
Assessing Student Retention in Online Learning Environments: A Longitudinal Study

Wallace Boston
American Public University System
wboston@apus.edu

Phil Ice
American Public University System
pice@apus.edu

Melissa Burgess
American Public University System
mburgess@apus.edu

Abstract

In their initial study, authors Boston, Ice, and Gibson (2011) explored the relationship between student demographics and interactions, and retention at a large online university. Participants in the preliminary study ($n = 20,569$) included degree-seeking undergraduate students who completed at least one course at the American Public University System (APUS) in 2007. Two notable findings from the study were (1) the importance of transfer credit, and (2) the consistency of activity in predicting continued enrollment. Interestingly, the latter finding was confirmed upon the analysis of longitudinal data from the current study. Further related to the latter finding-yet unexpected, was the existence of new literature that, although subtle, affirms the importance for online institutions to conduct ongoing research on these topics. Readers of the current study are encouraged to refer to the preliminary study toward a comprehensive understanding of these nuances. Though informative, the researchers wished to validate the original study findings through longitudinal evaluation of retention.

The Preliminary Study vs. Current Study: An Overview

Student enrollment and academic achievement data were analyzed using forward method linear regression resulting in the emergence of six predictors of student disenrollment. These predictors included: (1) no transfer credit received by the student; (2) the total number of registrations/courses previously taken; (3) last grade student received was an F; (4) last grade student received was a W; (5) student GPA (3.01 – 3.99); and (6) GPA (2.01 – 3.00.) Although the current study methodologically replicates the 2011 study, the literature review has been updated to reflect current research; and the study itself includes the analyses of data that spanned a five-year period (2006-2010); thereby increasing the number of participants ($n = 199,731$).

The Study and Its Context

Setting

The researchers in this study revisited current definitions of academic preparedness and non-traditional students within the context of online institutional programs, as suggested by Simpson (2003), toward gleaning a snapshot of their own institutional retention characteristics. Gilbert (2000) furthered this notion that policy and/or program improvement should be informed at the individual, course, program, institutional, or systems level. The institution from which the data in this study were collected, American Public University System (APUS), is an accredited, private, fully online university offering an extensive variety of academic programs that do not require students to physically attend classes.

In the preliminary study, student enrollment at APUS (2009) reached roughly over 65,000 students. Since that

time, student enrollment has spiked dramatically to a little over 100,000 (APUS, 2011). Current APUS demographic reports indicate that (a) the majority of students are seeking their undergraduate degree (64%); (b) most are working adults; (c) the median age of students is 30 years; (d) 64% of students are on active military duty from all branches of service, 36% civilian, other military; (e) most students have transfer credit from prior learning, other university coursework, or military experience averaging 35 semester hours (undergraduate), and 12 semester hours (graduate); and (f) student geographic distribution spans all fifty states in the U.S., and over 100 countries internationally.

Within the context of higher education as a whole, Allen and Seaman (2011) posited that student enrollment in online courses currently exceeds 6 million—making up nearly one-third of all students in higher education who are taking at least one online course. Also reported was that online enrollments have shown signs of slowing, but continue to exceed the rate for the entire higher education population. With these burgeoning numbers, institutions would be well-informed to closely, and continually, examine student characteristics. Student characteristics examined in the preliminary study, were also examined in the current study and included: academic preparedness, non-traditional age students, swirling, military students and demographic information including ethnicity and gender.

In a recent report, the National Center for Education Statistics (2011) examined how student participation in online courses varies with student characteristics. Pertinent findings include: (1) between 2000 and 2008, the percentage of undergraduates enrolled in at least one **distance education** class increased from 8 percent to 20 percent, and the percentage enrolled in a **distance education** degree program increased from 2 percent to 4 percent; (2) compared with all students, students studying computer science and those studying business enrolled at higher rates in both **distance education** classes (27 percent and 24 percent, respectively, vs. 20 percent) and **distance education** degree programs (8 percent and 6 percent, respectively, vs. 4 percent); (3) participation in a **distance education** course was most common among undergraduates attending public 2-year colleges; 22 percent were so enrolled. Participation in a **distance education** degree program was most common among undergraduates attending for-profit institutions; 12 percent were so enrolled; (4) older undergraduates and those with a dependent, a spouse, or full-time employment participated in both **distance education** classes and degree programs relatively more often than their counterparts; and (5) students with mobility disabilities enrolled in a **distance education** course more often than students with no disabilities (26 percent compared with 20 percent), but no other statistically significant difference between students with and without disabilities was detected.

Enrollment “Swirling” and Retention

As definitions for the terms “retention” and “nontraditional learner” evolve over time, so will the implications these changing definitions have for higher education. Implications and subsequent efforts toward improving student retention have focused primarily on institutional degree programs rather than the characteristics of the students themselves (Anderson, 2011). One implication, in terms of enrollment, that higher education institutions are currently observing relates to student “swirling.” Swirling is described as “the inconsistent flow in and out of college coursework from term-to-term, institution-to-institution” (Campbell & Mislevy, 2009, p.2-3). Further, while there are numerous institutional-and student-centric factors contributing to swirling and retention (Herzog, 2005; McCormick, 2003; and Porter, 2002), these factors focus on traditional students and/or traditional brick and mortar institutions (Pascarella & Terenzii, 2005; Tinto, 1975, 1993). Some of these factors consider the role of life challenges, academic-related skills, student background and commitment to succeed (Anderson, 2011).

Additionally, methodological models vary—some approaching swirling retention strategies in a linear fashion (Porter, 2002), as others view and approach it as circular (Campbell & Mislevy, 2009). Underpinning the latter model is swirl theory, which acknowledges the complex nature of college enrollment intertwined with students’ diversity in experiences. Concomitant to swirl theory, McCormick (2003) identified eight student swirling patterns: (1) trial enrollment; (2) special program enrollment; (3) supplemental enrollment; (4) rebounding enrollment; (5) concurrent enrollment; (6) consolidated enrollment; (7) serial transfer; and (8) independent enrollment. A student may *trial enroll* to determine the extent to which they are satisfied with the institution, but may transfer at a later time. *Special program enrollment* denotes a student who is enrolled at the home institution, but has the option to take courses at partner institutions. *Supplemental enrollment* refers to the ability for a student to enroll at another institution to accelerate their home institution program. *Rebounding enrollment* allows a student to alternate between two or more institutions. *Concurrent enrollment* allows a

student to take courses simultaneously at two institutions. *Consolidated enrollment* is the grouping of courses in a degree program that students may also take at other institutions. A student opts to *serial transfer* by transferring from one institution to another—although mindful of a final institution where they will complete their program. *Independent enrollment* refers to students who take courses at an institution that do not contribute to a degree program.

Given the variances in the above-mentioned areas, it is imperative that institutions of higher education examine these areas according to their own institutional characteristics. It is with this vision, APUS strives to continually examine both student and institutional factors that affect student retention and progression toward future strategic planning. Therefore, as a fully online university with a primarily nontraditional student population, APUS must consider possible limitations in social and academic interaction that may contribute to student disenrollment. Given the high percentage of students who dis-enroll early at APUS, examination of these factors against enrollment swirling patterns is imperative, and must further consider how these patterns relate to, or may impact other institutional areas (i.e., finances, financial aid, assessment and accountability, student advising, student assessment and curriculum).

Purpose of the Study

Since the publication of the Boston, Ice, and Gibson (2011) preliminary study, the overarching motivation for conducting the current study remains the same—to promote student achievement. Toward this goal, future research on student retention in online programs should build upon the most current research. Additionally, future longitudinal and replicated studies—with larger and more diverse participant pools, will establish clear-cut predictors for student retention in online learning environments, thus providing institutional administrators with the information needed to ensure student retention and achievement. Therefore, the same variables were examined in the current study toward further validating the results from the preliminary study (i.e., what type of student enrolls at an online institution, and what factors influence student retention in online courses). Additionally, the original study consisted of a point in time snapshot (students active in 2007), whereas the current study examined student enrollments over a five year period (2006 – 2010).

Research Questions

This study used descriptive statistics and multiple regressions to analyze the relationship between demographic and academic performance data and student retention at APUS from years 2006 through 2010, to answer the following research questions:

RQ1: What factors influence online student retention?

RQ2: Do the factors influencing online student retention change over time?

Method

Participants

Data from students' applications, enrollment (courses, degree programs), and academic achievement (grades) were extracted from the institution's data warehouse and aggregated in an Excel spreadsheet. Information such as age, military rank, military branch of service, academic credits transferred, GPA, degree program, etc., were evaluated as predictor variables in a regression analysis.

The American Public University System has both a sizeable amount of degree offerings (84), and enrolled students (over 100,000), therefore the primary motivator underlying this study, was to evaluate retention of undergraduate students toward minimizing differences in background characteristics between undergraduate and graduate students.

Given that the progression toward graduation takes years and not months, data was extracted for all degree-seeking (control variable) undergraduate students who completed at least one course (control variable) at APUS from 2006 through 2010. Data included enrollment and academic achievement data through December 31, 2010 with a total n of 199,731.

Procedure

The predictor variables included student background data downloaded from the APUS data warehouse. Specifically, the data were evaluated to determine if variables would be entered into the regression equation as either interval data or dummy categorical variables. The predictor variables included: Degree Program, Program Level (Associates' or Bachelor's degree), Cumulative GPA, Number of Registrations Taken in each of the cohort years (2006-2010), Gender, Race/Ethnicity, Cohort Age (age upon program entry), Military/Civilian Classification, Grade Received in Last Course, New Student/Returning Student Degree Program, Program Level, Gender, Ethnicity, and New/Returning Status were readily identified as categorical variables and entered into an Excel spreadsheet as such. The possible values for these variables did not imply a given order, but rather indicated nominal (categorical) values. To utilize these categorical variables as predictors in the regression model, "dummy variables" were created with a new variable for each possible level of the categorical variable. As an example, for predictor variable Race/Ethnicity an individual dummy variable was created for each ethnic classification (e.g. White, Black-non Hispanic, Hispanic, etc.) representing a students' classification recorded in a binary manner.

Grade in Last Course was recorded as a letter value with either a plus or minus modifier. Though the classification followed a clear, linear pattern, a precise numerical value was not present. As such, the variable was considered categorical in nature and entered as a dummy variable. Number of Registrations for years 2006-2010 clearly met the criteria for interval data. In addition, the range was small for this variable, lending statistical adequacy to all values. Based on these factors, Number of Registrations for years 2006-2010 was entered as interval data.

Though extracted as interval data from the data warehouse, determining the best method for entering Cohort Age, GPA, and Number of Transfer Credits was problematic. With respect to Cohort Age, it was decided to group the data in age bands that matched the age bands organized by Department of Education statisticians in the IPEDS surveys after visual inspection of a histogram of the data. Following the schema used for grade in last course, GPA was collapsed into the following buckets: 0.00, 0.01 – 1.00, 1.01 – 2.00, 2.01 – 3.00, 3.01 – 3.99, and 4.00.

Finally, a review of descriptive statistics and the histogram for Number of Transfer Credits revealed a clustering effect around certain thresholds (specifically multiples of 3 credit hours). One threshold, no transfer credit hours received, was the largest single value in the data set. Based on this evidence, it was determined that the best strategy would be to establish 15 credit hour interval dummy variable classifications for Number of Transfer Credits received with 0 Transfer Credit Hours received as a separate band. The 15 credit hour bands were selected as 15 credit hours represent the completion of the equivalent of a full-time semester.

The total number of predictor variables, including continuous variables and dummy variable categories was 116. The cumulative n for the study was 199,731 with the following breakdown by year: 2006 (12,975); 2007 (21,316); 2008 (33,166); 2009 (46,906); and 2010 (59,731). These data sets were regressed on the criterion variable, using suggestions from Cohen, Cohen, West, and Aiken (2002). The criterion variable was Enrollment Status, which was treated as a dichotomous variable. If a student was enrolled or had graduated at the end of 2009, a "0" was entered for Enrollment Status and a "1" was entered if the student was dis-enrolled.

Results and Discussion

For the 2006-2010 cohorts, the following table illustrates the participant sample, significant predictors and their associated r-squares and standardized coefficient betas for cohort years 2006, 2007, 2008, 2009, and 2010.

Table 1

Forward Regression Model for 2006-2010 Data Sets

Cohort Year	$n =$	Predictors	R-Square	Standardized Coefficient Beta
2006	12,975	a. No Transfer Credit	.176	.254
		b. GPA 3.01 to 3.99	.081	-.288
		c. Year Program Regs	.022	-.161
		d. GPA 2.01 to 3.00	.010	-.139

		e. W	.014	.137
		f. F		.146
2007	21,316	a. No Transfer Credit	.145	.251
		b. GPA 3.01 to 3.99	.215	-.376
		d. GPA 2.01 to 3.00	.248	-.211
		c. Year Program Regs	.274	-.162
		e. W	.286	.112
2008	33,166	a. No Transfer Credit	.149	.254
		b. GPA 3.01 to 3.99	.208	-.365
		d. GPA 2.01 to 3.00	.244	-.215
		c. Year Program Regs	.266	-.154
2009	46,906	a. No Transfer Credit	.101	.216
		b. GPA 3.01 to 3.99	.137	-.266
		d. GPA 2.01 to 3.00	.163	-.174
		c. Year Program Regs	.183	-.166
		g. N	.204	-.143
		e. W	.222	.166
2010	59,731	e. W	.090	.309

- a. Predictors: (Constant), No_Transfer_Credits
- b. Predictors: (Constant), GPA_3.01_to_3.99
- c. Predictors: (Constant), YearProgramRegs
- d. Predictors: (Constant), GPA_2.01_to_3.00
- e. Predictors: (Constant), W
- f. Predictors: (Constant), F
- g. Predictors: (Constant), N

2006 Cohort Data Set.

The forward entry method resulted in 29 of the predictor variables being significant and accounting for a combined 37.2% of variance. However, six of the predictors accounted for a combined 33.3% of variance, with none of the remaining predictors accounting for more than .01% of variance. Thus, even though the remaining

predictors were significant, the extremely low amount of variance accounted for should not be considered relevant in terms of predictive modeling.

2007 Cohort Data Set.

The forward regression model for the 2007 data set resulted in 25 of the predictor variables being significant and accounting for a combined 32.8% of variance. However, five of the predictors accounted for a combined 28.6% of variance, with none of the remaining predictors accounting for more than .01% of variance.

2008 Cohort Data Set.

The forward regression model for the 2008 data set resulted in 26 of the predictor variables being significant and accounting for a combined 31% of variance. However, four of the predictors accounted for a combined 26.6% of variance, with none of the remaining predictors accounting for more than .01% of variance.

2009 Cohort Data Set.

The forward regression model for the 2009 data set resulted in 30 of the predictor variables being significant and accounting for a combined 24.9% of variance. However, six of the predictors accounted for a combined 22.2% of variance, with none of the remaining predictors accounting for more than .01% of variance.

2010 Data Set.

The forward regression model for the 2010 data set resulted in 17 of the predictor variables being significant and accounting for a combined 32.1% of variance. However, only one predictor accounted for 30% of variance, with none of the remaining predictors accounting for more than .01% of variance.

Conclusion

As with the initial study, transfer credit remained the most meaningful predictor of student retention. From the authors' perspective, this further validates the original hypothesis that the high amount of variance accounted for by the presence of transfer credit, and the tendency of a significant number of students to disenroll after two courses, indicates that initial attempts at college enrollment online may be more exploratory than in the traditional university. Given the anytime, anywhere nature of online learning, this finding is not surprising; however, it should give pause to institutional administrators, educational leaders, and national bodies such as the U. S. Department of Education's Institute of Education Sciences and their Integrated Postsecondary Education Data System (IPEDS). As such, the nature of retention should be redefined to examine both non-exploratory students and those who migrate through a series of institutions to earn a degree.

Second, the ability to maintain an adequate GPA was, not surprisingly, found to be a meaningful predictor of retention. However, it is important to note that this factor did not present as a significant factor in the original study. As such, one may posit that as students' progress past initial "barrier courses" that the ability to maintain satisfactory progress is more important than simply progressing. In short, students may be internally differentiating between simply progressing and progressing in a manner that they believe will adequately prepare them for application of the knowledge acquired.

Third, as evidenced by the variance accounted for by annual enrollments, activity should be considered a primary catalyst for degree completion. While in the paragraph above it is noted that satisfactory progress is more meaningful than progress alone, the importance of variable speaks to the need to maintain academic trajectory and momentum.

Implications for Practice and Future Research

While this study validates the initial exploration of factors impacting retention at APUS, it remains a single institution research initiative. Thus, the findings should be viewed within this context. However, this work has received recognition within the last year, and a large scale research initiative, the Predictive Analytics Reporting Framework, was funded through the Bill and Melinda Gates Foundation and administered by WCET. This initiative applies methodologies similar to those described in this study to data sets from six institutional partners. At present the data is still being analyzed, however, it is believed that many of the factors

noted in this study will be applicable across other institutions. Dissemination of this data will provide further support and clarification to the field of retention and progression analysis. Within this context, this article can be viewed as a precursor to broader based inquiry in the field.

References

- Allen, I. E., & Seaman, J. (2011). *Going the distance: Online education in the United States 2011*. Babson Research Group. Retrieved from www.onlinelearningsurvey.com/reports/goingthedistance.pdf
- American Public University System (APUS). (2011). APUS Facts. Retrieved from <http://www.apus.edu/about-us/facts.htm>
- Anderson, K. (2011). Linking adult learner satisfaction with retention: The role of background characteristics, academic characteristics, and satisfaction upon retention (Doctoral dissertation Iowa State University, 2011). ProQuest, (UMI No. 3458241).
- Boston, W., Ice, P. & Gibson, A. (2011). Comprehensive assessment of student retention in online programs. *Online Journal of Distance Learning Administration*, 14(1).
- Campbell, C. M. & Mislevy, J. (2009, November). Students' perceptions matter: early signs of undergraduate student retention/attrition. Paper presented at the meeting of North East Association of Institutional Research, p. 2-3.
- DeAngelo, L., Franke, R., Hurtado, S., Pryor, J. & Tran, S. (2011). Completing college: Assessing graduation rates at four-year institutions. University of California, Los Angeles, CA: Higher Education Research Institute at UCLA.
- Herzog, S. (2005). Measuring determinants of student return vs. dropout/stopout vs. transfer: A first-to-second year analysis of new freshmen. *Research in Higher Education*, 46(8).
- McCormick, A. C. (2003). Swirling and double-dipping: New patterns of student attendance and their implications for higher education. *New Directions for Higher Education*, 121, 13-24.
- Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M. (2006). *Learning in adulthood: A comprehensive guide* (3rd ed.). Indianapolis, IN: Jossey-Bass.
- National Center for Education Statistics. (2011). *Learning at a distance: Undergraduate enrollment in distance education courses and degree programs*. NCES 2012154. October. Washington, DC:
- National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubs2012/2012154.pdf>
- Okpala, C. O., Hopson, L., Fort, E., & Chapman, B. S. (2010). Online preparation of adult learners in post-secondary education: A triangulated study. *Journal of College Teaching & Learning*, 7(5), 31-36.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students*. San Francisco, CA: Jossey-Bass.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89-125. Retrieved from <http://www.jstor.org/stable/1170024>
- Tinto, V. (1993). *Leaving college: rethinking the causes and cures of student attrition*. Chicago, IL: University of Chicago Press.
- Tinto, V. (1997). *Classrooms as communities: Exploring the educational character of student persistence*.
-

