THE RESEARCH AGENDA

Enrollment Projections: A Practitioner's Guide

By Kelly Perez-Vergara

Institutional staff such as enrollment managers, business officers, and institutional researchers are often asked to predict enrollments. Developing any predictive model can be intimidating, particularly when there is no textbook to follow. This paper provides a practical framework for generating enrollment projection options and for evaluating the accuracy of projections.

Enrollment Projection Theories

This paper communicates technical instructions for generating and evaluating the accuracy of enrollment projections, regardless of analytical methods selected. The context was the need for a small, private, non-profit business college to improve the accuracy of enrollment projections after almost a decade of decreasing enrollment. Given the institution's identity as an upper-division undergraduate institution, the undergraduate enrollment was primarily from community colleges; those institutions, too, had been experiencing enrollment decreases since the recession had ended. Graduate program enrollments were also waning in certain flagship areas such as the M.B.A. program. Yet the projections were overly optimistic; more accurate projections were needed for the annual budget cycle.

Enrollment projections are needed to plan for institutional budgets, including revenue and spending. Yet interpreting enrollment projections across an institution presents many challenges. The variables are the only factors accounted for in the model. If the projections show a decrease, they are called "pessimistic;" if they suggest flat enrollment, they are called "neutral;" and if the projections show increases in enrollment, they are called "optimistic." For example, if unemployment is a predictor in an enrollment model, then that factor is neutral because the rates in each year drive the estimates up or down. In this way, statistical models tend to be neutral unless the researcher selects a factor that only pushes enrollment estimates in a single direction that does not follow trends.

Similarly, once the actual term enrollment is known and compared to the projection, the model is perceived as a failure if enrollment is lower than the projection. If the enrollment matches projections, the model is perceived a success, and if the enrollment exceeds projections, then staff are credited for their effective initiatives. Of course, none of these is a fair interpretation of enrollment projections or their success. Rather, enrollment projections take specific factors and generate an estimate, but unaccounted factors—positive and negative—also impact enrollments. Evaluating enrollment projection accuracy by comparing it to actual enrollment in the projected years is not the best approach. Therefore, selecting the factors in enrollment and properly evaluating projections are critical to advancing an institution's perception and use of enrollment projections.

The best way to select factors is to start with a guiding theory. Bronfenbrenner's (1986) ecological systems theory is a useful framework for thinking about enrollment in higher education (Perez-Vergara 2019). According to this theory, a potential student is surrounded by influences ranging from the immediate network of family and close friends (microsystem) to the extended network (mesosystem), society (exosystem), cultural ideology (macrosystem), and generational impacts (chronosystem). Practically speaking, ecological systems theory considers realities such as how high school seniors are often told that college is the best pathway to future success.

Ecological systems theory also invites consideration of current generations of students and how they matriculate and suggests that how Generation Z experiences the world relates to information about and resources for college, which come from all parts of the ecological system. Nationally representative studies describe Generation Z in terms of its members' appetite for higher education, how they like to learn, what they feel about debt and cost, and more.

While it is important for researchers to review studies of Generation Z students, it is not practical for institutions to create enrollment projection models based on consumer sentiments. Moreover, ample research, regardless of the generation being studied, shows that labor market trends have an impact on college enrollments (see Perez-Vergara 2019). And while the magnitude of impact will change over time, it is unlikely that the use of labor market data in enrollment projections for Generation Z needs to be completely reconsidered.

As Perez-Vergara (2019) mentions, both human capital and generational trend theories fall within ecological systems theory. In other words, potential students' knowledge about college derives from the various systems outlined by Bronfenbrenner's theory. In this way, so long as no major societal, cultural, or generational changes occur, it is not unreasonable to look to enrollment pipeline data such as local high school enrollments (for incoming freshmen) or feeder school enrollments or completions (for transfer or graduate populations). Similarly, pipeline data make logical sense for predicting continuing student enrollments, even though the pipeline in this case is internal.

Enrollment Projection Methods

There is robust literature that reviews methods for projecting enrollment, and several recent articles have compared methods. Because the focus of this paper is on a framework to use for comparing enrollment projections and not for producing specific projections, various methods are reviewed only briefly. Allen (2013) compared Markov models, cohort flow models, grade progression ratio models, and trend analysis. Langston, Wyant, and Scheid (2016) described logistic and linear regression, big data analytics, curve fitting, and trend analysis approaches. Wiorkowski, Valcik, and Perry (2017) looked at the significance of enrollment changes (t-tests) with time series regressions and log linear regressions. Nakhkob and Khademi (2016) utilized neural networks and data mining approaches. Maldonado, Armelini, and Guevara (2017) as well as Saini and Jain (2013) used classification techniques. Tang and Yin (2012) used grey prediction. Wang, Wang, Guo, and Feng (2014) and Zang, Wang, and Wang (2015) used fuzzy time series. Zan, Yoon, Khasawneh, and Srihari (2013) used a ratio-based approach. The variety of techniques is staggering, so how do researchers approach enrollment projection modeling?

Framework for Enrollment Projections

While the literature describes different theories of enrollment as well as different methods for analyzing enrollment data to project enrollment, there was a lack in frameworks for how to approach this project. Therefore, this paper provides a step-by-step guide for staff who need to create accurate enrollment projections for their institutions. Whatever theories or methods are used, this framework facilitates comparison and presentation of the models and describes a way to select an accurate model from available options.

Step 1. Define Populations

The first step is to select the groups for which enrollment will be projected. Students among new and continuing categories are desirable at minimum because members of these groups respond to offers of enrollment differently. Mixing student segments would result in a less accurate projection. Undergraduate and graduate levels are also important categories given that undergraduate and graduate students respond differently to offers of enrollment. The reasons for the differences are clarified in the next step as theories of enrollment are specified.

Other populations that may need stand-alone projections include new and/or specialty programs that are affected differently by internal and external factors.

Example populations:

- New undergraduate
- Transfer undergraduate
- Continuing undergraduate
- New graduate
- Continuing graduate

Step 2. Define Output Data

The second step in projecting enrollment is to define the outputs the models will predict. Based on populations identified in step one, define what feature of the population will be predicted. Typically, models attempt to predict headcounts or credit hours. Recognize that while credit hour predictions are necessary for budget planning, headcount predictions are easier to align with theory. (In other words, the reasons students enroll differ from the reasons they take one or two classes.) In the example given, headcount projections would be used in a trend analysis to generate credit hour estimates.

After deciding upon headcounts or credit hours as an output, the timeframe needs to be selected. In the current example, the focus is on fall term projections, filling in the other term (*e.g.*, winter, summer) headcounts with a basic trend analysis once a model is selected for the projection group.

In addition to using headcount or credit hours as an output, it may be relevant (depending on the approach taken) to use the output of enrolled/not enrolled. In this scenario, the input data are not in summary form by timeframe but instead are student-level data for a specified timeframe.

Following is a list of possible outputs. It is recommended that researchers select either term-based or annual-based outputs. It is also recommended that researchers select an output that aligns with the funding of the institution. For example, if students pay flat-rate tuition, headcount might be the most relevant output whereas another institution may receive funding based on full-time equivalent students (FTEs). Models attempting to project retention could consider the last item below, using logistic regression.

Example outputs:

- Fall headcount
- Winter headcount
- Summer headcount
- Academic year headcount
- Fall credit hours
- Winter credit hours
- Summer credit hours
- Academic year credit hours
- Fall enrolled/not enrolled

Step 3. Define Theory

The third step in projecting enrollment is to determine an overarching list of major issues that affect the majority of students in each population selected for the particular outcome selected. For example, new undergraduate students may be more sensitive to entry-level wages regionally, a factor motivating them to seek higher education, compared to graduate students who have completed a bachelor's degree and may be more likely to be employed and earning a living wage. A review of the literature, as well as institutional context, should result in selection of a few key factors that influence enrollment (*i.e.* working theories). While it is ideal for individuals to carefully chart the majority of factors that impact enrollment, it is not practical to chart or collect data on all factors with the accuracy needed to add them to a model. Therefore, to practically create a model, enrollment theories thought to impact the majority of students provide the most reasonable avenue for investigation.

Example working theories:

- New undergraduate: K–12 pipelines drive new undergraduate enrollments
- Transfer undergraduate: community college enrollments drive transfer enrollments
- Continuing undergraduate: likely to remain consistent with prior year trends
- New graduate: feeder school bachelor's degree completions and labor market trends drive new graduate enrollments
- Continuing graduate: likely to remain consistent with prior year trends

Step 4. Define Input Data

The populations, working theories, and output data that are selected drive the input data used in the models. Ideally, each model will utilize one input variable to test the effect of that variable alone.

Example inputs:

- New undergraduate: K–12 pipelines
- Transfer undergraduate: community college enrollments
- Continuing undergraduate: prior year trends
- New graduate: feeder school bachelor's degree completions and unemployment rate
- Continuing graduate: prior year trends

Step 5. Select Projection Analyses

The following list of examples shows how each set of populations with its unique inputs and outputs could be studied using multiple approaches. For example, community college enrollments in a trend analysis are weighted by market share and then multiplied by an average capture rate to generate an estimate for the upcoming year, but these same data can also be used in a linear regression to generate an estimate. It would be ideal to test both versions to determine which set of estimates is most accurate over time.

Example analyses:

- Trend analysis
  - → Prior three-year average of subgroups (new, continuing, drop outs)
  - → Prior three-year average percent change in enrollment times prior year enrollment
  - → Community college enrollments times capture rate trend
- Linear regressions
  - → Unemployment rate
  - → Community college enrollments weighted by market share of feeder
  - → Number of recent bachelor's degree graduates weighted by market share of feeder

Step 5. Create Datasets

The fifth step in the process is to create datasets for each population that will be tested. The example dataset shown in Table 1 is for transfer students, fall headcount, using the community college pipeline input data. In this step, alternative data sources, cleaning, weighting, and lagging can occur. Be sure to keep notes about how data timeframes align when lagging is used (that is, the upcoming term may be predicted by a prior-year pipeline). If the input variables are not available in projected form, prior year values should be reviewed to see if they

TABLE 1 ➤ Transfer Undergraduate, Fall Headcount

Term	Transfer Headcount	Community College, Prior Year		
		Enrollments	Completions	Completions (Weighted)
Fall 2014	80	5,000	1,000	400
Fall 2015	90	5,500	550	450
Fall 2016	100	5,000	500	500
Fall 2017	110	5,500	550	550
Fall 2018	120	5,000	1,000	600
Fall 2018	Prediction	5,500	700	650

provide logical predictors of next year enrollments (that is, a trend analysis estimate within the larger projection).

Step 6. Run Models

Step six involves running the models using the selected data. For trend analyses, write the physical formulas used to compute the trends, and apply the formula to the dataset to generate the estimates for the upcoming as well as for the previous terms. In other words, create an estimate for the upcoming terms, and determine what the estimate would have been using this model in the previous years. Correlations may also be computed between the output variable and the trend rates that are applied to the input variable, and the significance of this relationship may be tested.

For linear regressions, run the data in the statistical software package, and write out the formula generated by the model. Note any correlations or significance. Then, compute the upcoming term estimate as well as prior year estimates using the regression formula.

Example formulas:

- *Trend Analysis—Three-Year Average Fall Percent Change*: The three-year average percent change from the same term in the prior year (fall 2018 uses fall 2017, fall 2016, and fall 2015 changes) (*see* Equation 1).
- *Linear Regression—Weighted Feeder School Enrollment, Lagged Two Years*: The line of best fit between (a) primary community college feeder school headcounts weighted by market share and (b) transfer headcount (*see* Equation 2).

TABLE 2 ➤ Calculating Estimates From Models, Trend Analysis Example

Fall Term	Input (fall head count)	Trend Analysis (3-year average change [%])	Output (estimated fall head count)
2016	102	-0.5	103
2017	108	0.1	102
2018	104	0.1	108
2019			104

TABLE 3 ➤ Calculating Estimates From Models, Linear Regression Example

Fall Term	Input (prior year fall headcount)	Calculation[1]	Output (estimated fall headcount)
2016	104	$(0.03 \times 104)+98$	101
2017	102	$(0.03 \times 102)+98$	102
2018	108	$(0.03 \times 108)+98$	101
2019	104		101

[1] (parameter estimate × input from prior year) + intercept

Example estimates using trend analysis and linear regression are shown in Tables 2 and 3.

Step 7. Evaluate Accuracy

Evaluating the accuracy of the model involves several steps. First, review correlations to ensure that there is a relationship between the variables used in the models. It is generally accepted that correlations greater than 0.5 are considered large (Aron, Coups and Aron 2013), but other thresholds could be selected by the researcher based on the practical meaning of the relationship of the variables. In this study, a correlation of 0.8 or higher

EQUATION 1.

$$\text{Fall 18 Transfer Head Count} = \left(A_{17} \times \left(\frac{\sum\left(\left(\frac{A_{17}-A_{16}}{A_{17}}\right)+\left(\frac{A_{16}-A_{15}}{A_{16}}\right)+\left(\frac{A_{15}-A_{14}}{A_{15}}\right)\right)}{3} \right) \right) + A_{17}$$

Where A_n = Cont FA_n HC

EQUATION 2.

$$\text{Fall 18 Transfer Head Count} = (\mathbf{B} \times \mathbf{X}) + \text{Intercept}$$

Where $\mathbf{B}$ = parameter estimate and $\mathbf{X}$ = Value associated with upcoming year

TABLE 4 ➤ Comparison of Model Estimates to Actual Enrollment With Accuracy Results

Term	Actual	Population, Timeframe, Adjustments		Population, Timeframe, Adjustments		Population, Timeframe, Adjustments		Population, Timeframe, Adjustments		Population, Timeframe, Adjustments	
		Est.	Error	Est.	Error	Est.	Error	Est.	Error	Est.	Error
Fall 2016											
Fall 2017											
Fall 2018											
Fall 2019[a]											
Correlation											
p-value											
$X Error		/3		/3		/3		/3		/3	

[a] Upcoming term

was considered large. It is also generally accepted that regression analyses with p-values of less than 0.05 are considered statistically significant (Aron, Coups, and Aron 2013).

Even if a model shows high correlations and statistical significance, the estimates the model produces and the error of those estimates year over year need to be considered from a budgetary standpoint. In particular, attention should be paid to debt covenant requirements as well as requirements from accrediting bodies, such as the Higher Learning Commission's Composite Financial Index. Ultimately, officers should determine a total dollar amount of error in the budget to guide enrollment projection evaluations. Because each population may differ in size, it is important to determine how the dollar amount relates to credit hour and student headcount error thresholds for each group.

Example error thresholds:

- $250,000 (or less) total annual error goal
- Total divided by percent of enrollment attributed to each population (*e.g.*, 20 percent of students are new undergraduates = $50,000 maximum error for new undergraduate annual credit hour estimates)
- Divided by percent of enrollment attributed to each timeframe (*e.g.*, 29 percent of enrollments occur in the fall term = $14,500)
- Divided by cost per credit hour (*e.g.*, $475/credit for new undergraduates = 30.5 credit hour error)
- Divided by average credit hour load of new undergraduate students (*e.g.*, 6.42 average load in fall = 5 student error)

Step 8. Present Results

In Table 4, the actual enrollment for the timeframe is presented. Compare each model for the population. The current approach should be included to ensure that the new models outperform the current models.

For each column in Table 4, fill in the model details, including population, timeframe, and adjustments made (such as lagging or weighting). Then, fill in the estimate generated by the model as well as the error (difference between estimate and actual). The correlations and p-values, as available, can also be added in the final summary rows. In the last summary row, when the absolute value of the difference (*i.e.*, ignoring whether the result is positive or negative) is less than the error threshold, the estimate is considered practically significant. The final row summarizes the number of times the estimate was practically significant during the years attempted. At our institution, for the model to be considered to have practical significance for projecting enrollment, we looked for models where the majority of estimates were successful.

Step 9. Compute Additional Estimates

This example describes a method to compute fall headcounts for each population (new/continuing, graduate/undergraduate). Yet enrollment projections are likely

TABLE 5 ➤ Summary of Selected Estimates for Enrollment Projections

Term	Headcounts				Credit Hours			
	Undergraduate		Graduate		Undergraduate		Graduate	
	New	Continuing	New	Continuing	New	Continuing	New	Continuing
20xx/Fall	X	X	X	X				
20xx/Winter								
20xx/Summer								

needed for credit hours as well as for all terms. Two approaches could be used: First, repeat the framework approach presented above for each term for headcount and again for each term for credit hours. Because four groups were tested each using four models, the 32 total groups at this institution needing four models each results in just over 100 models that needed to be tested. Given time constraints, it was desirable to use trend analysis to fill in the estimates for the remaining terms and outputs (*i.e.* credit hours). Table 5 shows an X in the cells in which steps one through eight (above) were used to generate estimates. Then, additional trend analyses were used to fill in the remaining blanks. The accuracy of this approach was then tested and summarized, as described in step seven.

Conclusions

Enrollment projections are complex. As Perez-Vergara (2019) shows in the review of some 200 articles, a few dozen theories are used to predict enrollments. Under these theories, numerous methods are used, from trend analysis to regressions to data mining approaches. The input variables also vary, from institutional factors like cost and ranking to regional factors like unemployment and entry-level wages, to personal factors about students, such as their test scores and parent income. That said, institutions need a practical way to create and compare enrollment projections each year. This paper provides a framework for comparing enrollment projections in light of theory and alternative methods using available data and simple approaches.

The framework outlined in this paper was utilized by a small, private non-profit business college to compare enrollment projection approaches and accuracy. Using the best performing model for each population, a projection for the upcoming year was selected. Finally, as teams met to review estimates with marketing and other units, it was relevant to discuss qualitative factors and other important trends; officers responsible for budgeting accept or adjust the offered projections.

Limitations

More advanced data and methods can be used to project enrollment. Ideally, researchers use the best data and methods to calculate broader higher education enrollment projections. For individual institutions, it is more likely that many staff will need to use available data and methods to generate reasonable enrollment estimates.

While cleaning output data, student enrollments related to deactivated programs may need to be removed from datasets. New program enrollments may also need to be reconsidered as unnatural shifts in enrollment may skew results. In other words, data may need to be smoothed to make models more accurate.

Finally, testing models can be accomplished using more sophisticated techniques than those outlined here. For example, when building a model, it is typically preferable to use more rather than fewer data points. But because of different impacts over time, this paper focuses on using only the most recent few years of data to predict enrollments. The ideal way to test model performance is to save the most recent data to use as test data but not to use them in building the model. Ultimately, the practical significance of these models lies in their meeting error thresholds for the budget, regardless of correlations or p-values.

Oddly, a test of a model's accuracy is not necessarily how close the projection was to actual enrollment. For example, a projection model may have been accurate in accounting for several forces shaping enrollment, but an internal recruitment strategy added the last two weeks of the enrollment period may have performed better than expected, causing enrollment to increase markedly such that enrollment projections based on all other unaccounted factors are still affecting enrollments.

References

Allen, D. M. 2013. Multi-purpose enrollment projections: A comparative analysis of four approaches. Unpublished dissertation. Orono: The University of Maine.

Aron, A., E. J. Coups, and E. N. Aron. 2013. *Statistics for Psychology.* Boston: Pearson.

Langston, R., R. Wyant, and J. Scheid. 2016. Strategic enrollment management for chief enrollment officers: Practical use of statistical and mathematical data in forecasting first year and transfer college enrollment. *Strategic Enrollment Management Quarterly*. 4(2): 74–89.

Nakhkob, B., and M. Khademi. 2016. Predicted increase enrollment in higher education using neural networks and data mining techniques. *Journal of Advances in Computer Research*. 7(4): 125–140.

Maldonado, S., G. Armelini, and C. A. Guevara. 2017. Assessing university enrollment and admission efforts via hierarchical classification and feature selection. *Intelligent Data Analysis*. 21(4): 945–962.

Perez-Vergara, K. 2019. Higher education enrollment theories: Setting context for enrollment projections. *Strategic Enrollment Management Quarterly.*

Saini, P., and A. K. Jain. 2013. Prediction using classification technique for the students' enrollment process in higher educational institutions. *International Journal of Computer Applications*. 84(14): 37–41.

Tang, H. W. V., and M. S. Yin. 2012. Forecasting performance of grey prediction for education expenditure and school enrollment. *Economics of Education Review*. 31(4): 452–462.

Wang, H., H. Wang, J. Guo, and J. Feng. 2014. A fuzzy time series forecasting model based on yearly difference of the student enrollment number. In *2nd International Conference on Soft Computing in Information Communication Technology*, edited by T. Chang and A. Hunter. Paris: Atlantis Press.

Wiorkowski, J., N. A. Valcik, and C. A. Perry. 2017. Comparison of two different enrollment projection models. In *Institutional Research Initiatives in Higher Education*, edited by N. A. Valcik and J. A. Johnson. New York: Routledge.

Zan, X. A., S. W. Yoon, M. Khasawneh, and K. Srihari. 2013. A comparison study of return ratio-based academic enrollment forecasting models. *The AIR Professional File*. Spring(Article 129): 1–12.

Zhang, K., Z. Li, H. F. Wang, and H. X. Wang. 2015. Fuzzy time series prediction model and application based on fuzzy inverse. *International Journal of Signal Processing, Image Processing and Pattern Recognition*. 8(10): 12–28.

About the Author

Kelly Perez-Vergara

Kelly Perez-Vergara is the Assistant Vice President of Institutional Research at Walsh College, an institution serving upper-division undergraduate and graduate students focusing on business and adjacent fields. Prior to Walsh, Kelly served as the Associate Executive Director of Institutional Effectiveness at Oakland Community College. Perez-Vergara is a current Ph.D. student in Higher Education Leadership at Oakland University, and she is an alumna of the M.S. in Survey Methodology program at the University of Michigan. Perez-Vergara serves in the higher education community as an editor for the Community College Journal of Research and Practice, and she recently completed training as a Higher Learning Commission peer reviewer for accreditation.