

Report on Key Performance Indicators for Student Achievement at University of New Mexico

Yingchen Wang

Background and Questions With No Child Left Behind Act (2001), the American public paid more and more attention to the academic performance of adolescents at secondary and post-secondary educational institutions. **Race to the Top (RTTT) (2009)** signaled the further commitment from government and general public to improve education. The public, policy makers, teachers and parents all want to provide better education to the young generation. The focus of No Child Left Behind was the standardized assessment of students and accountabilities of teachers and schools. As time went on, public and researchers turn attention to student achievement at higher education institutions.

At the University of New Mexico (UNM), the historical six year graduation rate has been around 40%, from 43.34% in 2001 to 45.8% in 2006 (UNM Factbook). While there are still rooms for improvement, we see steady increases over years. At the same time, university executive personnel, faculty and parents became more and more interested in what help(s) students succeed. Across university, efforts were delivered to help students succeed. For instance, mentoring program was established.

With the current report, we tried to explore the factors for student graduation and investigate any group differences. Specific questions include.

- What factors or variables contributed to student graduation within 6 years?
- Were there any significant group differences with respect to student UNM cumulative GPA across years?

Data and Method

Sample The student population in this study comprised of first time freshmen admitted into the different programs of UNM in academic periods spanning from summer, 2005 up to spring, 2006. In total, there were 11480 students in this cohort. Out of these 11480, 11.55% graduated within four years. 25.96% graduated within five years. Additional 13.82% graduated within six years. All together the six year graduation rate is 42.95%.

Table 1. Graduation Table

	Graduation Count and Rate
4 year	1211(10.55%)
5 year	2666(25.96%)
6 year	1051(13.82%)
Total	4928(42.93%)

Out of 11480 students, there were more female (6372) than male (5108). Comparing with each gender (column wise), more female students (46.45%) graduated within six years than male students (38.53).

Table 2 Percentage of Graduation With Respect to Gender

Degree	Gender		
	F	M	Total
No	3412	3140	6552
	52.08%	47.92%	
	53.55%	61.47%	
Yes	2960	1968	4928
	60.06%	39.94%	
	46.45%	38.53%	
Total	6372	5108	11480

Note: the first percentage is row percent and the second percent is column percent for each result of cross tabulation.

Data Data were collected longitudinally. The dependent variable is graduation within six years (Yes versus No). The independent variables include

- Student registration (1 or 0) per semester;
- Student GPA (on a continuous scale) per semester;
- Gender
- Race (American Indian, Black, Hispanic, Others, Asian, White)
- First generation on a likert scale (0=both receiving college education, 1=at least one receiving college education, 2=at least one receiving high school or below, 3=both receiving high school or below).

Method Because the data were collected longitudinally, student semester GPAs were correlated across these repeated measures. The same correlated data structure existed with student registration. To handle these correlated structure,

1. Generalized estimating equation (GEE) approach (Liang and Zeger, 1986) was applied. GEE, as a statistical method, was developed to model longitudinal dataset.
2. Principal component analysis (PCA) was applied. PCA is a statistical method to reduce data redundancy such that medium to highly correlated variables are transformed into a set of linearly uncorrelated new variables called 'principal components'. Theoretically, the number of the principal components can equal to the number of variables. However, in real life, researchers often seek to use less number of interpretable components that maximize the explained proportion of variance. After principal component analyses were performed, logistic regression was carried with principal component scores and other independent variables.
3. Repeated measures were performed to investigate if there were group differences with respect to student GPA across years.

In the next section, we are going to discuss the results.

Analysis and Result

This section discussed the results from different analyses. First is the question on prediction analysis on factors contributing to student graduation within 6 years. Second is about group difference.

Prediction Analyses

Results from Generalized Estimating Equation Table 3 listed out the different GEE models based on the number of semesters of semester GPA and semester registration with gender. From the table, we can find:

- As we increased the number of semesters into the model, the magnitude of Chisquare statistic decreased with semester GPA, but with all significance levels at less than 0.0001. However, the opposite occurred with registration—its value increased in

general with all significance levels also at 0.0001. What does this mean? This possibly indicates that student registration might be an indicator for student motivation to finish higher education. As the number of registration increased from four to twelve, the predictive power of GPA was not that strong as it was in the beginning.

- As we increased the number of semesters into the model, the significance level of gender increased from $p < 0.0001$ to $p > 0.60$. This might mean that gender, as an effect for graduation, may not have a strong effect on six year graduation, in comparison with registration.
- As we increased the number of semesters into the model, the significance level of race decreased from $p < 0.0001$ to $p = 0.456$. This might mean that race, as an independent variable in the model, may not have a strong effect on six year graduation, in comparison with semester registration and GPA.
- As we increased the number of semesters into the model, the significance level of first generation decreased from $p = 0.0003$ to $p = 0.383$. This might mean that first generation, as an independent variable in the model, may not have a strong effect on six year graduation, in comparison with semester registration and GPA. However, its predictive power was significant for 4 year graduation rate as indicated by significance level less than 0.05 for the models with less than or equal to 8 semesters.
- Registration and semester GPA were only two significant variables in the five models. As we increased the number of semesters into the model, the prediction accuracy with success increased from 87.47% to 94.99% (with 8 semesters). Then the prediction accuracy tended to drop a little bit with 10 semesters in the model (91.22% for success). The precision jumped up to 97.14% for failure when 12 semesters were entered into the model. This finding is very important because prediction of success and prediction of failure are two different things. Quite often, researchers found it difficult to predict failure. Our results suggest that we might want to consider collecting longitudinal data to study student failure for not graduating within six year.

Table 3. Test Statistics from GEE and Prediction Accuracy

# Semesters	Sample Size	Semester GPA		Registration		Gender		Race		First Generation		Prediction Accuracy
		Chisquare	Pvalue	Chisquare	Pvalue	Chisquare	Pvalue	Chisquare	Pvalue	Chisquare	Pvalue	
4	6301	994.89	<.0001	27.53	<.0001	22.65	<.0001	44.96	<.0001	18.48	0.0003	87.47% (success out of 4500) 51.48% (failure out of 1801)
6	5538	670.22	<.0001	79.48	<.0001	21.21	<.0001	38.01	<.0001	18.87	0.0003	92.84% (success out of 4654) 69.8% (failure out of 884)
8	5113	642.33	<.0001	118.6	<.0001	28.49	<.0001	38.35	<.0001	7.87	0.049	94.99% (success out of 4425) 72.97% (failure out of 668)
10	3131	470.84	<.0001	165.67	<.0001	9.59	0.002	17.29	0.004	3.38	0.337	91.22% (success out of 2426) 75.6% (failure out of 705)
12	1537	166.9	<.0001	174.94	<.0001	0.01	0.914	4.68	0.456	3.05	0.383	97.14% (failure out of 1458). 60.76% (success out of 79)

Results from Logistic Regression Using PCA

Before we started elaborating on the results from logistic regression, it is necessary to explain that we performed principal component analysis (PCA) on the datasets using semester GPA plus UNM cumulative GPA. The number of semester GPAs in the model was indicated by the number of semesters in the first column of Table 4. Based on the results from principal component analyses, we used the first component scores from PCA as one independent variable in the logistic regression for prediction analysis. We refer these component scores as [general academic ability](#) hereafter.

Table 4 demonstrated some similarities with Table 3.

- As we increased the number of semesters into the model, the magnitude of Chi-square statistic decreased with general academic ability, but with all significance levels at less than 0.0001.
- However, the test statistics fluctuated with registration in general with all significance levels also at less than 0.0001 (there was a dramatic drop in test statistic of registration when the number of semesters was six). Again, this indicates that student registration might be an indicator for student motivation to finish higher education.
- As we increased the number of semesters into the model, the significance level of gender increased from $p < 0.0001$ to $p > 0.699$. This result is consistent with the result from GEE method. Gender, as an effect for graduation, may not have a strong effect on six year graduation, in comparison with registration and general academic ability.
- As we increased the number of semesters into the model, the significance level of race decreased from $p < 0.0001$ to $p = 0.836$. This result is consistent with the result from GEE method. Race, as an independent variable in the model, may not have a strong effect on six year graduation, in comparison with semester registration and general academic ability.
- As we increased the number of semesters into the model, the significance level of first generation decreased from $p = 0.044$ to $p = 0.72$. This might mean that first generation, as

an independent variable in the model, may not have a strong effect on six year graduation, in comparison with semester registration and GPA.

- It is interesting to point out that general academic ability, as an unobserved ability, had a strong predictive power on student graduation in the different models. Registration and general academic ability were only two significant variables in the five models. We detected similar trends with prediction results from logistic regression with PCA scores as in the results from GEE. As we increased the number of semesters into the model, the prediction accuracy with success increased from 87.7% to 93.88% (with 8 semesters). Then the prediction accuracy tended to drop a little bit with 10 semesters in the model (90.66% for success). The precision were 86.98% for failure and 79.3% for success when 12 semesters were entered into the model. In these five models with general academic ability scores, the prediction accuracies for failure were higher than those from GEE method. This might indicate that general academic ability, although unobserved, might be more helpful with predicting failures. More evidence is necessary with different datasets and from different approaches.

Table 4. Test Statistics from Logistic Regression with PCA and Prediction Accuracy

# Semesters	Sample Size	General Ability		Registration		Gender		Race		First Generation		Prediction Accuracy
		Chisquare	Pvalue	Chisquare	Pvalue	Chisquare	Pvalue	Chisquare	Pvalue	Chisquare	Pvalue	
4	6069	997.624	<.0001	208.348	<.0001	18.387	<.0001	33.014	<.0001	8.102	0.044	87.7% (success out of 4174) 60.22% (failure out of 1922)
6	5091	838.428	<.0001	31.865	<.0001	6.568	0.01	27.059	<.0001	5.934	0.115	90.14% (success out of 4041) 66.38% (failure out of 1050)
8	4496	514.69	<.0001	394.459	<.0001	2.224	0.136	26.982	<.0001	6.258	0.10	93.38% (success out of 3723) 70.25% (failure out of 773)
10	2393	218.899	<.0001	470.038	<.0001	1.378	0.240	10.870	0.054	0.892	0.827	90.66% (success out of 1760) 74.88% (failure out of 633)
12	826	66.996	<.0001	161.509	<.0001	0.150	0.699	2.092	0.836	1.339	0.720	86.98% (failure out of 395). 79.3% (success out of 431)

Group Differences

Repeated measures were performed to investigate if there were group differences in student UNM cumulative GPA across years with respect to gender, race and first generation.

Table 5 presented the test statistics from repeated measure with the associated *p*-values.

Several important findings were:

- With the number of semesters less than or equal to ten, the effects from gender, race and first generation were significant at less than 0.0001 level except for first generation for 10 semesters in the model.
- When the number of semesters were twelve, the effects from gender and first generation were no long significant at alpha level of 0.05. Race was significant at alpha level of 0.05 with *p*-value of 0.0126.

Table 5. Test Statistics from Repeated Measures.

# of Semesters	Effect	DF	F Value	Pr > F
4	Gender	1	66.78	<.0001
	Race	5	29.22	<.0001
	First Generation	3	28.43	<.0001
6	Gender	1	71.65	<.0001
	Race	5	21.01	<.0001
	First Generation	3	12.4	<.0001
8	Gender	1	65.26	<.0001
	Race	5	23.55	<.0001
	First Generation	3	10.7	<.0001
10	Gender	1	20.72	<.0001
	Race	5	7.88	<.0001
	First Generation	3	3.68	0.0116
12	Gender	1	1.26	0.2627
	Race	5	2.92	0.0126
	First Generation	3	0.56	0.6411

References

Liang and Zeger (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13-22.

U.S. House of Representatives (2001). *Text of No Child Left Behind*

University of New Mexico Factbook 2012-2013. Retrieved from

http://oia.unm.edu/documents/factbook_docs/2012fb_updated.pdf