

Scholarship of Assessment Grant
Truman State University
Grant Report

**PROMOTING STUDENT SUCCESS: DEVELOPMENT AND ASSESSMENT OF AN
EARLY IDENTIFICATION MODEL**

Maria Di Stefano, Sue Pieper, Maureen Bell, Leena Phadke

Purpose

The purpose of this study is to develop a system for early identification of students at risk of academic failure that incorporates both cognitive and non-cognitive factors.

Theoretical Framework

Background

Over the past twenty-five years or so, educational researchers have documented the importance of the first year of college to student retention and success. Typically, advisors, instructors, and administrators responsible for students' first year experience do not yet have evidence of patterns of student performance at the college level. Intuitively, they may tend to rely on available cognitive factors such as high school grade point average (GPA) and standardized test scores to predict which students might need special support. However, such factors have been shown to have limited predictive value.

Over the past year, we have developed a preliminary predictive model to assist us in identifying students at academic risk before they encounter academic difficulty. We have used predictive discriminant analysis to determine which factors from a set of four variables (number of high school Cs, Ds and Fs, high school GPA, high school rank and ACT scores) would most effectively explain the differences between students in good academic standing and students placed on probation and predict risk for academic difficulty on the part of incoming first-year students. We are presently refining our model, seeking to incorporate non-cognitive factors that impact students' success in college. A diagnostic instrument, the College Success Factors Index (CSFI) (Hallberg, Hallberg, & Sauer, 1993), has been identified that focuses on eight non-cognitive factors: responsibility and control, competition and collaboration, task precision, expectations, wellness, time management, college involvement, and family involvement.

The development of an effective predictive model will result in the opportunity to develop concrete, replicable, and effective intervention strategies to address the identified cognitive and non-cognitive risk factors.

Literature review

A recent report from the Education Trust (Carey, 2004) shows that even though enrollment in higher education institutions and the proportion of high school graduates entering colleges and universities in the United States have been steadily increasing, graduation rates have not kept pace and have remained remarkably flat. Graduation rates are significantly lower than the average for minorities and low-income students, information that is particularly troublesome when coupled with the widening gap in average earnings between individuals with high school or some college studies and individuals with bachelor's or advanced degrees. It is imperative that institutions of higher learning assume real accountability by not only attracting and admitting students, but by providing realistic and efficient support systems that will offer all students equal opportunities for success. Academic failure at the undergraduate level, particularly in the first year of study, is one of the main reasons for non-degree-completion (Ting, 1998) and consequently needs to be addressed to reverse the trend in drop-out percentages (Gerald, 1992; Gose, 1996; Ting, 1998).

Some researchers now believe that traditional admissions criteria may not adequately gauge students' potential for academic success at the undergraduate level. Typically, admissions officers examine a combination of standardized test scores and high school GPA when making admissions decisions; however, research suggests that these variables alone do not have high predictive value (Schumacker & Saylor, 1995; Wesley, 1994; Baron & Norman, 1992). Many students admitted to college have fairly high test scores and high GPAs, and, yet, not all succeed. Therefore, it is critical, particularly for admissions officers and academic advisors, to be aware of potential risk factors that may aid in the prediction of academic success or failure.

In recent years, theorists have proposed additional factors that may be related to academic success or failure: level of self-efficacy; locus of control; traits related to a student's belief, behavioral and emotional systems; and motivational style, among others (Beck and Davidson, 2001; Larose, Robertson, Roy, & Legault, 1998; Dollinger, 2000). Therefore, the validation and subsequent use of an instrument designed to measure factors such as control, responsibility, and motivation would be beneficial to the prediction of academic success or failure.

The current study replicates and extends past research in the area of identifying students at academic risk. Similar to previous studies, our prediction variables include traditional cognitive indicators of student success (number of high school Cs, Ds and Fs, high school GPA, high school rank, and ACT scores). However, unlike previous studies, non-cognitive variables are also used for prediction. Specifically, the eight factors measured by the CSFI are included along with traditional indicators of student success to identify students at risk very early in their first year. Preliminary research with a small sample of students in good academic standing, students placed on probation (semester GPA between 1.00 and 1.99), and students placed on probation with contract (semester GPA below 1.00) showed that the number of Cs, Ds and Fs best differentiated among the three groups. For the current study, we have once again selected a statistical model, predictive discriminant analysis, which other researchers have demonstrated has the best success at classifying students into groups and predicting which students will have academic difficulties and which will not (Wilson & Hardgrave, 1998).

Objectives

The first objective of our study was to examine the psychometric properties of the CSFI in order to determine if it was sufficiently reliable and valid to include in our prediction formula. Although the reliability and validity of the CSFI total score have been supported in previous research, earlier studies have also indicated that some of the eight subscales may not possess an acceptable level of reliability for making individual decisions regarding students (Hallberg, Hallberg, & Sauer, 1993). Because we plan to eventually use the eight factors as measured by the CSFI for individual student advising, it was important for us to assess the reliability and validity of the subscale scores as well as the total score. The following research questions were examined:

- ***Will the internal consistency of the CSFI total scale score as well as the eight subscale scores as measured by Cronbach's alpha be at an acceptable level?***
- ***Will the correlations between the eight CSFI subscales be low to moderate, showing evidence of discriminant validity?***
- ***How well does the CSFI total score predict the end of the semester GPA, showing evidence of predictive validity?***

The second objective of our study was to replicate our findings from a preliminary small study. Like the original study, the current study used the four-variable model consisting of number of high school Cs, Ds, and Fs, high school GPA, high school rank, and ACT scores to explain the differences between students in good academic standing and students placed on probation and to predict risk of academic difficulty on the part of incoming first-year students. The size of the student data set was expanded, however, in order to validate the results of the earlier study. The following research question was investigated:

- ***Will the results of the original four-variable prediction model be replicated with a larger sample of students?***

The third objective of our study was to further develop a formula to explain the differences between students in good academic standing and students placed on academic probation and to predict early on which students may experience academic difficulty. The original four-variable prediction model of high school rank, high school GPA, number of Cs, Ds, and Fs, and ACT score was enhanced to include total score on the CSFI. The following research question was studied:

- ***Will adding the CSFI total score to the already validated four-variable model further enhance the prediction formula?***

Methods

Samples

To examine the psychometric properties of the CSFI, we piloted the instrument with a group of 151 freshmen in spring 2004. Nine students who had missing values on one or more of the variables were omitted from the analysis. The total usable sample consisted of 142 students.

In order to replicate our earlier study of the four-variable model with a larger data set, we used a combined data sample of first-time, full-time freshmen from the 2002-2003 and 2003-2004 classes. The 2002-2003 sample consisted of 99 students and the 2003-2004 sample consisted of 1,302 students for a total of 1,401 students. 152 students who had missing values on one or more of the variables were omitted from the analysis. The total usable sample consisted of 1,249 students.

To further develop our prediction formula by adding CSFI total scores to the four-variable model, we combined the two data sets above to make up a subset of freshmen from the 2003-2004 class. The sample consisted of 142 students. Twenty-two students who had missing values on one or more of the variables were omitted from the analysis. The total usable sample consisted of 120 students.

Measures

Students were administered the 80-item CSFI. The CSFI uses 8 subscales of 10 items each to measure eight factors known to be associated with college success: responsibility and control, competition and collaboration, task precision, expectations, wellness, time management, college involvement, and family involvement. In completing the CSFI, students used a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree) to respond to each of 80 items. Each item on the CSFI is positively worded (e.g., "I am in control of my academic life."), so a lower score indicates greater academic success. The CSFI instrument yields a total score and eight subscale scores corresponding to each of the eight factors previously described.

Analyses

The reliability and validity of the CSFI were assessed. Reliability of both the total scale and the eight subscales was assessed using Cronbach's coefficient alpha, a measure of the consistency of items within a scale or subscale. Discriminant validity was assessed by studying between-subscale correlations. Finally, predictive validity was assessed by correlating CSFI total scores with end-of-semester GPAs (end-of-semester GPA was chosen as an indicator of academic success).

Predictive discriminant analysis is a statistical procedure typically used to explain and predict group membership from a set of predictor variables. For this study, the first predictive discriminant analysis was conducted to validate our current formula consisting of high school rank, high school GPA, number of Cs, Ds, and Fs, and ACT score used to predict first year students' academic performance at Truman State University. Like our initial 2002-2003 pilot

study, our first performance group consisted of students in good standing. However, unlike the pilot study that examined two additional performance groups-- students who are placed on probation and students who are placed on probation with contract--for this analysis we collapsed these two groups in order to increase the overall size of the probation group. The current study, then, was conducted to determine how well high school rank, high school GPA, number of Cs, Ds, and Fs, and ACT score would differentiate between two groups of students: students in good standing (a 2.0 or above GPA) and students on probation (a 1.99 or below GPA) and how well the equation developed on this sample can be used to predict group membership.

A second predictive discriminant analysis was conducted in which we added students' total scores on the College Success Factors Index (CSFI) to the already validated set of predictors described above in order to further enhance our predictive formula. The second analysis, then, was conducted to determine how well CSFI total score, high school rank, high school GPA, number of Cs, Ds, and Fs, and ACT score would differentiate between two groups of students: students in good standing (a 2.0 or above GPA) and students on probation (a 1.99 or below GPA) and how well the equation developed on this sample can be used to predict group membership.

Results

In presenting the results of this study, the objectives and research questions will be restated, followed by a description of the pertinent results.

Objective 1

The first objective of our study was to examine the psychometric properties of the CSFI in order to determine if it was sufficiently reliable and valid to include in our prediction formula.

- ***Will the internal consistency of the CSFI total scale score as well as the eight subscale scores as measured by Cronbach's alpha be at an acceptable level?***

The internal consistency of the total scale measured by Cronbach's coefficient alpha was .93, as reported in Table 1, an excellent level for either group or individual interpretation of results.

Cronbach's coefficient alphas for the eight CSFI subscales ranged from .81 for both the Task Precision and the Time Management subscales to .52 for the College Involvement subscale. A total of seven of the eight subscales reached .70, an acceptable level for making group inferences but not acceptable for making individual decisions.

Table 1

Internal Consistency Reliability (Cronbach's coefficient alpha)

Variable	α
Total	.93
Control/Responsibility	.74

Competition	.73
Task Precision	.81
Expectations	.70
Wellness	.72
Time Management	.81
College Involvement	.52
Family Involvement	.72

N = 142

- ***Will the correlations between the eight CSFI subscales be low to moderate, showing evidence of discriminant validity?***

The correlations between the eight subscales were all statistically significant at $p \leq .05$ and ranged from .18 between Competition and Wellness to .76 between Competition and Task Precision, as illustrated in Table 2. Slightly over half of the total correlations (15 out of 28) are in the low to moderate range of less than .50. The Control/Responsibility subscale was highly correlated (greater than .50) with the greatest number of other subscales, a total of six of seven subscales.

Table 2

Between-Subscale Correlations

Variable	Control/ Responsibility	Competition	Task Precision	Expectations	Wellness	Time Mgmt.	College Invlmt.	Family Invlmt.
Control/Responsibility	1							
Competition	.450	1						
Task Precision	.759	.474	1					
Expectations	.655	.519	.652	1				
Wellness	.502	.183	.452	.349	1			
Time Management	.648	.350	.715	.542	.354	1		
College Involvement	.593	.503	.567	.620	.257	.442	1	
Family Involvement	.515	.373	.487	.413	.314	.414	.415	1

N = 142

- ***How well does the CSFI total score predict the end of the semester GPA, showing evidence of predictive validity?***

The correlation between the CSFI total score and the spring 2004 GPA was -.39, a moderate correlation. It should be noted that a negative correlation is desirable in this case because smaller CSFI scores are associated with academic success (higher GPAs) and larger scores are associated with academic difficulty (lower GPAs).

Objective 2

The second objective of our study was to replicate our findings from a previous small study using a larger data set in order to validate the results of our original study. We answered the following research question:

- ***Will the results of the original four-variable prediction model be replicated with a larger sample of students?***

A predictive discriminant analysis was used to determine 1) if students in two groups (students in good standing and students on probation) differed on a set of four cognitive variables (number of high school Cs, Ds, and Fs, high school GPA, high school rank, and ACT scores), and 2) how well the equation developed on this sample can be used in the future to predict group membership.

Table 3 presents a summary of means and standard deviations for the four predictor variables broken down by student academic performance groups.

Table 3

Means (Standard Deviations) for Each Academic Performance Group

Group	N	Rank	GPA	CDF	ACT
Not on Probation	1126	86.001 (12.103)	3.793 (.249)	2.189 (3.677)	27.488 (3.263)
Probation	123	72.236 (14.695)	3.442 (.361)	8.5772 (6.774)	26.203 (2.956)

Multivariate analysis revealed that the discriminant function reliably differentiated between the two student academic performance groups ($\Lambda = .816$, $\chi^2(4) = 252.465$, $p < .000$, $R^2c = .18$). Table 4 presents the standard coefficients and the structure coefficients for the discriminant function. From the standardized coefficients, we see that the number of high school C's, D's, and F's has the largest unique contribution to group separation, followed by high school GPA. The structure coefficients show us that the number of high school C's, D's, and F's also has the highest correlation with the composite. A cutoff of .30 was used to interpret the structure coefficients and label the discriminant function.

Table 4

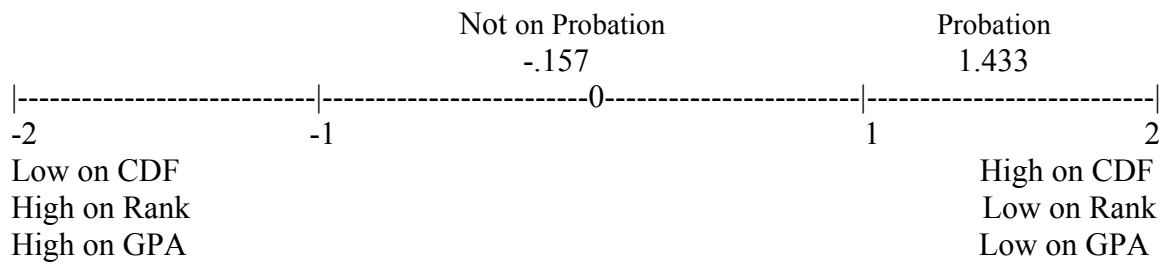
Standardized Canonical Coefficients and Structure Coefficients for the First Discriminant Function

Predictor	Standardized Coefficient	Structure Coefficient
CDF	.850	.984
Rank	.100	-.841
GPA	-.248	-.699
ACT	-.100	-.250

Figure 1 below gives a graphical depiction of the multivariate results. Specifically, the group centroids (means on the composite) are plotted on the labeled function to enable interpretation.

Figure 1

Graphical Depiction of the Discriminant Function



As Figure 1 demonstrates, students on probation had a significantly higher number of high school Cs, Ds, and Fs, were lower in high school rank, and had a lower GPA than students who were not on probation.

Table 5 below illustrates the percent correct classification for both functions. These results help us to know how well the equation developed on this sample can be used to predict group membership.

Table 5

Classification Analysis for Student Groups

Actual Group Membership	Predicted Group Membership			
	Not on Probation		Probation	
	N	%	N	%
Not on Probation	1126	968	158	14.0
Probation	123	44	79	64.2

Note: Overall the percentage of correctly classified cases is 83.8 %.

The overall percentage of cases correctly classified is 83.8% in the sample. Because this value is affected by chance agreement, kappa (κ), an index that corrects for chance agreement, was computed. A kappa value of .36 was obtained. Kappa values of .36 indicate moderately better than chance-level prediction. Kappa values range from -1 to +1. A value of 1 for kappa indicates perfect prediction, a value of 0 indicates chance-level prediction, and a value of less than 0 indicates poorer than chance prediction.

These findings validate the results of the pilot study conducted with a smaller student data set. Once again, the number of high school Cs, Ds, and Fs were found to be the largest contributor to

group separation, followed by high school GPA and high school rank. Additionally, as expected, the overall percentage of correctly classified cases increased from 68.3% to 83.8% when we classified two rather than three groups.

Objective 3

The third objective of our study was to further develop a formula to predict early on which students might experience academic difficulty. The following research question was answered:

- ***Will adding the CSFI total score to the already validated four-variable model further enhance the prediction formula?***

A predictive discriminant analysis was used to determine 1) if students in two groups (students in good standing and students on probation) differed on a set of five variables consisting of the original four cognitive variables (number of high school Cs, Ds, and Fs, high school GPA, high school rank, and ACT scores) as well as a non-cognitive variable (CSFI total score), and 2) how well the equation developed on this sample can be used in the future to predict group membership.

Table 6 presents a summary of means and standard deviations for the five predictor variables broken down by student academic performance groups.

Table 6

Means (Standard Deviations) for Each Academic Performance Group

Group	N	CSFI	Rank	GPA	CDF	ACT
Not on Probation	71	162.606 (28.629)	88.042 (12.419)	3.823 (.230)	1.704 (3.007)	27.887 (3.003)
Probation	49	187.633 (24.748)	71.327 (13.201)	3.461 (.355)	9.061 (6.728)	25.857 (3.035)

Multivariate analysis revealed that the discriminant function reliably differentiated between the two student academic performance groups ($\Lambda = .506$, $\chi^2(5) = 78.634$, $p < .000$, $R^2c = .49$). Table 7 presents the standard coefficients and the structure coefficients for the discriminant function. From the standardized coefficients, we see that the number of high school C's, D's, and F's has the largest unique contribution to group separation, followed by CSFI total score. The structure coefficients show us that the number of high school C's, D's, and F's also has the highest correlation with the composite. A cutoff of .30 was used to interpret the structure coefficients and label the discriminant function.

Table 8

Classification Analysis for Student Groups

Actual Group Membership	Predicted Group Membership				
	Not on Probation		Probation		
	N	%	N	%	
Not on Probation	71	59	83.1	12	16.9
Probation	49	8	16.3	41	83.7

Note: Overall the percentage of correctly classified cases is 83.3%.

The overall percentage of cases correctly classified is 83.3% in the sample. Because this value is affected by chance agreement, kappa (κ), an index that corrects for chance agreement, was computed. A kappa value of .66 was obtained. Kappa values of .66 indicate a solidly better than chance-level prediction. Kappa values range from -1 to +1. A value of 1 for kappa indicates perfect prediction, a value of 0 indicates chance-level prediction, and a value of less than 0 indicates poorer than chance prediction.

As expected, in this discriminant analysis the CSFI total score became a “major player,” second only to number of high school Cs, Ds, and Fs, as far as distinguishing between the two student groups. Furthermore, adding the CSFI total score to our prediction formula increased the amount of variance explained in the differences between the two groups, as indicated by the squared canonical correlation, from 18% to 49%. These are both large effect sizes, but adding the CSFI scores significantly boosted our explanatory power. Although the overall percent of correctly classified cases remains similar with both formulas, our ability to predict both students who will go on probation and students who will remain in good standing is equally good with the enhanced formula. In contrast, the original formula did a good job of predicting students who will remain in good standing but did less well at predicting students who will go on probation.

Discussion

The results of this study provide support for the psychometric soundness of the CSFI total scale scores. The internal consistency of the CSFI total scores is at an excellent level for making either individual or group inferences. Beyond a desirable level of internal consistency, the CSFI total scores were moderately correlated with students’ end-of-the-semester GPA. The results were not as encouraging, however, for the eight CSFI subscales. The analysis of the internal consistencies of the subscales as well as of the between-subscale correlations indicated less than desirable reliability for several of the subscales and questionable discriminant validity for one subscale. Although most of the subscales achieved an acceptable level of reliability for making group inferences, at least some of the eight CSFI subscales will need to be further refined before we will feel comfortable using subscale scores on our campus for individual student advising and/or intervention.

The current study successfully replicated the previous small study using a four-variable model of number of high school Cs, Ds, and Fs, high school GPA, high school rank, and ACT scores to predict student success or failure. As expected, the number of high school Cs, Ds, and Fs was the largest contributor to the separation between students who were in good academic standing and students placed on probation. As anticipated, the non-cognitive variables measured by the CSFI total score also became an important component of our model to predict college students' success. The addition of the CSFI total score to the predictive formula increased the amount of variance explained in the differences between the two groups and increased our ability to predict those students who will be placed on probation. Adding the non-cognitive variables to the traditional cognitive measures significantly enhanced our ability to separate students who experience academic success from students who are at risk for academic failure. Ultimately, the development of this prediction model will assist us in identifying students early on who are at risk for academic difficulty and allow us to provide needed academic as well as non-academic support.

Although the use of students' total scores on the CSFI significantly enhanced our prediction formula and the CSFI was found to be a well functioning instrument overall, we continue to have concerns regarding the CSFI subscales. We want to use the CSFI in advising individual students, but currently we cannot be confident regarding the inferences we might make based on the subscale scores. Further analysis of the CSFI conducted with a larger sample would allow us to determine if some of the items need to be revised. Given the costs involved, it remains to be determined if we should extend this investigation by administering the CSFI to all or a sample of first-year students.

Once the pending issues concerning the reliability of the CSFI subscale scores are resolved, we would like to see an intervention program that makes use of the CSFI. The predictive model enhanced with the CSFI total score would be used to identify students potentially at risk; advisors could devote particular attention to areas of concern within the eight factors in their discussions with those advisees and together develop a plan for addressing them. A structured framework for addressing those concerns would be provided through the Academic Planning course, and/or through specific workshops offered by the Residential College Program advisors organized around the eight CSFI factors. Following the Truman tradition of peer role modeling, study centers could be established where upperclassmen would model successful study habits for first-year students. The focus would be not on tutoring in specific disciplines (even though this could also be integrated in the center), but on general academic skills applicable to any area.

Taken as a whole, this study contributes uniquely to current theory and research regarding the promotion of academic success among first-year students. Like many previous studies, we use traditional indicators and the statistical technique of predictive discriminant analysis to develop a prediction formula that will help us to identify early on which students might be at risk of academic failure. Unlike previous studies, however, we added to the traditional measures a measure of non-cognitive factors known to contribute to student success in college. The addition of non-cognitive factors is highly relevant to the work many college professionals do and has been strongly supported and encouraged, especially among academic advisors. This research was presented at the National Academic Advisor Association Conference this past October and was very well received. Many advisors expressed interest in trying the prediction formula at

their own universities and in learning more about their students through the use of an assessment instrument such as the CSFI. Advisors would like to be able to predict students who may experience academic difficulty and also identify the areas these students struggle with so that they may work to help these students in their academic careers.

Although the current study was conducted at a highly selective Midwestern public liberal arts and sciences university, we feel confident that our findings could extend to other, similar student populations such as students typically found in college honors programs. The results of this study may also generalize to other, different student populations, for example older students and first-generation college students for whom the non-traditional factors may play a significant role in determining academic success. Further studies will be needed at other institutions to assess the generalizability of our findings to various student populations.

REFERENCES

- Baron, J. & Norman, M.F. (1992). `SATs, achievement tests, and high school class rank as predictors of college performance. *Educational and Psychological Measurement, 52*(4), 1047-1055.
- Beck, H.P. & Davidson, W.D. (2001). Establishing an early warning system: Predicting low grades in college students from Survey of Academic Orientations scores. *Research in Higher Education, 42*(6), 709-723.
- Carey, K. (2004). A matter of degrees: Improving graduation rates in four-year colleges and universities. Washington D.C.: The Education Trust.
- Dollinger, S.J. (2000). Locus of control and incidental learning: An application to college student success. *College Student Journal, 34*(4), 537-540.
- Gerald, D.E. (1992). Projections of education statistics to 2003. Washington D.C.: National Center for Education Statistics.
- Gose, B. (1996). New programs for freshmen smooth transition to college life. *The Chronicle of Higher Education-Academic Today*. Accessed from today@chronicle.com.

- Larose, S., Robertson, D.U., Roy, R., & Legault, F. (1998). Nonintellectual learning factors as determinants for success in college. *Research in Higher Education, 39*, 275-297.
- Schumaker, R.E., Sayler, M. (1995). Identifying at-risk gifted students in an early college entrance program. *Roeper Review, 18*, 126-129.
- Ting, S. (1998). First-year grades and academic progress of college students of first-generation and low-income families. *The Journal of College Admission, 158*, 15-23.
- Wesley, J.C. (1994). Effects of ability, high school achievement, and procrastinatory behavior on college performance. *Educational and Psychological Measurement, 54*(2), 404-408.
- Wilson, R.L. & Hardgrave, B.C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational & Psychological Measurement, 55*, 186-195.