The Federal Student Loan Default Cohort: A Case Study

By Thomas A. Flint

Over 1,000 Stafford borrowers who attended a private two-year college were studied. These borrowers represent students in the fiscal year 1990 government default cohort for the institution. In a small sample taken for a pilot study, withdrawal from college appeared to be a significant indicator of potential default status. However, a logistic regression model applied to the entire cohort shows instead that gender, race, age, high school rank of students and their cumulative grade point average at the college are significantly correlated to default status, and that students' amount of classwork taken at the institution reduces the impact of personal characteristics.

It is further shown that student enrollment choices (i.e., by campus or by division), their borrowing characteristics (number of loans and total amount borrowed), and their reasons for leaving college (withdrawal or graduation) are not significantly related to default status when statistically controlling for other variables. Some implications for recruitment and academic policies that can indirectly affect default status are discussed.

Introduction

In the summer of 1992 the author's institution received its notice from the U.S. Department of Education concerning its fiscal year 1990 (FY-90) default cohort on Stafford Loans. For FY-88 and FY-89 the institution's default rate hovered at 12.5%, while for FY-90 the default rate was reported at 17.7% (256 defaulters out of 1,450 borrowers). As a private two-year institution, this rate was slightly above the national average of 14% reported for FY-87 (U.S. Department of Education, 1990c, p. 11). (Note: The institution's FY-91 cohort default rate was 12.2%).

Because of the institutional penalties associated with higher Stafford default rates, the administration decided to take a proactive position by implementing a default management plan. To begin, the college would voluntarily implement the federal government's regulatory requirements for institutions exceeding a 20% default threshold. Beyond implementing these provisions, a sample of student borrowers was drawn from the cohort, including both defaulters and repayers, to identify characteristics which might distinguish these two groups and assist in the institution's goal.

The initial default study of August, 1992, sampled 95 students (35 defaulters and 60 repayers) from the FY-90 cohort. Student background, enrollment, and borrowing data were collected and tabulated on all students (Table 1). Viewing the differences between the groups, 83% of the repayers had graduated from college while only 49% of the defaulters had done so. Of those who graduated, 30% of the repayers had earned an Associate of Applied Science (A.A.S.) degree while only 17% of the defaulters had done so (the remaining graduates having earned a diploma of less than two years' work). Differences based on

Thomas A. Flint, Ph.D., is Vice President for Financial Services at Robert Morris College in Chicago, Illinois.

TABLE 1
Cohort Default Pilot Study, August 1992

Sample Size = 95		N = 35 % In Default	N = 60 % Nondefault
Background Data:			
Race/Ethnicity:	African American	69%	47%
	Hispanic	11%	32%
	Caucasian	20%	21%
Sex:	Male	29%	25%
	Female	71%	75%
Average Age (Years):		28	25
Average H.S. Percentile Rank:		57.26	52.87
Average College G.P.A. (4.0 scale)		2.38	2.95
Enrollment Data:			
Campus:	Carthage	3%	13%
	Chicago	80%	72%
	Springfield	17%	15%
Division:	Day	80%	78%
	Evening	20%	22%
Department:	Business Administration	57%	57%
	Secretarial Science	37%	42%
	Allied Health	6%	1%
Level:	Associate Degree	17%	30%
	Diploma	83%	70%
Completion:	Graduated	49%	83%
	Withdrew	51%	17%
Borrowing Data:			
Average Stafford Borrowed (Cumulative)		$3,189.00	$3,846.00
Number of Loans:	1	49%	17%
	2	51%	50%
	3	0%	33%

gender, amount borrowed, or number of loans appeared to be small. Disproportionate numbers of defaulters compared to repayers are African American, while disproportionate numbers of repayers compared to defaulters are Hispanic.

Paradoxically, the defaulters in this sample have somewhat lower cumulative grade point averages (CGPA) in college than students in repayment but have somewhat higher high school percentile rankings than repayers, suggesting possible differences in the academic rigor of the high schools attended.

This initial study is subject to several limitations. First, the sample is small and no tests of statistical significance were performed. Because the group of defaulters drawn is small, it is questionable how representative student subgroups are, especially when the subgroups comprise only a handful of students. The concerns limit confidence in any conclusions to be drawn from the pilot study. Finally, though students might default on their loans for multiple and complex reasons, the analysis done is simple in nature and fails to control for interrelationships among the variables. For example, no effort was made to identify reasons for withdrawal among the borrowers, which might have some bearing on their likelihood to default on their student loans at a later time. Though modest differences are reported in default rates by campus and by day

versus evening division, without controlling for student characteristics it is impossible to say whether the differences are due to the campus environment or to the kinds of students making different enrollment choices.

Purpose of this Study

The purposes of this study are several. First, the study includes as many student borrower records as possible so that the interpretation of even small subgroup differences can be made with greater confidence. Next, the study matches borrower data to the institutional registrar's enrollment data in much greater detail, including such variables as number of earned credit hours and reasons for withdrawal from college. Also, efforts are made in this study to control interrelationships which exist between variables, so that interpretations of the effects from variables may be made less equivocally. Finally, tests of statistical significance are done.

Method

Sample

In the fall of 1992 the institution obtained the government data tape containing all student borrowers in the cohort and attempted to match each record by student Social Security number with the registrar's record at the institution. The borrowers in the FY-1990 cohort are those whose Stafford Loans came into repayment between July 1, 1990, and June 30, 1991. Most of these borrowers had been enrolled the year before.

Defaulters in this group are those whose claims were paid by the federal government before September 30, 1991 (U.S. Department of Education, 1992). Each record on the data tape represented a separate loan (i.e., freshman, sophomore). The 2,385 loan records in the institutional default cohort, represent 1,313 borrowers. From this group, a total of 1,104 had Social Security numbers matched to academic records for use in this study. The final sample is 1,087. The loss of the remaining records is due either to the discrepant coding of the Social Security numbers between systems or to missing records on the registrar's system (the registrar's records were automated only in 1987). No effort was made to identify reasons for the unmatched records.

The final sample has 180 defaulters (16.5%) and 907 nondefaulters. This ratio closely approximates that in the Department of Education cohort, in which 17.7% are defaulters. Like many other two-year colleges, the institution has an unusually large percentage of minority students. In its 1991 annual report, the institution reports that its student body is 42.5% African American, 28.3% Hispanic, and 1.9% Asian American. The students are approximately 85% female, and about 95% study on a full-time basis, whether enrolled days or evenings. More than 80% of students starting academic programs at this institution graduate. In its 1992-93 aid applicant population, about half are dependent students, having an average family income (taxable and nontaxable) of $30,000, while the average family income of independent filers is $13,000. See Table 2 for descriptive characteristics of the final sample of borrowers. Table 3 describes characteristics by race/ethnicity of the borrowers.

TABLE 2
Descriptive Characteristics of the Sample (N = 1,087)

Variables	Frequency	Percent	Mean	Standard Deviation
Dependent Variable:				
0 = Not in Default	907	83.5%	—	—
1 = In Default	180	16.5%	—	—
Independent Variables:				
Gender: 0 = Males	163	15.0%	—	—
1 = Females	924	85.0%	—	—
Ethnicity: 1 = Asian American	14	1.3%	—	—
1 = African American	446	41.0%	—	—
1 = Hispanic	278	25.6%	—	—
0 = Caucasian	349	32.1%	—	—
Year of Birth (19):	—	—	67.68	3.97
Family Income ($):	—	—	$18,553.79	$16,062.99
High School Percentile Rank:	—	—	60.55	28.17
Transfer Credit Hours:	—	—	1.00	5.11
Campus: Carthage	109	10.0%	—	—
Chicago	826	76.0%	—	—
Springfield	152	14.0%	—	—
Division: 0 = Day	878	80.8%	—	—
1 = Evening	209	19.2%	—	—
Dept.: 1 = Allied Health	40	3.7%	—	—
0 = Business Administration	576	53.0%	—	—
1 = Secretarial	471	43.3%	—	—
Earned Credit Hours	—	—	71.17	22.88
Cumulative Grade Point Average	—	—	3.03	.64
Total Amount Borrowed	—	—	$3,926.04	$1,215.48
Number of Loans	—	—	1.85	.63
Exit Reason: 1 = Withdrawal	191	17.6%	—	—
0 = AAS Degree	296	27.2%	—	—
1 = Diploma	600	55.2%	—	—

See the Appendix for technical notes.

Why Do Students Default?

Stafford Loan defaults can occur for a variety of interconnected reasons. Differences in attitudes toward educational loans vary by family income. Students from low-income families are generally less receptive to borrowing (Mortenson, 1988, 1989). Enrollment patterns, type of academic program, student academic performance, credentials obtained, and the amount borrowed might also be related to student loan default (U.S. Department of Education, 1990c).

The study of defaulted loans described below is based on a model incorporating these possible influences. The purpose of the model is to estimate probabilities that particular students will default on their Stafford Loans. Variables will be introduced into the model in groups or blocks, in order to see which kinds of variables improve the model's predictive success and to observe changes in the influence of variables previously entered. Results from the model may aid in designing institutional default management strategies.

In the first step of analysis, only the pre-college characteristics of students are examined. The rationale for excluding other institutional variables and measures of student/college interactions is that it is possi-

TABLE 3
Descriptive Characteristics by Race/Ethnicity

Upper Row = Number, Lower Row = Row Percentage

	Asian American	African American	Hispanic	Caucasian	Total
Overall:	14	446	278	349	1087
	1.3	41.0	25.6	32.1	100%
Gender:					
Male	5	60	56	42	163
	3.0	36.8	34.4	25.8	100%
Female	9	386	222	307	924
	1.0	41.8	24.0	33.2	100%
H.S. Class Rank:					
0-25% Quartile	2	52	41	45	140
	1.4	37.1	29.3	32.2	100%
26–50% Quartile	5	112	73	93	283
	1.8	39.6	25.8	32.8	100%
51–75% Quartile	3	132	77	89	301
	1.0	43.8	25.6	29.6	100%
76–100% Quartile	4	150	87	122	363
	1.1	41.3	24.0	33.6	100%
Income Level:					
$00000–$10,000	5	235	85	83	408
	1.2	57.6	20.8	20.4	100%
$10,001–$20,000	3	93	62	78	236
	1.3	39.4	26.3	33.0	100%
$20,001–$30,000	1	43	64	80	188
	0.5	22.9	34.0	42.6	100%
$30,001–$40,000	3	38	41	56	138
	2.2	27.5	29.7	40.6	100%
$40,001–$50,000	2	23	24	31	80
	2.5	28.8	30.0	38.7	100%
$50,000 & up	0	14	2	21	37
	0.0	37.8	5.4	56.8	100%
Division:					
Day	12	318	230	318	878
	1.4	36.2	26.2	36.2	100%
Evening	2	128	48	31	209
	1.0	61.2	23.0	14.8	100%
Earned Credit Hours:					
0–16	0	22	8	16	46
	0.0	47.8	17.4	34.8	100%
17–32	1	34	7	5	47
	2.1	72.3	14.9	10.7	100%
33–48	1	18	6	6	31
	3.2	58.0	19.3	19.4	100%
49–64	5	137	108	117	367
	1.4	37.3	29.4	31.9	100%
65–80	3	114	43	121	281
	1.1	40.6	15.3	43.0	100%
81–96	0	27	12	14	53
	0.0	50.9	22.6	26.5	100%
97–112	4	91	92	70	257
	1.6	35.4	35.8	27.2	100%
113 or more	0	3	2	0	5
	0.0	60.0	40.0	0.0	100%

TABLE 3

TABLE 3
Descriptive Characteristics by Race/Ethnicity, *continued*

Upper Row = Number, Lower Row = Row Percentage

	Asian American	African American	Hispanic	Caucasian	Total
College GPA Level:					
0.00–1.00	0	5	1	5	11
	0.0	45.4	9.0	36.6	100%
1.01–2.00	2	22	8	6	38
	5.3	57.9	28.8	19.8	100%
2.01–3.00	5	227	130	89	451
	1.1	50.3	21.0	15.8	100%
3.01–4.00	7	192	139	249	587
	1.2	32.7	23.7	42.4	100%
Cumulative Loan:					
Loan < $1,751	0	21	5	15	41
	0.0	51.2	12.2	36.6	100%
Loan $1,750–$3,500	6	145	110	86	347
	1.7	41.8	31.7	24.8	100%
Loan > $3,500	8	280	163	248	699
	1.1	40.0	23.3	35.6	100%
Number of Loans:					
One	5	130	91	71	297
	1.7	43.8	30.6	23.9	100%
Two	7	268	142	249	666
	1.0	40.2	21.3	37.5	100%
Three	2	44	44	27	117
	1.7	37.6	37.6	23.1	100%
Four	0	3	0	2	5
	0.0	60.0	0.0	40.0	100%
Five	0	1	1	0	2
	0.0	50.0	50.0	0.0	100%
Reasons for Leaving:					
Missing	0	22	6	4	32
	0.0	68.8	18.7	12.5	100%
Voluntary	4	53	24	30	111
	3.6	47.7	21.7	27.0	100%
Academic	0	30	14	3	47
	0.0	63.8	29.8	6.4	100%
Disciplinary	0	1	0	1	2
	0.0	50.0	0.0	50.0	100%
Financial	0	5	3	0	8
	0.0	62.5	37.5	0.0	100%
Diploma Graduate	5	225	136	231	597
	0.8	37.7	22.8	38.7	100%
Degree Graduate	5	110	95	80	290
	1.7	37.9	32.8	27.6	100%
In Default?:					
No	12	329	243	323	907
	1.3	36.3	26.8	35.6	100%
Yes	2	117	35	26	180
	1.1	65.0	19.4	14.5	100%

ble that the tendency to default may be mostly related to background traits which students bring with them to the institution. If this is the case, then it may be erroneous to ascribe responsibility for defaulting to results of interactions between the student and the college environment. In this event the ability of the institution to reduce defaults through its structural or procedural changes may be quite limited (Emmert, 1978). On the other hand, if defaults are unrelated to background characteristics but significantly related to institutional variables, then the institution should expect larger reductions in defaults over time as a result of institutional changes. The pre-college characteristics which will be evaluated here include family income, sex, ethnicity, age, high school achievement, and prior college experience.

Since this institution had not studied its defaulters in depth before, the initial set of variables, beginning with the August 1992 pilot study, is based upon general studies done by others (St. John, 1991). The most recent and extensive research available is from the U.S. Department of Education 1987 National Postsecondary Student Aid Study (NPSAS), which found student demographic characteristics to be related to loan defaults (U.S. Department of Education, 1988). Preliminary results reported by NPSAS reveal that borrowers from families with incomes below $11,000 comprise 9.6% of all borrowers while they comprise 22.8% of defaulted borrowers. While borrowers from families with incomes over $50,000 comprise 9.3% of all borrowers, this high-income group comprises only 3.6% of defaulted borrowers (U.S. Department of Education, 1990a, p. 35). The NPSAS found that 28% of defaulters were older than 25 years at the time of enrollment, whereas only 19% of nondefaulters were of nontraditional college-going age.

NPSAS did not report gender, racial/ethnic background, or academic ability differences between defaulters and nondefaulters. However, it is known that such characteristics have been linked to institutional characteristics, such as selectivity, tuition and fee expenses, and institutional educational expenditures, all of which might plausibly influence post-college behavior (U.S. Department of Education, 1990b). Significant positive correlations have been found linking African American, Hispanic, and female defaulters with higher tuition and fees, and African American defaulters with lower institutional selectivity (Hearn, 1984, 1988, 1992).

It should be added that it is illegal and unethical to deny access to Stafford Loans or to vary the amount of loan awards to students based on personal traits like sex, race, age, or some of the other characteristics here. However, the government has historically limited the amount of Stafford Loan to students based upon financial need, which is in part a function of one of the variables here, family income. Personal characteristics could be used to identify students for whom the institution is eager to send its messages about default prevention. It is reasonable to expect that the government would encourage institutions to identify any kind of student at risk for default and to intervene with counseling and advising programs to reduce the government's default risks.

In the second step of this analysis, institutional characteristics of the program of study in which the student enrolled are entered into the model, in conjunction with the student pre-college characteristics already there. Thus, the question addressed in the second step is: Does the knowledge of students' enrollment patterns significantly add to our ability to correctly predict default status, controlling for existing student characteristics? The institutional variables in this study include campus location, division (day or evening), and department (business administration, administrative support, and allied health).

This institution has had uniquely diverse campuses, all within Illinois. Carthage students attend a residential branch campus in a very small farm community; Chicago students commute to the institution's main campus in the heart of a major urban business district; Springfield students either reside in or commute to a branch in a suburban-like residential/commercial area. Do feelings about the campus environment contribute to eventual loan status? Do the larger size and greater resources of a main campus positively influence repayment behavior compared to the institution's branches? In addition to these questions, the differences in the day and evening divisions will be measured. Student satisfaction with the greater availability of full-time staff and faculty during daytime hours may conceivably influence later repayment behavior.

Students' major field of study has been linked to default status outcomes. Disproportionate numbers of vocational students default on Stafford Loans compared to students in more traditional degree programs (U.S. Department of Education, 1990a). In this study, three programs of study are included: business administration, secretarial science, and allied health. These three represent the only academic departments available at the institution.

In the third step, academic performance data are added to existing variables. Only 10 of the 1087 students had no reported earned credit hours; virtually all borrowers had at least 12 credits or more (one FTE quarter). Students' earned credits and cumulative grade average are a direct index of exposure to and performance within the institution. One might expect those borrowers with more credit and higher grades to be more satisfied with the college and more inclined to repay student loans.

The academic data are entered separately and prior to the addition of loan data in the fourth step for the following reasons. Students are permitted to obtain second and subsequent loans (thereby increasing amounts borrowed) only by virtue of completing earned credits (usually three full-time quarters) and by maintaining satisfactory academic progress (a 2.0 CGPA). Thus, there is a temporal and causal relationship between having more credits and obtaining more loans, so the model imitates this sequence. Even in the case of students who do not complete one full-time academic year, the government's refund policies require that very short attendance periods result in very small outstanding loan balances.

Numbers and amounts of loans are generally related to defaults. The government has noted that "Defaulters had fewer and smaller loans,

"A significantly larger percentage of defaulters than nondefaulters attended proprietary institutions."

which indicates that defaulters attended postsecondary institutions for fewer years or went to relatively inexpensive schools. This statistic may be explained in part by the fact that a significantly larger percentage of defaulters than nondefaulters attended proprietary institutions—schools with programs usually lasting no more than two years." (U.S. Department of Education, 1990c, p. 5). Specifically, about 70% of defaulters had only one loan or borrowed less than $2,500, whereas this was true of only about 45% of nondefaulters.

"The most common reason for default is simply the inability of the borrower to repay the loan for lack of sufficient income."

Again, the questions addressed in the third and fourth steps should be framed with a cumulative understanding of variables contributing to the prediction of default status. Does the knowledge of students' academic performances contribute to the prediction of default, controlling for their pre-college characteristics and their enrollment choices? Does the knowledge of their borrowing behavior add to the ability to correctly predict default status, controlling for the preceding variables?

In the fifth and final step of this analysis, student exit characteristics are combined with the preceding variables in the model. The exit characteristics include the reasons for student withdrawal or the kind of educational credential the graduate earned, having either a diploma or an Associate of Applied Science degree. The circumstances of students' withdrawal may affect prospects for employment and the likelihood of an economic return on investment in college. The level of the educational credentials for those who complete may also affect future employment. The addition of this final variable asks: Does the reason for leaving college contribute to the success of predicting default, controlling for all other variables simultaneously?

Analyses

Regression analyses are the method of choice for questions such as those described above. Several excellent nontechnical descriptions of the theory and methodology of regression analyses applied to higher education topics can be found in books by Astin (1977, 1992) and Pascarella and Terenzini (1991). Readers interested in the technical aspects of the current study are directed to the explanations in the Appendix.

Since the prediction of default status involves a dichotomous outcome (default or not in default), the proper analytic tool is logistic regression (Cabrera, in press). The data were analyzed using the Statistical Package for the Social Sciences (Norusis, 1990). Statistical probabilities reported below refer to the likelihood that the effects observed are due to chance.

Results

First it is important to note that in making errors in predicting default, in *individual cases* one can only err 'one way' since the outcome is one of two possibilities. However, over *multiple predictions* there is more than one way to err: one can too frequently predict default when the truth is otherwise, and similarly, one can too frequently predict repayment. Put in other words, one can 'overpredict' defaults (false positives) or 'underpredict' defaults (false negatives). The latter occurs

here. The underprediction is evident from the first step and is still prevalent in a lesser degree in the final step of the analysis. It is probable that the underprediction could be improved if additional data were collected, such as post-college income and expense data, since these other more immediate and practical factors may dictate the borrower's ability to repay.

Given only student background data, this is how the model performs:

Predicted:	No Default	Default		
Observed:				
No Default	902	5		99.45% Correct
Default	174	6		3.33% Correct
			Overall	83.53% Correct

In the final (fifth) stage of the model, the results are virtually the same:

Predicted:	No Default	Default		
Observed:				
No Default	890	17		98.13% Correct
Default	162	18		10% Correct
			Overall	83.53% Correct

"Withdrawers default more often than graduates."

Viewing the cumulative results over the five stages, it is apparent first that one's knowledge of these variables generally contributes more to the successful prediction of repayment as opposed to the prediction of default. Also, across the five steps the accuracy of the model in predicting defaults improves from about 3% to 10 percent, yet the model still seriously underpredicts actual defaults. What this underprediction suggests is that while the variables may be useful to such prediction, they are insufficient by themselves. What is absent from the model may be most important: knowledge of post-college experiences of the students, such as whether or not the students are employed, how much they earn, etc. It seems plausible to believe such data would contribute to more frequent predictions of default without seriously undermining the existing high accuracy in predictions of repayment.

Tables 4 and 5 present the results of the regression in a format exemplified by Cabrera (in press), with technical notes and references for the tables' entries found in the Appendix. For the analysis, defaulters were coded *1* and repayers coded *0*. Interpretation of the beta coefficients is different for continuous variables such as grade point average versus categorical variables such as race/ethnicity. For categorical variables in this study, the interpretation of the sign of the beta weight is made in reference to students who are Caucasian males, having enrolled at the main campus in the business administration department, day division, and having earned the A.A.S. degree. The Appendix provides more explanation.

In Table 4 each column after the first represents the results from each of the five steps described above, with the bottom row labeled X squared (chi-square) indicating by the level of statistical significance shown whether the addition of a step's variables adds to the prediction of default. The first column's statistical significance represents the

TABLE 4
Logistic Regression Predicting Defaulted Stafford Loans: Effects of Background, Enrollment Choices, Academic Performance, Borrowing Levels, and College Exit Reasons

(Default status coded = 1; See the Appendix for technical notes).

Sample Size = 1,087 Independent Variables:	Step 1: Background Variables Only (Beta)	Step 2: Enrollment Choices Added (Beta)	Step 3: Academic Record Added (Beta)	Step 4: Loan Variables Added (Beta)	Step 5: Exit Reasons Added (Beta)
Female (=1):	.379***	.365**	.307*	.309*	.299*
Ethnicity:					
Asian American (=1)	-.265	-.249	-.158	-.163	-.138
African American (=1)	-.722***	-.696***	-.588***	-.584***	-.574***
Hispanic (=1)	-.267	-.258	-.190	-.190	-.181
Year of Birth	-.054**	-.041*	-.038	-.037	-.038
Family Income	.000	.000	.000	.000	.000
H.S. Percentile Rank	-.009**	-.009**	-.007*	-.007*	-.007*
Transfer Hours	-.065*	-.064*	-.052	-.052	-.050
Campus:					
Carthage		-.006	.077	.100	.101
Springfield		-.056	.035	.038	.042
Division (Evening)		-.209*	-.200	-.195	-.177
Department:					
Secretarial		-.001	-.035	-.041	-.046
Allied Health		.022	-.063	-.058	-.076
Earned Credits			-.009*	-.010*	-.006
Cumulative GPA			-.544***	-.543***	-.507***
Amount Borrowed				.000	.000
Number of Loans				.081	-.056
Exit Reasons:					
Withdrawal					-.217
Diploma					-.055
Intercept	3.390	2.654	4.271	4.148	3.863
G squared	883.136	879.134	847.896	847.595	845.978
d.f.	1078	1073	1071	1069	1067
G squared / d.f.	.819	.819	.792	.793	.793
Pseudo R squared	.079	.004	.028	.000	.001
PCP	83.53	83.81	83.53	83.44	83.53
X squared	92.620***	4.001	31.238***	.301	1.618
d.f.	8	5	2	2	2

Probabilities: * = $p<.05$; ** = $p<.01$; *** = $p<.001$

TABLE 5
Change in Probabilities in the Final (Fifth) Model

Variable	Beta	St. Error	Change in Probability
Female	.299*	.122	.045
Ethnicity:			
Asian American	−.138	.423	−.018
African American	−.574***	.138	−.065
Hispanic	−.181	.161	−.023
Year of Birth	−.038	.022	−.005
Family Income	.000	.000	.000
H.S. percentile Rank	−.007*	.003	−.001
Transfer Hours	−.051	.028	−.007
Campus:			
Carthage	.101	.194	.014
Springfield	.042	.159	.006
Division (Evening)	−.177	.110	−.022
Department:			
Secretarial	−.046	.100	−.006
Allied Health	−.076	.276	−.010
Earned Credits	−.006	.007	−.001
Cumulative GPA	−.508***	.158	−.059
Amount Borrowed	.000	.000	.000
Number of Loans	−.056	.252	−.007
Exit Reasons:			
Withdrawal	−.217	.206	−.028
Diploma	−.055	.154	−.007

Probabilities: * = $p<.05$; ** = $p<.01$; *** = $p<.001$

See the Appendix for technical notes.

improvement over chance alone (that is, no variables), while each succeeding column's significance is contrasted with the column to its immediate left.

In the first stage, gender and ethnicity show significant relationships to loan status ($p<.001$), as do age and high school academic achievement ($p<.01$). Also, in step one prior postsecondary experience is significantly related to repayment behavior ($p<.05$), but family income is not.

As one reads across the rows in Table 4, following the contribution of each variable over several steps, some levels of significance diminish. This happens in part because the effects of students' background characteristics are mediated by their subsequent experiences at the institution. The strongest sign of that fact is the negative coefficient in step three with the entry of the student's cumulative grade average: higher CGPA's mean less likelihood of defaults. Higher CGPA's mitigate the tendency of gender and ethnicity to predict default.

Table 5 presents changes in probabilities as represented by the full model (stage 5). The probability of default decreases by 4.5% for female borrowers compared to male borrowers, while it increases by 6.5% for African American borrowers compared to Caucasian borrowers.

In summary, the results of the regression show that student characteristics before any contact with the institution are significantly associated with the likelihood to default, and that the additional consideration

of student academic performance variables (especially cumulative grade average) both diminishes the significance of the role of demographic factors and entirely removes the significance of choice of division (day or evening) and prior college experience. Enrollment choices, loan variables, and reasons for leaving the institution do not contribute significantly to the ability to predict default.

Limitations of the Model

Notably absent from this model are post-college characteristics of students having loans. Indeed, other research has shown that the most common reason for default is simply the inability of the borrower to repay the loan for lack of sufficient income. Seventy percent of defaulters from two-year colleges and 50% of defaulters from four-year colleges cited unemployment as the cause of their default in the 1987 NPSAS. The second most commonly cited reason after unemployment was underemployment: "working but had insufficient income," claimed by about 50% of former two-year students and by 60% of former four-year students. 30 to 40% cited interference from personal problems, while less than 20% cited dissatisfaction with the academic program, as reasons for their default (U.S. Department of Education, 1990c, p. 6). Certainly it is reasonable to hypothesize that these same problems haunt former students of this college who default.

The job placement records for these borrowers are no longer retained at the institution, and employment records on students who withdrew before completing an academic program have never been obtained. If the purpose of the study were to determine the general reasons for default, including those beyond the scope of practical control by the college, then this omission would be significant. Since the purpose of the study is to determine what the institution can do to reduce defaults, this absence is not critical. Short of hiring all its former students, the college can do little to *directly* effect the economic fortunes of students no longer enrolled. *Indirectly*, the default rate might be improved by changes in recruitment, academic, and financing strategies which improve former students' lives after college.

Other institutional variables missing from the model would conceivably be of value. Perhaps those borrowers employed part-time by the college while they are enrolled here exhibit a sense of gratitude for that experience by being more likely to repay their loans. Again, this is an unmeasured and therefore unknown effect which possibly could improve default prediction.

One pre-college variable which was to be included in the model had to be dropped because of missing data. Although the data tape from the government specified the student filing status (dependent or independent), only one record of the 1,087 contained a non-blank status indicator. Consequently, one should exercise caution in interpreting the family income variable, since income was not calculated identically for all students. Some cases include parent income, others do not. While cases of modestly high family incomes (above $25,000) are almost certainly indicative of dependent students whose parents' income is included in family income, the interpretation of lower incomes is more problematic. A case in which $10,000 or less is reported as family

income could be either an independent student of modest means or a dependent student from a poor family. The author performed separate regressions (not reported here) on this model for dependents and assumed independents (student year of birth prior to 1965). These regressions yielded virtually the same results, including nonsignificance for the family income variable.

Conclusions and Implications

This study is based on a single sample at a single institution. As with any statistical model of this kind, the model works best only for this sample. Consequently, other institutions cannot apply these findings as an analysis of defaults at other campuses. Nonetheless, these findings are generally consistent with some studies done at other institutions (Gray, 1985; Myers & Siera, 1980). In addition, the model may be adapted by other institutions needing to research the importance of variables which contribute to the prediction of defaulted loans.

Several things can be learned from this study. First, regarding the predictive model, it is shown that students' pre-college characteristics are associated with the likelihood to default. Also, the additional consideration of grade point average improves such prediction. What is somewhat surprising is what does *not* contribute significantly: enrollment choices (including day/evening), amounts borrowed, number of loans, and reasons for leaving the college (whether through graduation or withdrawal).

The model underpredicts default status, probably because some relevant data (post-college income and expenses) were not collected and analyzed. Other studies have shown similar limitations (Gray, 1985; Myers & Siera, 1980). Since the student identities are known, the institution could conduct a follow-up survey on all 1087 to obtain this data from a sample of these borrowers.

Regarding changes to institutional practices, it makes more sense to address causes rather than effects. For example, though it seems on the surface that withdrawers default more often than graduates, it appears that mere withdrawal is not the problem so much as low grade point average which contributes to withdrawals. In other words, improving retention without improving student academic achievement may only hold out false hope for reducing the institution's student default rate. Given a choice between higher retention with weaker students, and more withdrawals by talented students, it may even be that the latter may be preferable! High-achievement borrowers probably expect to return and may not want to jeopardize future aid by default, even in the absence of any specific plan to continue their educations.

Based on this evidence, the institution determined that the following directions might be worthwhile:

Financial Strategies
Logically, one way to prevent default is to have fewer borrowers. However, the trend over the last few years is towards more borrowing, not less (U.S. Department of Education, 1988). This study indicates that more borrowing, in itself, is not necessarily a problem. As of January, 1993, though, the institution is likely to reduce the pool of borrowers

by refunding all loan monies received in their first and second terms for students who withdraw. Approximately 75% of all withdrawals at the college occur in terms one and two, and the additional refund expense is manageable at existing withdrawal rates. Of course, this throws out the potentially 'good' future repayers with the potentially 'bad' future defaulters. A more conservative strategy based on this study might call for refunding all loan monies received through the end of the second term only for students below a given CGPA who withdraw. Low achievement borrowers are most likely to be dissatisfied with their college experience, less likely to want to return, and less likely to be in a future financial position for repayment.

Awarding scholarships to mitigate default would seem to be questionable. Scholarships would tend to reduce borrowing, but in this study, amounts borrowed do not seem to be a factor in default. Moreover, default seems likely to be dependent upon students' *post*-college circumstances which are nearly impossible to predict. Race-based or gender-based awards would seem to address the personal traits this study found associated with default, but until recently the U.S. Department of Education has been involved in highly publicized disputes about their legality.

Counseling efforts by financial staff are certainly appropriate, if of unknown effectiveness. At this writing, financial managers at the institution were preparing default awareness workshops to begin in summer 1993. As the college retains existing students for other academic programs, financial staff will work out a model of an ideal monthly loan payment level, expressed as a percentage of after-tax monthly income, for use in setting an institutional recommendation for total borrowing levels for its students.

Academic Strategies

Since CGPA and earned hours are significantly related to default, the obvious implication is to identify and enact programs to improve student achievement. Since the institution has no research or public service mission, this is already the foremost goal of the academic staff. Insofar as the default rate is concerned, it would seem to make sense to continue to admit only those students whose prior academic record makes them good prospects for continuing and succeeding at the college. Future research should focus upon monitoring those initiatives (remedial coursework, support services, student activities, tutoring, on-campus work study jobs, experimental class schedules and term length, etc.) which are correlated to students' greater academic achievement and progress.

Recruitment Strategies

College and curricular choices may become intertwined with decisions about borrowing, which in turn might impact loan repayment. Despite concerns that students' needs for borrowing may influence curricular choices (Kramer and Van Dusen, 1986), there is limited evidence suggesting that no such effect occurs (Pedalino et al, 1992). This study found no evidence that choices of enrollment significantly influence

> *"This study indicates that more borrowing, in itself, is not necessarily a problem."*

default status, so recruitment-related strategies appear unwarranted. For keeping the institution's future default rates under control, it is less risky but increasingly difficult to recruit students with higher levels of high school academic achievement. Success in recruiting higher achieving students, though, should have a double pay-off: they are not only more disposed to repay student loans in the first place, but are also more likely to demonstrate higher achievement at the college, further improving their propensity to repay.

References

Astin, Alexander W. (1977). *Four Critical Years: Effects of College on Beliefs, Attitude, and Knowledge.* San Francisco: Jossey-Bass.

Astin, Alexander W. (1992). What Matters in College: *Four Critical Years Revisited.* San Francisco: Jossey-Bass.

Cabrera, Alberto F. (in press). Logistic regression analysis in higher education: an applied perspective. In J.C. Smart (Ed.). *Higher Education: Handbook of Theory and Research.* New York: Agathon Press.

Emmert, Mark A. (1978). National Direct Student Loan default rates: a measure of administrative quality or something else. *Journal of Student Financial Aid, 8* (3), 43-47.

Gray, Kevin S. (1985). Can student loan default be forecast accurately? *Journal of Student Financial Aid, 15* (1), 31-41.

Hearn, James C. (1984). The relative roles of academic, ascribed, and socio-economic characteristics in college destination. *Sociology of Education, 57,* 22-30.

Hearn, James C. (1988). Attendance at higher-cost colleges: ascribed, socioeconomic and academic influences on student enrollment patterns. *Economics of Education Review, 7* (1), 65-76.

Hearn, James C. (1992). Emerging variations in postsecondary attendance patterns: an investigation of part-time, delayed, and nondegree enrollment. *Research in Higher Education, 33* (6), 657-688.

Kramer, Martin A. & Van Dusen, William D. (1986). Living on credit. *Change, 18* (3), 10-19.

Mortenson, Thomas G. (1988). Attitudes of Americans toward borrowing to finance educational expenses 1959-1983. ACT Student Financial Aid Research Report Series 88-2. Iowa City: American College Testing Program.

Mortenson, Thomas G. (1989). Attitudes towards education loans. *Journal of Student Financial Aid, 19* (3), 38-51.

Myers, Greeley and Siera, Steven. (1980). Development and validation of discriminant analysis models for student loan defaultees and non-defaultees. *Journal of Student Financial Aid, 10* (1), 9-17.

Norusis, Marija J. (1990). *SPSS Advanced Statistics User's Guide.* Chicago: SPSS, Inc.

Pascarella, Ernest T. & Terenzini, Patrick T. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research.* San Francisco: Jossey-Bass.

Pedalino, Marilyn; Chopick, Cynthia; Saunders, Diane; & McHugh, Sean. (1992). The New England Student Loan Survey II. *Journal of Student Financial Aid, 22* (2), 51-60.

St. John, Edward P. (1991). The impact of student financial aid: a review of recent research. *Journal of Student Financial Aid, 21* (1), 118-132.

St. John, Edward P. (1992). Workable models for institutional research on the impact of student financial aid. *Journal of Student Financial Aid, 22* (3), 13-26.

Stage, Frances K. (1990). LISREL: an introduction and applications in higher education. In J.C. Smart (Ed.). *Higher Education: Handbook of Theory and Research.* Vol. VI. (427-466). New York: Agathon Press.

U.S. Department of Education. (1988). Undergraduate Financing of Postsecondary Education: A Report of the 1987 National Postsecondary Student Aid Study. Washington, D.C.: National Center for Education Statistics.

U.S. Department of Education. (1990a). Characteristics of Stafford Loan Recipients, 1988. Washington, D.C.: National Center for Education Statistics.

U.S. Department of Education. (1990b). Profile of Undergraduates in American Postsecondary Institutions. Washington, D.C.: National Center for Education Statistics.

U.S. Department of Education. (1990c). Reducing Student Loan Defaults: A Plan for Action. Washington, D.C.: Author.

U.S. Department of Education. (1992). *Enclosure B: Information on the FY-1990 Cohort Default Rate.* Washington, D.C.: Division of Audit and Program Review, Default Management Section.

Appendix
Technical Notes on Table 2 (Descriptive Characteristics):

Categorical variables (such as gender, ethnicity) were dummy-coded for the regression with the values shown in the column of variable labels. See also the notes below.

Technical Notes on Table 4 (Logistic Regression Results):

The beta values shown carry a sign indicating the direction of the effect which an independent variable has upon the dependent variable. In the case of categorical variables (such as gender or ethnicity), the interpretation of the sign is in reference to an excluded category. For this study, the reference categories are: male, Caucasian, main campus, business administration department, day division, associate degree graduate. Thus, for example, the positive coefficient on beta weights for females means that they are more likely to repay than males, while the negative beta weights for the three ethnic categories shown means each is less likely than whites to repay loans. While the beta weights are used in a regression equation of the standard form, the underlying metric is in logit units, not in the original scale of measurement; see Cabrera (in press) for further information.

The row of values labeled as G squared (also known as the maximum likelihood function) represents a comparison between two alternative models. In Table 4 the model to which a column is compared is that in the column to its immediate left, with the leftmost column being compared to a model with no independent variables. Reduction of values in this function across columns with the associated probability levels which exceed chance indicate significant improvements in the 'fit' between the model and actual outcomes.

The ratio of G squared to its degress of freedom (d.f.) is another indicator of 'fit.' A rule of thumb offered by Stage (1990) is that a hypothesized model be accepted when this ratio is less than 2.5, which is the case for all of the models here.

'Pseudo R squared' represents the proportion of error variance that an alternative model has in relation to the model containing only the intercept (no independent variables); see Cabrera (in press).

The row labeled as PCP (the proportion of cases correctly predicted) is yet another indicator of the validity of a model. Entries in this row indicate the overall success of each model. Details of two of the row's entries, for the first and fifth model, are given in the section on Results.

The row labeled X squared refers to the chi-square statistical test. Degrees of freedom are shown on the line below it. Again, probability levels for each column are assigned based on comparisons made to the column to the immediate left.

Notes on Table 5 (Changes in Probabilities):

This table indicates changes in probability level based on the delta-p statistic. In the case of categorical independent variables, the reference to change in probability is made with respect to an excluded group. Thus, the table shows that the probability of default decreases by 4.5% for female borrowers compared to male borrowers, while it increases by 6.5% for African American borrowers compared to Caucasian borrowers. For continuous variables (such as income or CGPA), the delta-

p represents the change in predicted outcome for each unit change in an independent variable. For example, each unit decrease in cumulative grade point average increases the probability of default by 5.9%. More information on delta-p may be found in St. John (1992).