March 2005

Multivariate Analysis of Student Loan Defaulters at the University of South Florida

Conducted by TG Research and Analytical Services

Matt Steiner
Carmen Tym

Multivariate Analysis of Student Loan Defaulters at the University of South Florida

TG Research and Analytical Services
Matt Steiner
Carmen Tym

Table of Contents

Multivariate Analysis of Student Loan Defaulters at the University of South Florida[1]

TG Research and Analytical Services
Matt Steiner
Carmen Tym

Executive Summary and Highlights

This study on default analyzes 17,036 undergraduate borrowers who attended the University of South Florida (USF), and who entered repayment on Federal Family Education Loan Program (FFELP) loans during federal fiscal years 1999, 2000, 2001, and 2002. Of the 17,036 USF undergraduate borrowers, 766 defaulted (4.5 percent). The report employs a technique known as logistic regression that isolates the independent relationship of each variable to the probability of default after accounting for the relationships of other relevant variables. The key findings are:

- Student borrowers who graduate are three percentage points less likely to default than those who do not graduate.
- Having no course hours failed reduces a borrower's chance of default, while the borrower's likelihood of default gains four percentage points when he or she fails between two and nine course hours, and inflates by nine percentage points when he or she fails 10 or more hours.
- The higher the student's total family income, the lower the student's risk of default.
- Borrowers who enter repayment at age 31 or older are two percentage points more likely to default than those who are age 22 to 30.
- Females are one percentage point less likely to default than males.
- University of South Florida students who are Hispanic have a default rate that is not statistically different from USF students who are White. Black USF students have a default probability that is two percentage points higher than White USF students.
- Students who have a grade point average (GPA) of less than 2.00 have a risk of default that is three percentage points greater than students who have a GPA between 3.00 and 4.00.
- Having between one to four course hours incomplete raises a student's chance of default by one percentage point as compared to having no hours incomplete.
- Parent marital status affects a student's probability of default. Students who have parents who are married are two percentage points less likely to default as compared to those whose parents are not married (i.e., a single or widowed parent, or parents who are separated or divorced).
- A borrower whose father's highest education level is high school, college, or beyond is two percentage points less likely to default as compared to a borrower who has a father with a middle school or junior high education.
- The more total hours a student attempts, the less likely the student is to default.

[1] The authors of this study would like to thank Sandra Barone for help throughout, Aiju Men for testing the logit linearity of continuous variables, and Jeff Webster for assistance with writing and editing.

Multivariate Analysis of Student Loan Defaulters at the University of South Florida

Introduction

Higher education is an investment that, on average, pays handsomely. Over the course of a lifetime of work a person with a four-year college degree can expect to earn $1 million more than someone with only a high school degree. The non-monetary enrichment associated with a college education adds to the total value of the experience. Some students rely on federally subsidized, low interest loans to finance their education, deferring payments until after they leave school. While most student borrowers repay their loans without problem, some face difficulties. To prevent these difficulties from leading to a defaulted loan, schools, lenders, and guarantors engage in various default aversion activities.

The Higher Education Act allows guaranty agencies to use interest earned on their reserve account for default aversion projects, provided that the U.S. Department of Education (ED) approves the efforts as constructive. TG[2] successfully petitioned ED for approval for research efforts to learn more about the factors that contribute to default focusing, in particular, on what happens at the campus level. Each study has a similar structure but produces somewhat different results that are specific to the university being analyzed. Though TG collects approximately the same campus variables from each institution and applies the same research methodologies, the results for each university reflect the distinctive institutional policies, student body, and campus culture associated with them.

In an effort to better understand student loan default behavior at the University of South Florida (USF), at the request of USF, the research staff at TG conducted a study of the relationship between loan default and student characteristics. The study examines the default behavior of 17,036 undergraduate borrowers who attended USF and who entered repayment on Federal Family Education Loan Program (FFELP) loans during federal fiscal years 1999, 2000, 2001, and 2002. The study regards a borrower as being in default if the borrower defaulted within the fiscal year the borrower entered repayment or within the following fiscal year, which corresponds to the official definition used by the Department of Education for cohort default rates. Based on this definition, 766 of the 17,036 USF undergraduate borrowers defaulted (4.5 percent). University of South Florida staff supplied information describing graduation status, college grade point average, course hours, total family income, age of borrower, father's highest education level, gender, race, marital status, and many other aspects of students' backgrounds and college experiences. This information was merged for each student with data from the National Student Loan Data System (NSLDS), which indicated whether or not the borrowers had student loan defaults.

In analyzing these data, TG used a dynamic statistical modeling technique called multiple logistic regression. The advantage of logistic regression over a two-variable analysis is that it can reveal the separate relationship of each variable to default after accounting for the independent relationships to default of all the other variables within the model.

This analysis will attempt to increase our understanding of the factors that are related to default at the University of South Florida. Hopefully, USF, TG, and others will be able to use this better understanding

---

[2] TG is a public, nonprofit corporation that helps create access to higher education for millions of families and students through its role as an administrator of the Federal Family Education Loan Program (FFELP). Its vision is to be the premier source of information, financing, and assistance to help all families and students realize their educational and career dreams. Additional information about TG can be found online at www.tgslc.org.

of default behavior to identify at-risk groups of borrowers, craft intervention or default mitigation strategies, and allocate scarce default aversion resources to the situations in which they will have the greatest chances of making a difference.

Prior Research on the Factors Relating to Student Loan Default[3]

Early default studies commented on the (then) new federal policy of holding schools responsible for borrower defaults. Therefore, many prior studies have concerned themselves with evaluating the relative importance of borrower and institutional characteristics. Several have found that institutional characteristics have little or no association to loan repayment behavior and that borrower variables are much more important predictors of default (Knapp & Seaks, 1990; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987). Since the present analysis of the University of South Florida concerns the default behavior of students at one institution, prior work on the influence of institutional characteristics is of little relevance.

Nevertheless, in their endeavor to find the factors related to default, researchers evaluated many borrower characteristics that are relevant to the present study. These factors include demographic descriptors (such as ethnicity or race, gender, age, and income), financial aid-related variables (like financial need and expected family contribution) and some high school-related variables (like ACT scores and whether the borrower has a high school diploma).

Importance of Graduation
The most consistent finding of past studies is that borrowers who graduate (or who earn a degree or who do not withdraw) have a much lower probability of defaulting on their loans, as compared to borrowers who do not graduate (Dynarksi, 1994; Knapp & Seaks, 1990; Meyer, 1998; Podgursky et. al., 2000; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). For many of these studies, graduation status was the single most important variable.

Ethnicity/Race
The second most prominent finding of multivariate default studies has been that ethnicity/race is strongly related to default (Dynarksi, 1994; Knapp & Seaks, 1990; Podgursky et. al., 2000; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). In particular, being Black greatly increases the probability of default. In three of the studies (Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002), being Black had the largest effect of all variables, and in the remainder of the cited studies, being Black was the second most influential factor.

College Performance
Prior studies have tested only a few variables that measure the borrower's performance in college. Volkwein et. al. (1995) found that the borrower's GPA in college and whether the borrower was a science or technology major produced significant but relatively small decreases in the probability of default. They also determined that a variable signifying that the borrower was a transfer student did not have a significant relationship to default. A related study by Volkwein and Szelest (1995) uncovered similar results with respect to college GPA, majoring in science or technology, and transfer status. Woo (2002) found that attainment of a graduate or professional degree greatly reduces the chances of default. She further established that borrowers who attended more than one school were also less likely to default. (Woo noted that this variable partially reflects the fact that borrowers who go to graduate school

[3] For a more comprehensive review of student loan default research, see TG's Student Loan Default Literature Review, McMillion (2004) available at http://www.tgslc.org/schools/index.cfm under Default Aversion.

frequently have attended more than one school.) Whether or not a borrower studied a business or computer curriculum did not have a significant association to default in Woo's study. Meyer (1998) found that as the academic level attained by a borrower increases, the probability of default decreases.

Income, Age, Gender, and Other Demographic Variables

Previous research has determined that demographic characteristics other than ethnicity have significant, though mostly smaller, associations to default. After ethnicity, parental income appears to be the most commonly-tested demographic variable, and studies have found higher income levels to be associated with decreases in the probability of default (Dynarksi, 1994; Knapp & Seaks 1992; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). Gender is also routinely analyzed, and researchers usually conclude that being female is related to a substantial reduction in the likelihood of defaulting (Podgursky et. al., 2000; Volkwein et. al., 1995; Woo, 2002). Podgursky et. al., Woo, and Meyer (1998) examined the age of the borrower and determined it to have a significant but small effect on default behavior, with increases in age related to higher probabilities of defaulting. In contrast, Knapp & Seaks (1992) did not detect a statistically significant relationship for either the gender or age of the borrower. Volkwein and Szelest (1995) also did not find an association between gender and default behavior. Among the other demographic variables that researchers have found to have significant relationships to default are the marital status of parents, (Knapp & Seaks, 1990), U.S. citizenship (Wilms, Moore & Bolus, 1987), the parents' educational level (Volkwein et. al., 1995), being Hispanic (Dynarksi, 1994; Woo, 2002), having dependents (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002), the marital status of the borrower (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995), the borrower's income (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002) and several others.

Pre-college Experience

To a very limited extent, researchers have evaluated characteristics reflecting the borrower's experience before college. Several studies have found that graduation from high school reduces the likelihood of default (Dynarksi, 1994; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). However, Volkwein and Szelest (1995) did not detect a significant relationship between having a high school diploma and default behavior. Podgursky et. al. (2000) also examined ACT scores and identified a small negative effect on default.

Financial Aid and Cost of Education

Studies have generally paid scant attention to financial aid-related variables. Nevertheless, it is important to test whether financial assistance mitigates the probability of default in ways that are independent of income. Among the studies reviewed here, only a couple reviewed variables other than family income and family assets. Volkwein et. al. (1995) tested several financial aid-related variables – such as the receipt of scholarships/grants, whether the borrower participated in work study, and whether the borrower had other employment – but found none of them to be significant. Meyer (1998), however, determined that the probability of default declined with increases in the cost of attendance, controlling for type of institution. He further discovered that the likelihood of default increased substantially for borrowers who received more than $1,000 from non-loan aid sources. He noted a small decrease in the chances of defaulting as the expected family contribution of borrowers increased.

Loan Patterns

Several of the studies have also included loan-related variables. Four of the analyses determined that there was not a statistically significant relationship between the amount of loans borrowed and default behavior (Knapp & Seaks, 1990; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002). Meyer (1998), however, found that each $1,000 of total debt increases the probability of default by about one percentage point. Dynarski (1994) determined that the probability of default rose with increases in the size of borrowers' monthly loan payments. Furthermore, Woo detected a small increase in the likelihood of default associated with an increase in the number of loans a borrower has. Meyer also examined the types of federal loans that borrowers received and showed that borrowers with only subsidized Stafford loans had the highest probability of default. In his study, he further demonstrated that borrowers who utilized deferments had a somewhat smaller chance of defaulting.

Studies for Texas A&M University by TG

In January 2005, TG published a multivariate analysis of student loan default for Texas A&M University – College Station.[4] This study found that college grade point average was most strongly associated with whether or not borrowers defaulted on student loans. In addition, it determined that in-person exit counseling, graduation indicator, college/school last attended, age of borrower entering repayment, number of hours failed, race/ethnicity of borrower, mother's highest education level, Expected Family Contribution, number of hours transferred, adjusted gross income of student, and gender were all statistically significant predictors of default. Overall, the study for College Station concluded that college success variables were more strongly related to default than student background variables.

Methodology for Multivariate Analysis of Defaulters at the University of South Florida

TG uses logistic regression for conducting multivariate analyses of behaviors, such as repayment behavior, in which outcomes can assume one of two classes, like defaulting or not defaulting. The statistical analysis proceeds by determining the relationships between borrower characteristics and default behavior within a past population of borrowers. The known outcomes (i.e. default behaviors) of this population serve as the basis for statistical estimation. The result of the analysis is a set of coefficients or weights. The logistic regression method chooses the set of weights that would produce predictions of default that match as closely as possible to the known outcomes of default.

As is true of any statistical modeling approach, the results should be interpreted with care. All models are subject to error, which the modeler can attempt to minimize but cannot completely eliminate. One source of error stems from the likelihood that the modeler has failed to include all the variables that are relevant to the phenomenon being studied. Another reason to be careful when generalizing a study's findings derives from the tendency of results to better describe the sample from which they were produced than any other sample or group. All studies are also susceptible to a certain amount of measurement error. Despite these limitations, however, the statistical techniques have been well-tested and widely-used to produce reasonable results in many different applications.

Variable Selection Process

One goal of this analysis is to find, among all possible relevant variables, the subset of variables that best explains default behavior. This subset of variables is likely to be much smaller in number than the total number of variables that were gathered for the study. In fact, statistical analysis showed that many variables explain very little about the likelihood of borrowers to default. Therefore, variables that have no statistically significant relationship to default were dropped from inclusion in a final default model. In addition, some groups of variables tend to provide similar explanations of default behavior and are, therefore, redundant with each other; in many such cases it is possible to select one variable to represent the other variables. However, sometimes a variable is so important from a theoretical or practical

[4] Multivariate Analysis of Student Loan Defaulters at Texas A&M University is available at http://www.tgslc.org/schools/index.cfm TG will release two additional reports in 2005 with one looking at Texas A&M – Kingsville and the other at Texas A&M – Prairie View.

standpoint that the modeler must include it, even if it overlaps with other variables. Incorporating all of these considerations, the final default model is the combined result of statistical relevance, theoretical importance, organizational requirements, and human judgment.

Results of the Multivariate Analysis

The multivariate analysis produced a default model containing the variables listed in the table below. The table lists each variable, its reference variable, the coefficient, and the delta-p (change in probability), each of which will be explained below. Variables are listed in order of their strength in predicting default, from strongest to weakest.

The multivariate estimation process produces a coefficient for each variable. The sign (positive or negative) of a coefficient indicates whether the presence of the variable increases or decreases the likelihood of default, and the size of a coefficient generally reflects the strength of the relationship between the variable and the occurrence of default. For example, any Grade Point Average (GPA) below 3.00 is associated with an increase in a borrower's chances of defaulting (since the coefficients are all positive). Moreover, as GPA decreases, the probability of default increases (since the coefficients are larger for lower GPA categories). In contrast, an increase in the borrower's total family income is associated with a decrease in a borrower's chances of defaulting (since the coefficients for these categories are all negative).

The presence of an asterisk next to a coefficient indicates that the variable has a statistically significant relationship to default behavior. Statistical significance means that there is a relatively high confidence that a relationship really exists – that the size of the coefficient did not result from the peculiarities of the sample that we analyzed. The more asterisks there are, the higher the level of confidence that a true relationship exists between a variable and default behavior. Thus, whereas all the variables based upon college GPA have a statistically significant relationship to default, the variable indicating that a borrower has a Total Family Income of '$1 to $10,000' does not have a statistically significant relationship.

Unfortunately, the coefficients are difficult to interpret in their raw form. In order to more easily understand their meaning it is necessary to convert them to another form. The delta-p represents the percentage point change in the probability of default given the presence of a characteristic. For example, a borrower who 'Graduated' has a probability of default that is three percentage points lower than a borrower who 'Did not graduate'. This means that if all other variables in the model besides the graduation variable indicated that a borrower had a 4.5 percent probability of defaulting, the fact that the borrower graduated would lower the likelihood of default to 1.5 percent (4.5 percent minus 3.0 percentage points). The default rate of 4.5 percent is the overall average for the study sample, and the delta-p statistics in the following table always reflect a change relative to this default rate. Change in Probability percentages (delta-p statistics) are only provided for variables that are statistically significant.

A reference group is required for interpreting the variables used in this analysis. Consider the group of age variables. Borrowers who entered repayment at '31 or more' years of age have a likelihood of default that is two percentage points higher than borrowers between the ages of 22 and 30 years (who belong to the reference group). In most cases, the analysis would produce an equivalent result regardless of which category served as the reference group. However, in some cases, the desire to test a prior hypothesis has suggested a useful reference category. For example, the belief that students of traditional college-graduating age differ in repayment behavior from other students might lead a researcher to select the '22 to 30' year category as the reference group.

No single statistic – whether the coefficient, the level of significance, or the change in probability (delta-p) – provides an unambiguous way of ranking variables in terms of how adequately they explain default behavior. Each statistic in the table below has its drawbacks in depicting the strength of the relationship between these variables and whether or not borrowers default. As a consequence, the subsequent discussion of the variables relies upon a composite picture of the various statistics. In general, groups of variables with large coefficients, whether positive or negative, and high levels of significance (more asterisks) have stronger relationships to default behavior; groups of variables with smaller coefficients and low levels of significance have weaker associations to whether or not borrowers default. The variable groups are discussed in the order of their strength of association to default.

Table: Results of Multivariate Analysis for the University of South Florida

Variable Group	Variable	Reference Variable	Coefficient	Change in Probability
Graduation Status	Graduated	Did not graduate	-0.93 ***	-3%
Number of Hours Failed	1	0	0.05	
	2 to 9	0	0.68 ***	4%
	10 or more	0	1.20 ***	9%
Total Family Income	$1 - $10,000	Up to and including $0	-0.30	
	$10,001 - $20,000	Up to and including $0	-0.49 **	-2%
	$20,001 - $30,000	Up to and including $0	-0.67 ***	-2%
	$30,001 - $50,000	Up to and including $0	-0.74 ***	-2%
	$50,001 - $70,000	Up to and including $0	-0.99 ***	-3%
	$70,001 - $720,000	Up to and including $0	-1.01 ***	-3%
Age of Borrower at Time of Entering Repayment	18 to 21	22 to 30	0.21	
	31 or more	22 to 30	0.45 ***	2%
Gender	Female	Male	-0.32 ***	-1%
Race	Black	White	0.33 ***	2%
	Hispanic	White	-0.17	
	Other	White	0.23	
Grade Point Average (GPA)	1.00 - 1.99	3.00 - 4.00	0.55 ***	3%
	2.00 - 2.99	3.00 - 4.00	0.25 *	1%
Number of Hours Incomplete	1 to 4	0	0.30 ***	1%
	5 or more	0	-0.17	
Parent Marital Status	Married	Not married	-0.44 ***	-2%
	Missing	Not married	-0.13	
Father's Highest Education Level	High School	Middle School/Junior High	-0.46 **	-2%
	College or Beyond	Middle School/Junior High	-0.53 **	-2%
	Missing	Middle School/Junior High	-0.60 ***	-2%
Total Hours	60 - 119	0 - 59	-0.26 *	-1%
	120 - 149	0 - 59	-0.22	
	150 - 189	0 - 59	-0.41 **	-1%
	190 - 900	0 - 59	-0.65 **	-2%
Intercept			-2.03 ***	

Sample Size: 17,036
Defaulters: 766 (4.5%)
- 2 Log Likelihood for Intercept and Covariates: 5527
Likelihood Ratio Chi-Square: 723 with
29 degrees of freedom (Pr > ChiSq = <.0001)

C Statistic: 76.5
* Statistically significant at a 0.05 level
** Statistically significant at a 0.01 level
*** Statistically significant at a 0.001 level

Graduation Status

Consistent with previous studies, this model confirms the importance of graduating from college in lowering student loan defaults. Students who 'Graduated' are three percentage points less likely to default than students who 'Did not graduate', all other factors being equal. By meeting their educational goal, these graduates are more likely to pay back their loans as compared to those who never graduated. This suggests that educational achievement is related to repayment success. Upon graduating the student has a degree, which can lead to job opportunities with higher earnings potential. Degree completion demonstrates admirable personal qualities such as persistence, intellectual accomplishment, and reliability that often translate into career success. After graduating, a student has a better chance of finding a job in his or her field, or finding any job for that matter, which in the student's mind increases the value of education, as well as the value of the loans taken. High student completion rates are desirable for the university. Having a high graduation rate makes an impact on how the university is perceived in Florida and in neighboring states. By increasing the graduation rate, USF not only creates a favorable impression with prospective students, but also decreases the risk of defaulting on student loans. These USF graduates become part of the alumni base and go on to contribute to the good reputation of the university and enter the pool of likely donors to the school. Maintaining low default rates helps the college maintain eligibility for federal student aid programs and may qualify the school for regulatory relief in some instances.

Number of Hours Failed

Number of Hours Failed is a strong predictor of default. Failing '10 or more' hours hikes the probability of default nine percentage points as compared to failing '0' hours, provided all other things are the same. A student who fails '2 to 9' hours has a four percentage point higher probability of default than a student who does not fail any hours. This study shows that failing one hour is not any different from failing zero hours, meaning that a student who fails one hour does not have a statistically significantly increased chance of defaulting on his or her loan as compared to a student who has no hours failed. Of the variables included in the model, Number of Hours Failed is one of the strongest variables to affect default and can serve as a useful early warning of potential default.

The reasons that a student fails hours at a university may be similar to the reasons a student fails to make payments on his or her student loan. When difficulties arise in meeting a commitment a resourceful person takes steps to deal with them. If such a person is facing difficulty in school that person may withdraw from a class before the deadline without penalty for extenuating circumstances, or take an incomplete as an extension. Parallel to this, a person experiencing financial difficulties may explore options such as deferment or forbearance on a loan instead of defaulting. In a study of California borrowers, Dr. Jennie H. Woo found that deferment or forbearance was negatively related to default. 'It could be that borrowers who are organized enough to follow through on using deferments (forbearances are comparatively rare) are also better able to handle repayments in general' (Woo, 2002: 16). The ability to follow through on an obligation, whether scholastic or financial, involves being well-informed and being able to access the support necessary to find a solution to problems that arise.

Based on the strength of this finding, we would recommend that USF utilize the academic information it has on students to monitor the number of hours failed. If a student fails more than one hour, USF could require that student to attend academic and/or financial counseling. The counseling session provides the opportunity for USF to convey the message that there is a connection between academic achievement and the student's ability to meet loan obligations. Talking with the student may also help the counselor understand what difficulties the individual student is encountering, and how these might be dealt with.

Total Family Income
In this study we found that the higher a student's total family income, the lower the probability of default, holding all other factors constant. For every category of Total Family Income of $10,000 and above, the probability of default is lower than the reference category, which is 'Up to and including $0'. However, a Total Family Income of '$1 to $10,000' does not statistically significantly lower the chance for default in comparison to a Total Family Income of 'Up to and including $0'. Total Family Income for a dependent student is equal to the parent's adjusted gross income; whereas, for an independent student, it is equal to the student's adjusted gross income. A negative Total Family Income is possible, representing a loss of income. The coefficients indicate the effect of each variable on the probability of default. We see that the higher the coefficient, the stronger the effect. Since the coefficients for Total Family Income increase as the amount increases, the effect of Total Family Income on the probability of default is not uniform; rather, it becomes stronger as the Total Family Income gets larger.

It is important to note that Total Family Income does not measure family income *after* college and is therefore not a direct measure of the resources available to a borrower for repaying student loans. Moreover, even if these incomes did reflect available resources for loan repayment, the parents of these students do not have responsibility for repaying the loans analyzed by this study. Nonetheless, in the case of dependent students, parent resources may create a buffer, which helps the student avoid default if the student experiences financial hardship after leaving school. Though the income of a student's parents is not necessarily accessible to the student as a source for repaying student loans, students whose families have higher total family incomes might have more financial resources available to them in times of repayment difficulties. For an independent student, having a higher total family income during college may forecast a similar or higher level of total family income after college with money available to pay his or her loan.

A second aspect represented by Total Family Income is the level of financial knowledge and skills that a student possesses. Money management skills are learned, either by being taught or by example. Students who come from families with low income may have much less experience in borrowing and repaying loans, or with credit, than do students from families with higher incomes. It is also possible that students from some families may have had a negative experience with borrowing, rather than a lack of experience, that dissuades them from taking available student loan aid. In addition to offering financial aid, if USF identifies students that would benefit from counseling on personal finances, it may lower the default rates on student loans. In a personal finance class, USF might teach students skills such as understanding credit cards and interest, managing debt, budgeting for irregular expenses, regularly setting aside savings, and how to write a check.

Age
We conducted a careful evaluation of the Age of the Borrower at Time of Entering Repayment to determine the categories that best depict the associated differences in default behavior. The resulting age categories are '18 to 21', '22 to 30', and '31 or more'. We found that students who enter repayment at age '31 or more' have a probability of defaulting on their loans that is two percentage points higher than students who are '22 to 30' years of age, assuming that the borrowers are alike in other respects. Another way of saying this is that despite having the same value on everything else, those 31 years or older are still more likely to default than students who are age 30 or younger. The youngest age group, '18 to 21', is not statistically different from the reference group '22 to 30' once all other factors are accounted for. The non-traditional students – 31 or older – might have families and jobs that result in a weaker integration to the campus, and therefore a weaker sense of obligation to repay student loans. They also might have established non-educational financial commitments, like home mortgages, that make it relatively difficult for

them to manage their student loans debts. At this stage in life a student may already have multiple financial obligations such as a mortgage, car payments, child expenses, and credit card debt. The person may be juggling payments and trying to keep track of bills. It may seem as if all these other responsibilities are in competition with the student loan payments to be made.

The University of South Florida may want to design a counseling program that specifically addresses the financial challenges that a student over age 30 may face. Increasing awareness of the options available in repaying a loan may help — such as deferment or forbearance if the borrower is temporarily unable to pay. Special consideration might be given to non-traditional age students such as by offering assistance over the telephone or through the Internet, or in after hours counseling sessions.

Gender

A 'Female' student is one percentage point less likely to default on her student loan than a 'Male' student is, all other things being equal. This finding is consistent with the findings in other studies.

Race

Broad categories of race are included in the analysis. These broad categories do not depict the rich variation in ethnicity of each group. In this study, Hispanics are no more likely to default on their student loans than Whites are. Hispanics may include Cuban and Puerto Rican ethnicities, as well as Mexican-American. A student who is 'Black' has a probability of default that is two percentage points higher than a student who is 'White', if they were the same on all other variables. However, while this result supports the general finding of past studies that being Black is associated with a higher likelihood of default, it indicates nowhere near the strength of relationship that prior studies revealed. The 'Other' category consists of a wide array of groups including Asian, Pacific Islander, American Indian, Alaskan Native, Non-resident Alien, missing, and unknown. Of course, there is nothing about race itself that results in differences in default behavior; rather, this acknowledges that there are factors associated with these ethnic or racial categories – such as the legacy of discrimination – that may affect their ability to repay their student loans. Differences in socio-economic conditions, as well as differences in history, culture, and traditions are all assumed under the umbrella of race. Including the variable acknowledges that there are unexplained aspects of default.

Grade Point Average (GPA)

This study shows that the higher a student's grade point average, the lower the probability the student will default. A student with a GPA that falls between 0.00 – 1.99 is three percentage points more likely to default than a student whose GPA is 3.00 – 4.00, controlling for other variables.

This empirical evidence supports the policy of placing an undergraduate student on Academic Probation (AP) if the student's cumulative grade point average falls below 2.00. USF may consider using this 'below 2.00 grade point average cut-off' as an intervention point at which to require the student to receive academic or financial counseling. Knowing that a student whose GPA falls below 2.00 is at greater risk of defaulting on his or her student loans provides an available measure by which USF can strategically intervene. A student who has a grade point average of 2.00 – 2.99 is one percentage point more likely to default as compared to one with a GPA of 3.00 – 4.00, all else considered. Although these students in the 2.00 – 2.99 GPA bracket are at a higher risk of defaulting, USF may want to first focus limited resources on the students with a GPA below 2.00 because of their greater risk.

Incomplete Course Hours
Taking ‘1 to 4’ incomplete course hours raises the probability of default one percentage point for a student compared to having zero hours incomplete, for borrowers who otherwise share the same characteristics. However, taking ‘5 or more’ hours incomplete does not increase the student’s chance of default with the reference being no hours incomplete; from a statistical standpoint these two categories are not different from each other. It is possible that the Number of Hours Incomplete variable is a proxy for length of attendance, and might account for the sign of the coefficient for the ‘5 or more’ hours category.

The Hours Incomplete variable includes incomplete, withdrawn, and unsatisfactory hours. An incomplete may be awarded to an undergraduate student only when a small portion of the student’s work is incomplete and only when the student is otherwise earning a passing grade.

A suggestion is to have all faculty members put a note on their syllabus saying, ‘If you need to drop this class, please get the financial aid office to help you understand the implications that dropping a class might have for your student loans and other financial aid.’ This is an easy, low-cost way of reminding students that withdrawing from a class may affect their loan.

Parent Marital Status
Parent marital status affects the likelihood that a student will default. A student who has parents who are ‘Married’ or remarried is two percentage points less likely to default on his or her loan than a student whose parents are ‘Not married’, holding all other factors constant. The variable group ‘Not married’ includes single, divorced or separated, and widowed statuses. While there are differences among these ‘Not married’ statuses, all of the ‘Not married’ categories are statistically significantly different from the ‘Married’ status, and so were grouped together.[5] On average, a parent who is not married has fewer total resources to draw on than a parent who is married. To the extent that a parent helps the student financially, parental resources might affect the student borrower’s ability to repay loans. In this study, 60 percent of students had a missing value on Parent Marital Status. For this reason, ‘Missing’ was included in the model as its own category. The results show that ‘Missing’ is not statistically significantly different from the ‘Not married’ reference, possibly indicating among other things, that some of the ‘Missing’ values would fall into the ‘Not married’ status.

Father’s Highest Education Level
The education level of a borrower’s father is related to default behavior but only to a certain extent. A student whose father has a high school or college education is two percentage points less likely to default on his or her student loan than a student whose father went as far as middle school or junior high, after all other things have been considered. There is no statistically significant difference between the categories ‘High School’ and ‘College or Beyond’ (results not shown) in predicting the probability of default. The category ‘$1 - $10,000’ on the Total Family Income variable lost its statistical significance once father’s highest level of education was introduced into the model. This means that Father’s Highest Education Level is related to Total Family Income. A father with more education may have a higher earning job. We also tested mother’s highest education level in the model and found it to be not statistically significant.

Researchers include father’s highest education level – and the comparable variable for the mother’s educational attainment – in the analysis of college attendance, retention, and completion because they believe that it reflects the student’s ‘inherited’ valuation of higher education. The theory is that students whose parents did not attend college are less likely to attend or complete

[5] In order of strength of effect on default from highest to lowest: widowed, divorced or separated, single.

college themselves because higher education is not a traditional value of their families. First-generation students may also lack a general awareness of college and its many rules and conventions.

Total Hours

Generally speaking, the longer a student has been in college, the lower the likelihood of defaulting on his or her loan, all other things being equal. This means that the farther a student progresses in college the better, even if the student does not graduate. We created this Total Hours variable by adding Transfer Hours to USF Hours Attempted. Number of Transfer Hours by itself was not statistically significant in the model.

Number of Children

Although this measure has not been available to TG in models that we have done, other research indicates that number of children a borrower has is related to default. Having this measure might increase the performance of the model. One study found that having dependent children increases the probability of default by four percentage points per child (Volkwein & Szelest, 1995). The number of dependent children that a borrower has may be part of the missing explanation for default that is being picked up by the set of race categories in this model. In one study, African American and Hispanic families had almost twice the number of children as White families (Volkwein et. al., 1995).

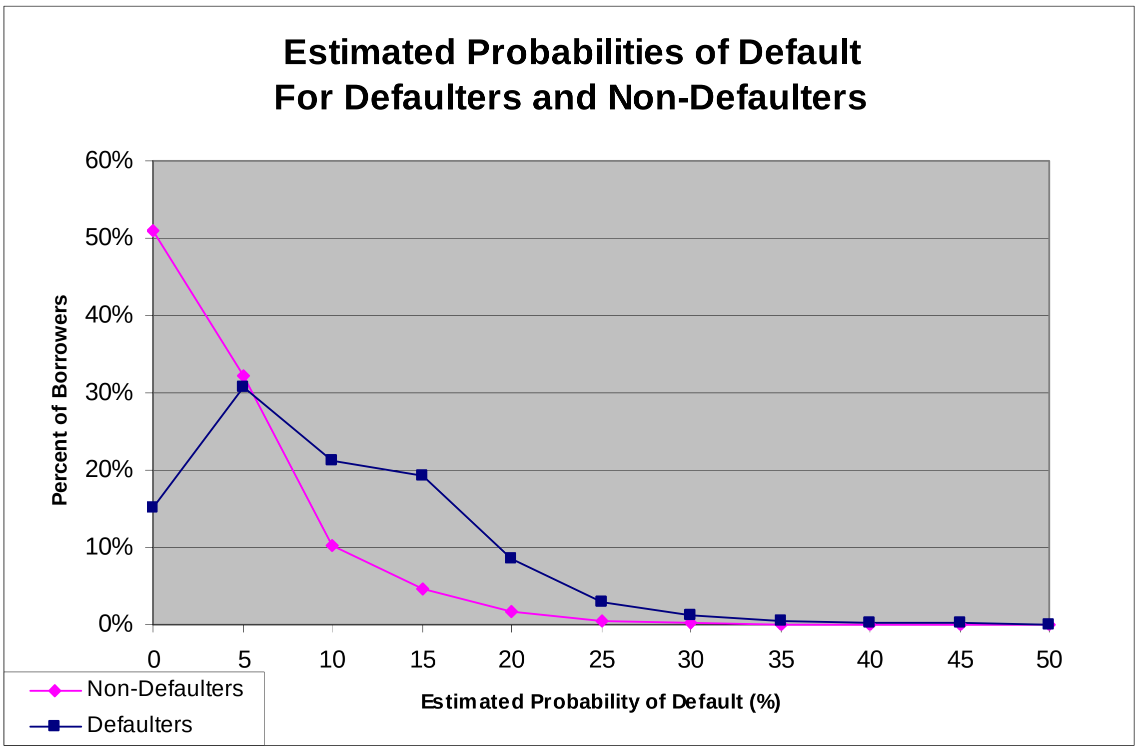
Model Performance

Based upon the characteristics of a borrower, it is possible to sum the coefficients for the variables in the prior section and to convert that sum to a probability that the borrower will default. The estimated probability can then be compared to the known outcome for the borrower. This comparison can be made for all borrowers in the study in order to gauge the performance of the multivariate model. In general, the performance measures in this section assess how well the statistical model correctly classifies defaulters and non-defaulters.

The performance measures indicate that this statistical model performs very well. It does a very good job in assigning high probabilities of default to borrowers who actually defaulted and low likelihoods of default to borrowers who did not actually default.

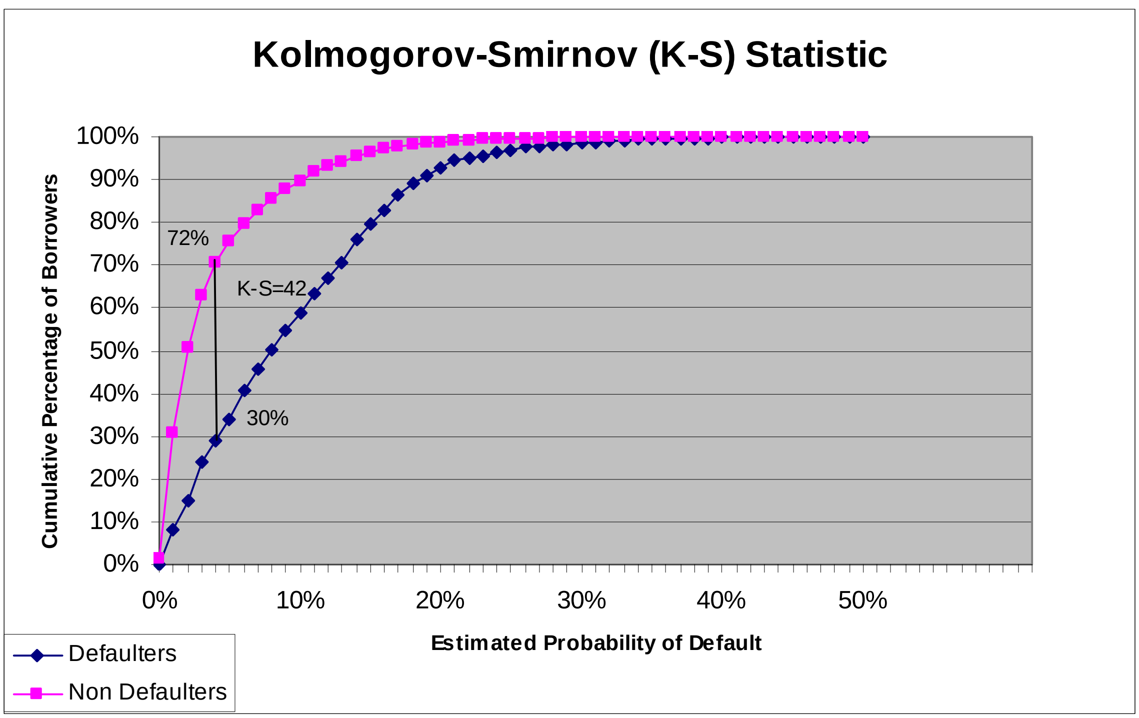
Distribution of Probabilities

The following chart shows the default probabilities assigned by the multivariate model to borrowers in the study. The chart provides a separate distribution of probabilities for actual defaulters and actual non-defaulters. (Each borrower's estimated probability of default was rounded to the nearest five percent.) The vertical axis shows the percentage of borrowers who were assigned each probability. Thus, whereas the model assigned estimates of a zero percent (rounded) probability of default to 51 percent of actual non-defaulters, it assigned a zero percent (rounded) probability of default to only 15 percent of actual defaulters. In general, if the model is performing well, the curve for the non-defaulters should be higher than the curve for the defaulters on the left side of the chart. Similarly, the curve for the defaulters should be higher than the curve for the non-defaulters on the right side of the chart. The visual impression of this chart is that the model appears to have performed well.



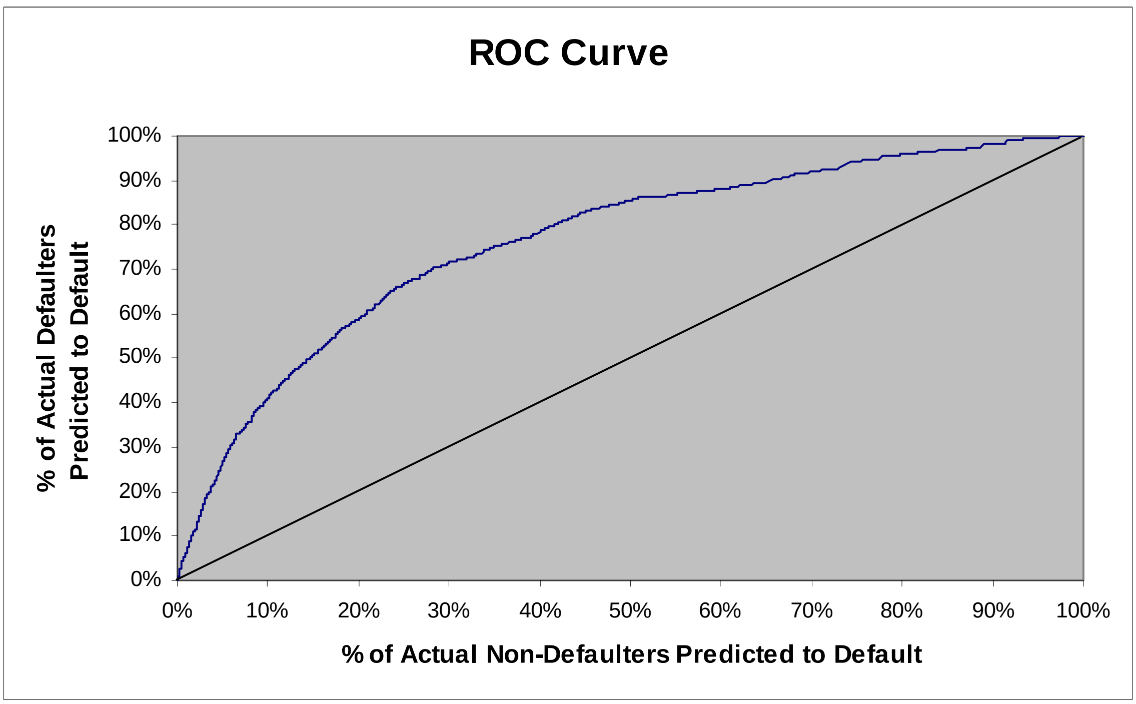
Kolmogorov-Smirnov (K-S) Statistic
The previous distributions can be transformed into a set of cumulative distributions. Cumulative distributions give the percentage of borrowers who have an estimated probability that is equal to, or less than, a given point along the horizontal axis. For example, the chart below shows that 72 percent of actual non-defaulters have an estimated probability of default that is less than or equal to five percent and that only 30 percent of actual defaulters have an estimated probability of default in that range. As it turns out, at five percent (along the horizontal axis), the curves for defaulters and non-defaulters are separated by the greatest distance. This distance is known as the Kolmogorov-Smirnov (K-S) statistic. For the present model, the K-S statistic is 42 percent (72 percent minus 30 percent). Models with large K-S statistics are said to have done a good job of distinguishing between defaulters and non-defaulters. Forty-two percent is a high K-S statistic and indicates that the model does well in separating defaulters and non-defaulters.

A high K-S means that a model will predict default outcomes for a much higher percentage of actual defaulters than non-defaulters. Suppose we predicted default for borrowers who the model assigned a default probability greater than five percent. The K-S of 42 percent indicates that using five percent as the prediction cutoff means that we will predict default 42 percent more frequently for defaulters than for non-defaulters. At five percent, the model would predict 70 percent of actual defaulters to default (that is, one minus the 30 percent with probabilities less than or equal to five percent), but it would only predict 28 percent of actual non-defaulters to default (one minus the 72 percent with probabilities less than or equal to five percent).



C Statistic
The c statistic measures how consistently a model assigns higher probabilities to actual defaulters than it does to actual non-defaulters. It compares each defaulter with each non-defaulter. In the present analysis, there are therefore 12,463,586 pairings (766 defaulters multiplied by 16,271 non-defaulters). The c statistic indicates the proportion of these cases for which the model assigns a higher probability of defaulting to the defaulter than it assigns to the non-defaulter. For the present model, the c statistic is 77 percent – a very high value for this statistic.

Receiver Operating Characteristic (ROC) Curve
The c statistic is represented graphically in the chart below. The area under the curve – called a Receiver Operating Characteristic (ROC) curve – is the c statistic: 77 percent of the chart is below the curve. A statistical model that assigned the same probabilities to defaulters and non-defaulters – a model that does no better than chance – would have an ROC curve that formed a diagonal running from the lower left corner of the chart to the upper right corner. To the extent that an ROC curve bows above the diagonal, the performance of the model increases. A model that perfectly separates defaulters and non-defaulters would have an ROC curve that hugged the left-hand side and top of the chart. The ROC curve for this model ranges well above a diagonal and indicates a high level of performance.



Classification Matrix and Misclassification Rate

Constructing a classification matrix provides an easy way to assess how well the statistical model classifies defaulters and non-defaulters. In the following example, the matrix employs a classification rule: if the model assigns a probability of default of five percent or more, the borrower is classified as a defaulter; a borrower with less than a five percent probability of default is predicted to be a non-defaulter. The matrix shows the numbers of actual defaulters that the classification rule predicts to be defaulters and non-defaulters. It also provides the same information for actual non-defaulters.[6]

		Predicted Outcome	
	N=17,036	Default	Non Default
Actual Outcome	Default	521	245
	Non Default	4,338	11,932

It is possible to derive a misclassification rate from the classification matrix. When the predicted outcome does not align with the actual outcome, the classification rule resulted in a misclassification. The total number of misclassifications (4,583) is the sum of the defaulters who the model predicted to be non-defaulters (4,338) and the non-defaulters who the model predicted to be defaulters (245). The misclassification rate is 27 percent (4,583 divided by 17,036).

Whether or not this misclassification rate is good depends upon the frame of reference. If the school's alternative to using the model is to treat all borrowers as if they are potential defaulters, then a misclassification rate of 27 percent is very good. Treating all borrowers as potential defaulters will misclassify all 16,270 non-defaulters and result in a misclassification rate of 95.5 percent. In this comparison, use of the model produces a three and a half times reduction in the misclassification rate.

Uses of the Findings and the Model

The results of the statistical analyses used in this report provide insight into the nature of student borrower behavior at the University of South Florida. This insight can supplement the professional knowledge of financial aid administrators who work closely with students every day and who have valuable experience administering student assistance programs. This section of the paper offers our suggestions for how to apply the lessons learned from the findings, with the acknowledgment that they are speculative in nature and are shared in the spirit of starting discussions, not ending them.

In mitigating potential default, USF faces a common challenge of how to best allocate limited resources to many students.[7] By identifying USF students who are most likely to default, this study allows the university to strategically direct efforts and effectively target resources toward at-risk students for special intervention. We recognize that any actions implemented must be economical, and be in alignment with existing policies and infrastructure. The authors hope this study will help the University of South Florida use resources in new ways in order to prevent defaults. The greatest success in preventing default may be achieved through campus wide involvement, and through the integration of faculty and administration efforts.

[6] For results from interviews TG did with borrowers based on their predicted classification see Telephone Surveys in Meyer (1998) http://www.tgslc.org/publications/reports/defaults_texas/ins_four.cfm

[7] For an example of how one university provides financial education and counseling to students see the Texas Tech University website 'Red to Black' http://www.orgs.ttu.edu/r2b/R2B.htm

The need for coordination of efforts between offices and colleges across the university is lent strong support by this study's finding on graduation status. Graduation of a borrower has the strongest relationship to default behavior of any variable in the study. Yet no one part of the institution can guarantee that an individual student will graduate. It takes the entire university to graduate a student (assuming the student is willing and able). By analogy, it takes the complete campus to help the financial aid office to prevent student loan defaults, if in fact the student's success in college is so strongly associated with whether or not the student defaults. While offices outside of financial aid might not be able to instruct the student in the fine details of student loan rules, any effort that increases retention and completion will likely reduce the risk of student loan default.

Although default occurs after a student has left school, this study shows there are several important factors during school that affect the student's probability of default. USF can use some of the findings of this study to target intervention strategies to subpopulations of borrowers who appear to be at the highest risk of default, even before the students have left the university. A couple of the significant variables in this study can be measured as the student borrower is attempting coursework at USF. The University of South Florida can use college performance measures such as Number of Hours Failed and Grade Point Average as flags to create an early warning system. If a student fails more than one course hour, that student could be identified for mandatory counseling. Similarly, if a student's GPA drops below 2.00, the student could be singled out for intervention. The goal of monitoring these performance measures is to help the student get on track with his or her studies, and ultimately graduate. It is in the interest of both the student and the school for the student to successfully complete his or her studies.

Ideally, monitoring key factors in real time requires the cooperation of the academic affairs office, and faculty, with the financial aid office. If under the current Satisfactory Academic Progress (SAP) policy, grade point average is monitored by academic affairs, and the financial aid office monitors progress toward degree, then this information could be shared to identify students at risk of defaulting. Good communication between offices will help coordinate efforts. As a large university, USF might incorporate an automated procedure utilizing electronic communication. In addition, faculty members are the first line of defense in addressing student academic performance. At final exams, when the student hands in his or her exam, someone could hand the student a card that explains what to do if the student does not pass the class. Since the student is physically present to write the exam, this might be the last chance the university has of contacting the student in person if that student leaves and never returns.

The counseling offered for students who are not performing well would address the student's loan, as well as other areas as needed. By no means do interventions with high-risk students have to be limited to financial counseling. A low GPA or a failed class might indicate the need for remediation in a certain subject area. Alternatively, low college performance might be a sign that the borrower is in need of tutoring, social services, or health services. The counseling should emphasize that the borrower's ability to repay loans is dependent upon the borrower's ability to command an adequate salary in the labor market, which is turn is very likely connected to whether the borrower succeeds in college.

Counseling may be done by a variety of persons, depending on the type and level of counseling required, to overcome fiscal constraints. Financial aid offices may best do some counseling, and this may call for professional development of financial aid staff. Another way to lower the expense of counseling is to utilize a peer counseling program, which provides credit or practice to students who want to specialize in financial or other types of counseling. Peer counseling may prove most effective in some instances, because the ability of the students to relate to each other creates necessary empathy. Trained graduate students may carry out peer counseling. It is also possible to have certain classes or programs involved in counseling, for example, psychology and sociology teaming up with accounting and economics. The University of South Florida or a community college might offer continuing education credits in financial planning as a way for faculty and alumni to become informed and qualified. Since all of these interventions are potentially expensive, the need for targeting services, and having variables with which to do it, is all the more important.

Further, based on this statistical model, USF can identify at-risk groups of students who would most benefit from personal finance classes or counseling. The research indicates that borrowers who have low total family incomes are at a higher risk of default as compared to borrowers with higher incomes. This measure of Total Family Income appears promising as a means of targeting students before they attempt coursework and before they borrow student loans. As such, USF should be able to identify these higher risk borrowers during the application process for student financial aid, when this information becomes first available. Since lower incomes might be associated with limited knowledge of, and experience with, loans and loan repayment, borrowers who are targeted in this manner might be particularly suited for financial or debt counseling that addresses these issues.

The better a student's money management skills, the greater the student's chances of paying back his or her loan. To view finances holistically, we must consider financial literacy. Knowing how to stretch an income may come, in part, from being financially literate. A very practical approach to finances in a personal finances class could, for example, teach students skills on things like how to get married on a budget. Possibly a personal finance class could be part of the remedial course curriculum, or somehow incorporated into existing core classes. The University of South Florida might consider creative ways to offer rewards to students who complete personal financial counseling or a class, such as a coupon for free coffee or food from an on campus vendor, who might be interested in marketing their product. If not already doing so, USF could partner with lenders and request that they offer training and workshops for students. The University of South Florida might also contact alumni to challenge them to come up with ways to become involved in student financial literacy.

Other types of interventions might enhance in-person counseling or stand by themselves as good solutions. USF could require at-risk borrowers to complete online counseling or instruction on student loans between semesters, supplemental to the regular entrance and exit counseling requirements that are in place. The university could send postcards, letters, or e-mail notifications to targeted borrowers. The university could also leverage points of contact outside the financial aid office in order reinforce the sense of obligation to student loan repayment. Social service or career counseling offices could convey to students that the financial aid office is willing to talk to borrowers about student loans. One idea is to have professors place a message on their course syllabus that says, 'If you need to drop this class, please get the financial aid office to help you understand the implications that dropping a class might have for your student loans and other financial aid.' The possibilities for low-cost interventions that use simple forms of messaging, such as postcards, bulletin boards, posters, e-mail notifications, and content embedded within other communications are limited only by imagination (and, of course, budgets).

As USF experiences a shift in focus from growing in size to increasing quality of the student body, enrollment management plays a heightened role. Academic preparedness helps students successfully navigate college. A limitation of this study is that is does not include measures such as high school class rank or SAT scores (due to the large number of missing values for these cohorts). Recruitment of students who have performed well before college may cause the school's default rate to go down over the long-term. However, if a school has a sophisticated way of identifying who is at risk, such as the findings from this model, it is possible to mitigate that risk by providing help to students who need it, while preserving access to higher education for those students. For a guide on integrating default aversion strategies into institutional enrollment management plans see *A Clear and Present Danger to Institutional and Student Success*. This report shows how campuses lower default rates by engaging with students from pre-admission through graduation.[8]

After a borrower leaves school and the grace period has ended, the borrower enters the repayment phase on a student loan. If a borrower is delinquent on a payment for more than 270 days, the loan goes into default. In a national study, borrowers indicated the most important reasons for default were being unemployed (59 percent) and working at low wages (49 percent) (Volkwein et. al., 1995). Clearly, finding employment is a key element enabling a borrower to pay a student loan, and working for good wages helps as well. These elements can be addressed through the USF Career Center on campus. Helping students make a quick and successful transition from school to work helps in preventing default.

Integrated Default Assistant (IDA)

TG has been an innovative leader in guarantor-provided default aversion programs. The latest tool will be released during the first quarter of 2005. The Integrated Default Assistant (IDA) is a series of reports, self-serve query systems, and letter writing applications that help schools and lenders track default aversion performance on a weekly basis. The tool provides ways for customers to target at-risk borrowers and a method for forecasting future cohort default rates. IDA also produces a report that monitors the default performance of lenders doing business at a particular campus.[9]

Conclusion

A borrower whom the university retains, who persists to graduation, who has a high GPA, and who did not fail courses along the way has a much lower chance of defaulting than an otherwise similar borrower who did not exhibit these signs of success. While such success indicators might say as much about the character and strengths of the students who embody them, the university also makes a difference in the quality of the students who leave its institution. To the extent that this is true, efforts that increase persistence, improve the quality of education, and place students on rewarding career paths — worthy goals in themselves — will incidentally lower default rates. Since so many diverse factors influence whether these goals are achieved, no one campus office can be expected to carry the burden for their achievement. Greater success in these areas requires a systematic and coordinated effort that brings together several administrative and academic offices.

---

[8] http://www.tgslc.org/pdf/default_model.pdf

[9] Given the time lag in entering a cohort and tracking that for two years, relatively new customers with TG will have to rely on a work around solution that uses NSLDS data to supplement TG data.

Appendix

Notes on Model Development
We began model development by testing a group of variables that had been shown by prior, similar studies to be important from both a statistical and theoretical point of view (see the section 'Prior Research on the Factors Relating to Student Loan Default'). We entered these variables into the model simultaneously and tested for statistical significance. After dropping some variables, we introduced other candidate variables. We examined some variables multiple times within the presence of different sets of variables. We placed emphasis upon the theoretical importance of variables and kept a close eye on the interrelationships of variables whether they were included or excluded from the model.

Close attention was paid to how independent variables related to each other. We tested tolerances, examined correlation statistics, and performed variable clustering in order to determine the overlap or collinearity of independent variables. Whenever variables were introduced to or removed from the model, we examined the impacts upon the coefficients and delta-p values of the variables that remained in the model. Each variable was also cross-tabulated with default so we could look for differences in default rates and better understand each variable. All of these efforts helped us refine the model and prepared us for explaining the results.

We transformed many continuous variables into sets of dummy variables, with each dummy variable defining a particular range of the original continuous variable. This approach provided a means of modeling non-uniform or nonlinear relationships between the continuous independent variable and the dependent outcome. For example, the Age variable is statistically significant as both a continuous variable and as a set of dummy variables. However, only the oldest age group from the set of dummy variables is statistically significantly different from the other age categories. In a sense, the coefficient of the original continuous variable represents an 'averaging out' of the power of the highest age category across the continuum of the Age variable. For this reason, the use of the continuous version of the Age variable would be potentially misleading.

There are a couple of other interesting properties associated with producing sets of dummy variables from continuous variables. The categories that contain the lowest and highest values of the continuous variable dampen the effects of outliers by treating them merely as members of groups that also contain many non-outliers. In fact, this 'dampening' effect occurs within each dummy category. For the categorized version of Total Family Income, there is no difference between $10,001 and $20,000. Depending upon their predispositions, modeling practitioners will view this property as either desirable or undesirable. There is undeniably a loss of information in moving from the continuous variable to the set of dummy variables, but there is also a loss of irrelevant information. After all, for Total Family Income, there probably is not a real difference (in terms of impact upon default behavior) between incomes of $10,001 and $10,002 or, for that matter, between $10,001 and $10,500. (The widths of the dummy ranges and the statistical relationships between adjacent categories determine the trade-off between losing relevant and irrelevant information.) Incidentally, we have found in practice that the choice between continuous variables and groups of dummy variables typically has very little effect upon overall model performance (K-S statistic, c statistic, etc.), with dummy variables slightly outperforming their continuous, parent variables at least as often as not.

We defined the ranges of continuous variables for conversion to dummy variables by using a variety of guides. For a given variable, we began by establishing many categories of equal or approximately equal size. We then collapsed adjacent categories that appeared to have similar relationships to default. Often, we could collapse categories because their default rates appeared to be very close. Sometimes, adjacent categories were collapsed after we observed that their coefficients were nearly identical after entering them into the model. In other cases, categories had to be combined because one or more of them contained very low numbers of observations (this occurred with the nominal variables as well). In a couple of instances, we made range cutoffs correspond to institutional practices. For example, the 2.00 grade point cutoff corresponds to the GPA cutoff for USF's policy on Satisfactory Academic Progress.

Cross Validation

Whereas TG normally prefers to validate its models against samples that were randomly held out of the model development process, it is not always feasible to do so. Sometimes breaking the original sample into development and validation samples will result in samples that are too small to support robust statistical estimation. The sample size of interest for the logistic regression analysis of a dichotomous dependent variable is the smaller of the two outcome categories. In default studies, the smaller category is typically the group of defaulters. For the University of South Florida study, there were 766 defaulters in the overall sample (4.5 percent of all borrowers). Based upon experience and training, we prefer not to allow the sample size of the smallest category to drop below 600 observations, since falling below this threshold could result in model instability. With this in mind, we chose to preserve the original sample size and forgo a test against a validation sample. Instead, we have attempted to support validity by assessing the general reasonability of the results, as checked by comparison to prior research as well as our own working hypotheses concerning the variables.

Notes on Variables in the Model

Graduation Status variable is based on Term Graduated USF.

Number of Hours Failed captures over 70 percent of those who defaulted, based on the combined categories of '2 to 9' and '10 or more' hours failed.

Total Family Income had 25.6 percent missing (imputed value $5,000).

Age of Borrower at Time of Entering Repayment is constructed from Student Date of Birth and Repayment Date, both from the National Student Loan Data System (NSLDS).

In this model, 63 students (0.4 percent) had a missing value on Gender, and they were combined with the 'Male' category.

In Race, the 'American Indian, Alaskan Native' group also shows a higher percentage of defaults; however, due to the small number of students in this category, they were combined with the 'Other' category.

The Grade Point Average less than 2.00 captures 46 percent of those who defaulted; GPA had 1.2 percent missing (imputed value 3.10).

Tables for Variables in the Model

Graduation Status	Total		Defaulters	
	N	% of cases	N	% of row
Graduated	10,237	60.1	198	1.9
Did not graduate	6,799	39.9	568	8.4
All Undergraduates	17,036	100	766	4.5

Number of Hours Failed	Total		Defaulters	
	N	% of cases	N	% of row
0	7,210	42.3	148	2.1
1	2,750	16.1	75	2.7
2 to 9	6,446	37.8	462	7.2
10 or more	630	3.7	81	12.9
All Undergraduates	17,036	100	766	4.5

Total Family Income	Total		Defaulters	
	N	% of cases	N	% of row
Up to and including $0	720	4.2	59	8.2
$1 - $10,000	7,118	41.8	349	4.9
$10,001 - $20,000	2,709	15.9	129	4.8
$20,001 - $30,000	2540	14.9	104	4.1
$30,001 - $50,000	1,503	8.8	56	3.7
$50,001 - $70,000	1,302	7.6	38	2.9
$70,001 - $720,000	1,144	6.7	31	2.7
All Undergraduates	17,036	100	766	4.5

Age of Borrower at Time of Entering Repayment	Total		Defaulters	
	N	% of cases	N	% of row
18 to 21	2,359	13.8	186	7.9
22 to 30	10,950	64.3	395	3.6
31 or more	3,727	21.9	185	5.0
All Undergraduates	17,036	100	766	4.5

Gender	Total		Defaulters	
	N	% of cases	N	% of row
Female	10,400	61.0	392	3.8
Male	6,636	39.0	374	5.6
All Undergraduates	17,036	100	766	4.5

Race	Total		Defaulters	
	N	% of cases	N	% of row
Black	2,572	15.1	198	7.7
Hispanic	1,726	10.1	67	3.9
Other	1,114	6.5	56	5.0
White	11,624	68.2	445	3.8
All Undergraduates	17,036	100	766	4.5

Grade Point Average (GPA)	Total		Defaulters	
	N	% of cases	N	% of row
0.00 - 1.99	2,977	17.5	350	11.8
2.00 - 2.99	6,717	39.4	275	4.1
3.00 - 4.00	7,342	43.1	141	1.9
All Undergraduates	17,036	100	766	4.5

Number of Hours Incomplete	Total		Defaulters	
	N	% of cases	N	% of row
0	12,293	72.2	462	3.8
1 to 4	4,122	24.2	269	6.5
5 or more	621	3.6	35	5.6
All Undergraduates	17,036	100	766	4.5

Parent Marital Status	Total		Defaulters	
	N	% of cases	N	% of row
Married	3,199	18.8	110	3.4
Not married	3,600	21.1	208	5.8
Missing	10,237	60.1	448	4.4
All Undergraduates	17,036	100	766	4.5

Father's Highest Education Level	Total		Defaulters	
	N	% of cases	N	% of row
Middle School/Junior High	779	4.6	55	7.1
High School	5,226	30.7	243	4.6
College or Beyond	5,409	31.8	211	3.9
Missing	5,622	33.0	257	4.6
All Undergraduates	17,036	100	766	4.5

Total Hours	Total		Defaulters	
	N	% of cases	N	% of row
0-59	1,942	11.4	162	8.3
60-119	3,503	20.6	223	6.4
120-149	5,754	33.8	182	3.2
150-189	4,327	25.4	144	3.3
190-900	1,510	8.9	55	3.6
All Undergraduates	17,036	100	766	4.5

Additional Tables

Dependency Status	Total		Defaulters	
	N	% of cases	N	% of row
Dependent	5,587	32.8	275	4.9
Independent	7,628	44.8	335	4.4
Missing	3,821	22.4	156	4.1
All Undergraduates	17,036	100	766	4.5

Expected Family Contribution	Total		Defaulters	
	N	% of cases	N	% of row
$0	6,950	40.8	373	5.4
$1-$500	722	4.2	35	4.8
$501-$1,000	762	4.5	47	6.2
$1,001-$2,000	1,404	8.2	70	5.0
$2,001-$3,000	1,184	6.9	42	3.5
$3,001-$5,000	1,789	10.5	75	4.2
$5,001-$7,000	1,240	7.3	51	4.1
$7,001-$10,000	1,140	6.7	19	1.7
$10,001 or more	1,845	10.8	54	2.9
All Undergraduates	17,036	100	766	4.5

Mother's Highest Education Level	Total		Defaulters	
	N	% of cases	N	% of row
Middle School/Junior High	686	4.0	34	5.0
High School	6,206	36.4	280	4.5
College or Beyond	5,068	29.7	220	4.3
Missing	5,076	29.8	232	4.6
All Undergraduates	17,036	100	766	4.5

Number of Major Changes	Total		Defaulters	
	N	% of cases	N	% of row
Missing	1	0	0	0
0	32	0.2	2	6.3
1	6,449	37.9	417	6.5
2	6,096	35.8	219	3.6
3	3,122	18.3	95	3.0
4	974	5.7	28	2.9
5	271	1.6	5	1.8
6 to 9	91	0.5	0	0
All Undergraduates	17,036	100	766	4.5

Number of Summer Terms Enrolled	Total		Defaulters	
	N	% of cases	N	% of row
0	3,342	19.6	240	7.2
1	4,394	25.8	240	5.5
2	4,985	29.3	158	3.2
3	2,630	15.4	84	3.2
4	1,130	6.6	28	2.5
5	372	2.2	12	3.2
6	133	0.8	4	3.0
7 to 11	50	0.3	0	0
All Undergraduates	17,036	100	766	4.5

Number of Terms Enrolled	Total		Defaulters	
	N	% of cases	N	% of row
1	699	4.1	47	6.7
2	1,046	6.1	82	7.8
3	880	5.2	74	8.4
4	991	5.8	73	7.4
5	1,474	8.7	71	4.8
6	1,587	9.3	78	4.9
7	1,509	8.9	65	4.3
8	1,358	8.0	44	3.2
9	1,221	7.2	37	3.0
10	1,143	6.7	45	3.9
11	1,041	6.1	33	3.2
12	907	5.3	33	3.6
13	750	4.4	17	2.3
14	591	3.5	16	2.7
15	426	2.5	15	3.5
16	328	1.9	12	3.7
17	262	1.5	7	2.7
18	200	1.2	3	1.5
19	156	0.9	6	3.8
20	105	0.6	0	0
21	86	0.5	5	5.8
20 to 50	276	1.6	3	1.1
All Undergraduates	17,036	100	766	4.5

Number of Terms Enrolled Less Than Half	Total		Defaulters	
	N	% of cases	N	% of row
0	2,096	12.3	125	6
1	2,739	16.1	160	5.8
2	2,701	15.9	133	4.9
3	2,357	13.8	110	4.7
4	1,845	10.8	77	4.2
5	1,327	7.8	46	3.5
6	1,027	6.0	32	3.1
7	790	4.6	22	2.8
8	572	3.4	14	2.4
9	425	2.5	20	4.7
10 to 41	1,157	6.8	27	2.3
All Undergraduates	17,036	100	766	4.5

Number of Terms Withdrew	Total		Defaulters	
	N	% of cases	N	% of row
0	11,350	66.6	395	3.5
Withdrew One or more Terms	5,686	33.4	371	6.5
All Undergraduates	17,036	100	766	4.5

Number of Transfer Hours	Total		Defaulters	
	N	% of cases	N	% of row
0	2,312	13.6	177	7.7
1 to 30	3,254	19.1	159	4.9
31 to 60	9,308	54.6	345	3.7
91 to 120	1,452	8.5	63	4.3
121 or more	710	4.2	22	3.1
All Undergraduates	17,036	100	766	4.5

Parent Marital Status	Total		Defaulters	
	N	% of cases	N	% of row
Married/Remarried	3,199	18.8	110	3.4
Single	1,420	8.3	70	4.9
Divorced/Separated	1,582	9.3	102	6.4
Widowed	598	3.5	36	6.0
Missing	10,237	60.1	448	4.4
All Undergraduates	17,036	100	766	4.5

Student Marital Status	Total		Defaulters	
	N	% of cases	N	% of row
Not Married	10,726	63.0	514	4.8
Married or Remarried	2,208	13.0	72	3.3
Separated	254	1.5	21	8.3
Missing	3,848	22.6	159	4.1
All Undergraduates	17,036	100	766	4.5

Total Loan Aid	Total		Defaulters	
	N	% of cases	N	% of row
Missing	59	0.3	2	3.4
$0	247	1.4	9	3.6
$1 - $2,500	1,125	6.6	80	7.1
$2,501 - $5,000	2,184	12.8	129	5.9
$5,001 - $7,000	1,782	10.5	93	5.2
$7,001 - $10,000	1,736	10.2	72	4.1
$10,001 - $12,500	1,612	9.5	61	3.8
$12,501 - $16,000	1,885	11.1	66	3.5
$16,001 - $20,000	1,684	9.9	70	4.2
$20,001 - $25,000	1,693	9.9	61	3.6
$25,001 - $33,500	1,606	9.4	74	4.6
$33,501 - $153,000	1,423	8.4	49	3.4
All Undergraduates	17,036	100	766	4.5

Table: 95% Confidence Interval for Coefficients and Standard Errors

Variable Group	Reference Variable	Variable	Coefficient	95% Confidence Interval	Standard Error
Graduation Status	Did not graduate	Graduated	-0.93	-1.15 to -0.71	0.11
Number of Hours Failed	0	1	0.05	-0.25 to 0.35	0.15
	0	2 to 9	0.68	0.44 to 0.92	0.12
	0	10 or more	1.20	0.83 to 1.57	0.19
Total Family Income	Up to and including $0	$1 - $10,000	-0.30	-0.61 to 0.02	0.16
	Up to and including $0	$10,001 - $20,000	-0.49	-0.82 to -0.15	0.17
	Up to and including $0	$20,001 - $30,000	-0.67	-1.02 to -0.32	0.18
	Up to and including $0	$30,001 - $50,000	-0.74	-1.13 to -0.34	0.20
	Up to and including $0	$50,001 - $70,000	-0.99	-1.44 to -0.55	0.23
	Up to and including $0	$70,001 - $720,000	-1.01	-1.50 to -0.54	0.24
Age of Borrower at Time of Entering Repayment	22 to 30	18 to 21	0.21	-0.03 to 0.45	0.12
	22 to 30	31 or more	0.45	0.25 to 0.65	0.10
Gender	Male	Female	-0.32	-0.47 to -0.16	0.08
Race	White	Black	0.33	0.14 to 0.52	0.10
	White	Hispanic	-0.17	-0.45 to 0.09	0.14
	White	Other	0.23	-0.07 to 0.52	0.15
Grade Point Average (GPA)	3.00 - 4.00	1.00 - 1.99	0.55	0.27 to 0.84	0.15
	3.00 - 4.00	2.00 - 2.99	0.25	0.01 to 0.49	0.12
Number of Hours Incomplete	0	1 to 4	0.30	0.12 to 0.47	0.09
	0	5 or more	-0.17	-0.57 to 0.20	0.20
Parent Marital Status	Not married	Married	-0.44	-0.70 to -0.19	0.13
	Not married	Missing	-0.13	-0.35 to 0.09	0.11
Father's Highest Education Level	Middle School/Junior High	High School	-0.46	-0.77 to -0.13	0.16
	Middle School/Junior High	College or Beyond	-0.53	-0.85 to -0.20	0.17
	Middle School/Junior High	Missing	-0.60	-0.92 to -0.26	0.17
Total Hours	0 - 59	60 - 119	-0.26	-0.49 to -0.02	0.12
	0 - 59	120 - 149	-0.22	-0.50 to 0.05	0.14
	0 - 59	150 - 189	-0.41	-0.72 to -0.11	0.15
	0 - 59	190 - 900	-0.65	-1.06 to -0.26	0.20
Intercept			-2.03	-2.58 to -1.51	0.27

Table: 95% Confidence Interval for Change in Probabilities

Variable Group	Reference Variable	Variable	Change in Probability	95% Confidence Interval
Graduation Status	Did not graduate	Graduated	-2.7%	-3.0% to -2.2%
Number of Hours Failed	0	1	0.2%	-0.9% to 1.7%
	0	2 to 9	4.0%	2.3% to 6.1%
	0	10 or more	9.0%	5.3% to 13.9%
Total Family Income	Up to and including $0	$1 - $10,000	-1.1%	-2.0% to 0.1%
	Up to and including $0	$10,001 - $20,000	-1.7%	-2.5% to -0.6%
	Up to and including $0	$20,001 - $30,000	-2.2%	-2.8% to -1.2%
	Up to and including $0	$30,001 - $50,000	-2.3%	-3.0% to -1.3%
	Up to and including $0	$50,001 - $70,000	-2.8%	-3.4% to -1.9%
	Up to and including $0	$70,001 - $720,000	-2.8%	-3.5% to -1.8%
Age of Borrower at Time of Entering Repayment	22 to 30	18 to 21	1.0%	0.1% to 2.4%
	22 to 30	31 or more	2.4%	1.2% to 3.8%
Gender	Male	Female	-1.2%	-1.6% to -0.7%
Race	White	Black	1.7%	0.7% to 2.9%
	White	Hispanic	-0.7%	-1.6% to 0.4%
	White	Other	1.1%	-0.3% to 2.8%
Grade Point Average (GPA)	3.00 - 4.00	1.00 - 1.99	3.1%	1.3% to 5.4%
	3.00 - 4.00	2.00 - 2.99	1.2%	0.1% to 2.7%
Number of Hours Incomplete	0	1 to 4	1.5%	0.6% to 2.5%
	0	5 or more	-0.7%	-1.9% to 1.0%
Parent Marital Status	Not married	Married	-1.6%	-2.2% to -0.8%
	Not married	Missing	-0.5%	-1.3% to 0.4%
Father's Highest Education Level	Middle School/Junior High	High School	-1.6%	-2.4% to -0.5%
	Middle School/Junior High	College or Beyond	-1.8%	-2.5% to -0.8%
	Middle School/Junior High	Missing	-2.0%	-2.7% to -1.0%
Total Hours	0 - 59	60 - 119	-1.0%	-1.7% to -0.1%
	0 - 59	120 - 149	-0.9%	-1.7% to 0.2%
	0 - 59	150 - 189	-1.5%	-2.2% to -0.5%
	0 - 59	190 - 900	-2.1%	-2.9% to -1.0%

Bibliography

Dynarski, M. (1994). Who Defaults on Student Loans? Findings from the National Postsecondary Student Aid Study. *Economics of Education Review*, 13(1), 55-68.

Knapp, L.G., and Seaks, T.G. (1992). An Analysis of the Probability of Default on Federally Guaranteed Student Loans. *Review of Economics and Statistics,* 74(3), 404-411.

Meyer, D. (1998). Predicting which Borrowers are Most Likely to Default. In: Webster, J., Meyer, D. & Arnold, A. Student Loan Defaults in Texas: Yesterday, Today, and Tomorrow. Austin, TX: Texas Guaranteed Student Loan Corporation.

Monteverde, K. (2000). Managing Student Loan Default Risk: Evidence from a Privately Guaranteed Portfolio. *Research in Higher Education*, 41(3), 331-352.

Podgursky, M., Ehlert, M., Monroe, R., Watson, D., and Wittstruck J. (2000). Student Loan Defaults and Enrollment Persistence. *NASFAA Journal of Student Financial Aid*, 32(3), 27-42.

Steiner, M. and Teszler, N. (2005) Multivariate Analysis of Student Loan Defaulters at Texas A&M University. Austin, TX: Texas Guaranteed Student Loan Corporation. January.

Steiner, M. and Teszler, N. (2003). The Characteristics Associated with Student Loan Default at Texas A&M University. Austin, TX: Texas Guaranteed Student Loan Corporation. January.

Volkwein, J. F., Cabrera, A. F., Szelest, B. P., and Napierski, M. (1995). *Characteristics of Student Loan Defaulters Among Different Racial and Ethnic Groups.* Boston, MA: AIR 1995 Annual Forum Paper. (ERIC Document Reproduction Service No. ED 386 972).

Volkwein, J. F., and Szelest, B. P. (1995). Individual and Campus Characteristics Associated with Student Loan Default. *Research in Higher Education*, 36(1), 41-72.

Wilms, W. W., Moore, R. W., and Bolus, R. E. (1987). Whose Fault Is Default? A Study of the Impact of Student Characteristics and Institutional Practices on Guaranteed Student Loan Default Rates in California. *Educational Evaluation and Policy Analysis*, 9(1), 41-54.

Woo, J. H. (2002). Factors Affecting the Probability of Default: Student Loans in California. *NASFAA Journal of Student Financial Aid*, 32(2), 5-23.