Digging Deeper into the Data: The Role of Gateway Courses in Online Student Retention

William Bloemer, Karen Swan, Scott Day, and Leonard Bogle
University of Illinois Springfield

Abstract
Improvement in undergraduate retention and progression is a priority at many US postsecondary institutions. A number of institutions address this issue by identifying gateway courses (foundational courses in which a large number of students fail or withdraw) and concentrating on "fixing" them. This paper argues that may not be the best use of limited resources. No matter what we do, there will always be courses with too many students receiving grades of D, F and W simply because of the nature of their content and the preparation of the students who must take them. Our research suggests that student type and academic stage affect student success and that gateway courses (courses which block student progression) can be found at all undergraduate levels. Specifically, we have found that one can use binary logistic regression with student type, academic stage, cumulative GPA, and prior withdrawals as predictor variables to predict success in undergraduate courses at our institution. Moreover, relating predictions to observed DFW rates can highlight courses exceeding expectations, as well as those which fail to meet them, to support a more nuanced understanding of where and what attention is needed. In this paper, we describe our "gap analysis" procedure and illustrate the utility of such approach by examining issues surrounding success in online courses at our institution.

Keywords: gateway courses, retention, student success, learning analytics

Bloemer, W., Swan, K., Day, S., & Bogle, L. (2018). Digging deeper into the data: The role of gateway courses in online student retention. *Online Learning, 22*(4), 109-127. doi:10.24059/olj.v22i4.1515

Digging Deeper into the Data: The Role of Gateway Courses in Online Student Retention

In recent years, the success of undergraduate post-secondary students has become a major issue in the United States. One promising step in addressing this issue has been identifying gateway courses, entry level courses which are not passed by a large number of students, impeding their progression to degree (Koch & Pistilli, 2015; Gardner Institute, 2017). Our work takes the notion of gateway courses one step farther to consider courses at all academic levels and the potential of the students enrolled in them. Specifically, we have developed a methodology which uses the characteristics of individuals enrolled to predict the numbers of students who will receive grades of D, F, or W in individual courses. We then explore the gaps between the predicted DFW rates and the actual ones in each course. DFW rates which are lower than predicted indicate courses in which students are exceeding expectations; DFW rates which are higher than expected indicate courses in which students are failing to meet expectations. This approach allows us to isolate courses for further (qualitative) investigation with the goal of identifying issues impeding students' progression to degree.

In the sections which follow, we provide a context for our work and describe the gap analysis approach we have developed. We then use worked examples from our institution to explore and explain the general application of gap analysis, and its specific use in investigating the efficacy of online courses. The final sections of the paper discuss the implications of our work and revisit our conclusions.

Review of Related Literature

Increasing the numbers of young Americans achieving a postsecondary degree has been a national priority for over a quarter of a century (Arnold, 1999; Shapiro, Dundar, Yuan, Harrell & Wakhungu, 2014), with little improvement seen. Indeed, a myriad of studies in the last decades of the 20th century tested the assumptions of theories concerning the reasons why students drop out of higher education institutions (Bean & Metzner, 1985; Mallette & Cabrera, 1991; Munro, 1981; Tinto, 1987) to develop models of student progression. More recently, educators have built on these models to address issues of working students (Falcone, 2011) and online learners Vignare, Wagner, & Swan, 2017).

Likewise, there is a substantial body of literature that has examined determinants of course non-completion (Juhong & Maloney, 2006; Ishitani, 2006; Jia, 2014; Montmarquette, Mahseredjian, & Houle, 2001; Wetzel, O'Toole, & Peterson, 1999), especially as regards online learning (Boston et al., 2009; Clay, Rowland, & Packard, 2008; Morris, Wu, & Finnegan, 2005; Rovai, 2003). Most recently, predictive analytics are being applied to help online educators in particular address undergraduate attrition (Baepler & Murdoch, 2010; Barber & Sharkey, 2012: Campbell & Oblinger, 2007). Predictive analytics applies analytic and statistical analyses of both historical and current data to forecast behavior and trends by creating predictive models that place a numerical value on the likelihood of a particular event happening. Most of these efforts have focused on the analysis of student data from learning management systems, software applications for the administration and delivery of educational courses and training programs. Institutional data represents another source of data that can be mined to inform student success initiatives.

An emerging strategy for enhancing postsecondary outcomes is to measure the patterns by which students reach and move through intermediate stages of degree completion. One of the issues identified as contributing to attrition is poor performance in gateway courses (Koch & Pistilli, 2015). The Gardner Institute (2017) has identified "gateway courses" as foundational, credit-bearing, lower division courses, for which large numbers of students are at risk of failure and which accordingly stand as "gatekeepers" to further study and degree completion. Indeed, researchers have found that retention in these courses is strongly correlated with successful degree completion (Cabrera, Burkum, & La Nasa, 2005; Herzog, 2005; Lewis & Terry, 2016; Moore & Shulock, 2009; Offenstein & Shulock, 2010). Koch and Pisitilli (2015) add that "courses with high rates of unsuccessful outcomes [as identified by DFW rates] 'kill' a student's grade point average (GPA), motivation, and academic progress" (p. 3).

The problem of gateway courses is especially challenging in online environments, and online educators are attempting to address the issue, primarily through course redesign (Education Advisory Board, 2016). At DePaul University, for example, course redesign focuses on approaches that will help students learn more effectively, by looking systematically at the structure of the courses themselves. Rather than focusing solely on the content of particular courses, DePaul faculty involved in course redesign explored pedagogical approaches that might help students learn more effectively, including the use of technology to provide students with greater opportunities to practice course concepts. Other strategies for improving success in gateway courses include providing extra support for faculty teaching such courses, in the form of professional development and instructional design support (Nogaj & Kans, 2017), and/or peer support for students taking the form of peer tutoring and student-led orientations (Arendale, 2004).

Traditional approaches to identifying gateway courses sort courses by DFW rates and then target those courses with the highest percentages of DFWs for intervention. The Gardner Institute (2017), for example, suggests targeting all courses with DFW rates of 30% or more, and hundreds of postsecondary institutions are using this criterion or one like it. Such approaches, however, assume that all gateway courses have the same impact on student success. Our investigations indicate that this is not always the case. In this paper we describe the gap analysis procedure we have developed to distinguish between courses that are really impeding student success and those that are not. Such distinction is important so that institutional resources available for improving courses, student support, advisement, and placement practices can focus on those areas in which the problems are most critical and identifying non-existent problems and applying misplaced or even damaging “solutions” can be avoided.

Methods

A more careful approach to identifying gateway courses than simply finding courses with the highest DFW rates would recognize that some courses are simply harder than others for a majority of students. There is no single standard for DFW rates that is universally applicable within an institution. For example, the DFW rates for upper-division classes tend to be lower than lower-division DFW rates. There are also significant differences in students that need to be considered when looking at the effectiveness of courses. Students enter with widely different backgrounds and learning goals. Individual students change significantly over the course of their studies. The effectiveness of a course needs to be judged in the context of the students it serves. Simply put, it is not reasonable to expect all courses to serve all students equally well. Efforts to do so are futile and may actually do harm. In this section, the methodology used to explore gateway courses’ effects on retention and progression in a more nuanced way is described.

The premise of this work is that taking into account the academic characteristics of students in a given course will help when interpreting the numbers of students who receive grades of D, F, or Win it. Towards that end, historical grade data is used to develop a model that predicts the probability that a particular student will get a D, F, or W grade in an average course, based on what is known about that student’s academic history.

Whether or not a student will get a grade of D, F, or W in a course is a binary outcome. Binary Logistic Regression (Hosmer, Lemeshow, & Sturdivant, 2013) is a common method for using predictor variables to make such an assessment. What can be predicted is the probability or odds of a particular outcome, p. For purposes of the regression, that probability is converted to the log of an odds ratio, logit(p). That transformed function is then predicted from a linear combination of predictor variables according to the equation:

$$\text{logit}(p) = \square_0 + \square_1 x_1 + \square_2 x_2 + \ldots\ldots + \square_n x_n$$

Where logit(p) = log (p/(1-p)), x_n are the predictor variables, and $\square_n$ are the regression coefficients.

For these purposes, and because the focus of our research is to account for the academic abilities and preparation students bring to the courses they take, categorical predictor variables were used to describe different types of students at different points in their academic life cycles. For student type, the traditional distinction between Native freshmen and Transfer students was the first level of classification. Within those categories, Native Freshmen were further classified according to whether they were in the Honors program or not. Transfer students were further classified according to whether or not they were enrolled in programs that were being delivered completely online. The four Student Types, then, were Native, Honors, Transfer, and Online Transfer.

The National Student Clearinghouse (2018) collects and reports data on the progress and success of students through the first six years of their degree pursuit. Their annual snapshots can be used as a starting point for defining significant stages in the Student Life Cycle. For full-time, first-time Freshman, the first two years are significant as the transitions from first year to second year and from second year to third year have the greatest attrition. Additionally, the greatest attrition for online transfer students occurs between the first and second semesters (Bloemer, Day, & Swan, 2017). Since those students are an important part of our student population, we have classified separately the student's two semesters in their first year. The stages used here for student life cycle were: First Semester, Second Semester, Second Year, Third Year and Later. Different definitions of these categorical variables may serve to make this model more appropriate for other institutions.

Beyond these categorical variables, major predictors of individual student performance can be found in students' previous academic records (Sweeney, Rangwala, Lester & John, 2016; Jayaprakash, Moody, Eitel, Regan, & Baron, 2014; Bloemer, et al., 2017). At our institution, prior cumulative GPA is the most important of these. Adding the fraction of previous courses in which the student received a D, F, or W grade improved the model enough to merit inclusion. High W rates will not show up in the student's prior GPA, but are clearly a factor in predicting future D, F, or W grades.

Subjects and Setting

The data from which our observations are drawn included all undergraduate degree-seeking students enrolled in a small, Midwestern public university over a three-year period from Fall 2014 to Spring 2017. The unit of analysis was the course, and all courses taken by undergraduate degree seeking students with total registrations of 20 or more across the entire period were included, resulting in 549 courses; 270 of which were taught exclusively on-ground, 119 of which were taught exclusively online, with 160 courses taught in both formats. Courses examined came from 39 different subject areas as indicated by course prefixes. Twenty-seven percent were lower division courses.

The academic calendar at our institution is semester-based. Course grades were those awarded at the end of each term, not the final transcript grade which is occasionally different for any number of reasons, such as incompletes completed, or grades changed after courses are retaken. DFW rates, the percentage of students receiving grades of D, F, or Ws, were used to measure students' success in the courses studied. An end of term grade of D or F or prior Withdrawal indicates that the student failed to complete the course successfully.

Analyses

A binary logistic regression with D, F, or W grade (yes/no) as the dependent variable was used to create a model predicting the probability that a particular student would get a D, F, or W in the average course he or she takes. The predictor variables were the student's prior cumulative GPA, the student's prior cumulative DFW rate, Student Type, and the current point the Student's academic Life Cycle as described above. The relative importance for each predictor variable in shown in Figure 1. as estimated by the effect of their individual removal from the complete model.

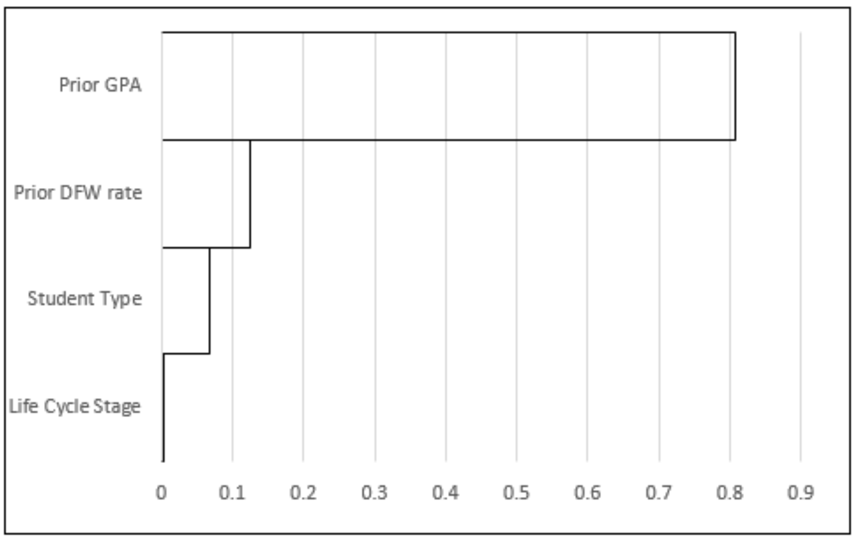


Figure 1. Relative strength of predictor variables

The student's prior GPA has the greatest effect on the predicted DFW rates, followed by the prior DFW rate. Each of the predictor variables is statically significant at the 0.05 level, with the exception of the Life Cycle Stage. That term was kept in the model as one category within it, First Semester, was significant.

SPSS was used to perform the binary logistic regression described here. The predicted probabilities of receiving D, F, or W grades for each record in the dataset were saved when the regression was calculated. Those predictions permit the creation of a classifier, which turns the predicted probabilities into a binary prediction. Predicted probabilities greater than a cut point, typically 0.5, are classified as predicting a DFW grade. Lesser probabilities would predict the grade would not be DFW.

Comparing the predicted categorization of cases to the observed values is often done to judge the effectiveness of the model. We provide one such comparison here: the ROC (Receiver Operating Characteristic) chart. It relates the true positive rates to false positive rates for the entire range of possible cutoff points. The area under the ROC curve measures the model's ability to discriminate between the binary outcomes. The area ranges from 0.5 (no discrimination) to 1 (perfect discrimination). The ROC curve for this regression is shown in Figure 2.

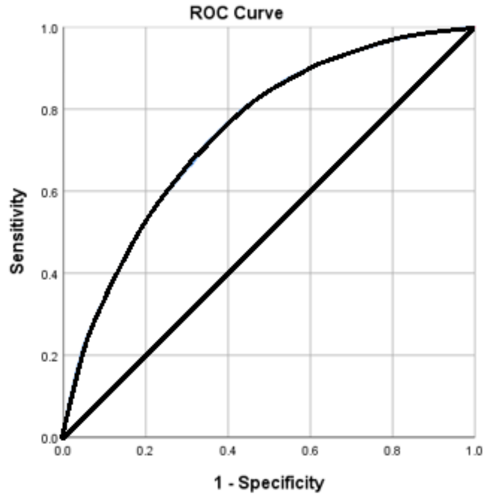


Figure 2. ROC curve for regression model.

The area under the curve for this regression is 0.748, which places it squarely in the region of acceptable discrimination (Hosmer et al., 2013).

From our previous work (Bloemer, et al., 2017), we know that the discrimination ability of the regression for predictions for individual students can be improved by including categorical variables for at least some courses to reflect differences between courses. That step would be inappropriate here as the purpose of this regression equation is to control for student characteristics so that those very differences between courses can be examined more accurately. On the basis of these results, the model is judged to be adequate for that purpose.

In order to set a reference level for a particular course, then, these individual predicted DFW probabilities are averaged for the students who were actually registered in the course. Since the goal here is to prioritize courses for further investigation, courses were ranked by two measures: the actual DFW rates, and the average predicted DFW rates for the students registered in the course. The difference between the predicted and actual DFW rates in a course establishes the size and direction of the gap in it. A positive gap indicates DFW rates that are greater than predicted; a negative gap indicates DFW rates that are less than predicted.

Results

Gap Analyses: Worked Examples

"Gap analysis" is the name we have applied to the procedure we have developed for identifying courses that may impede student success. Application of the 30% criterion, or any absolute criterion for identifying gateway courses (Gardner Institute, 2017) needs only the observed DFW rates. The assertion here is that the gap between actual and predicted DFW rates adds considerable information to the search for such courses and is a better measure for the identification of courses that should be further investigated. In this section, we provide worked examples from our institution to support that assertion.

In our earlier work (Bloemer, Day, & Swan, 2017), the difference between the actual DFW rate and the predicted DFW was used to produce a quantitative measure of course effectiveness.

This approach is supplemented here with a graphical depiction. To show the merit of comparing the observed DFW rates to the predicted ones, a scatterplot was created, representing the predicted and observed DFW rankings for each course (Figure 3). For the purposes of the graph, courses with the lowest DFW rates were placed nearest the origin. If consideration of the predicted DFW rates were of no value, the predicted and observed rankings for the courses would be identical and the scatterplot would be a straight line through the chart at a 45^{o} angle.

It is clear from a glance at Figure 3 that observed course DFW rankings often differ from the rankings that would be predicted based on the characteristics of the students involved.

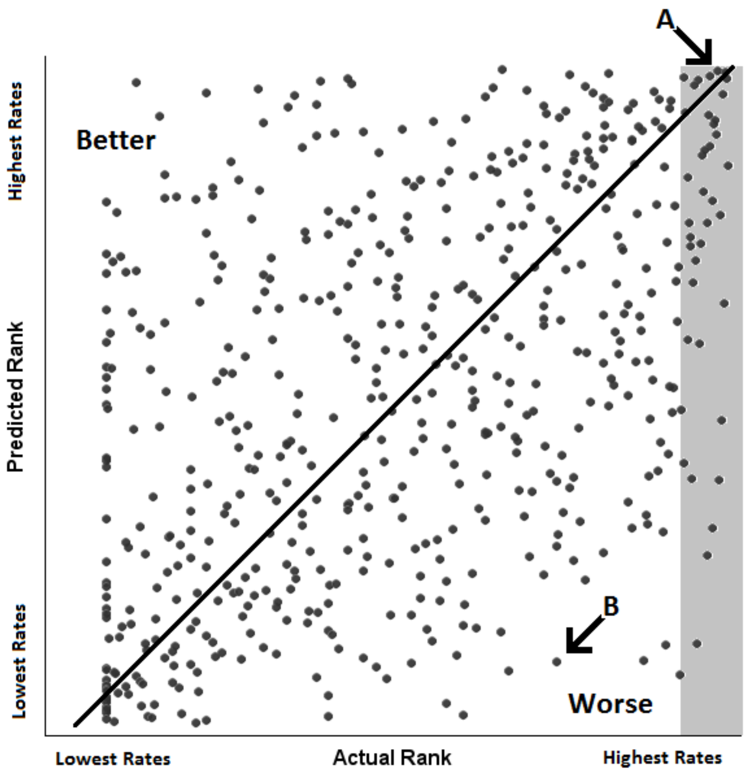


Figure 3. Predicted course DFW rankings vs. actual course DFW rankings.

The diagonal line through the chart shows where identical actual and predicted course rankings would fall. Courses to the bottom and the right of that line have higher DFW rates than would be predicted. They are performing worse than expected. To the top and left of the line, courses with DFW rates better than predicted are shown.

If the 30% observed DFW rate were used to identify gateway courses in this dataset, they would include all the courses in the grey bar on the right of the chart. In such a scenario these courses would be marked for further attention, without regard for the actual students enrolled. The reader is invited to consider two particular courses, each marked with an arrow, one at the top (and within the grey bar) and the other towards the bottom of the chart (but outside the grey bar). The issue would be which of these two courses is more worthy of attention.

The course at the top has the higher DFW rate, but it actually is doing a little bit better for the students enrolled in it than other courses they are taking; that is the actual DFW rate is less than the predicted one. The course at the bottom has an actual DFW rate that would not bring it to anyone's attention for some time. Unfortunately, on average, the students in this latter course are underperforming relative to the other courses they are taking; the actual DFW rate is considerably larger than the predicted rate. They should be doing better. Arguably, the second course is much more problematic than the first because students enrolled in it did not succeed at the expected rate. Thus, we argue that considering the characteristics of the students enrolled in a course can add valuable perspective to the evaluation of its effectiveness.

Rather than simply looking within that grey bar for actionable information, the scatterplot makes it possible to provide a more nuanced way to prioritize the use of limited resources, as resources are always limited. Figure 4 suggests possible approaches.

The most obvious potential problem courses are identified in region A on the chart which highlights courses in which students are not doing as well as would be expected from their performance in all their courses. The bar is wider towards the bottom of the chart as expectations for students in the classes are getting higher, but their actual performance is not. Courses in Region B might also be of interest, not for being problematic, but because the performance of the enrolled students is better than their average performance across the classes for which they are registered. These courses arguably should be examined for best practices that might be used elsewhere. This region is intentionally extended into areas in which the actual DFW rates are high, but not as high as expected for the students enrolled. Finally, Region C on the top left raises other considerations. Although there may very well be good reasons for them, the actual DFW rates for these courses are so much better than predicted that issues of course credibility can be raised (Burton, & Peachey, 2014).

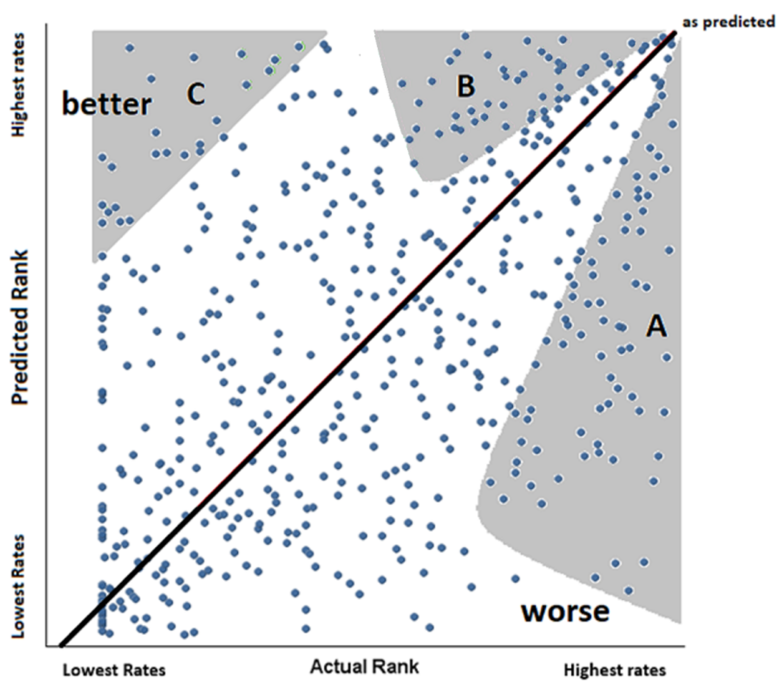


Figure 4. Potential courses of interest identified by gap analysis.

Charts such as Figure 4 might help focus remedial efforts. Of course, such charts only identify courses that require further investigation. It doesn't make sense to think one remedy fits all courses. Indeed, our research suggests that issues such as student placement, hidden prerequisites, and course delivery formats can be as impactful as course design and faculty behaviors. The former issues are also much easier to address. Thus, it is critical that courses identified through gap analysis are individually investigated to determine factors leading to the underperformance of students in them.

Courses in the top right extremes of the chart are certainly worthy of attention, although course-specific efforts in such cases are probably futile because the students in these courses are doing poorly in other classes as evidenced by high predicted DFW rates. High predicted rates result when multiple courses have high numbers of students receiving grades of D, F, and W. Even if the DFW rates in these courses could be drastically reduced, the fact would remain that those students would still have high DFW rates in other courses. Effective solutions to progression and retention issues in such cases are likely to involve remedies that extend beyond the limits of individual courses to more general factors, such as tutoring and other student supports. On the other hand, it is much more likely that courses which fall far from the diagonal line are manifesting issues that are course-specific, and not attributable to more general issues. This and similar charts can be used not only to identify courses worthy of further attention, but also to suggest the general sorts of corrective measures that are likely to be most effective.

Exploration of the courses that populate the various areas of interest quickly reveals patterns among various categories of courses: by discipline, or by curricular role. For example, it is not particularly surprising that STEM courses often show up in the areas of highest concern in the bottom right of the chart. Filtering the course population to include only STEM courses confirms that many are in that area, but not all. While one might expect that introductory majors' courses would be the most arduous and so have bigger gaps between actual and expected DFW rates and that non-majors' general education courses might be easier, this isn't always the case. As the number of courses is now substantially reduced, a cursory review of the courses which are falling well away from the average quickly raises interesting questions about how effective courses are in serving their intended audiences.

Capstone courses in various majors comprise a group of courses that are similar in their curricular role. This much smaller set of courses cluster near the average band, as one would hope, because students must necessarily pass these courses to graduate. There are, however, instances which fall well into a problematic range and clearly should garner further attention. As bad as attrition is early on in a program of study, attrition near the end is obviously much worse.

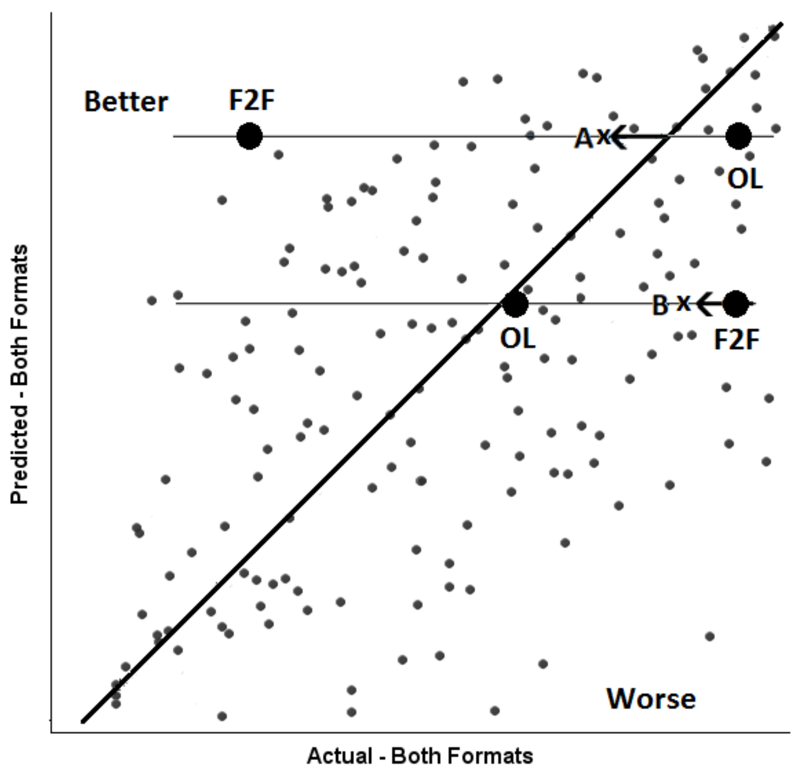
Institutionally specific general education courses are another group with similar curricular roles. Their overall pattern is similar to the capstone courses, but again a group of problematic courses can be identified. As important as these courses may be, they probably shouldn't be the gateway courses (in the sense of higher than expected DFW rates) and generally they are not, but those that are so identified definitely should be investigated further.

Finally, there are lower division general education requirements in various disciplinary categories. Every effort is made to assess the preparation level of entering students and place them in the appropriate level of these general education courses, if not dictated by the requirements for their chosen majors. Looking at the small set of courses that fulfill the mathematics requirement, for example, shows each of them is doing as well or slightly better than expected by these measures. Confirmation of the effectiveness of the existing curricula and student placement is clearly also a useful result.

Using Gap Analysis to Explore the Effects of Delivery Mode on Student Success

In this section we use examples to show how gap analysis might be used to explore differences between on-ground and online courses in terms of their impact on student success. The institution at which this study was conducted offers a full mix of online and on-ground courses to an amalgam of online and on-campus students. The curricular requirements for the majors are identical. The content and requirements for the courses are similar as well, taking into consideration the instructional delivery format used. Ideally, then, the results of measures of course performance should not depend on the delivery format.

Figure 5 shows courses offered in both delivery modes and the effect of course format on two specific courses. The combined results for course A show it to be unremarkable. The course has a relatively high DFW rate, but actually not quite as high as might be expected for the students enrolled. However, separating out the online and on-ground sections produces a very wide divergence. The online section has DFW rates noticeably higher than expected, while the on-campus sections have remarkably low DFW rates, particularly in regard to the characteristics of the students enrolled. Course B shows a less drastic divergence, but in the opposite direction. The course, overall, has a higher DFW rate than predicted, enough so to be a matter of concern. When separated out by course format, however, one can see that the on-campus section is problematic, while the online sections are performing almost exactly as expected. While these data cannot reveal the exact nature of the issues involved, they certainly suggest that the online sections of this course are not the problem.



.

Figure 5. Plot of actual vs. predicted DFW ranks for course offered with both online and on-ground sections.

When instructional delivery format is explored as another potential source of high DFW rates, the differentiation used may simplify things a bit by showing that potentially problematic courses may not be problematic in all instructional formats. Courses which appear to present no exceptional issues may turn out to have issues in one or more instructional formats when the data is parsed. Given that the course format does sometimes effect course performance, the issue then becomes whether this is an undesirable effect that needs to be remedied, or an advantageous effect that needs to be capitalized on. Are there some courses that are simply better delivered in the online or the on-campus format?

Exclusively Online Courses

When there is a difference in DFW rates associated with course modality, the next question is whether the issue is the course modality or the population of students in the course. Students in online degree programs only rarely take F2F courses, as most of them live well beyond reasonable commuting distances and, at the institution where this study took place, the tuition and fee structure are also barriers. Therefore, to begin to sort issues related to online learning, one of the variables can be removed by looking at courses which have only been offered online. The modality of the course is no longer an issue, then, but the student population includes students in both online degree programs and on-ground degree programs. As the institution under study has no online programs available to native Freshman the differentiation of majors as online vs on-ground is only meaningful for the transfer population.

Starting with those courses that are only offered online, the only factor related to delivery format involves the format of the student's major. How does the performance of online majors (OLMs) enrolled in online courses compare to that of on-ground majors (OGMs) enrolled in online courses? As expected, this goes both ways in different cases. The histogram in Figure 6. shows the distribution of results for all the courses in this sample.

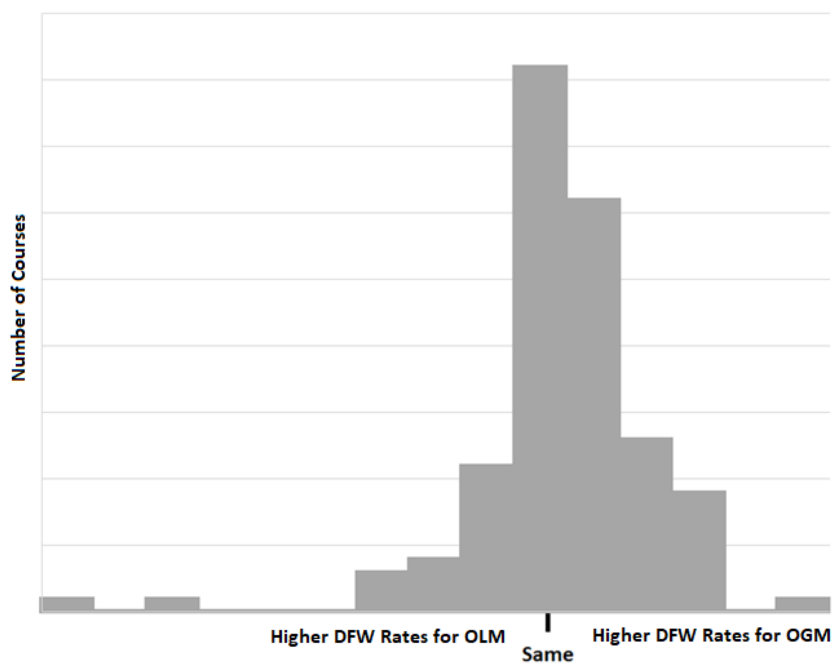


Figure 6. Distribution of Online Only courses by differences in DFW rates between online majors (OLM) and on-ground majors (OGM).

Disregarding the obvious outliers, the results are roughly normally distributed around a "no significant difference" mean, although they do show that there are more courses in which there are higher DFW rates for on-ground students than online students, suggesting that online majors tend to do better than their on-campus counterparts in solely online courses. For the median course in the distribution, on-ground majors have 4.7% higher DFW rates than online majors. These results are consistent with our earlier work (Bloemer & Swan, 2014) which also showed that on-ground students, on average, do not perform as well in online courses as they do in on-ground courses. The issue is especially true for on-ground majors who have not yet successfully completed an online course. The implications here for placement and for student support are evident.

For courses offered online and on-ground, both the modality of the student's degree program and the delivery format of the course are factors influencing DFW rates. The effects of the two variables are shown in Figure 6. As before, the effect of the modality of the major is shown on the horizontal axis. The DFW rates in these online courses are slightly higher for the on-ground majors than their online counterparts as shown. The vertical axis now shows the differences between the DFW rates in the online and on-ground sections of the course. The DFW rates in the online sections are slightly higher than those in the on-ground equivalent courses.

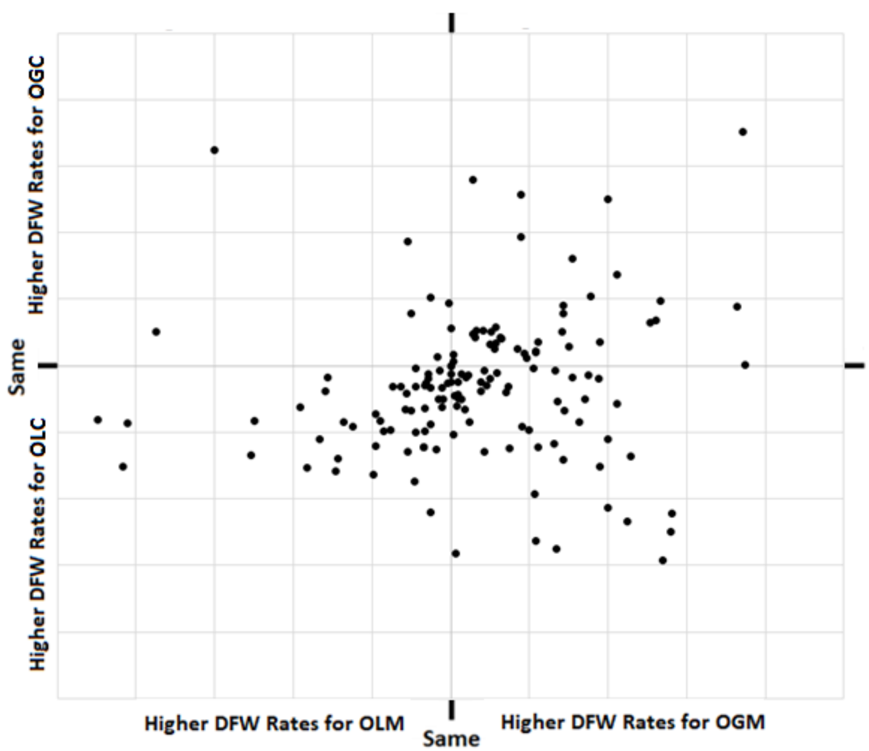


Figure 6. Dispersion of courses taught both online and on-ground by differences in DFW rates between online majors (OLM) and on-ground majors (OGM), and by differences in DFW rates by course section format: online courses (OLC) and on-ground courses (OGC).

Ideally, neither the modality of the student's program nor the delivery format of the course would be a contributing factor to the DFW rate. In that event, this scatterplot would be a single dot at the origin. Given that sampling errors will exist, the plot should then be normally distributed around the origin in two dimensions. The courses would be clustered around the origin, symmetrically left to right and top to bottom, getting more diffuse as the distance from the origin increases. The anomalous cases, then, would be those at greater distances from origin, whether that is caused by the student's major, the delivery format of the course, or some combination of the two. The radius from the origin on this chart could become the operational parameter for identifying cases for further examination, cases in which modality issues may be the source of elevated DFW rates.

A bubble chart can be used to incorporate this information into the original scatterplot. Figure 7 does that for the courses in this sample that were offered in both online and on-ground formats. The size of the "bubbles" is proportional to the distance from the origin as described above, incorporating differences in DFW rates associated either with online courses or online majors or both. While many of the courses associated with larger dots which fall in the bottom right area of the chart might already have been targeted by initial plotting of all courses, it is noteworthy that some of the courses that fall in the safer areas, in their composite measures, might be worth a second look, as relatively large differences in DFW rates associated with course or program modality are present. The earlier question arises again. Are such differences acceptable, and if not, should steps be taken to minimize them?

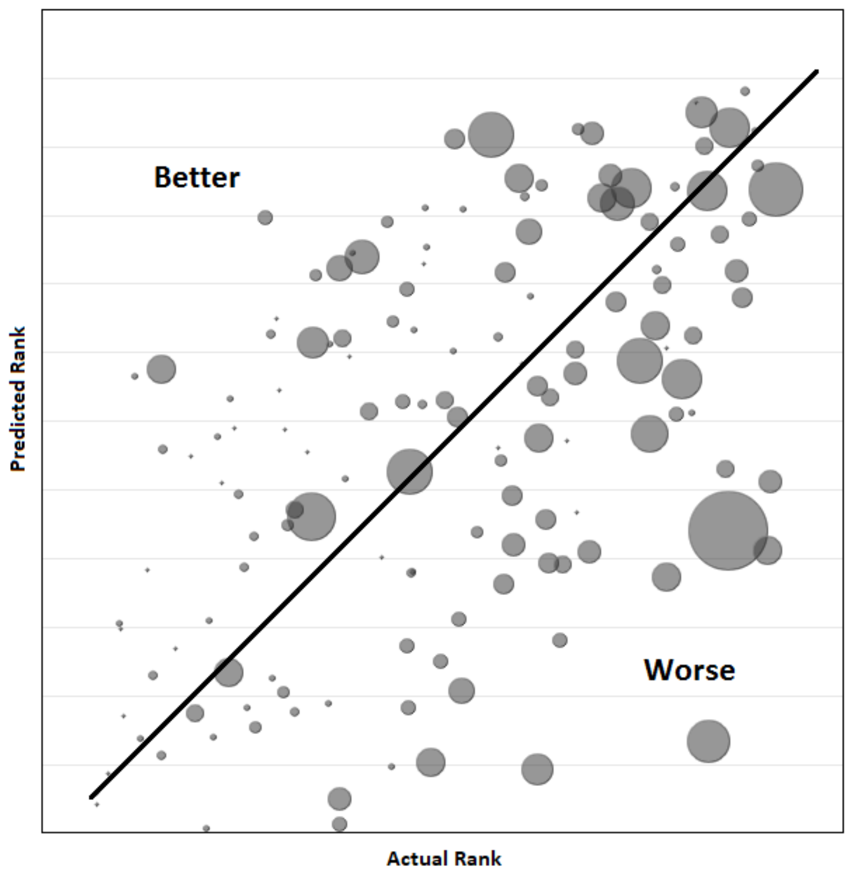


Figure 7. Plot of predicted DFW ranks vs actual DFW ranks for courses offered with both online and on-ground sections. Size of the bubbles represents differences in DFW rates associated with different course formats (OLC vs OGC) and/or different student majors (OLM vs OGM).

Discussion

Gateway courses with high DFW rates are known to contribute to student attrition (Koch & Pistilli, 2015), which makes them attractive targets for the use of data analytics. However, as the amount of available data increases, it becomes possible to examine such courses in a more nuanced way and be more selective in identifying and prioritizing courses for improvement.

Sorting courses in descending order of DFW rates is a good start. This often turns up introductory courses in areas of the core general education curriculum. That makes sense for an institution in which most students enter at the same point and pursue, for at least a while, a common curriculum. Like many similar institutions, the one involved in this study does not fit that description well. It serves a variety of students who enter the institution at many different stages in their degree pursuits and who bring diverse academic backgrounds with them.

Moreover, the DFW rates produced in a given course depend on many factors and, in themselves, provide no information about the underlying reasons for student performance. Remediation strategies, however, need to address such reasons. In addition, gateway courses are not gateways for all students, at all times in their academic journeys, nor are the hazards to student degree completion limited to the early stages of a student's studies. Thus, the preferred method for identifying courses for further attention should take into account the characteristics of the students enrolled as well as student outcomes, and it should produce a list that includes a distribution of courses across various stages of the curriculum.

Using a multiyear sample of grades earned by students and incorporating other readily available data, we are able to predict a student's probability of getting a DFW grade in an "average" course. Those individual predictions can be combined across students registered for a course to predict its DFW rate. The difference between the actual DFW rates and these predicted ones, the gap, becomes a useful measure for prioritizing courses for further attention.

The graphical approaches shown above not only illustrate the great diversity in gaps between predicted and observed DFW rates, but suggest ways for identifying both courses that need further investigation and those which are performing exceptionally well. The location of individual courses on the graph can also be used to show whether the performance issues are affecting students in several courses, in which case systemic corrective strategies would seem justified, or in particular courses, in which case course specific efforts would be appropriate.

Finally, the continued growth of online instruction adds issues of course modality, as well as the modality of entire degree programs into the list of factors that could influence student success. Described here is a technique to incorporate differences in DFW rates related to course and program modality into the analysis. This addition contributes a few more courses to the targeted list, but more importantly, helps identify those courses whose DFW issues do not apply equally to online and on-ground courses.

Once a prioritized list of courses for further attention has been developed, the real analytical work can begin, using quantitative and qualitative information as available. Existing institutional policies and practices are usually well established and founded in sound curricular judgements, refined by years of experiences. As most of this has happened long before the emergence of feasible data analytics, it is difficult to fault them for not being "data-based." However, sound curricular principles are not always perfect, and students, faculty, pedagogy and knowledge change. Bureaucracies, and the creative efforts of the people who navigate them, can produce unintended results. Given the current level of available analytics, is it reasonable to look at course performance data to see whether policies and procedures have lived up to expectations and to require that modifications to them be "data-based?"

For a given course, we look first for things that can most easily and quickly be changed. Are existing advising guides and prerequisite/co-requisite structures effective, and more importantly are they correlated with student success? Are there other student characteristics associated with student success that should be incorporated into those guides? Are there beneficial course sequences that ought to be included? Schedules and registration policies obviously impact student course selections. Addressing scheduling issues is often simpler and more cost effective than attempting to rescue students from unnecessarily difficult circumstances that they may have been forced into, as is addressing modality issues before they manifest. Should faculty strengths be matched to instructional modalities rather than trying to develop all faculty to do all things equally well?

Once these simpler and quicker issues have been addressed, it is possible to move on to the more substantive and more challenging issues of comprehensive course redesign, effective student support, and meaningful faculty development. This is not only possible, but necessary for preventing unnecessary distractions from the real work at hand, or, worse yet, wasting valuable resources addressing problems which really don't exist.

Limitations

The findings reported in this paper are, of course, unique to our institution. They certainly are not generalizable to other institutions. However, we believe the gap analysis procedure described herein is generalizable. Given that predictive models will most likely differ from institution to institution, the differences between predicted and actual outcomes will most always still yield more information at the course level than simple rankings that can be factored into student success efforts.

It is also the case that the predictive model we used to calculate gaps which categorized courses as problematic or not involves a paired down representation of actual factors impacting student success. We believe it works well enough for our purposes, but other institutions interested in using a gap analysis will surely want to adopt a predictive model that fits their circumstances.

Conclusions

In this paper, we presented a strategy for identifying courses which might impede undergraduate students' progress to degree which we call "gap analysis." Gap analysis involves using students' types, academic stages, prior GPA, and prior DFW rate to predict the DFW rates for the courses in which they are enrolled. Of course, different institutions might have differing predictive models, but we believe the idea is generalizable. Comparing predicted DFW rates to actual ones uncovers gaps that point to underperforming courses as well as those which are doing better than expected considering the students enrolled in them. We argue that applying gap analysis not only affords a more nuanced approach to identifying gateway courses, but it reveals courses encumbering student progression at all levels which is important when considering today's non-traditional student population.

We provided graphical examples illustrating the efficacy of this approach. In particular, we used gap analysis and the graphing of courses based on their predicted and actual DFW rates to explore issues related to online courses. These investigations demonstrated that the DFW rates for courses offered in both online and on-ground formats can obscure real differences between delivery modes. They also illuminated the potential for interaction between the modality of the programs in which students are enrolled and the delivery format of the courses they are taking.

The specific findings in the illustrations provided in this paper are, of course, limited to the institution under investigation. Moreover, that institution is a small, master's granting university and its student success issues are, if not unique, surely quite different from many other colleges and universities. However, the gap analysis strategy described herein utilizes a heuristic, not an algorithm. We believe that colleges and universities concerned with student progression issues can adapt it to their unique circumstances through the choice of predictors used to model DFW rates and the specific distinctions made in subsequent investigations. In this vein, it should also be noted that the predictive model we used in our work is a somewhat simplistic one which utilizes only four predictive variables.

It should also be noted that we believe that ultimately the remediation of problems with underperforming courses and students requires a qualitative examination of the specific problematic courses themselves. Our future work will take this direction. In particular, we will explore those courses identified through gap analysis and attempt to classify them by their underlying cause(s) and remediate the problems accordingly.

References

Arendale, D. (2004). Pathways of persistence: A review of postsecondary peer cooperative learning programs. In I. Duranczyk, J. L. Higbee, & D. B. Lundell (Eds.), *Best practices for access and retention in higher education.* Minneapolis, MN: University of Minnesota, Center for Research on Developmental Education and Urban Literacy, 27-40.

Arnold, A. (1999, March). *Retention and persistence in postsecondary education: A summation of research studies.* Texas Guaranteed Student Loan Corporation.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, *4*(2). Retrieved from: https://dx.doi.org/10.20429/ijsotl.2010.040217

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 259-262. Retrieved from: http://dl.acm.org/citation.cfm?id=2330664

Bean, J. P., & Metzner, B. S. (1985). College student retention—defining student retention, a profile of successful institutions and students, theories of student departure. *National Center for Educational Statistics.* Retrieved from: http://education.stateuniversity.com/pages/1863/College-Student-Retention.html

Bloemer, W., Day, S. & Swan, K. (2017). Gap analysis: An innovative look at gateway courses and student retention. *Online Learning, 21*(3), 5-14. doi: 10.24059/olj.v21i3.1233

Bloemer, B., & Swan, K, (2012, August,). Predicting student retention and success in online programs. Paper presented at the 28th Annual Conference on Distance Teaching & Learning, Madison, WI.

Bloemer, W. & Swan, K. (2014). Investigating informal blending at the University of Illinois Springfield. In A. G. Picciano, C. D. Dziuban, & C. R. Graham (Eds), *Blended Learning Research Perspectives, Volume 2.* New York: Routledge, 52-70.

Boston, W., Diaz, S. R., Gibson, A. M., Ice, P., Richardson, J., & Swan, K. (2009). An exploration of the relationship between indicators of the community of inquiry framework and retention in online programs. *Journal of Asynchronous Learning Networks, 13*(3).

Burton L., & Peachey, J. (2014). Ethical leadership in intercollegiate sport: Challenges, opportunities, future directions. *Journal of Intercollegiate Sport, 7*, 1-10.

Cabrera, A. F., Burkum, K. R. & La Nasa, S. M. (2005). Pathways to a four-year degree: Determinants of transfer and degree completion. In Alan Seidman (Ed.). *College Student Retention: A formula for success*. ACE/Prager Series on Higher Education, 155-214.

Campbell, J. P., & Oblinger, D. J. (2007). Academic analytics. *EDUCAUSE*. Retrieved from: http://net.educause.edu/ir/library/pdf/PUB6101.pdf

Clay, M. N., Rowland, S., & Packard, A. (2008). Improving undergraduate online retention through gated advisement and redundant communication. *Journal of College Student Retention, 10*(1), 93-102.

Cochran, J. D., Campbell, S. M., Baker, H. M., & Leeds, E. M. (2014). The role of student characteristics in predicting retention in online courses. *Research in Higher Education*, *55*(1), 27–48. https://doi.org/10.1007/s11162-013-9305-8

DePaul University. (nd.). Gateway Courses Redesigned. Retrieved from: https://offices.depaul.edu/enrollment-management-marketing/student-retention/Documents/8_GatewayCourseRedesign_2012.pdf

Education Advisory Board. (2016). Bottleneck course redesign. Retrieved from: https://www.eab.com/research-and-insights/online-course-prioritization-guide/bottleneck-course

Falcone, T. M. (2011, November). Toward a new model of student persistence in higher education. Paper presented at the annual meeting of the Association for the Study of Higher Education, Charlotte, NC

Gardner Institute. (2017). *Roll Out of Gateways to Completion.* John N. Gardner Institute for Excellence in Undergraduate Education. Retrieved from: http://www.jngi.org/roll-out-of-gateways-to-completion/

Hachey, A. C., Wladis, C. W., & Conway, K. M. (2014). Do prior online course outcomes provide more information than G.P.A. alone in predicting subsequent online course grades and retention? An observational study at an urban community college. *Computers & Education*, *72*, 59–67. https://doi.org/10.1016/j.compedu.2013.10.012

Herzog, S. (2005). Measuring determinants of student return vs. dropout/stopout vs. transfer: A first-to-second year analysis of new freshmen. *Research in Higher Education*, *46*(8), 883-928.

Hosmer, D., Lemeshow, S., & Sturdivant, R., (2013). *Applied Logistic Regression, Third Edition*. New York: John Wiley & Sons, Inc.

Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *Journal of Higher Education*, 77(5), 861-885.

Jayaprakash, S., Moody, E., Eitel, L., Regan, J., & Baron, J. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6-47.

Jia, P. (2014). *Using predictive risk modelling to identify students at high risk of paper non-completion and programme non-retention at university.* Doctoral dissertation. Retrieved on July 13, 2015.

Juhong, B. & Maloney, T. (2006). Ethnicity and academic success at university. *New Zealand Economic Papers, 40*(2), 181-213.

Koch, D., & Pistilli, M. (2015). Analytics and Gateway Courses: Understanding and Overcoming Roadblocks to College Completion. Retrieved from https://www.insidehighered.com/sites/default/server_files/files/Analytics%20and%20Gateway%20Courses%20PPt.pdf

Lee, Y., & Choi, J. (2011). A review of online course dropout research: implications for practice and future research. *Educational Technology Research and Development*, *59*(5), 593–618. https://doi.org/10.1007/s11423-010-9177-y

Lewis, M., & Terry, R. (2016). Registering risk: Understanding the impact of course-taking decisions on retention. Norfolk, VA: *Proceedings of the 12th Annual National Symposium on Student Retention*, 364-371.

Mallette, B. L., & Cabrera, A. F. (1991). Determinants of withdrawal behavior: An exploratory study. *Research in Higher Education, 32*(2), 179-194.

Montmarquette, C., Mahseredjian, S., & Houle, R. (September, 2001) The determinants of university dropouts: A bivariate probability model with sample selection. *Economics of Education Review 20*(5): 475-484.

Moore, C. & Shulock, N. (2009, September). *Degree completion: Lessons from the research literature.* Sacramento, CA: Sacramento State University Institute for Higher Education Leadership & Policy.

Morris, L. V., Wu, S-S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *American Journal of Distance Education, 17*(1), 23-36.

Munro**,** B. (June 20, 1981). Dropouts from higher education: Path analysis of a national sample. *American Educational Research Journal, 18,* 133-141

National Student Clearinghouse (2018). *Yearly Success & Progress Rates – 2018*. Retrieved from https://nscresearchcenter.org/snapshot-report-yearly-success-and-progress-rates/

Nogaj, A., & Kons, E. (2016). Beyond tutoring: Intrusive academic assistance to increase student success and retention. Norfolk, VA: *Proceedings of the 12th Annual National Symposium on Student Retention*, 55-66.

Offenstein, J., & Shulock, N. (2010, 2010). Taking the next step: The promise of intermediate measures for meeting postsecondary completion goals. Sacramento, CA: Sacramento State University Institute for Higher Education Leadership & Policy.

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education, 6*(1), 1-16.

Shapiro, D., Dundar, A., Yuan, X., Harrell, A., & Wakhungu, P. K. (2014, November). *Completing College: A National View of Student Attainment Rates – Fall 2008 Cohort* (Signature Report No. 8). Herndon, VA: National Student Clearinghouse Research Center.

Sweeney, M., Rangwala, H., Lester, J., & Johri, A. (2016). Next-term student performance prediction: A recommender systems approach. *Journal of Educational Data Mining*, *8*(1), 22-50.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* Chicago: University of Chicago Press.

Vignare, K., Wagner, E. D., & Swan, K. (2017). The value of common definitions in student success research: Setting the stage for adoption and scale. *Internet Learning Journal, 6*(1), 7-23. Retrieved from: http://www.ipsonet.org/publications/open-access/internet-learning/volume-6-number-1-spring-2017-summer-2017

Wetzel, J., O'Toole, D., & Peterson, S. (1999). Factors affecting student retention probabilities: A case study. *Journal of Economics and Finance, 23*(1), 45-55.