Predicted versus Actual GPA: Do Large Differences Signal an Early Warning for Student Attrition?

Nivedita Bhaktha[1], Justine Radunzel[2], and Krista Mattern[2]

[1] The Ohio State University

[2] ACT

Abstract

Postsecondary institutions often implement early alert systems to identify students at risk of dropping out. This study examines whether a measure of under-/over-performance given a student's first-term grade point average (GPA) contributes to identifying students with greater chances of leaving the institution in year two, while controlling for other student and institutional characteristics. Study results are based on nearly 120,000 students from 37 four-year institutions. Students were tracked across institutions using National Student Clearinghouse data. Over-/under-performance is calculated as the difference between the actual GPA and that predicted from ACT scores and high school grades. Results suggest that under- and over-performance help to identify students who may be more likely to leave the institution and where transfer students may go.

Predicted versus Actual GPA: Do Large Differences Signal an Early Warning for Student Attrition?

Objectives

To help improve their retention and degree completion rates, postsecondary institutions are interested in student information that can be incorporated into their early alert systems. These systems are used for early identification of students who are at-risk of leaving their institution, either due to poor academic performance or other issues such as academic disengagement or having trouble making the transition from high school to college (Faulconer, Geissler, Majewski, & Trifilo, 2014; Hudson, 2006; Tampke, 2013). The goal of such systems is to get students connected with academic and student support services when the interventions might be most beneficial and effective, before it is too late and students find themselves in academic jeopardy and unable to adjust to college life and expectations (Beck & Davidson, 2002).

Given that institutions are interested in supplementing their early alert systems with meaningful, diagnostic information that helps to identify students who are likely to leave their institution, this study sought to explore the usefulness of comparing students' expected and actual performance during their first term for this purpose. Such information is likely readily available for use as institutions often develop admissions models to estimate first-year grade point average (GPA) from students' standardized test scores and high school GPAs (Clinedinst, 2015). With this in mind, the primary objective of the current study was to examine whether the difference between the actual and predicted first-term GPA was related to student attrition at year two, while statistically controlling for other relevant student and institutional characteristics and

differentiating between attrition due to drop out versus transfer. A secondary objective was to evaluate whether over-/under-performance was related to the type of institution transferred to in year two.

Theoretical Framework

According to Tinto (1975; 1993), student retention is multifaceted and influenced by various factors related to students' pre-entry attributes, academic goals and commitments, institutional experiences, and academic and social integration into the college environment. In terms of pre-entry characteristics, students entering college better prepared academically typically have higher retention rates than those entering underprepared (Lotkowski, Robbins, & Noeth, 2004; Mattern & Patterson, 2009; Radunzel & Noble, 2012). Additionally, certain student demographic groups typically have lower retention rates than their peers. For example, males, racial/ethnic minority students, economically disadvantaged students, and first-generation students generally have lower retention rates than their corresponding peers (Ishitani, 2016; Kopp & Shaw, 2016).

The extent to which students are committed to attaining a college degree is another factor positively related to students' chances of being retained (Allen, Robbins, Casillas, & Oh, 2008), as are various measures of academic and social integration (Ishitani, 2016). Some choices that can help foster such integration include: living on campus, participating in campus activities, and enrolling full-time (Kuh, Kinzie, Buckley, Bridges, & Hayek, 2006; Radunzel, 2017). Moreover, having to work many hours off-campus and attending a college farther from home have been found to be negatively related to social integration and student retention (Kuh et al., 2006; Mattern, Wyatt, & Shaw, 2013; Tognoli, 2003). Institutional characteristics can also play a role in social integration (Tinto, 1993) and students' chances of being retained (Bowen, Chingos, & McPherson, 2009; Kopp & Shaw, 2016).

Given that some of the noncognitive factors related to retention can be difficult for an institution to measure and define, Shaw & Mattern (2013) proposed that institutions consider using the difference between a students' observed first-year GPA and that predicted from students' standardized test scores and high school grades. The study found that students who earned a higher or lower GPA than expected were more likely than their peers to leave their institution after their second, third, and fourth year of college, even after controlling for other relevant student and institutional characteristics. Given that institutions are interested in intervening with at-risk students before the end of the first year, the current study seeks to replicate and extend the results from the Shaw & Mattern (2013) study by using the first-term residual for earlier identification of those with greater chances of leaving their institution in year two, as well as evaluating whether the residual helps to inform where transfer students are likely to go.

Data

Sample

The data consisted of over 120,000 ACT-tested students who had enrolled as first-time entering students in fall 2012 through 2014 in one of thirty-seven public four-year postsecondary institutions from three state systems. Data for students' demographic characteristics, high school coursework taken, high school GPA (HSGPA), educational goals, and college plans were provided by students at the time they registered to take the ACT test. The state systems provided students' term and cumulative GPA, credit hours attempted and earned, and an indicator for subsequent fall enrollment. Subsequent fall enrollment was augmented with enrollment information from the National Student Clearinghouse to track students across institutions. Complete data were available for 100,007 students.

Measures

Primary outcome: The primary outcome was whether a student returned to their initial institution in the fall of year 2. This variable was coded into the following three distinct categories to differentiate between two types of attrition:

- Returned to initial institution (retention)
- Transferred to another institution (transfer)
- Dropped out – not enrolled in college (dropout)

Secondary outcome: The secondary outcome was a binary outcome for the type of institution transferred to in year two (1 for reverse transfer (transferred down to a two-year institution); 0 for lateral transfer (transferred to another four-year institution)).

FTGPA: First-term GPA ranged from 0.00 to 4.00 with a mean of 2.73 and SD of 1.03.

HSGPA: Self-reported high school GPA ranged from 0.00 to 4.00 with a mean of 3.37 and SD of 0.51. It has been reported that students report high school coursework and grades accurately relative to information provided in their official high school transcripts (Sanchez & Buddin, 2016; Shaw & Mattern, 2009).

ACT Composite score: The ACT Composite score is the rounded average of the four subject area scores. Students' most recent official ACT score was used in the analyses. ACT scores ranged from 8.00 to 36.00 with a mean of 22.61 and SD of 4.57.

Other student and institutional characteristics: Table 1 in the Appendix provides a description of the other student and institutional characteristics that were statistically controlled for in the multiple-predictor retention models. These characteristics have been found to be associated with

retention and transfer in other studies (Mattern, Wyatt, & Shaw, 2013; Radunzel, 2017; Shaw & Mattern, 2013).

Method

Hierarchical regression models were developed to account for the nested structure of the data given that students are clustered within institutions.

FTGPA Prediction Model

FTGPA was estimated from a hierarchical linear regression model that included HSGPA, ACT score, and their interaction and allowed the intercept and slopes to vary randomly across institutions. The residual was defined as follows:

Residual = $\mathrm{FTGPA}_{\mathrm{observed}} - \mathrm{FTGPA}_{\mathrm{predicted}}$.

A student was said to have underperformed if their residual was negative (the student earned a lower FTGPA than was predicted) and overperformed if their residual was positive (the student earned a higher FTGPA than was predicted). The predicted values were restricted to be in the range of [0, 4].

Retention Model

A hierarchical multinomial regression model was estimated for the three-category retention outcome based on student and institutional characteristics, including the residual. A quadratic and cubic residual term was used in the model to account for the curvilinear relationship suggested by the data. The intercepts were allowed to vary randomly across institutions.

Transfer Model

For the secondary outcome, a hierarchical logistic regression analysis was conducted to model the probability of a student transferring down to a two-year institution. The same set of predictors that were employed in the retention model were used, including the quadratic and cubic residual terms.

Model Fit

An index to flag students at-risk of leaving based on certain criteria (such as bottom and top 5%) and corresponding classification accuracies are currently being explored.

Results

Regression coefficients for the FTGPA prediction model are provided in Table 2. The resulting FTGPA residuals are shown in Figure 1. Given there was a high correlation between the residuals and FTGPA (r=0.86; Table 3), it was decided to not include FTGPA as a predictor in the retention/attrition models to avoid issues of multicollinearity. Residuals were used in the model instead of FTGPA as it was believed that they not only contained information about students' academic performance but also other noncognitive information that may be related to reasons why students leave the institution (Shaw & Mattern, 2013).

Retention

Figure 2 shows the curvilinear relationship between the residuals and observed retention/attrition rates. Compared to students who performed as expected based on their pre-college achievement measures, students who underperformed as well as students who overperformed tended to be less likely to return to their initial institution in year two. The correlation between the predicted retention probabilities based on the curvilinear relationship with the residual and retention is

0.46, which is greater than the linear correlation coefficient of 0.39 between the residual and retention.

Results from the multiple-predictor retention/attrition models are summarized in Table 4. Even after controlling for other relevant student and institutional characteristics, the FTGPA residual terms were significantly related to both dropout and transfer outcomes (Figure 4). The directionality and significance of the other predictors included in the model were consistent with those found in an earlier study (Radunzel, 2017).

Transfer

The curvilinear relationships between students' residuals and observed transfer type rates are plotted in Figure 3. As initially hypothesized, these results suggest that transfer students who had underperformed during their first term (i.e., those with larger negative residuals) were more likely to reverse transfer down to a two-year institution, whereas those who overperformed (i.e., those with larger positive residuals) were more likely to laterally transfer to another four-year institution.

Results from the multiple-predictor transfer models are summarized in Table 5. After controlling for other student and institutional characteristics, the FTGPA residual terms remained related to the type of institution transferred to in year two in a similar manner as suggested by the descriptive analyses (Figure 5). The significance and directionality of the other predictors included in the model were generally consistent with those found in an earlier study (Radunzel, 2017).

Scholarly Significance

This study contributes to the existing literature on student retention in several ways. First, this study found for a recent cohort of students that the differences in the predicted and actual FTGPA helps to inform students' likelihood of returning to the initial institution in year two, as well as helps to differentiate between two types of attrition – dropout versus transfer. The study also found the residuals to be related to reverse versus lateral transfer. Lastly, evaluating students' retention probabilities using grades from the end of the first term instead of waiting until the end of the first year would seem to provide an earlier opportunity for connecting students with institutional services and supports that are aimed at getting students back on track to succeed.

References

Beck, H. P., & Davidson, W. D. (2001). Establishing an early warning system: Predicting low grades in college students from survey of academic orientations scores. *Research in Higher Education, 2*(6), 709–723.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

Clinedinst, M. (2015). *2014 State of college admission*. Alexandria, VA: National Association for College Admissions Counseling.

Faulconer, J., Geissler, J., Majewski, D., & Trifilo, J. (2014). Adoption of an early-alert system to support university student success. *Delta Kappa Gamma Bulletin, 80*(2)*,* 45–48.

Hudson, W. E. (2006). Can an early alert excessive absenteeism warning system be effective in retaining freshman students? *Journal of College Student Retention: Research, Theory and Practice* 7, 217-226.

Ishitani, T. T. (2016). Time-varying effects of academic and social integration on student persistence for first and second years in college: National data approach. *Journal of College Student Retention: Research, Theory & Practice, 18*(3), 263–286.

Kopp, J. P., & Shaw, E. J. (2016). How final is leaving college while in academic jeopardy? Examining the utility of differentiating college leavers by academic standing. *Journal of College Student Retention: Research, Theory, & Practice, 18*(1), 2–30.

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature* (Report commissioned for the National Symposium on Postsecondary Student Success: Spearheading a Dialog on Student Success). Washington, DC: National Postsecondary Education Cooperative.

Lotkowski, V. A., Robbins, S. B., & Noeth, R. J. (2004). *The role of academic and non-academic factors in improving college retention*. Iowa City, IA: ACT.

Mattern, K. D., & Patterson, B. F. (2009). *Is performance on the SAT related to college retention?* (College Board Research Report No. 2009-7). New York, NY: The College Board.

Mattern, K. D., Wyatt, J. N., & Shaw, E. J. (2013). College distance from home: Implications for

student transfer behavior. *Journal of the First-Year Experience & Students in Transition, 25*(1), 77–92.

Radunzel, J. (2017). *Using incoming student information to identify students at-risk of not returning to their initial institution in year two.* Iowa City, IA: ACT.

Radunzel, J., & Noble, J. (2012). *Tracking 2003 ACT-tested high school graduates: College readiness, enrollment, and long-term success.* Iowa City, IA: ACT.

Sanchez, E., & Buddin, R. (2016). *How accurate are self-reported high school courses, course grades, and grade point average?* Iowa City, IA: ACT.

Shaw, E. J., & Mattern, K. D. (2013). Examining student under- and overperformance in college to identify risk of attrition. *Educational Assessment, 18*(4), 251–268.

Shaw, E. J., & Mattern, K. D. (2009). *Examining the accuracy of self-reported high school grade point average.* (College Board Research Report 2009-5). New York, NY: The College Board.

Tampke, D. R. (2013). Developing, implementing, and assessing an early alert system. *Journal of College Student Retention: Research, Theory, and Practice, 14*(4), 523–532.

Tognoli, J. (2003). Leaving home: Homesickness, place attachment, and transition among residential college students. *Journal of College Student Psychotherapy, 18*(1), 35–48.

Appendix

Figure 1

Histogram of the residuals from the predicted FTGPA model

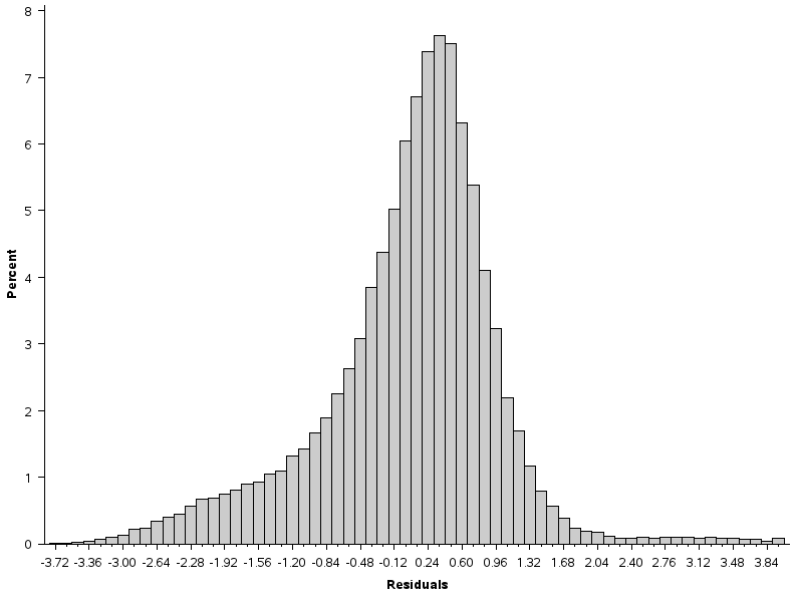

Figure 2

Plot of residuals versus the retention, transfer and dropout rates

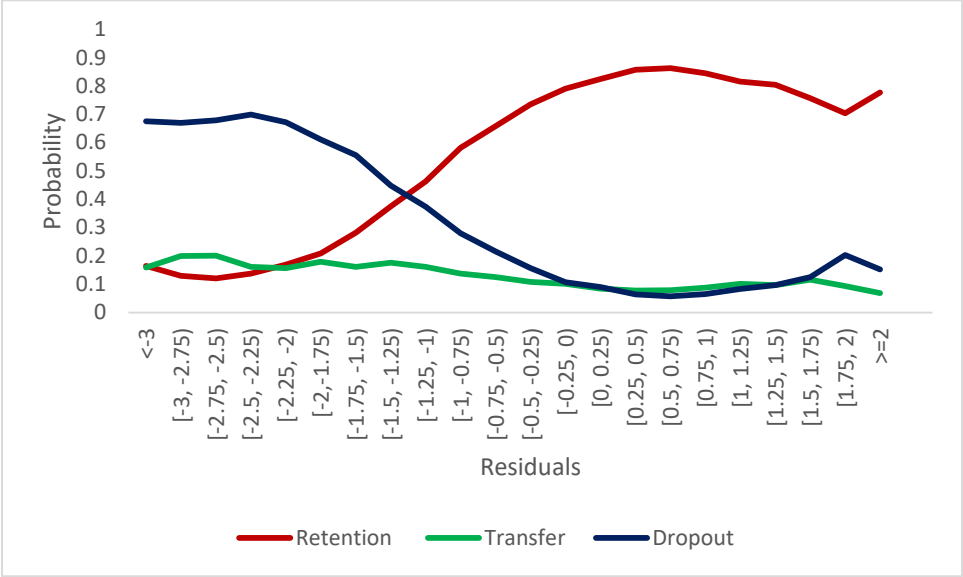

Figure 3

Plot of residuals versus the transfer rate

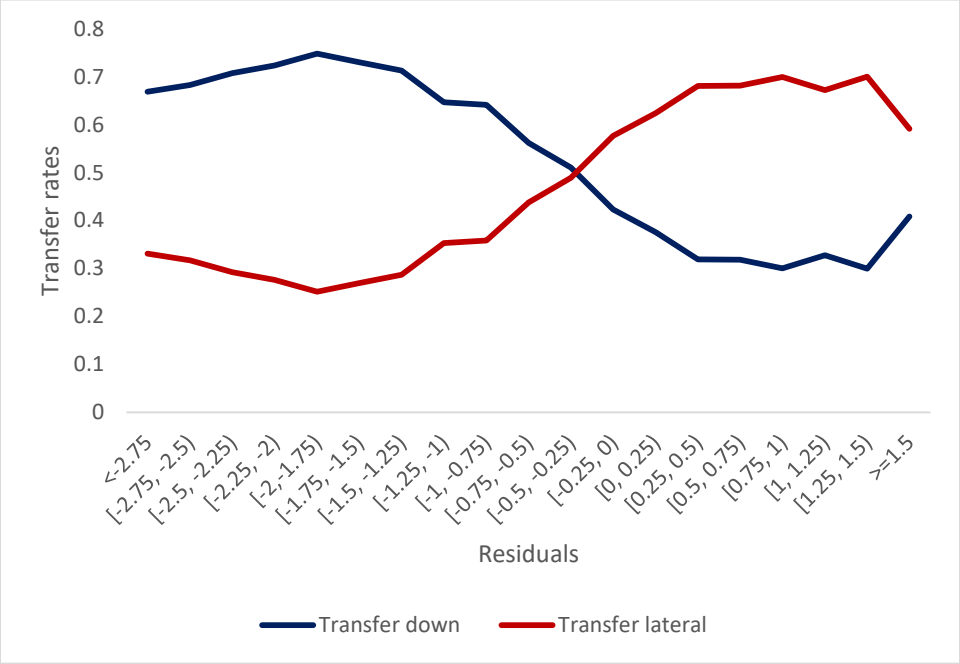

Figure 4

Plot of modeled retention, transfer and dropout probabilities
with all other predictors held constant at sample mean values

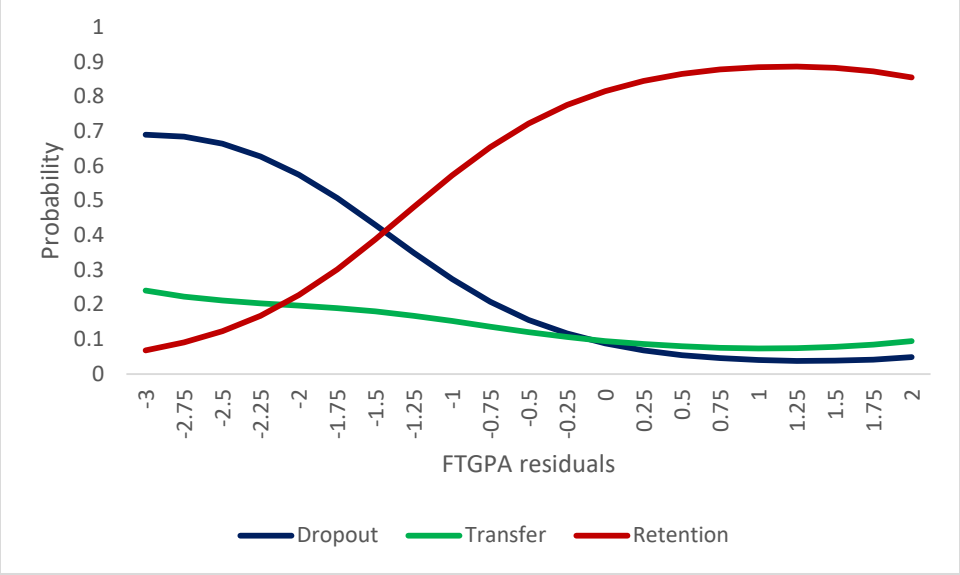

Figure 5

Plot of modeled transfer down and transfer lateral probabilities with all other predictors held constant at sample mean values

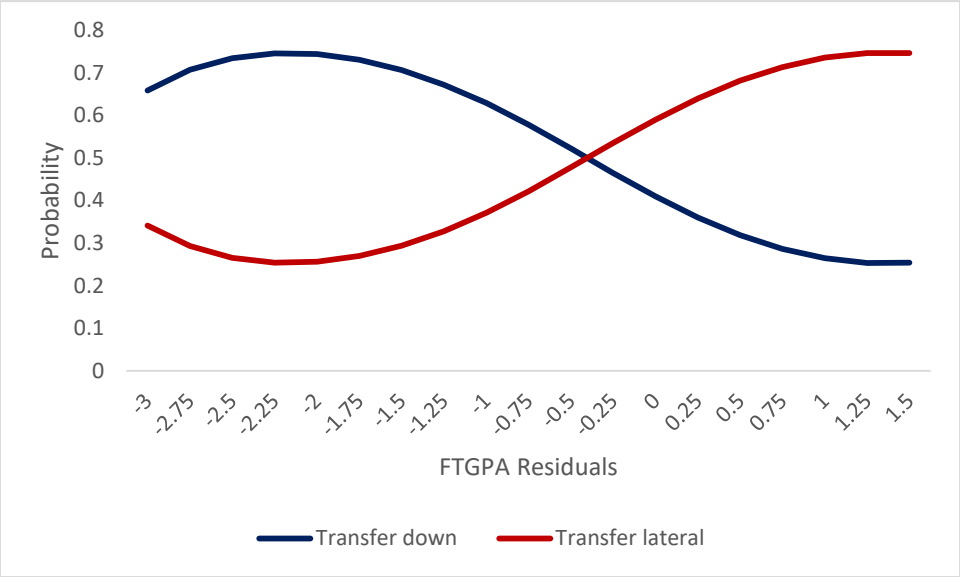

Table 1

Description of the student and institutional characteristics that were included in the model

Variables	Categories	Description
Gender	Male Female (Reference)	Student's gender
Race/Ethnicity	White (Reference) Black/African-American American Indian/Alaskan Native Hispanic/LatinX Asian Other	Student's race/ethnicity
Log Distance		log (distance from home to college attended in miles + 1) Based on student's residence zip code.
Planned Work Hours	None 1-10 11-20 21-30 31 or more (Reference)	Number of hours the student plans to work per week while enrolled in college
Median Family Income (in $)	0 - 43,315 43,315 - 61,580 >61,580 (Reference)	Median family household income based on the zip code of the student's residence
Educational Plans	Bachelor's degree Graduate degree Other Less than a Bachelor's degree (Reference)	The highest level of education that the student expects to complete
Highest Parental Education	No college education Some college education Bachelor's degree Degree beyond a Bachelor's (Reference)	Highest education level attained by the students' parents
Instate	Out of state student In state student (Reference)	Location of the student's state residence with respect to the state location of postsecondary institution
Highest Math Course	Calculus Trig/Other Advanced Math Algebra II Below Algebra II (Reference)	Highest math course taken by the student in high school

Institution Size	<10,000 >10,000 (Reference)	Institution size based on total students enrolled for credit - Source: The Integrated Postsecondary Education Data System (IPEDS) 2010
Policy	Highly Selective Selective Traditional Liberal Open (Reference)	Self-reported by institution as part of the ACT Institutional Data Questionnaire

Table 2
Regression coefficients of the FTGPA prediction model

Predictors	Coefficient	SE	p-value
Intercept	1.5576	0.1309	<.0001
ACT Score	-0.0608	0.0064	<.0001
HSGPA	0.0661	0.0361	<.0001
ACT Score*HSGPA	0.0308	0.0018	<.0001

Covariance Parameter estimates			
	Estimate	SE	p-value
Intercept	0.2800	0.0964	0.0018
ACT Score	0.0006	0.0003	0.0067
HSGPA	0.0161	0.0065	0.0064
ACT Score*HSGPA	0.0000	0.0000	0.0065

AIC=301050.8, R^2 = 0.7396

Table 3
Pearson correlations of the study variables

Variables	1	2	3	4	5	6	7	8
1.Retention	1.00							
2.Transfer	-0.58	1.00						
3.Dropout	-0.72	-0.15	1.00					
4.Residuals	0.39	-0.09	-0.40	1.00				
5.FTGPA	0.49	-0.12	-0.49	0.86	1.00			
6.HSGPA	0.27	-0.07	-0.27	0.00	0.46	1.00		
7.ACT score	0.26	-0.09	-0.24	0.00	0.37	0.55	1.00	
8.Log distance	0.03	0.07	-0.09	0.04	0.09	0.10	0.10	1.00

*All the correlations are significant at $p \leq 0.01$.

Table 4
Regression coefficients of the retention model
Level 1

	Dropout vs. Return			Transfer vs. Return		
Predictors	Coefficient	SE	p-value	Coefficient	SE	p-value
Intercept	-0.2575	0.2865	0.3722	-2.4242	0.2194	<.0001
Residuals	-1.2488	0.0181	<.0001	-0.6022	0.0187	<.0001
Residual sq.	0.3142	0.0095	<.0001	0.2444	0.0100	<.0001
Residual cu.	0.0755	0.0046	<.0001	0.0223	0.0050	<.0001
Gender - Male	0.0561	0.0216	0.0093	-0.1633	0.0222	<.0001
Race/Ethnicity						
Black/African-American	0.3387	0.0320	<.0001	0.2207	0.0344	<.0001
American Indian/Alaskan Native	0.2195	0.0948	0.0206	0.1147	0.1057	0.2778
Hispanic/Latinx	-0.0266	0.0505	0.5983	-0.0048	0.0536	0.9290
Asian	-0.6428	0.1041	<.0001	-0.2364	0.0938	0.0117
Others	0.2590	0.0480	<.0001	0.0809	0.0538	0.1329
Log Distance	-0.1095	0.0220	<.0001	0.5732	0.0253	<.0001
Planned work hours						
None	-0.8984	0.0663	<.0001	-0.2608	0.0809	0.0013
10-Jan	-0.7739	0.0637	<.0001	-0.1609	0.0795	0.0429
20-Nov	-0.4741	0.0606	<.0001	-0.1059	0.0777	0.1729
21-30	-0.1591	0.0635	0.0123	0.0034	0.0816	0.9664
Median Family Income						
0 - 43,315	0.2770	0.0329	<.0001	-0.0338	0.0322	0.2935
43,315 - 61,580	0.1793	0.0305	<.0001	0.0175	0.0283	0.5370
Educational Plans						
Bachelor's degree	-0.6391	0.0645	<.0001	-0.0631	0.0900	0.4837
Graduate degree	-1.0648	0.0667	<.0001	-0.2381	0.0914	0.0091
Others	-0.3601	0.1206	0.0028	0.0412	0.1518	0.7860
Highest Parental Education						
No college education	0.6850	0.0377	<.0001	0.1986	0.0389	<.0001
Some college education	0.4883	0.0335	<.0001	0.1989	0.0319	<.0001
Bachelor's degree	0.1143	0.0348	0.0010	0.0576	0.0311	0.0637
Instate	-0.0380	0.0473	0.4225	0.0835	0.0427	0.0505
Highest Math Course						
Calculus	-1.3653	0.0963	<.0001	-0.8122	0.1163	<.0001
Trig/Other Advanced Math	-0.9168	0.0956	<.0001	-0.4959	0.1157	<.0001
Algebra II	-0.4272	0.0978	<.0001	-0.3004	0.1187	0.0114

Level 2

	Dropout vs. Return				Transfer vs. Return			
	Estimate	OR	SE	p-value	Estimate	OR	SE	p-value
Policy								
Highly Selective	-0.2939	0.7454	0.5625	0.6013	-0.2758	0.7590	0.3002	0.3584
Selective	-0.0328	0.9677	0.3399	0.9230	0.0237	1.0240	0.1873	0.8994
Traditional	0.1620	1.1759	0.2297	0.4807	0.2408	1.2723	0.1328	0.0699
Liberal	0.2530	1.2879	0.3406	0.4576	0.5383	1.7131	0.1923	0.0051
Institution Size								
<10,000	0.6873	1.9883	0.2033	0.0007	0.2855	1.3304	0.1145	0.0127
Covariance parameter	0.2502		0.0632	<.0001	0.0680		0.0221	0.0010

-2 Log Likelihood = 121798.3

Table 5
Regression coefficients of the transfer down (vs. lateral transfer) model
Level 1

Independent Variables	Coefficient	OR	SE	p-value
Intercept	1.2791	3.5934	0.4061	0.0036
Residuals	-0.8738	0.4174	0.0358	<.0001
Residual sq.	0.1176	1.1248	0.0232	<.0001
Residual cu.	0.0984	1.1034	0.0101	<.0001
Gender – Male	-0.0685	0.9338	0.0447	0.1256
Race/Ethnicity				
Black/African-American	0.4027	1.4959	0.0667	<.0001
American Indian/Alaskan Native	0.3845	1.4689	0.2059	0.0619
Hispanic/Latinx	0.3598	1.4330	0.1058	0.0007
Asian	-0.3262	0.7217	0.1983	0.0999
Others	0.0072	1.0072	0.1061	0.9462
Log Distance	-0.4966	0.6086	0.0520	<.0001
Planned Work Hours				
None	-0.7132	0.4901	0.1646	<.0001
1-10	-0.5386	0.5836	0.1616	0.0009
11-20	-0.3160	0.7291	0.1584	0.0460
21-30	-0.2200	0.8025	0.1656	0.1840
Median Family Income				
0 – 43,315	-0.0971	0.9075	0.0634	0.1259
43,315 – 61,580	-0.1687	0.8448	0.0568	0.0030
Educational Plans				
Bachelor's degree	-0.7767	0.4599	0.1889	<.0001
Graduate degree	-1.0258	0.3585	0.1915	<.0001
Others	-0.7747	0.4608	0.3069	0.0116
Highest Parental Education				
No college education	0.5234	1.6878	0.0772	<.0001
Some college education	0.3818	1.4649	0.0640	<.0001
Bachelor's degree	0.2196	1.2456	0.0638	0.0006
Instate	0.1886	1.2076	0.0832	0.0234
Highest Math Course				
Calculus	-0.3492	0.7053	0.2212	0.1145
Trig/Other Advanced Math	-0.1287	0.8792	0.2201	0.5588
Algebra II	0.1106	1.1169	0.2262	0.6249

Level 2

	Estimate	OR	SE	p-value
Policy				
Highly Selective	0.6351	1.8872	0.4570	0.1646
Selective	0.0847	1.0884	0.2924	0.7721
Traditional	0.4793	1.6149	0.2120	0.0238
Liberal	0.5676	1.7640	0.3016	0.0599
Institution Size				
<10,000	0.309	1.3621	0.1801	0.0863
Covariance parameter	0.1511		0.0470	0.0006

-2 Log Likelihood = 12660.09