

**RESEARCH BRIEF**

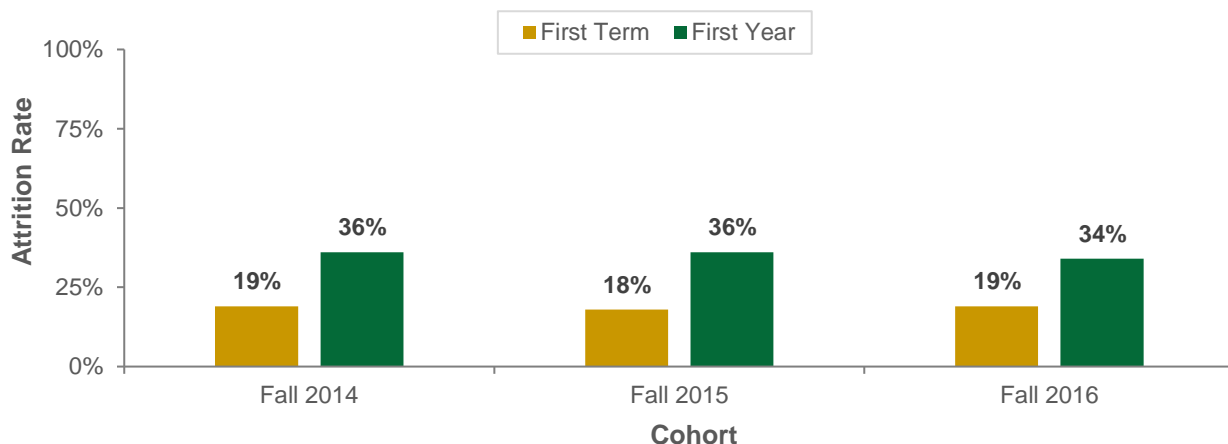
**Early Identification of “At-Risk” Students at NOVA:  
Volume 1**

This Research Brief is a part of the ongoing institutional effectiveness program at Northern Virginia Community College (NOVA).

- The **Persistence Rate** (retention, graduation, and transfer-out) is a measure of NOVA’s institutional effectiveness, student success, and commitment to its students.
- The **Attrition Rate** (stop-out) is the percentage of students who did not return to NOVA or graduate from NOVA or transfer-out. The most significant loss of students, as a result of attrition, occurs during the first term and the first year.

On average, 19 percent of the NOVA first-time in college (FTIC), program-placed students stop-out (do not return, graduate, or transfer) after the first term, and an additional 16 percent of the students stop-out by the end of the first year (35 percent). Therefore, intervening early in the first year to retain students past the first term and first year is an efficient way to increase the Institution’s persistence rates.

**Figure 1. First-Term and First-Year Attrition Rates of First-Time in College, Program-Placed Students: Fall 2014, Fall 2015, and Fall 2016 Cohorts<sup>1</sup>**



<sup>1</sup> The **first-term attrition rate** is calculated as the number of students who did not return to NOVA for the second term of study nor did they graduate from NOVA or transfer to a four-year institution by the end of the first term, divided by the number of students who started at NOVA. The **first-year attrition rate** is calculated as the number of students who did not return to NOVA in the following Fall term of study nor did they graduate from NOVA or transfer to a four-year institution by the end of the first year, divided by the number of students who started at NOVA.

To enable early identification of students who are at risk for attrition, three logit models were developed by analyzing the following:

1. The predictive relationship of entering student characteristics with persistence after first term.
2. The predictive relationship of entering student characteristics and first-term momentum metrics with persistence after first year.
3. The predictive relationship of entering student characteristics and first year momentum metrics with persistence after third year.

This volume (Volume 1) presents the results of the first logit model (persistence after first term). Volumes 2 and 3 will present the results of second and third logit models. The study population for the first logit model consisted of 22,805 first-time in college (FTIC), program-placed students who entered NOVA in the fall terms of 2014, 2015, and 2016. To increase the generalizability of the model beyond the data on which the model is fit, three years of data were used in this analysis.

## Key Findings

- An odds ratio of greater than one indicates that the likelihood of a target event occurring is more likely in the focus group than in the non-focus group. An odds ratio of one implies that the target event is equally likely for the two groups. An odds ratio of less than one suggests that the target event is less likely to occur in the focus group than in non-focus group (Meyers et al., 2017).

In this study population, the odds ratios indicate the following (controlling for all other predictors).

- **Gender:** The odds of a male student persisting after the first term were less than (0.70 times) the odds of a female student persisting.
- **Race/Ethnicity:**
  - The odds of an Asian student persisting after the first term were 1.62 times higher than the odds of a White student persisting.
  - Conversely, the odds of a Hispanic student (0.96 times) or a student of other/unknown ethnicity (0.98) persisting after the first term were slightly lower than the odds of a White student persisting.
  - The odds of a Black student (0.99 times) persisting after the first term were nearly equal to the odds of a White student persisting.
- **Degree Program:**
  - The odds of persisting after the first term for a student pursuing an A.A.A. or A.A.S. degree were 0.76 times the odds of persisting for a student pursuing an A.A. or A.S. degree.
  - The odds of persisting after the first term for a student pursuing a certificate were 0.69 times the odds of persisting for a student pursuing an A.A. or A.S. degree.
- **Major:** The odds of persisting after the first term for a student enrolled in Science (A.S.) (1.37 times) or Business Administration (A.S.) (1.11 times) or Social Sciences (A.S.) (1.06 times) were higher than the odds of persisting for a student enrolled in General Studies (A.S.).
- **SDV Course Enrollment:** The odds of persisting after the first term for a student who enrolled in an SDV course in the first term were 1.38 times higher than the odds of persisting for a student who did not enroll in an SDV course.

- **Developmental Course Enrollment:** The odds of persisting after the first term for a student who enrolled in a developmental course in the first term were 0.86 times the odds of persisting for a student who did not enroll in a developmental course.
- **Day Course Enrollment:**
  - The odds of persisting after the first term for a student who enrolled in only day courses were 1.24 times higher than the odds of persisting for a student who enrolled in a combination of both day and night courses.
  - The odds of persisting after the first term for a student who enrolled in only night courses were about the same (0.99 times) as the odds of a student enrolled in a combination of both day and night courses.
- **College-Level Course Enrollment:**
  - The odds of persisting after the first term for a student who enrolled in a college-level English course in the first term were 1.22 times higher than the odds of persisting for a student who did not enroll in college-level English.
  - Likewise, the odds of persisting after the first term for a student who enrolled in a college-level math course in the first term were 1.17 times higher than the odds of persisting for a student who did not enroll in college-level math.
- **Credits Enrolled:** For every one credit increase in the number of credits enrolled during the first term, there was a 17 percent increase in the odds of persisting after the first term.
- **Instruction Mode:**
  - The odds of persisting after the first term for a student enrolling only in online courses were 0.67 times the odds of persisting for students enrolling in online and on-campus courses.
  - The odds of persisting after the first term for students enrolling in only on-campus courses were 0.94 times the odds of persisting for students enrolling in online and on-campus courses.

To identify students who are at risk for attrition, a binary logit model was developed by analyzing the predictive relationship between student persistence after first term and each of the following entering student characteristics (Table 1):

**Table 1. Entering Student Characteristics**

Gender	Distance from home to campus
Race/ethnicity	Residence (in-state or out-of-state)
Age at initial enrollment	Enrollment in developmental courses
Entering major	Enrollment in an SDV course
Degree (A.A./A.S., A.A.A./A.A.S., or certificate)	Enrollment in college-level math
Recent high school graduate	Enrollment in college-level English
Day/Night Status	Enrollment in NOVA Online courses
First-Generation status	Credits enrolled in the first term

Data availability and previous research on student success in community colleges provided the rationale for the selection of variables. Due to large sample size, the significance level cutoff was set at  $p < 0.05$  for a variable to stay in the model. This allowed for the development of a more parsimonious model without losing explanatory power. Listwise deletion was used wherever there were missing data. In essence, any student who had a missing value on any of the predictors was

excluded from the study. Thirty-six students were excluded from regression analyses because of missing data.

Table 2 presents the statistics assessing the model fit. The results indicate that the model provided a statistically significant prediction of persistence,  $\chi^2(20) = 2,286.37$ ,  $p < 0.001$ . The Max-rescaled R-Square indicated that the model accounted for approximately 15 percent of the total variance. Classification accuracy for the cases based on a classification cutoff value of 0.65 for predicting membership in the persisted group was high. The overall rate of correct classification was 80 percent, with 92 percent of persisted (sensitivity), and 31 percent of not persisted (specificity) being correctly classified. In the opinion of Hosmer and Lemeshow (2000), “the classification table is most appropriate when classification is a stated goal of the analysis; otherwise it should only supplement more rigorous methods of assessment of fit” (p. 160). The Hosmer and Lemeshow goodness of fit test which yields a test statistic significance level above 0.05 leads to a failure to reject the null hypothesis for the model that there is no difference between the observed and model predicted values of the dependent variable; therefore, the model’s estimates are acceptable (Menard, 1995). The Hosmer and Lemeshow test in this model resulted in a goodness-of-fit statistic of 15.21, distributed as a chi square value, and is associated with a p value of 0.055, indicating an acceptable match between predicted and observed probabilities. The area under the curve (AUC) was 0.723 suggesting that the fit of the logistic regression model is acceptable (Meyers et al., 2017).

**Table 2. Model Fit Statistics (N = 22,769, Persisted = 18,510)**

Model Fit Statistics		
Likelihood Ratio $\chi^2$	2,286.37	$p < 0.001$
Max-rescaled R-Square	0.154	
-2 Log likelihood	21945.711	
AIC	21947.711	
SC	21955.744	
Hosmer and Lemeshow Goodness-of-Fit Test	15.208	$p=0.055$
Correctly Classified %	80.3	
Sensitivity	91.8	
Specificity	30.5	
Area Under Curve (AUC)	0.723	

Table 3 (next page) presents the standard results produced by logistic regression using the study population (coefficient estimates and their respective standard errors, Wald statistics, p-values, and odds ratios).

- The Wald test indicated that gender, race/ethnicity, degree type, major, credits enrolled in first term, enrollment in developmental, SDV, college-level math and English courses, enrollment in only day courses, and instruction mode were statistically significant predictors of persistence after the first term in the study population.

**Table 3. Logistic Regression Predicting First Term Persistence**

Predictors	B	S.E.	Wald	p	OR	95% CI	
Intercept	-0.537	0.094	32.466	<0.001			
Male vs Female	-0.363	0.038	90.946	<0.001	0.696	0.646	0.750
Black vs White	-0.008	0.052	0.025	0.875	0.992	0.896	1.098
Hispanic vs White	-0.045	0.046	0.968	0.325	0.956	0.873	1.046
Asian vs White	0.481	0.065	54.889	<0.001	1.617	1.424	1.837
Other Unknown vs White	-0.021	0.073	0.083	0.773	0.979	0.848	1.130
Cert vs A.A./A.S.	-0.371	0.107	11.988	0.001	0.690	0.559	0.851
A.A.A./A.A.S. vs A.A./A.S.	-0.273	0.062	19.587	<0.001	0.761	0.674	0.859
Online Only vs Mixed	-0.403	0.100	16.173	<0.001	0.668	0.549	0.813
On-campus Only vs Mixed	-0.064	0.074	0.748	0.387	0.938	0.812	1.084
Day Only vs Day and Night	0.218	0.044	24.426	<0.001	1.243	1.140	1.355
Night Only vs Day and Night	-0.008	0.102	0.007	0.935	0.992	0.812	1.211
Business Admin vs General Studies	0.100	0.063	2.530	0.112	1.105	0.977	1.249
Social Sciences vs General Studies	0.059	0.064	0.852	0.356	1.061	0.935	1.204
Science vs General Studies	0.316	0.076	17.454	<0.001	1.372	1.183	1.592
All Other Majors vs General Studies	0.231	0.052	20.121	<0.001	1.260	1.139	1.394
Credits Enrolled in Fall	0.156	0.006	776.669	<0.001	1.169	1.156	1.181
Enrolled in Developmental Course	-0.148	0.043	11.647	0.001	0.862	0.792	0.939
Enrolled in SDV Course	0.322	0.039	67.275	<0.001	1.380	1.278	1.491
Enrolled in College-level Math	0.158	0.048	10.705	0.001	1.171	1.066	1.288
Enrolled in College-level English	0.202	0.042	23.144	<0.001	1.224	1.127	1.329

Notes: B=coefficient estimates; SE=standard error; Wald=Wald statistic; p=p-value; OR=odds ratio; CI=confidence interval for odds ratio. All Other Majors: Includes all majors other than General Studies (A.S.), Business Administration (A.S.), Social Sciences (A.S.), and Science (A.S.).

The odds ratio is known as odds change, which describes the proportionate change in the odds for one-unit difference in the explanatory variable (Hosmer and Lemeshow, 2000; and Menard, 1995). Logistic regression models the log odds of a positive response (probability modeled is persisted=1) as a linear combination of the predictor variables. In logistic regression, the dependent variable (persisted) is a logit, exponentiating logit gives us odds, dividing the odds by odds+1 gives probability.

In this model, the estimated logit is given by the following expression:

$$\text{Log} [p / (1-p)] = - 0.5366 - (0.3629*\text{male}) + (0*\text{White}) - (0.0082*\text{Black}) - (0.0454*\text{Hispanic}) + (0.4808*\text{Asian}) - (0.0211*\text{Other Race/Ethnicity}) + (0*\text{AA/AS}) - (0.2730*\text{AAA/AAS}) - (0.3708*\text{Certificate}) + (0.2310*\text{Other Major}) + (0.0997*\text{Business Admin}) + (0.0595*\text{Social Sciences}) + (0.3163*\text{Science}) + (0*\text{General Studies}) - (0.4034*\text{Online Only}) - (0.0636*\text{On-campus Only}) + (0*\text{On-Campus and Online}) + (0.2176*\text{Day Only}) - (0.0083*\text{Night Only}) + (0*\text{Day and Night}) + (0.1558*\text{Credits Enrolled in fall}) - (0.1480*\text{Enrolled in Developmental Course}) + (0.3224*\text{Enrolled in SDV Course}) + (0.1583*\text{Enrolled in College-level Math}) + (0.2022*\text{Enrolled in College-level English})$$

Estimates produced by the model were validated by computing the predicted probability of persisting past the first term for each of the Fall 2017 FTIC program-placed students (Table 4, next page). Error rate and accuracy are the most common and intuitive measures derived from the confusion matrix.

- Error rate is calculated as the number of all incorrect predictions (cells in yellow) divided by the total number of the dataset. The best error rate is 0.0 (or 0 percent), whereas the worst is 1.0 (or 100 percent).

- Accuracy is calculated as the number of all correct predictions (cells in green) divided by the total number of the dataset. The best accuracy is 1.0 (or 100 percent), whereas the worst is 0.0 (or 0 percent). It can also be calculated by  $1 - \text{Error Rate}$ .

Table 4 present the results of model performance on the Fall 2017 cohort.

**Table 4. Confusion Matrix: Fall 2017 Cohort**

Predicted by the Model	Observed (Actual)		Total
	Persisted past 1 <sup>st</sup> Term	Did not persist	
Persisted past 1 <sup>st</sup> Term	5,124	975	6,099
Did not persist	442	394	836
<b>Total</b>	<b>5,566</b>	<b>1,369</b>	<b>6,935</b>

Note: First-term persistence is calculated as the number of students who graduated from NOVA or transferred to a four-year institution by the end of the first term, or who returned to NOVA for the second term of study.

The error rate based on Fall 2017 test data was 20 percent (cells in yellow) and accuracy was 80 percent (cells in green). When compared to the training data (i.e., the data with which the model was developed), the overall rate of correct classification (80 percent) and sensitivity (i.e., percent correctly classified as persisted) (92 percent) were same with test data. The specificity (i.e., percent correctly classified as not persisted) decreased marginally by 2 percentage points from 31 percent to 29 percent. Therefore, we can conclude that model performed reasonably well with the new data.

### Use of Regression Results

Using the estimates produced by the model, the Office of Institutional Research will compute the predicted probability of persisting after first term for the Fall 2019 FTIC, program-placed cohort of students. Students with low predicted probability (i.e., lower than the classification cutoff value) will be categorized as “at-risk.” The list of at-risk students will be provided to the College for targeting intervention strategies.

### Limitations

Many variables affect a student’s decision to remain at or leave an institution of higher education. Variables that affect persistence are generally categorized as cognitive (intellectual), non-cognitive, and environmental. The challenge of student retention is complex. Not all students are alike, nor are programs and institutions. This analysis did not examine the relationship between student support services and student persistence. The type and quality of counseling, advising, orientation, and learning assistance services can have positive effect on student persistence. Additionally, this analysis examined only the data that was available to the Office of Institutional Research. For example, a preliminary review of the data revealed that 35 percent of the study population did not submit the Free Application for Federal Student Aid (FAFSA). Therefore, after reviewing the various techniques for handling missing data (listwise and pairwise deletions, mean substitution, last observation carried forward, maximum likelihood, expectation-maximization, and multiple imputation), it was decided not to include financial aid variables. High school GPA and admission test scores were not available; therefore, these variables were not included.

## Conclusion

Retaining students is fundamental to an institution's ability to carry out its mission. Institutions of higher education have a distinct responsibility to help students complete their goals. That responsibility can be fulfilled effectively when institutions use available information to properly counsel students and to implement programs and policies that work for student persistence. Literature on student persistence suggests that institutions use research specific to their institution and to their student populations as a basis for designing and implementing student persistence programs. Early identification of at-risk students is crucial. If not identified, institutions will notice students who are struggling academically only after they start to fail and interventions may only be attempted after several failures, which reduces the likelihood of persistence. Therefore, in order to improve academic performance and, thus, improve student persistence, an early warning system for at-risk students is recommended.

The Office of Institutional Research welcomes comments and suggestions regarding this report.

## References

- Allison, P. D. (2012). *Logistic regression using SAS: Theory and application* (2nd ed.). Cary, NC: SAS Institute Inc.
- Bailey, T., & Jenkins, D. (2017). *Early Momentum Metrics: Why They Matter for Higher Education Reform*. Research Brief 65. New York: Columbia University, Community College Research Center.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons, Inc.
- Levitz, R. S., Noel, L., & Richter, B. J. (1999). Strategic moves for retention success. In G.H. Gaither (Ed.), *Promising practices in recruitment, remediation, and retention* (pp. 31–50). (New Directions for Higher Education, n. 108) San Francisco: Jossey-Bass.
- Menard, S. (1995). Applied logistic regression analysis. *Sage university papers: Quantitative applications in the social Sciences, 07-106*. Newbury Park, CA: Sage.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2017). *Applied multivariate research: Design and interpretation* (3rd ed.). Thousand Oaks, CA: Sage.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon.

## Appendix

**Table A1. First-Term Persistence by Student Characteristics:  
Fall 2014 through Fall 2016 Cohorts**

	Fall 2014			Fall 2015			Fall 2016		
	N	Persisted past 1 <sup>st</sup> Term		N	Persisted past 1 <sup>st</sup> Term		N	Persisted past 1 <sup>st</sup> Term	
		#	%		#	%		#	%
<b>Gender</b>									
Male	4,059	3,187	78.5%	4,224	3,390	80.3%	3,883	3,100	79.8%
Female	3,670	3,051	83.1%	3,523	2,952	83.8%	3,446	2,859	83.0%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>
<b>Race/Ethnicity</b>									
White	2,482	2,003	80.7%	2,470	1,975	80.0%	2,334	1,882	80.6%
Black/African American	1,382	1,059	76.6%	1,364	1,071	78.5%	1,255	1,015	80.9%
Hispanic/Latino	2,151	1,721	80.0%	2,151	1,745	81.1%	2,107	1,669	79.2%
Asian	1,175	1,026	87.3%	1,160	1,055	90.9%	1,152	1,018	88.4%
American Indian/Alaska Native	14	13	92.9%	14	10	71.4%	18	11	61.1%
Hawaiian/Pacific Islander	42	34	81.0%	32	27	84.4%	39	27	69.2%
Not Specified	2	2	100.0%	13	13	100.0%	6	6	100.0%
Two or More Races	375	291	77.6%	430	346	80.5%	339	262	77.3%
Unknown	106	89	84.0%	113	100	88.5%	79	69	87.3%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>
<b>Program Placement</b>									
AAA/AAS	942	689	73.1%	874	651	74.5%	727	525	72.2%
AA/AS	6,550	5,395	82.4%	6,720	5,582	83.1%	6,468	5,348	82.7%
Certificate	237	154	65.0%	153	109	71.2%	134	86	64.2%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>
<b>Developmental Course Enrollment</b>									
Enrolled	2,773	2,247	81.0%	2,692	2,200	81.7%	2,531	2,045	80.8%
Did not Enroll	4,956	3,991	80.5%	5,055	4,142	81.9%	4,798	3,914	81.6%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>
<b>SDV Course Enrollment</b>									
Enrolled	5,370	4,515	84.1%	5,193	4,402	84.8%	5,334	4,558	85.5%
Did not Enroll	2,359	1,723	73.0%	2,554	1,940	76.0%	1,995	1,401	70.2%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>
<b>College-level English Course Enrollment</b>									
Enrolled	5,464	4,631	84.8%	5,610	4,798	85.5%	5,356	4,513	84.3%
Did not Enroll	2,265	1,607	70.9%	2,137	1,544	72.3%	1,973	1,446	73.3%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>
<b>College-level Math Course Enrollment</b>									
Enrolled	3,100	2,727	88.0%	3,191	2,836	88.9%	3,056	2,683	87.8%
Did not Enroll	4,629	3,511	75.8%	4,556	3,506	77.0%	4,273	3,276	76.7%
<b>Total</b>	<b>7,729</b>	<b>6,238</b>	<b>80.7%</b>	<b>7,747</b>	<b>6,342</b>	<b>81.9%</b>	<b>7,329</b>	<b>5,959</b>	<b>81.3%</b>

Note: First-term persistence is calculated as the number of students who graduated from NOVA or transferred to a four-year institution by the end of the first term, or who returned to NOVA for the second term of study.