

FA10-FA13 Freshman Retention Analysis, CSU Admissions Variables

Report Date: 6/23/16

This study seeks to identify variables that predict freshman student retention across different levels of pre-college preparation. Unlike predictive studies that are buoyed by demographics, this study omits financial aid variables that are unavailable to CSU Admissions and omits demographic variables like ethnicity whose use by Admissions may be in opposition to CSU's mission to provide equal access. New predictor variables are substituted that are available to Admissions, including the graduating High School's CDE score, student remedial status in 1+ subject, and income and population density variables associated with a student's home and graduating High School ZIP codes.

Executive Summary

- A full [executive summary](#) for the project is available.

Study Population

- First-Time Full-Time Freshman cohorts from the FA10, FA11, FA12, and FA13 terms.
- International students are omitted from the analysis, along with domestic students who did not provide full application information (HS GPA, ACT/SAT).
- Regression models that include CDE high school performance variables omit private school students, home-schooled students, and other students whose high schools are not assessed by the CDE. These students represent 8.1% of the Colorado resident sample.

Dependent Variable

- 3rd Fall Persistence. This outcome provides a more accurate view of student success than 2nd Fall persistence without extending the cohort frame to the Six-Year Graduation cohorts FA09 and earlier. Per past data, around 84% of students who retain to the 3rd Fall also graduate from CSU within six years, so 3rd Fall persistence serves as a relatively accurate proxy for likely graduation with a bachelor's degree.

Independent variables used in the models:

• High School GPA (z-score)	• Remedial X STEM Major Interaction: (binary = 1 if STEM and Remedial = 1)
• ACT (z-score; maximum of ACT or converted SAT)	• Colorado Residency (binary = 1 if student residence in Colorado)
• # Credits Transferred to CSU in First Academic Year	• HS CDE Math & Reading Score (z-score of factors; CO residents only)
• Home ZIP Mean Household Income (z-score)	• Gender (binary; Male = 1)
• HS ZIP Mean Household Income (z-score; CO residents only)	• First Generation (binary)
• Home ZIP Population Density (z-score)	• Gender X Cohort STEM Major (binary)
• STEM Major Status (binary)	• Gender X First Generation Status (binary)
• Remedial Status (binary = 1 if any Remedial Flag = 1)	

Figure 1: Modeled Independent Variables

In this study, student record demographic variables have been intentionally omitted from the list of predictor variables. Gender has been included because it is a variable of relatively less concern in terms of equal access, and more importantly because it serves as a leveling variable for student success. At CSU similar proportions of in-state male (77.8%) and female (78.6%) FTFT freshmen persist to Fall Semester 3, despite a substantially lower mean for males in the most strongly predictive pre-college academic preparedness variable, High School GPA (3.49 males, 3.64 females, $p < .001$). Including Gender helps account for that effect and allows for exploration of interactions between Gender and other independent variables like Remedial status and Cohort STEM major.

Method

- Logistic regression models are used to assess the contribution of independent variables toward 3rd Fall persistence.
- Prior research has shown that student persistence varies across certain dimensions. Models are constructed to individually model variance across the following variables:
 - CCHE Index \leq 100 versus CCHE Index $>$ 100, by CO Residency
 - Index 100 was selected as a dichotomous breakpoint because it represents the value where retention rates to the second fall semester drop below the CSU average retention rate of ~86%.
 - CCHE Index \leq 95, Full Population and by CO Residency
 - Index 95 was selected as an additional breakpoint, due to its status as a demarcator of interest among some on-campus decision makers.
 - Remedial Status, by Residency
 - Models that are limited to students with a positive remedial status are also investigated to determine whether there are predictive effects that are specific to these remedial students.

Population Independent Variable Summary

Variable	Mean	(StDev)
ACT	24.82	(3.61)
HS GPA	3.59	(0.42)
Transfer Credits	7.5	(11.2)
Home ZIP Population Density	1985.0	(2017.8)
Home ZIP Mean HH Income	\$90,019	(30829)
HS ZIP Mean HH Income	\$87,833	(26562)
HS CDE Reading + Math	1277.3	(38.3)

(Min 1077.8, Max 1374.1)

Figure 2: Independent Variable Mean and SD

- Z-scored variables were created for each of the above IVs, except for Transfer Credits.

Relationship between 3rd Fall Retention and Colorado High School CDE Math & Reading Scale Scores

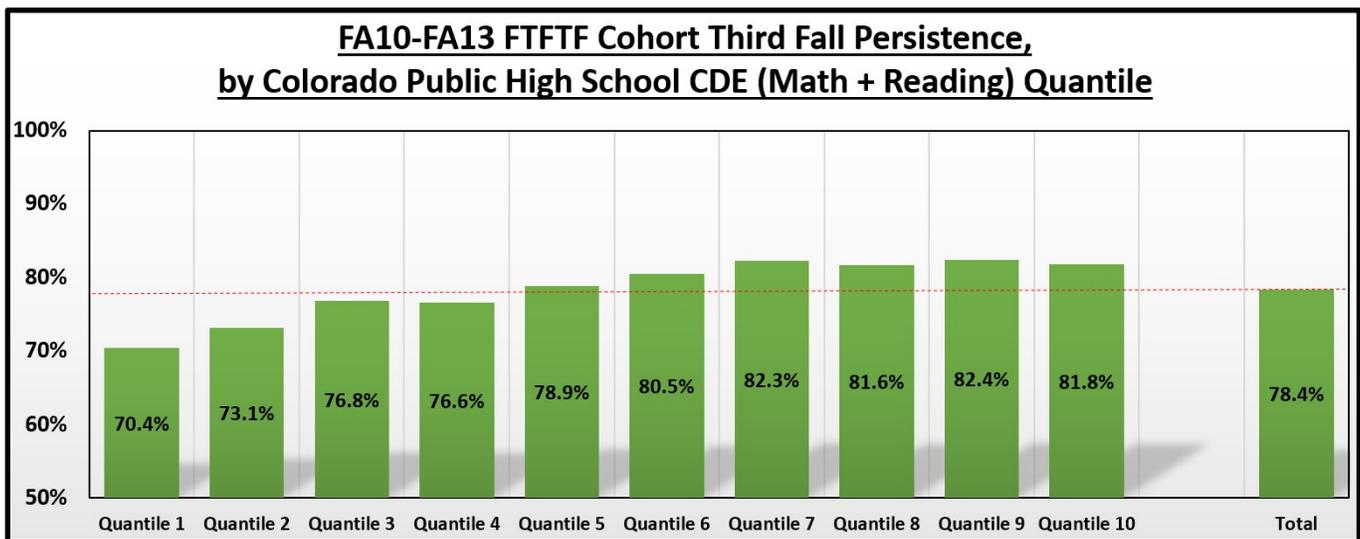


Figure 3: Retention by CDE Math + Reading Decile

- We observe a significant positive relationship between graduating high school CDE (Math + Reading) score and 3rd Fall persistence. These school-level values appear to be valid predictors for individual students.

Index <= 95 Students, 3rd Fall Persistence

Colorado Residents		
Observations:		517
Prob > chi ²	=	0.044
Pseudo R ²	=	0.020
Predictor	Odds Ratio	p-value
Gender (M=1)	1.23	0.317
HS GPA (z)	1.40	0.114
Transfer Credits	1.02	0.498
Remedial Status	0.76	0.181
HS ZIP HHI (z)	1.28	0.011
Home ZIP Density (z)	1.18	0.100
(constant)	3.75	0.000

Non-Residents		
Observations:		160
Prob > chi ²	=	0.046
Pseudo R ²	=	0.067
Predictor	Odds Ratio	p-value
Gender (M=1)	0.70	0.335
Cohort STEM Major	1.36	0.461
HS GPA (z)	1.81	0.122
Transfer Credits	0.93	0.090
Remedial Status	0.34	0.011
HS ZIP HHI (z)	0.84	0.143
Home ZIP Density (z)	0.86	0.171
(constant)	4.82	0.034

Figure 4: Index <= 95 Logistic Regression Models

Students with CCHE Index <=95 comprise 3.9% of the overall study population. CDE High School Reading + Math score is not a significant predictor for this population and thus has been omitted; therefore this Colorado model includes all resident students at Index <=95. Both models are significant at the $p < .05$ level. Interaction terms were omitted from these models due to binary variable sample size. While the overall models are significant, few predictors are statistically significant at the $p < .05$ level (highlighted in red). Given the sample size, predictors meeting or approaching traditionally marginal levels of significance ($p < .10$; highlighted in gold) may also be of interest. Predictors in the next significance band ($.10 < p < .20$) are marked in yellow.

We observe that High School ZIP mean household income is a significant predictor of retention for Colorado residents with Index <=95. A student in this group who attended high school in a ZIP with a mean household income one standard deviation (SD) below the population mean (\$61,271) experiences 78% relative odds of third fall persistence compared to a student attending school in a ZIP at the population mean (\$87,833). For non-residents with Index <=95, the effect of Home ZIP household income trends toward the opposite direction ($p = .143$); a non-resident one SD above the population mean in home ZIP HHI (\$120,848) experiences 84% odds of third fall persistence compared to a similar non-resident living in a ZIP at the population mean (\$90,019). Other findings include:

- High school GPA approaches marginal significance as a positive predictor of persistence in both models ($p = .114$ and $.122$). For Colorado residents, a one SD decrease in GPA (0.42) yields 71% relative odds of persistence, while a similar decrease in GPA for non-residents yields 55% relative odds.
- Remedial status is directionally negative but misses marginal significance ($p = .181$) for Colorado students, and is a statistically significant negative predictor of persistence for non-residents ($p < .05$). Carrying any remedial status is tentatively associated with 76% relative odds of persistence for Index <=95 residents, while remedial status carries 34% odds of persistence for non-residents in this low Index group.
- Living in a less densely populated home ZIP is a marginally significant ($p = .100$) negative predictor of retention for CO residents, wherein a one SD decrease of 2017.8 people/square mile is associated with 84% relative odds of 3rd fall persistence. For non-residents, the population density effect approaches marginal significance and is in the opposite direction ($p = .171$), wherein an increase of 2017.8 people per square mile is tentatively associated with 87% relative odds of third fall persistence.
- Number of initial transfer credits is a marginally significant negative predictor of third fall persistence for non-residents only. Among non-residents, each credit transferred to CSU in the first year corresponds with 93% relative odds of 3rd fall persistence.

Index <= 100 Students, 3rd Fall Persistence

Colorado Residents			Non-Residents		
Observations:		1340	Observations:		394
Prob > chi ²	=	0.000	Prob > chi ²	=	0.017
Pseudo R ²	=	0.032	Pseudo R ²	=	0.039
Predictor	Odds Ratio	p-value	Predictor	Odds Ratio	p-value
First Gen Status	0.70	0.005	Gender (M=1)	0.81	0.362
GenderM * STEM	0.83	0.482	First Gen Status	0.63	0.070
GenderF * STEM	0.52	0.018	Cohort STEM Major	1.40	0.222
GenderF * Not STEM	0.97	0.818	ACT Score (z)	1.12	0.701
GenderM * Not STEM	(reference)		HS GPA (z)	1.67	0.067
ACT Score (z)	0.74	0.093	Transfer Credits	0.98	0.332
HS GPA (z)	1.11	0.511	Remedial Status	0.55	0.043
Transfer Credits	1.01	0.266	Home ZIP HHI (z)	1.03	0.685
Remedial * Not STEM	0.90	0.460	Home ZIP Density (z)	0.86	0.072
Remedial * STEM	0.71	0.344	(constant)	5.71	0.008
HS ZIP HHI (z)	1.20	0.009			
Home ZIP Density (z)	1.22	0.003			
HS CDE R+M score (z)	1.15	0.060			
(constant)	1.96	0.089			

Figure 5: Index <= 100 Logistic Regression Models

Among the FA10-FA13 cohorts, students with CCHE Index <= 100 comprise 10.0% of the study population. Note that this group encompasses the Index <= 95 sample explored in the first model. Separate models were applied to resident and non-residents. For the Colorado resident logistic regression, interaction terms were included in the model if they approached significance individually or if inclusion increased estimated explanatory capacity. For non-residents, interaction terms were excluded from the model due to small sample size. Both models are significant at $p < .05$.

For Colorado residents with Index <= 100, we observe similar positive effects of High School ZIP Mean Household Income and Home ZIP population density as observed in the Index <= 95 model; these effects are significant at $p < .05$. The quality of education at the student's high school, as proxied by the school's 2011-2014 CDE Reading and Math scale scores, is also a marginal predictor of 3rd Fall Persistence ($p < .05$); a 38.3 point decrease in the student's graduating High School's CDE math and reading scale scores is associated with 87% relative odds of 3rd Fall persistence.

Effects of remedial status appears to vary for Index <=100 residents compared to Index <= 95 residents. Neither the main effect of remedial status nor the (remedial status * STEM major) interaction are significant for this population, although both display negative coefficients and inclusion improves estimated explanatory capacity. For males, cohort STEM major status results in a statistically non-significant decrease in persistence odds (83%). On the other hand Females STEM majors in this model exhibit 52% relative odds of 3rd Fall persistence ($p < .05$).

ACT score appears in this resident Index <= 100 model after being excluded from the resident Index <= 95 model due to lack of predictive capacity. For this model, we observe that ACT score generally reaches marginal significance ($p = .093$) with a negative direction; among Index <=100 Colorado students, an increase of 3.6 points on the ACT corresponds to 74% relative odds of 3rd Fall Persistence. This finding suggests that low-Index residents with low GPA but high ACT are less likely to persist than those with high GPA and low ACT.

Assessing the non-resident Index <=100 model, we observe that there are three marginal predictors and one significant predictor at $p < .05$. For each one SD decrease of 0.42 in a non-resident student's high school GPA, that student experiences 60.8% odds of 3rd Fall Persistence ($p = .067$). Remedial status is a negative predictor of persistence for Index <= 100 non-residents ($p < .05$) associated with 55% relative odds of persistence. Home ZIP density is also a marginally significant inverse predictor of non-resident 3rd Fall persistence ($p = .072$), with a one SD increase yielding 86% relative odds. As with Colorado residents (70%), First Generation status is also a negative predictor of 3rd Fall persistence for non-residents ($p = .070$) associated with 63% relative odds of persistence.

Index > 100 Students, 3rd Fall Persistence

Colorado Residents			Non-Residents		
Observations:		10839	Observations:		3505
Prob > chi ²	=	0.000	Prob > chi ²	=	0.000
Pseudo R ²	=	0.055	Pseudo R ²	=	0.044
Predictor	Odds Ratio	p-value	Predictor	Odds Ratio	p-value
First Gen Status	0.73	0.000	First Gen Status	0.56	0.000
GenderM * STEM	0.84	0.025	Gender (M=1)	1.05	0.595
GenderF * STEM	0.91	0.269	Cohort STEM Major	0.99	0.937
GenderF * Not STEM	0.84	0.004	ACT Score (z)	1.13	0.027
GenderM * Not STEM	(reference)		HS GPA (z)	1.45	0.000
ACT Score (z)	0.94	0.096	Transfer Credits	1.00	0.370
HS GPA (z)	1.62	0.000	Remedial * Not STEM	0.92	0.798
Transfer Credits	1.02	0.000	Remedial * STEM	0.51	0.352
Remedial * Not STEM	0.83	0.122	Home ZIP HHI (z)	1.16	0.000
Remedial * STEM	0.34	0.000	Home ZIP Density (z)	1.02	0.635
HS ZIP HHI (z)	1.11	0.001	(constant)	2.61	0.000
Home ZIP Density (z)	0.98	0.495			
HS CDE R+M score (z)	1.22	0.000			
(constant)	3.67	0.000			

Figure 6: Index > 100 Logistic Regression Models

Sample sizes are much larger for the Index>100 sub-population, including 10,839 residents and 3,505 non-residents, and thus the models' ability to detect predictive effects is substantially enhanced compared to the small sample lower-Index models. Both logistic regression models are significant at the $p < .05$ level.

For Colorado residents with Index > 100, we observe significant and positive effects of High School Home ZIP income and High School's CDE HS Reading + Math score that are similar to the findings for residents with Index < 100 ($p < .05$). Student High School GPA, while non-significant in the Index <= 100 model, also is substantially predictive for higher Index residents ($p < .05$), with a one SD decrease in HS GPA of 0.42 associated with 62% relative odds of 3rd Fall Persistence. Transfer credit total is also a significant positive predictor for higher Index residents ($p < .05$), with each one credit reduction in transfer credit associated with 98% odds of persistence.

We also observe two significant Gender by STEM major interactions ($p < .05$), as non-STEM females retain at rates that are 84% of the rate shown by non-STEM males. On the other hand, STEM males also exhibit 84% relative odds of persistence compared to non-STEM males, representing a statistically significant difference, while STEM females retain at rates that are 7% greater (directionally; $p = .269$) than their male STEM counterparts.

Effects of remedial status on resident Index > 100 students also varies based across cohort STEM major status. For non-STEM majors, remedial status is tentatively associated with 83% odds of 3rd Fall persistence, though the effect is not statistically significant ($p = .122$). On the other hand, for STEM majors a remedial status is a strong negative predictor of persistence ($p < .01$), with individuals in that group exhibiting 34% relative odds of 3rd Fall persistence.

For Index > 100 non-residents, both ACT score ($p < .05$) and HS GPA ($p < .05$) are positively and significantly associated with 3rd fall persistence. Among the six tested Residency X Index models, this is the only model for which ACT score predicts a significant portion of variance at the $p < .05$ level, although the relative effect of HS GPA remains greater than ACT; a one SD (3.6 point) decrease in ACT corresponds to 85% relative odds of 3rd fall persistence, while a one SD (0.42) decrease in HS GPA corresponds to 69% relative odds.

(Any Remedial Status = 1) Students, 3rd Fall Persistence

Colorado Residents			Non-Residents		
Observations:		978	Observations:		135
Prob > χ^2	=	0.000	Prob > χ^2	=	0.019
Pseudo R ²	=	0.035	Pseudo R ²	=	0.082
Predictor	Odds Ratio	p-value	Predictor	Odds Ratio	p-value
GenderM * STEM	0.64	0.172	Cohort STEM Major	0.66	0.396
GenderF * STEM	0.33	0.000	ACT Score (z)	1.35	0.380
GenderF * Not STEM	0.86	0.357	HS GPA (z)	2.13	0.005
GenderM * Not STEM	(reference)		Transfer Credits	0.94	0.090
ACT Score (z)	1.03	0.803	Home ZIP HHI (z)	1.12	0.512
HS GPA (z)	1.36	0.002	Home ZIP Density (z)	0.88	0.190
Transfer Credits	1.01	0.323	_cons	2.84	0.077
HS ZIP HHI (z)	1.16	0.088			
Home ZIP Density (z)	1.13	0.067			
HS CDE R+M score (z)	1.16	0.041			
(constant)	2.85	0.000			

Figure 7: (Remedial Status = 1) Logistic Regression Models

Logistic regression models limited to students with at least one incoming remedial status are also explored. These models allow for detection of independent variables that may not be predictive for the general freshman cohort population, but may be predictive for remedial status students in isolation. Significant logistic regression models ($p < .05$) predict 3rd fall persistence among Colorado resident and non-resident students with one or more remedial flag in the areas of reading, writing, or math.

For remedial Colorado residents, ZIP-level variables are marginally predictive of student success. A one SD decrease in high school ZIP household income (\$26,562) is marginally associated ($p = .086$) with 86% odds of 3rd fall persistence, while a one SD decrease in home ZIP density (2017.8 people/square mile) is marginally associated ($p = .067$) with 88% odds of retention. High school performance, in terms of CDE Reading + Math Scale achievement, is a significant predictor of remedial student persistence, wherein a one SD decrease in CDE Reading + Math score (38.3 points) corresponds to 86% odds of 3rd fall persistence. High School GPA remains the strongest significant predictor of persistence for resident remedial students, with a 1 SD decrease associated with 73% relative odds.

Colorado remedial students show evidence of differential effects by Gender and STEM status. Among non-STEM majors, female remedial students retain at rates that are directionally lesser than non-STEM males (86% relative odds, $p = .357$). The odds deficit is significant ($p < .01$) and far greater for STEM females, who carry 33% relative odds of persistence compared to non-STEM remedial males. Male STEM remedial students also exhibit directionally lower retention odds than non-STEM males (64%) although this effect is not statistically significant ($p = .172$).

For out-of-state students, we observe a similar non-significant directional effect wherein Cohort STEM majors persist to the 3rd Fall at 66% the rate of non-STEM out-of-state remedial students ($p = .396$). High School GPA is even more strongly predictive in the non-resident remedial model than other studied models, with a 1 SD decrease associated with 47% odds of persistence. Transfer credits ($p = .090$) and Home ZIP population density ($p = .190$) are negative directional predictors in this model, with each credit transferred in associated with 94% relative odds and a 1 SD decrease in Home ZIP density associated with 88% relative odds of persistence.

Summary of Findings:

- The logistic regression models used in this study omit financial aid and ethnicity variables in favor of ZIP-level and High School-level variables that are available to CSU Admissions during the application evaluation process.
 - This change in independent variable set produces statistically significant models with similar levels of estimated predictive capacity as previously studied CSU persistence models.
- High School GPA and ACT score represent the primary components in the CCHE Index. Throughout the models, we find that High School GPA tends to be a strong positive predictor of persistence, while ACT score has a less strong effect overall and tends to be a negative predictor of persistence in the smaller Index-limited models (i.e., in the Index ≤ 100 model where students with a higher HS GPA will necessarily have a lower ACT score).
- Freshmen carrying a remedial status of any type with a cohort STEM major tend to have very low odds of 3rd Fall persistence. Even among Index > 100 residents we observe severe declines (34% relative odds) in retention for students with both remedial and STEM status.
- In the absence of student financial aid data, Graduating High School ZIP Mean Household Income serves to explain a significant portion of variance in persistence for Colorado resident students.
- Home ZIP Population Density also predicts a portion of the variance in 3rd Fall persistence in the Index ≤ 95 , Index ≤ 100 , and Remedial = 1 models. The effects of population density vary by residency.
 - Low-preparation Colorado residents are more likely to persist to 3rd Fall if they are from a more densely populated area.
 - Low-preparation non-residents are more likely to persist to 3rd Fall if they live in a less densely populated area.
- Graduating High School's CDE Math + Reading Scale scores are a positive predictor of persistence in each Colorado resident model except for Index ≤ 95 . A high school's CDE scores appear to serve as a proxy for quality of education delivered, except perhaps in the case of the lowest Index students.
- Among resident Remedial students, carrying a STEM major is strongly related to failure to persist. For all Remedial students, High School GPA is strongly predictive of persistence.