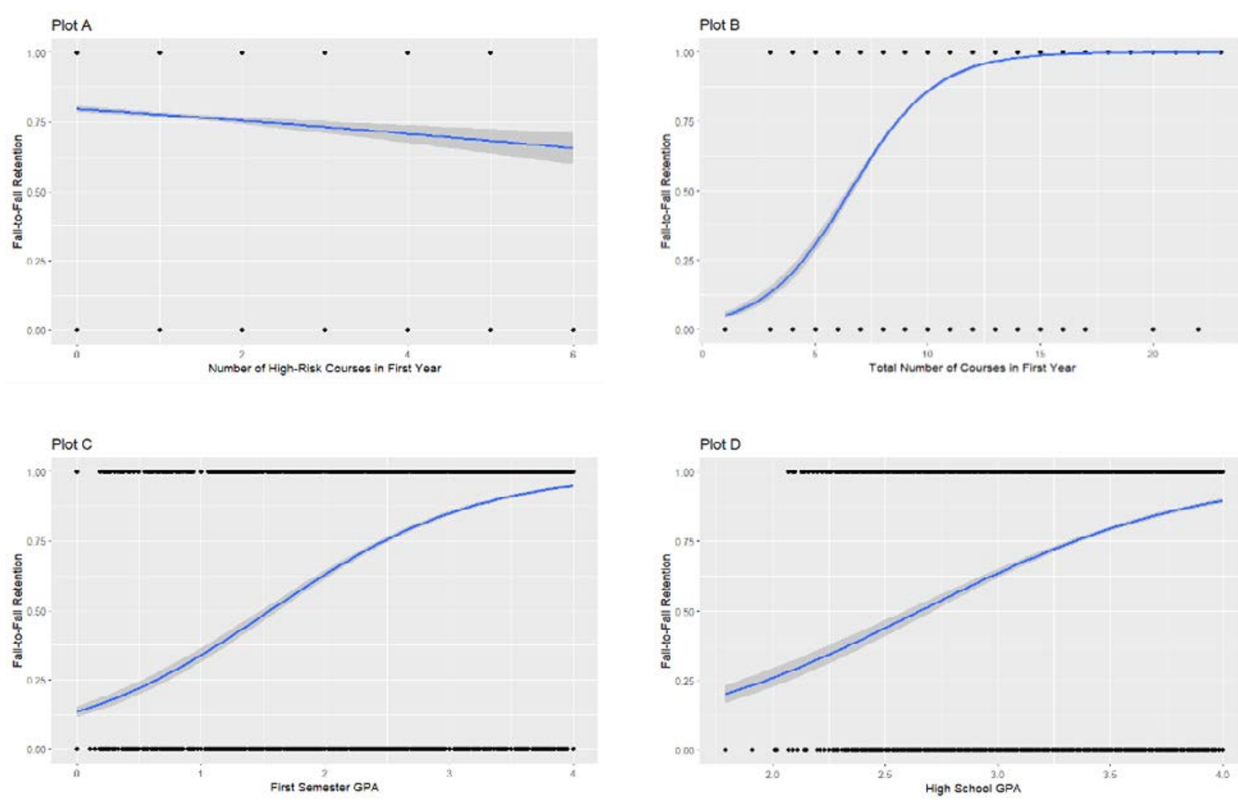
Correlation effect size: small $r = 0.10$, moderate $r = 0.30$, large $r = 0.50$ (Cohen, 1988).

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

The total number of high-risk courses taken in their first year (Risk Total Y1) had a negative correlation with Fall-to-Fall retention ($r = -0.05$, $p < .001$), but the effect size was very small according to Cohen (1988), which limits its practical significance. First-semester GPA had the strongest positive correlation with Fall-to-Fall retention, $r = 0.48$, $p < .001$, which is considered a moderate-to-large effect size (Cohen, 1988). This means that students who had relatively high first-semester GPAs were more likely to retain after 1 year. The total number of any courses taken in their first year ($r = 0.39$, $p < .001$) and high school GPA ($r = 0.30$, $p < .001$) also had moderate positive correlations with Fall-to-Fall retention.

The relationships between Fall-to-Fall retention and the total number of high-risk courses taken in the students' first year, the total number of any courses taken in their first year, first-semester GPA, and high school GPA were visualized using logistic regression curve plots (Figure 1). Figure 1 shows that students who take none to two high-risk courses tend to have a 75% or higher probability to retain; when students take more than three high-risk courses the probability of retaining decreases (Plot A). Students who take more than nine courses in their first year tend to have a 75% or higher probability of retaining; when students take ten or more courses the probability of retaining can be 87% or higher (Plot B). Students who have a first-semester GPA higher than 2.5 tend to have a 75% or higher probability to retain (Plot C). Students who have a high school GPA higher than 3.3 tend to have a 75% or higher probability to retain (Plot D).

**Figure 1. Logistic Regression Curve Plots for Fall-to-Fall Retention**



### RQ3: HOW WELL DOES A COMBINATION OF STUDENT DEMOGRAPHICS, HIGH SCHOOL ACADEMIC BACKGROUND, UNIVERSITY ACADEMIC EXPERIENCE, AND FIRST-YEAR COURSE ENROLLMENTS PREDICT FALL-TO-FALL RETENTION?

Logistic regression was conducted to investigate the best model using the selected 13 predictor variables to predict whether a first-time, full-time student retained after 1 year (Fall-to-Fall retention).

The method of model selection is used to simplify the logistic regression model by removing variables (Dey et al., 2025; Starbuck, 2023). All 13 predictor variables were entered as independent variables in the base model/full model, and Fall-to-Fall retention was entered as the dependent variable. The base model was run with the binomial logistic regression analysis in R. Then the backward elimination method was used to simplify the model by removing the least significant variables iteratively based on *p*-value (MedCalc, 2024; Starbuck, 2023). Two methods for the model simplification were used, and the results were compared for cross-validation: (a) Only one variable with the largest *p*-value ($> = .05$) was deleted in each step, and the revised model was rerun until all variables were statistically significant ($p < .05$); (b) If $p > = .05$, the variable with the smallest log odds value (estimate) was deleted, and the revised model was rerun until all variables were statistically significant ($p < .05$). The results showed that the final models were the same after using both methods.

Table 7 presents the results of the logistic regression model predicting Fall-to-Fall retention of first-time, full-time students. When all seven predictor variables are considered together, they significantly predict whether a student is retained after 1 year. The coefficient values in Table 7 (i.e., log odds) indicate the unstandardized effect size of each predictor. It tells us the direction (i.e., positive or negative) and the strength of the relationship between the predictor and how likely that a student would retain. The results suggest that the odds of Fall-to-Fall retention are increasingly greater as first-semester GPA, high school GPA, tuition residency, and the number of total courses in their first year increase; the odds of retention decrease for first-generation students and students with an undeclared major. In addition, the first-semester GPA has the largest effect size, and the number of high-risk courses taken has the smallest effect size.

**Table 7. Significant Predictors of First-Time, Full-Time Students' Fall-to-Fall Retention**

| Predictor | Coefficient | Std. Error | z value | Pr (>\|z\|) |
|---|---|---|---|---|
| (Intercept) | –6.60 | 0.33 | –19.92 | <0.001 |
| First-Gen | –0.33 | 0.07 | –4.47 | <0.001 |
| Tuition Residency | 0.43 | 0.07 | 6.25 | <0.001 |
| HS GPA | 0.43 | 0.09 | 4.91 | <0.001 |
| Undeclared Major | –0.23 | 0.12 | –1.98 | 0.048 |
| First-Semester GPA | 0.99 | 0.04 | 23.63 | <0.001 |
| Total Courses Y1 | 0.40 | 0.02 | 21.15 | <0.001 |
| Risk Total Y1 | 0.07 | 0.03 | 2.13 | 0.033 |

It is noticeable that the number of high-risk courses in their first year (risk total Y1) had a positive logistic regression coefficient in the logistic model (coefficient = 0.07), but a negative correlation coefficient with Fall-to-Fall retention in the correlation matrix ($r$ = –0.05). This might be because "the original relationship between the two variables is so close to zero that the difference in the signs simply reflects random variation around zero" (Falk & Miller, 1992, pp. 75–76).

Finally, the likelihood ratio for logistic regression was calculated in R to compare the likelihoods of two models: the null model (with only the intercept) and the final model (with seven predictors). The results showed that the difference in deviance between the two models was statistically significant ($\chi^2$ = 2393.9, $df$ = 7, $p$ < .001), indicating that the final model provides a significantly better fit to the data than the null model. The model explained 39.9% (Nagelkerke $R^2$) of the variance in retention.

# CONCLUSION

This research report examined the relationship between high-risk courses of Fall-to-Fall retention of the first-time, full-time students at a large public research university. The findings of this study reveal significant insights into the impact of high-risk courses on student retention rates. In this study, 21 courses are identified as high-risk courses, including Biology, Algebra, Trigonometry, Business Calculus, and Computer Science, among others. Hybrid and face-to-face courses are more likely to be high risk compared with online courses. STEM core courses such as Quantitative Reasoning, and Physical and Natural World are more likely to be high risk compared to non-STEM core courses like Communication and Human Culture. Math gateway courses are more likely to be high risk compared to English gateway courses.

## Students in High-Risk Courses

The average number of high-risk courses taken in the first year was 1.08, with students who did not retain after 1 year taking more high-risk courses (1.17) than those who retained (1.05). This correlation between the number of high-risk courses and retention rates underscores the importance of managing academic risk to improve student outcomes. Pierre (2015) emphasizes the significance of academic risk-taking and its impact on adult learners, suggesting that strategic learning activities can support students who struggle with academic achievement.

Furthermore, the study highlights disparities among different demographic groups. American Indian or Alaska Native students had the highest average number of high-risk courses in their first year (1.38) and the lowest retention rate (52.4%) among all racial/ethnic groups. The age group of 25+ had the highest average number of high-risk courses (1.67) and the lowest retention rate (63.0%) compared to all other age groups. This suggests that older students may face additional challenges in managing academic risk, possibly due to balancing education with other responsibilities such as work and family.

Students who took three or more high-risk courses in their first year had the lowest Fall-to-Fall retention rate (72.2%), while those who did not take any high-risk courses had the highest retention rate (79.4%). This difference suggests that there is a critical need for institutions to carefully consider the academic load and support mechanisms for students enrolled in high-risk courses. For example, advisors can recommend students not taking more than three high-risk courses in their first semester or first year. Previous studies also support the importance of early predictors and early intervention for improving student success and retention rates (e.g., Baker et al., 2015; Daniel, 2022).

## High-Risk Courses and Student Retention

The findings suggest that the number of high-risk courses taken in the first year is significantly negatively associated with student retention, meaning that students who enroll in fewer high-risk courses during their first year are more likely to retain after 1 year. When a student takes two or fewer high-risk courses, the probability of retention is predicted to be 75% or higher. This aligns with previous research by Haynes Stewart et al. (2011) and Salazar-Fernandez et al. (2021), which confirmed that course failure or high-failure rate courses negatively affect first-year university students' dropout or retention rates. This study further demonstrated that the negative correlation between high-risk courses and retention is statistically significant, even when considering other demographic information and previous academic performance. However, while the negative correlation between high-risk courses and retention is statistically significant, the effect size ($r$ = –0.05) is very small, which limits its practical significance.

Among the other predictors of student demographics, high school academic background, and university academic experience, first-semester GPA, and high school GPA emerged as the strongest indicators of retention. Students with a higher first-semester GPA are more likely to retain after 1 year, with a GPA of 2.5 or higher predicting a retention probability of 75% or more. Similarly, students with a high school GPA of 3.3 or higher also have a retention probability of 75% or higher. These results are consistent with previous research that underscores the importance of academic performance in predicting student retention. For example, Estepp et al. (2019) found that high school GPA and first-semester GPA were highly correlated with freshman academic outcomes and retention. Their study demonstrated that first-semester GPA was moderately correlated with sophomore retention ($r$ = .45), explaining 29.1% of the variance in retention. Additionally, the study by Hosch (2008) examined the predictive relationship between first-semester GPA and retention rate, suggesting "institutions trying to improve their one-year retention rates and subsequent graduation rates should continue to focus on student success in the first-semester" (p. 9). However, Hosch (2008) argues that graduation and retention rates alone are insufficient measures of educational effectiveness because these statistics do not account for differences in student effort or motivation to succeed. This perspective supports our findings, indicating that, while first-semester GPA is a strong predictor, other factors (e.g., course engagement, advising experiences) must also be considered to fully understand student retention. Furthermore, Westrick et al. (2015) conducted a meta-analysis examining the predictive validity of high school GPA, ACT scores, and socioeconomic status (SES) on college performance and retention. Their study found that high school GPA was a robust predictor of first-year academic performance and subsequent retention, reinforcing the importance of pre-college academic preparation. This meta-analysis supports our conclusion that high school GPA is a critical factor in predicting student retention.

In addition to GPA, other significant predictors identified in our study include tuition residency, total number of courses taken in the first year, first-generation status, undeclared major, and the number of high-risk courses taken in the first year. These factors contribute to a comprehensive understanding of student retention and highlight the need for targeted interventions to support students who are enrolled in high-risk courses.

## Practical Implications

Based on the research findings, there are some practical recommendations for improving student success in the high-risk courses and their retention rate. High-risk courses present significant barriers to student success, but targeted interventions and instructional methods specifically designed for high-risk courses have been implemented in previous studies. For example, Norton et al. (2018) introduced a new instructional method called SCALE-UP (student-centered activities for large enrollment undergraduate programs) that "supports student collaboration and active learning by minimizing lecture time and focusing on hands-on problem solving in the classroom" (p. 42). They examined the impact of the method on the trend in DFW proportions for an introductory calculus course, and found the positive influence of SCALE-UP on reducing DFW proportions. Active learning techniques, which emphasize student engagement and participation, have also been effective in improving performance in high-risk courses (Higgs et al., 2021). Roberts et al. (2018) suggested that implementation of active-learning practices (e.g., in-class assignments, group work, problem solving, and discussion) into STEM courses demonstrated benefits, including better student learning and performance, and smaller achievement gaps among different student populations when compared to lecture-based approaches. In addition, collaborating with the tutoring center on the identified high-risk courses, especially in STEM majors, would ensure tutoring resources are available.

The findings of this study provide valuable insights into the impact of high-risk courses on student retention and the predictors of retention. While the negative correlation between high-risk courses and retention is statistically significant, the small effect size suggests that other factors also play a crucial role in student retention. Institutions should continue to explore comprehensive strategies that address the diverse needs of students and that provide targeted support to those enrolled in high-risk courses. First-semester GPA and high school GPA are the strongest indicators, but other factors such as tuition residency, course load, and first-generation status also play significant roles. Institutions should consider these predictors when developing strategies to improve retention rates and support student success. For example, consideration of student characteristics needs to be part of advising first-time students on whether they should take high-risk courses in their first year. Factors such as ethnicity, age, tuition residency, high school GPA, first-generation status, and undeclared majors should be balanced against the number and type of high-risk courses that students are advised to take in their first year. In addition, advisors and faculty members should offer support and resources for the undeclared students to determine which high-risk core courses they should take, and when they should take them.

## Limitations and Recommendations

First, the results might not be generalizable because only one institution's data were used in this study. Certain subgroups, such as students aged 25 and older ($n$ = 27) or racial/ethnic subgroups, are underrepresented, making it difficult to draw generalizable conclusions for these populations. Future research can use different data sources from additional institutions or different student populations (e.g., full time vs. part time, institution type, pre-COVID vs. post-COVID samples) to identify effective strategies for supporting diverse students in high-risk courses.

Second, the selection of variables is limited by the availability of a database for this study. The use of a single threshold (pass rate < 80%) to define high-risk courses may oversimplify the complexity of course performance. Additional factors, such as student engagement or instructor effectiveness, could provide a more holistic view. Future research should examine more variables related to high-risk courses for predicting retention using different data collection techniques such as student perspectives and experiences on high-risk course learning, advising, tutoring, and faculty/instructors, and so on.

Finally, the approach for model simplification or model selection has limitations, since it could exclude variables that might be important in combination with others. Future research should continue to refine the retention model, and should consider using other statistical methods (e.g., non-parametric tests, causal models) or qualitative methods to evaluate the correlations and investigate how the high-risk course experiences or performance differed by student backgrounds.

### Significance

This study could be interesting for institutional research professionals and other higher education researchers, particularly those at large, public institutions. The topic is highly relevant to educators, administrators, and policymakers who are focused on improving student retention and success. The findings can contribute meaningfully to the growing body of research on student retention and offer practical recommendations for improving student outcomes.

Methodologically, the study uses a robust data set of 8,220 first-time, full-time students spanning five cohorts, allowing for meaningful longitudinal insights. The methods to identify high-risk courses and student characteristics, and the research process to develop a retention model using high-risk courses, can be helpful to similar institutions and could provide an example for reproducing similar studies at their own institutions.

Institutions frequently attribute students' underperformance to inadequate preparation. However, this study suggests there is an optimal number of high-risk courses that first-year students should enroll in for the highest chance of success. It also raises critical questions about the existence of high-risk courses and emphasizes the necessity for faculty, advisors, and administrators to prioritize the curriculums and delivery of these courses to improve student success.

## FUNDING

The authors received no financial support for the research, authorship, and/or publication of this article.

## DECLARATION OF CONFLICTING INTERESTS

The authors declare that there is no conflict of interest.

# REFERENCES

Baker, R. S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing early at-risk factors in higher education e-learning courses. In *Proceedings of the 8th International Conference on Educational Data Mining*, 150–155. https://files.eric.ed.gov/fulltext/ED560553.pdf

Bambara, C. S., Harbour, C. P., Davies, T. G., & Athey, S. (2009). Delicate engagement: The lived experience of community college students enrolled in high-risk online courses. *Community College Review, 36*(3), 219–238.

Bass, L. H., & Ballard, A. S. (2012). Student engagement and course registration methods as possible predictors of freshman retention. *Research in Higher Education Journal, 18*, 1–13. https://files.eric.ed.gov/fulltext/EJ1064635.pdf

Bloemer, W., Day, S., & Swan, K. (2017). Gap analysis: An innovative look at gateway courses and student retention. *Online Learning, 21*(3), 5–14.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Daniel, A. M. (2022). Identification of skill-appropriate courses to improve retention of at-risk college freshmen. *Journal of College Student Retention: Research, Theory & Practice, 24*(1), 126–143.

DeNicco, J., Harrington, P., & Fogg, N. (2015). Factors of one-year college retention in a public state college system. *Research in Higher Education Journal, 27*, 1–13. https://files.eric.ed.gov/fulltext/EJ1056244.pdf

Dey, D., Haque, M. S., Islam, M. M., Aishi, U. I., Shammy, S. S., Mayen, M. S. A., Noor, S. T. A., & Uddin, M. J. (2025). The proper application of logistic regression model in complex survey data: A systematic review. *BMC Medical Research Methodology, 25*(15), 1–18.

Djulovic, A., & Li, D. (2013). Towards freshman retention prediction: A comparative study. *International Journal of Information and Educational Technology, 3*(5), 494–500. http://www.ijiet.org/papers/324-K045.pdf

Estepp, C. M., Shoulders, C. W., & Johnson, D. M. (2019). Predicting freshmen academic outcomes and sophomore retention in a college of agriculture. *NACTA Journal, 64*, 466–475.

Falk, R., & Miller, N. (1992). *A primer for soft modeling*. The University of Akron Press.

Gajewski, A., & Mather, M. (2015). Remediation strategies for learners at risk of failure: A course-based retention model. *College Quarterly, 18*(1), 1.

Haynes Stewart, T. L., Clifton, R. A., Daniels, L. M., Perry, R. P., Chipperfield, J. G., & Ruthig, J. C. (2011). Attributional retraining: Reducing the likelihood of failure. *Social Psychology of Education, 14*, 75–92.

Higgs, M., Cobb, C., & Morris, P. (2021). Active learning, students who are academically at-risk, and institutional classification. *Journal of College Academic Support Programs, 4*(1), 26–36.

Hosch, B. J. (2008, May 26). *The tension between student persistence and institutional retention: An examination of the relationship between first-semester GPA and student progression rates of first-time students* [Conference presentation]. Association for Institutional Research Annual Forum, Seattle.

Integrated Postsecondary Education Data System (IPEDS). (2025, February 27). *Fall enrollment (EF).* https://nces.ed.gov/ipeds/survey-components/8

Johnson, L., Coder, D., & Jodi Harris. (2022). *Look before you leap into retention: The journey through successful retention roadmap planning* [Webinar]. The National Symposium on Student Retention, the Consortium for Student Retention Data Exchange at the University of Oklahoma.

Laskey, M. L., & Hetzel, C. J. (2011). Investigating factors related to retention of at-risk college students. *Learning Assistance Review, 16*(1), 31–43.

Martin, D. C., & Arendale, D. R. (1992). *Supplemental instruction: Improving first-year student success in high-risk courses.* Monograph. National Resource Center for the First-Year Experience. https://conservancy.umn.edu/bitstream/handle/11299/200471/SIFYEmonograph.pdf?sequenc

MedCalc. (2024, May 7). *Logistic regression.* https://www.medcalc.org/manual/logistic-regression.php

Nash, R. D. (2005). Course completion rates among distance learners: Identifying possible methods to improve retention. *Online Journal of Distance Learning Administration, 8*(4), 1–26.

Norton, P., Bridges, W., & High, K. (2018). Impact of course policy changes on calculus I DFW rates. *Journal of STEM Education, 19*(1), 41–47.

Paterson, K., & Guerrero, A. (2023). Predictive analytics in education: Considerations in predicting versus explaining college student retention. *Research in Higher Education Journal, 44*, 1–12.

Pierre, Y. (2015). *Academic risk-taking in higher education.* Florida International University. https://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1427&context=sferc

Ram, S., Wang, Y., Currim, F., & Currim, S. (2015). Using big data for predicting freshman retention. *Proceedings of the Thirty-Sixth International Conference on Information Systems, 2015*, 1–16. https://core.ac.uk/download/pdf/301370007.pdf

Roberts, J. A., Olcott, A. N., McLean, N. M., Baker, G. S., & Möller, A. (2018). Demonstrating the impact of classroom transformation on the inequality in DFW rates (“D” or “F” grade or withdraw) for first-time freshmen, females, and underrepresented minorities through a decadal study of introductory geology courses. *Journal of Geoscience Education, 66*(4), 304–318.

Salazar-Fernandez, J. P., Sepúlveda, M., Munoz-Gama, J., & Nussbaum, M. (2021). Curricular analytics to characterize educational trajectories in high-failure rate courses that lead to late dropout. *Applied Sciences, 11*(4), 1436.

Sanabria, T., Penner, A., & Domina, T. (2020). Failing at remediation? College remedial coursetaking, failure and long-term student outcomes. *Research in Higher Education, 61*(4), 459–484. https://doi.org/10.1007/s11162-020-09590-z

Sargent, C. S., Sullivan, T., & McAlum, H. (2022). DFW in gateway courses not always a graduation problem: A study in Intermediate Accounting I from 2007 to 2018. *Journal of Accounting Education, 60*, 100795.

Simpson, O. (2013). Student retention in distance education: are we failing our students? *Journal of Open, Distance and e-learning, 28*(2), 105–119.

Starbuck, C. (2023). *The fundamentals of people analytics, with applications in R.* Springer. https://library.oapen.org/handle/20.500.12657/63962

Stone, M. E., & Jacobs, G. (2008). *Supplemental instruction: Improving first-year student success in high-risk courses* (3rd ed.). National Resource Center for The First-Year Experience and Students in Transition. https://files.eric.ed.gov/fulltext/ED559247.pdf

Valentine, J. C., Hirschy, A. S., Bremer, C. D., Novillo, W., Castellano, M., & Banister, A. (2011). Keeping at-risk students in school: A systematic review of college retention programs. *Educational Evaluation and Policy Analysis, 3*3(2), 214–234.

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment, 20*(1), 23–45.