ACT Research Report Series 92-4

Factors Affecting Student Persistence: A Longitudinal Study

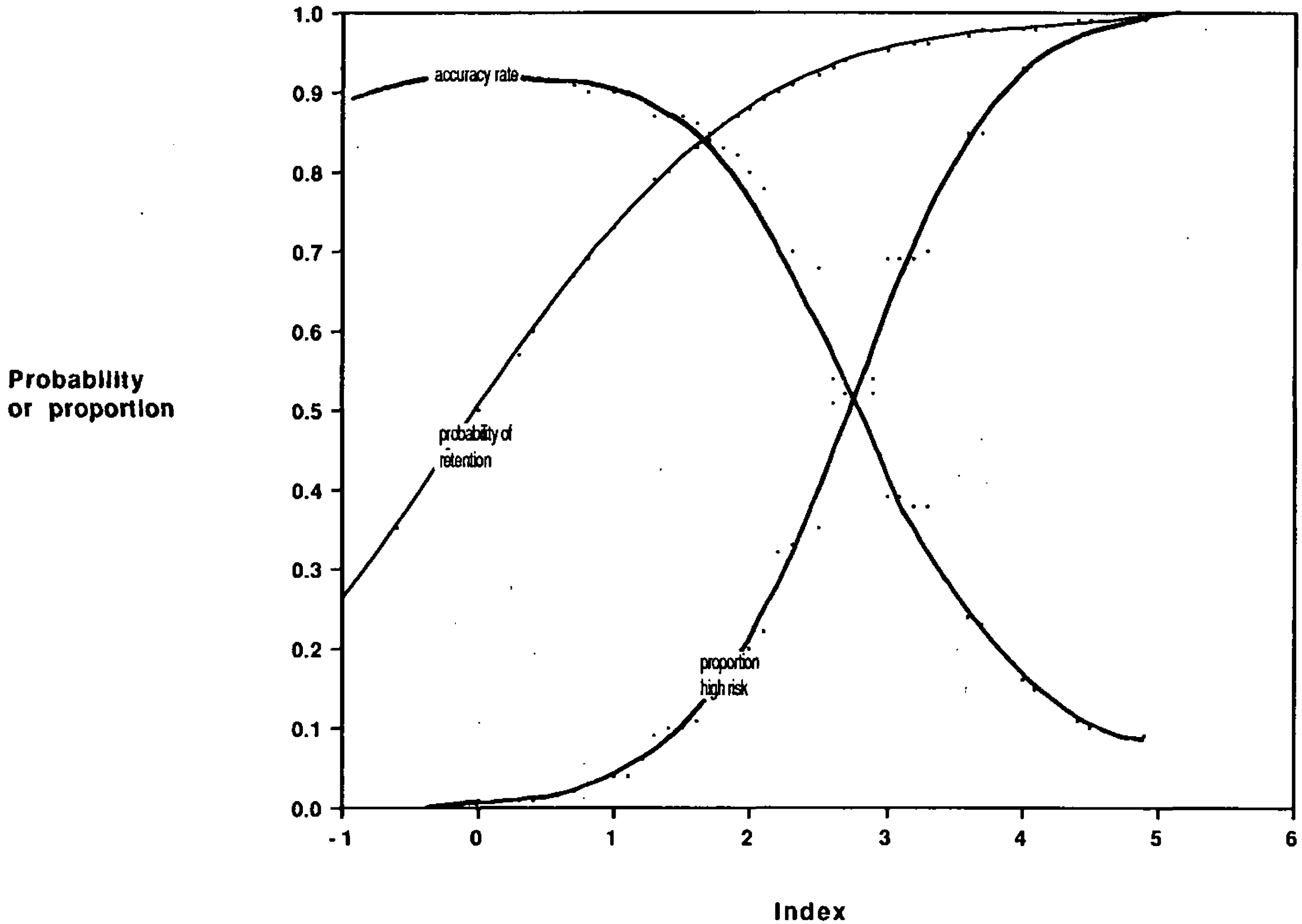
Maggie Gillespie
Julie Noble

November 1992

ACT

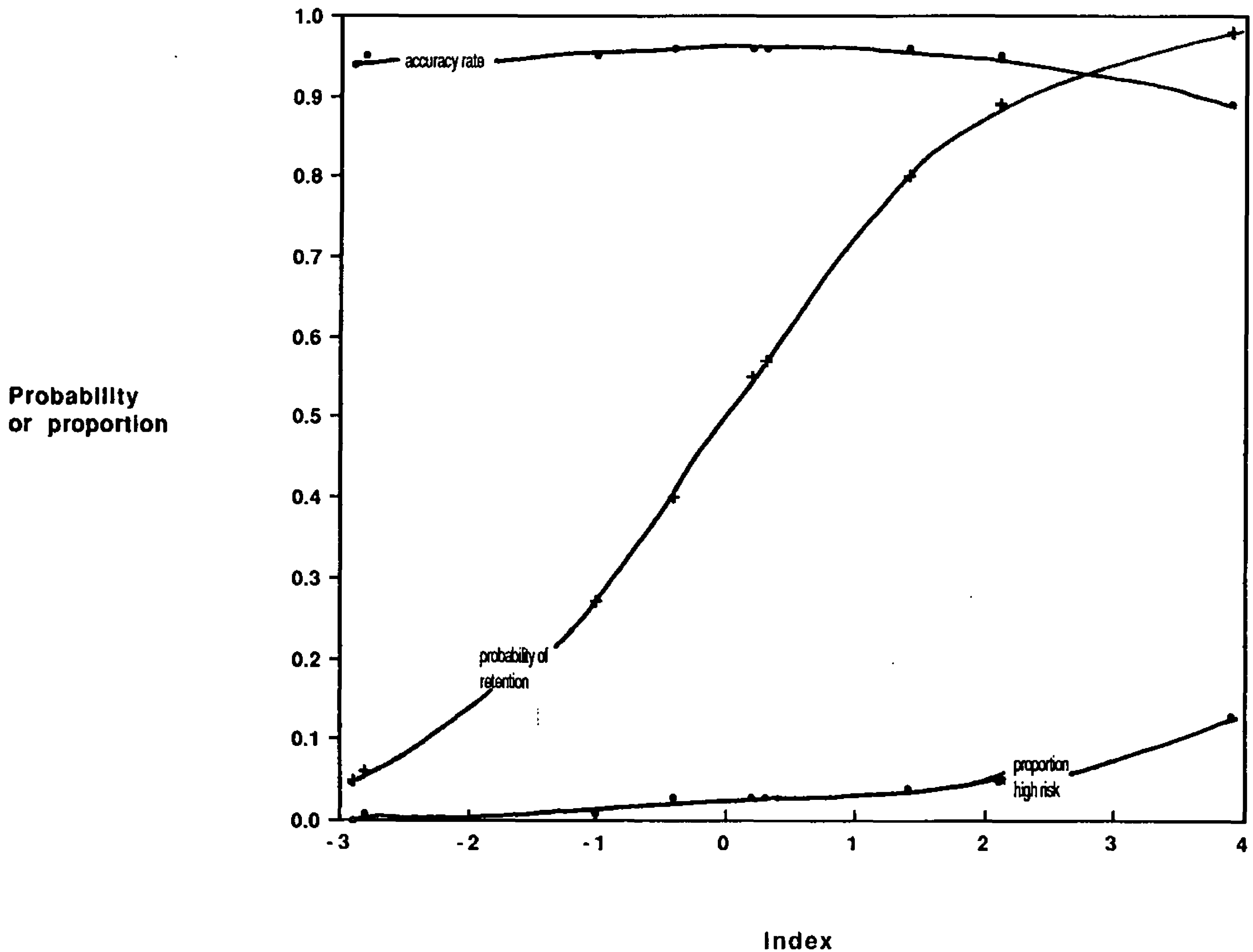
For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

Figure 1
Statistics Associated with First-term Retention at Institution 2 Using Pre-enrollment Data



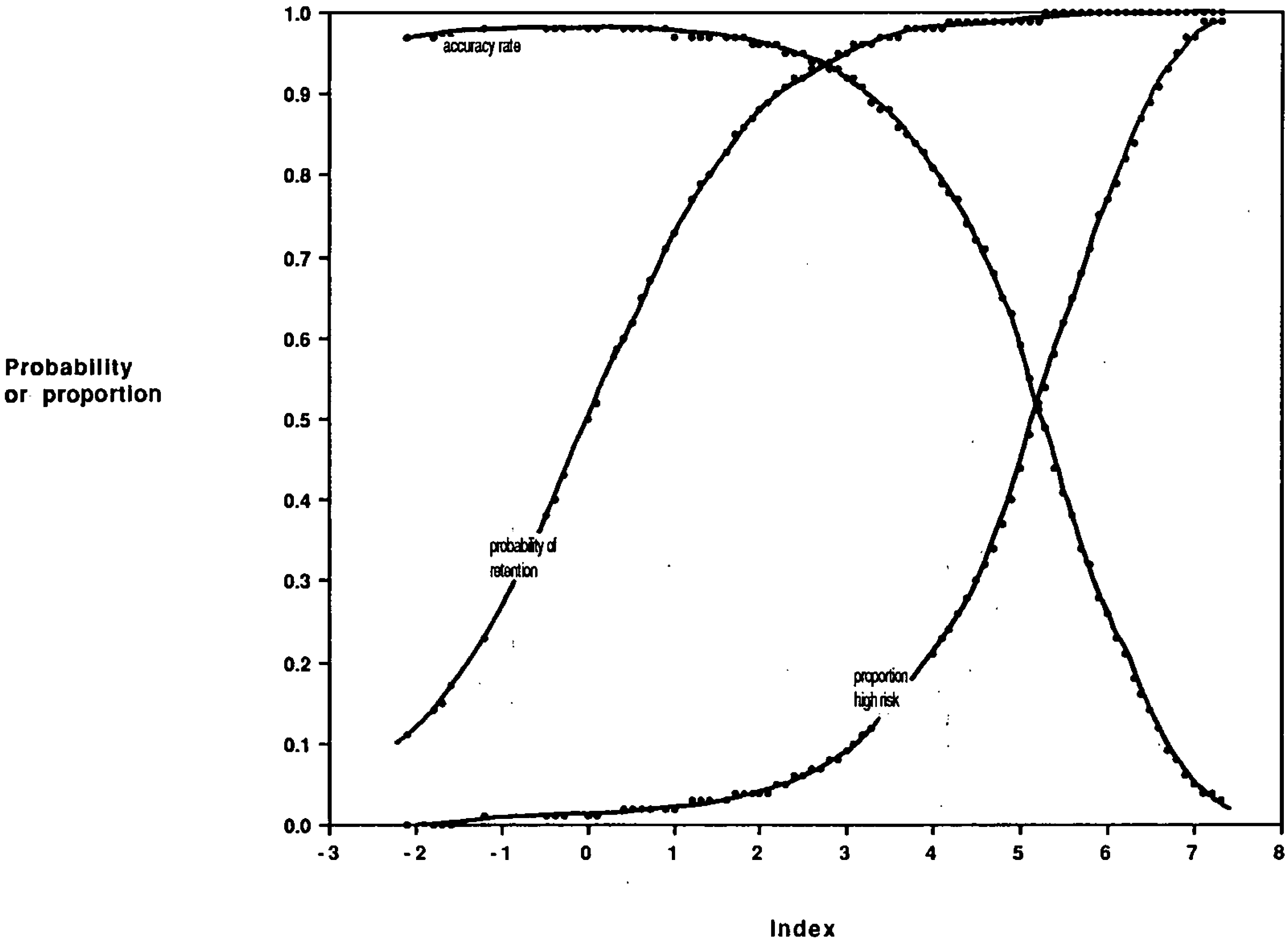
Note: "Index" is a weighted combination of the predictor variables

Figure 2
Statistics Associated with First term Retention at Institution 2, Using Pre-enrollment and First term Data



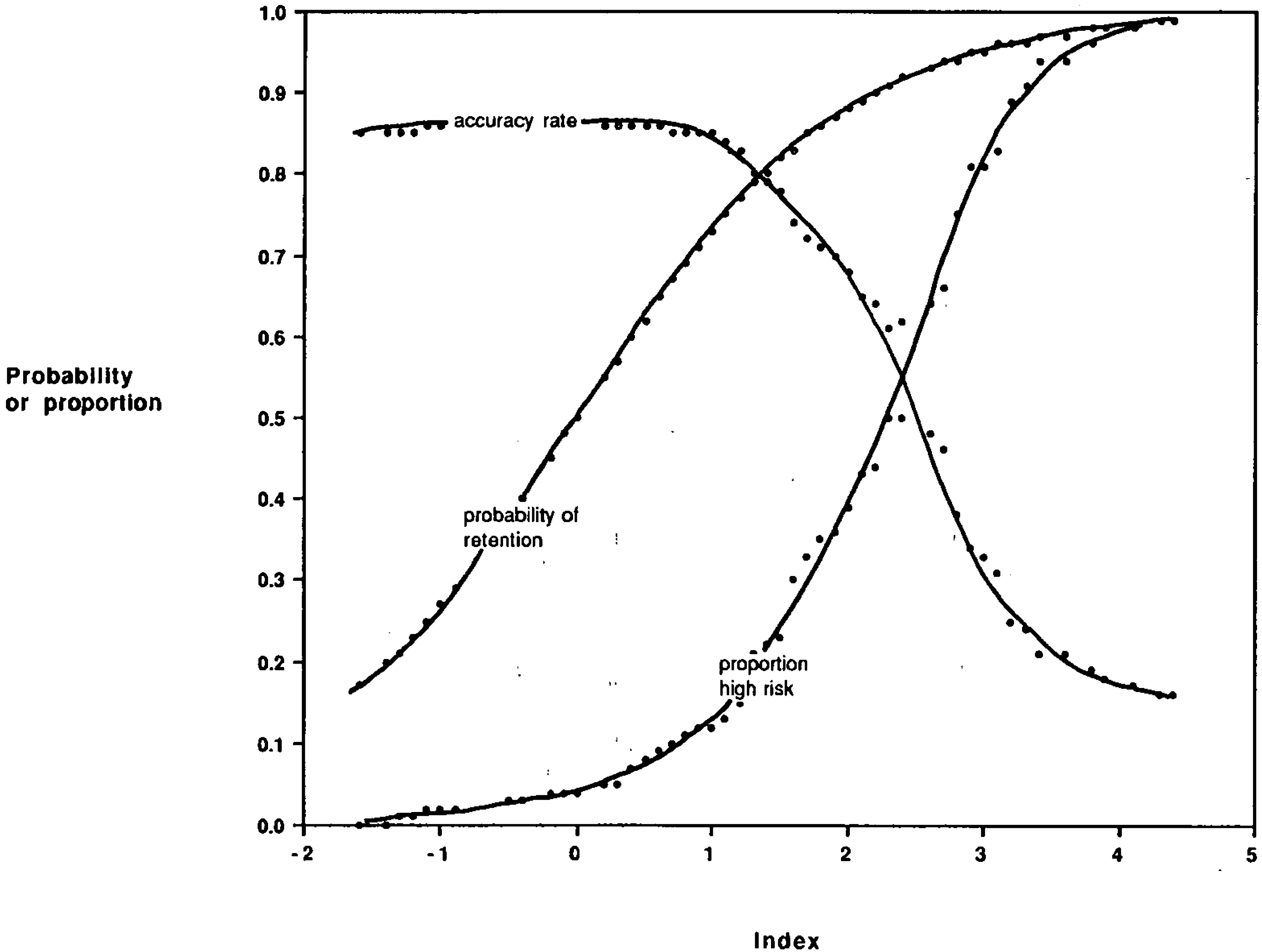
Note: "Index" is a weighted combination of the predictor variables

Figure 3
Statistics Associated with Spring term
Retention at Institution 2



Note: "Index" is a weighted combination of the predictor variables

Figure 4
Statistics Associated with Sophomore Re-enrollment at Institution 2



Note: "Index" is a weighted combination of the predictor variables

FACTORS AFFECTING STUDENT PERSISTENCE: A LONGITUDINAL STUDY

Maggie Gillespie
Julie Noble

ABSTRACT

The purpose of this study was to identify student and institutional characteristics related to college freshman persistence. Persistence was examined for five institutions at four points in time: end of first term, re-enrollment in the spring term, end of spring term, and re-enrollment in the fall of the sophomore year. Data from a variety of sources were used; predictor variables were selected to represent the components of Tinto's model of persistence. Both linear and logistic regression were used to develop separate prediction models for each institution. Estimated success rates and accuracy rates of the models for identifying high-risk students were calculated from the probabilities generated by logistic regression.

The results supported Tinto's view that persistence models are specific to individual institutions and to the time period being examined. An example is provided for using such results for identifying high-risk students and for developing intervention strategies targeting key factors related to student retention. A discussion of the practical issues involved in collecting retention data is also provided.

FACTORS AFFECTING STUDENT PERSISTENCE: A LONGITUDINAL STUDY

Student retention is a major and on-going concern at postsecondary institutions. With restrictions on financial resources and decreases in the traditional-age college freshman pool, institutions are striving to find ways to identify and retain potential dropouts. An effective identification process would enable an institution to assist potential dropouts through advising, counseling, or other interventions. This could result in a more productive and satisfying educational experience for students, and an improved retention rate for the institution.

Earlier Research

Numerous studies have been conducted to examine student retention; several propose theoretical models to explain student retention (e.g., Spady, 1971; Pascarella, 1980; Bean, 1986). The model developed by Tinto (1975) is probably the one most widely tested (e.g., Terenzini & Pascarella, 1977; Munro, 1981; Pascarella & Terenzini, 1983; Noble, 1988; Halpin, 1990). Tinto's model emphasizes integration and commitment: students' background characteristics (family background, individual attributes, and precollege schooling) interact and influence students' initial commitment to the goal of college completion (goal commitment) and initial commitment to the institution (institutional commitment). These commitments in turn influence students' intellectual development and academic performance, which determine academic integration. Goal and institutional commitment also influence peer group and faculty interactions, which determine one aspect of social integration. Increased academic integration and social integration lead to greater goal commitment and institutional commitment, which reduce the probability of the student dropping out.

Though many studies have examined student persistence, they have often not included one or more important factors: a focus on the relationship between student characteristics and institutional environment, an adequate definition of "dropout," a theoretical framework to explain the attrition process, appropriate multivariate statistical techniques, and the use of several institutions or a representative sample to support generalizable results (Pallett, 1984). In addition, the majority of the studies do not include all of the major components that research has shown to be related to persistence (background characteristics, initial and subsequent institutional and goal commitment, and social and academic integration). The few studies that have included all of these components typically used a limited sample of students from a single institution and obtained relatively low response rates.

Tinto (1975) argued that his model was an institutional model, not a model for general use across all institutions. He emphasized the importance of the interactions of the individual student with the institution. This argument was supported by Bean (1986), who stated that the heterogeneity of students and institutions "indicates that a single model of attrition will tend to work poorly in explaining the dropout process for individual students at particular institutions" (p. 49). These views support the practice of developing attrition models and measures of student/institution fit for specific institutions or types of institutions.

Cope (1978) emphasized the need for an early identification of potential dropouts, with the use of "readily ascertainable" characteristics of students (e.g., high school background, demographic variables). Lenning (1982) also supported the use of easily-obtained data, with the view that this type of data might provide comparable accuracy to other more costly and difficult measures. The data currently available from college admissions tests like the ACT Assessment are easily obtained, and may provide a means to identify potential dropouts. In addition, questionnaires administered to students during the freshman year would provide relevant information to measure institutional and goal commitment and academic and social integration, which are important components of Tinto's model.

Noble (1988) studied Tinto's model of attrition using ACT Assessment and ACT's Evaluation Survey Services survey data as predictors of freshman persistence. The data for the study consisted of matched ACT Assessment and Entering Student Survey records, and course credit and GPA data provided by 14 institutions. The results showed that, in combination, ACT student background information, Entering Student Survey data, cumulative GPA, and full-time/part-time enrollment were moderately effective in predicting freshman persistence (median multiple R=.52). However, a crossvalidation analysis at three institutions revealed a large reduction in prediction accuracy. Furthermore, linear regression was used to develop the prediction models. Linear regression is not well suited to modelling dichotomous variables. A more appropriate method involves the use of logistic regression.

The present study was intended to overcome these limitations and problems. The variables included in the study emphasize those issues that are important elements of Tinto's model, but that many research studies have found difficult to measure adequately--students' expectations of the institution and of themselves, and social and academic integration. In addition, the data for the study were collected throughout the freshman year, during the critical time periods for student persistence. Approximately three-fourths of all dropouts leave sometime during the freshman year (Tinto, 1987); most of these dropouts leave during the first six weeks of the fall term (Blanc, DeBuhr, & Martin, 1983). Persistence studies typically lack information that would allow one to identify these at-risk students because they do not assess students early in the term. Furthermore, whereas most persistence studies are based on outcome data collected at one point in time, this study involved multiple data collection points occuring before college entry and throughout the freshman year. Because of the longitudinal nature of this study and the fact that data were collected at critical points in time, it was expected that the findings would be more comprehensive than those of previous studies.

The definition of dropout in this study included any student who left the institution during a specified time period. Because the students were all entering freshmen and the study followed them only through the beginning of the sophomore year, it was assumed that any student who left during this time did not complete a degree program. All dropouts, regardless of their reasons for leaving, are generally of interest to institutions. It may be desirable, however, to use information collected prior to enrollment or early in the year regarding individual enrollment plans and goals in deciding whether a student is a candidate for intervention.

The overall purpose of the study was to determine the validity of using students' ACT Assessment scores, background characteristics, perceptions of the college, expectations, and college credit and GPA information for predicting college freshman persistence. The focus of this paper is on the utility of persistence models developed during the first year of college, and on potential ways of using the results in developing interventions for high-risk students.

This study was primarily designed as a statistical exploration and identification of variables that are highly related to student persistence. The anticipated result, however, is a practical one: the knowledge gained from the study will enable institutions to develop efficient methods for identifying high-risk students and recommending interventions for keeping them in school.

Data

The predictor variables identified for the study were drawn from current research on persistence. They included the following.

A. Background information

1. Demographic characteristics (sex, race, age, etc.)
2. Academic development/ability (ACT test scores, high school GPA, courses taken, etc.)
3. Nature of high school preparation (self-reported adequacy, curriculum type, SES level of the school)
4. Extracurricular participation
5. Financial (ACT Assessment and ACT Student Needs Analysis System)
6. Family attitudes towards education (interest and perceived expectations of parents, financial support, parents' level of education)
7. Academic and personal needs (needs for help with writing, study skills, personal counseling, etc.)
8. Self-reported physical health
9. Self-reported personality characteristics

B. Initial commitment to institution

1. Purpose for enrolling
2. Institutional choice (was the institution their first choice, second, etc.)
3. Importance of selected institutional characteristics for attending the institution (social, academic reputation, physical characteristics)
4. Full-time/part-time enrollment

C. Initial and subsequent academic goal commitment

1. Expected degree and strength of expectations
2. Certainty of career aspirations
3. Commitment to and value placed on college education (academic motivation, academic relevance)
4. Actual versus expected progress in reaching academic goals

5. Satisfaction with academic progress and services
6. Absenteeism

D. Student/institution academic fit

1. Does the institution meet the academic expectations of the student
2. Course enrollment, completion and grades
3. Need for remediation or advanced (honors) course work
4. Perception of relationships with faculty, advisors, and staff

E. Student/institution social fit

1. Amount of friendship, peer support
2. Social relationships with faculty and staff
3. Comfort and satisfaction with the environment
4. Extracurricular activities

F. Student/institution financial fit

1. Amount of immediate family contribution required to meet expenses
2. Hours per week spent working
3. Loans required to meet expenses

The criterion variable was student persistence. Four separate time periods were examined: completion of the fall term, re-enrollment in the spring term, completion of the spring term, and re-enrollment in the following fall term. For institutions on the quarter system, re-enrollment and completion were also examined for the winter term.

Twenty-three ACT user institutions were contacted in May, 1989 and invited to participate in the study. Of that number, six agreed to participate for all three years, beginning in Fall, 1989. One school withdrew from the study during the first term because of data collection difficulties.

The data for the study were drawn from several sources: the ACT Assessment, the ACT Institutional Data Questionnaire, the Market Data Retrieval public and private high school files, and three questionnaires developed specifically for this study.

The ACT Assessment is a comprehensive evaluative, guidance, and placement program used by more than a million college-bound students each year. It consists of four academic tests, self-reported high school course work and grades, the Student Profile Section (SPS), and the ACT Interest Inventory. The SPS, which collects information related to family and high school background and preferences with regard to college characteristics, was the source of many background characteristics items, as well as early indicators of institutional and goal commitment.

The ACT Institutional Data Questionnaire (IDQ) collects descriptive information about two- and four-year postsecondary institutions; the information is used to develop the College Planning/Search Book (ACT, 1990). The instrument includes information related to enrollment size, tuition, degrees offered, control (public/private), and selectivity. Items from the questionnaire were used to develop the student/institution fit variables for measuring institutional commitment.

The Market Data Retrieval (MDR) files provide descriptive information from public and private secondary schools throughout the United States. The variables taken from these files were per-pupil expenditure, availability of special education, and percent of students in the district with family incomes below the federal poverty level.

Three survey instruments were developed for the study: The Entering Freshman Survey, administered immediately before or after fall enrollment; the Enrolled Freshman Survey, administered in the middle of the fall and spring terms; and the Withdrawing Student Survey, administered to students who withdrew. The Entering Freshman Survey was designed to assess students' initial perceived needs, expectations, and perceptions related to college in general and to the specific college attended. The Enrolled Freshman Survey was designed to measure similar variables and issues, with particular emphasis on whether and how students' initial needs and expectations were being met. The Withdrawing Student Survey was also designed to measure students' perceptions and attitudes, and included students' reasons for leaving. This questionnaire was administered to students at the time they withdrew from college.

The Entering and Enrolled survey forms were designed so that it was possible to compute discrepancies between a student's responses to comparable items on the two instruments. Discrepancies were computed by subtracting a student's response on an Enrolled Freshman Survey item from his or her response on the comparable item from the Entering Freshman Survey. When spring Enrolled Freshman Survey data became available, the discrepancies were recomputed using Entering and spring Enrolled responses. The objective in using discrepancies was to measure changes in students' attitudes, expectations, and goals from the time they entered college to the time period of interest.

Course credit and grade information for each student was obtained from the institution in which the student was enrolled. Participating institutions were asked to provide the number of credit hours attempted, midterm status, number of credit hours earned, and GPA for each term the student was enrolled.

Method

Data Collection

The Entering Freshman Survey was administered to randomly selected entering freshmen at each institution during the freshmen orientation/registration period or in intact classes during the first week of school. The Enrolled Freshman Survey was administered shortly after fall midterm and again shortly after spring midterm. The Withdrawing Student Survey was administered in exit interviews or by mail to students who left the college at any point between freshman enrollment and re-enrollment at the beginning of the sophomore year.

Institutions were also asked to provide specific enrollment and completion information about their freshman class at several points during the freshman year. This information included name, social security number, credit hours enrolled, midterm performance, credit hours earned, and GPA for the fall and spring term and fall 1990 re-enrollment. Midterm information was obtained from only those institutions that routinely collected it.

Analysis

Persistence was modelled at four separate points in time: end of fall term, spring term re-enrollment, end of spring term, and re-enrollment the following fall term. For institutions on the quarter system, winter term re-enrollment and completion of winter term were also modelled. Persistence through the end of the fall or spring terms was examined in two ways: Initially, for each term, models were developed using only information received prior to the term in question and enrollment information from the beginning of that term. Subsequently, models were developed that also considered information received during the term of interest: the Enrolled Freshmen Survey and midterm performance data.

Several steps were necessary to reduce the number of predictors to a manageable set for each institution. First, simple correlations were computed between each predictor and the dichotomous persistence variable. Any variable that had a correlation coefficient of .10 or higher with persistence, and that was statistically significant ($p \leq .05$), was considered for inclusion in the prediction model. These variables were then examined with respect to the number of respondents and content redundancy. Any variable for which there were relatively few observations or that was clearly redundant with other predictors was eliminated. The remaining predictors were then entered into a multiple linear regression model of persistence. Collinearity diagnostics were examined, and variables with high variance-decomposition proportions at high condition numbers were dropped from the model (see Belsley, Kuh, and Welsch, 1980). The remaining variables were evaluated in terms of their contribution to R^2 and statistical significance ($p \leq .05$).

Multiple linear regression is the most commonly used statistical method for predicting outcomes. However, because linear regression assumes that the criterion, or dependent variable, has multiple values that are interval in scale, it is not the most appropriate method for modelling a dichotomous criterion variable such as persistence. Because a dichotomous criterion variable is bounded, a linear regression might result in impossible values. Although polynomial models might be constructed to fit the data, they would be complex and difficult to interpret.

Another commonly used method for this type of study is discriminant function analysis. This method uses one or more continuous, or metric, predictor variables and a categorical, or nonmetric, criterion variable. In discriminant function analysis, individuals are assigned to a group, or category, on the basis of their weighted scores on the predictor variables. It could be argued that discriminant function analysis would be a suitable method for analyzing these data. However, discriminant function analysis assumes a linear relationship between predictors and criterion; when the criterion is a dichotomous variable, the problems of linear regression also apply here. In fact, it should be noted that in the case of a dichotomous criterion variable, such as persist/dropout, discriminant analysis is equivalent to linear regression (Tatsuoka, 1971).

An alternative method, logistic regression, was developed specifically to deal with dichotomous dependent variables. Logistic regression assumes curvilinear relationships between the independent and dependent variables; hence, a predictor variable's influence on the outcome is more likely to be reflected in the model if curvilinearity is present. Finally, logistic regression computer routines directly estimate the probability of the outcome variable (in this case, persistence) for each student on the basis of his or her values on the predictor variables. This is very practical and useful information when the goal is to identify students who are likely to drop out.

The Logistic Model

The prediction model for logistic regression looks similar to the prediction model for linear regression; the same terms appear in both models, but have different meanings. In a linear regression model,

$$\hat{Y} = a_0 + a_1 x_1 + ... + a_n x_n,$$

$\hat{Y}$ is the predicted value on the criterion variable, x_1, ..., x_n are the observed values on the predictor variables, a_0 is the intercept term, and a_1, ..., a_n are the regression coefficients associated with the predictor variables. For a one-unit change in x_i there is an expected change in the criterion variable equal to the size of a_i.

A logistic regression prediction model is formulated as

$$\text{Index} = a_0 + a_1 x_1 + ... + a_n x_n,$$

where the criterion variable is a dichotomous variable, such as success/failure on some criterion, and the predictors are metric variables. The probability of persistence is:

$$p = 1/(1+e) - \text{Index},$$

where e = 2.718 is the base of natural logarithms. Index is the logarithm of $\hat{p}/(1\text{-}p)$, the odds of persistence. The symbols x_1, ..., x_n again denote the observed values of the predictor variables, and a_0 denotes the intercept. In the logistic prediction model, a_1, ..., a_n are called the regression coefficients for the predictor variables; however, since the Index represents the log odds of persistence, not the predicted value of the criterion, a_i represents the degree of change in the log odds of persistence (the Index) given a one-unit change in x_i.

The independent variables from the final linear regression models were entered into a logistic regression analysis to predict persistence. Any of these predictors that were not found to be statistically significant at $p \leq .05$ were removed from the model.

Results

Table 1 provides some descriptive information about the institutions participating in the study. Four of the five schools were 4-year institutions. The enrollments of the five schools ranged from below 5,000 to over 15,000, and four of the schools were public institutions. Table 2 provides selected descriptive data for the freshman sample at each of the five schools. The ratio of females to males ranged from 50:50 to 68:32. Virtually all of the freshmen at all five schools were below the age of 22; a large majority were 18 or younger. At four of the five schools, 90% or more of the freshman were white, and almost all were unmarried. The percentage of students who had a high school GPA of 2.50 or lower ranged from 1% to 33%. Fifteen percent to 43% of the freshmen's fathers had gone no farther than high school, and the percentages were quite similar for the mothers.

Table 3 shows the total numbers of students in the samples, by school and time period, and the actual numbers of dropouts. Table 4, in contrast, shows the number of student records included in the final model for each institution and the number of students who dropped out at each time period. As shown by comparing the numbers in Table 3 to those in Table 4, there was a considerable amount of data loss over time. The percent of data lost throughout the first year ranged from 26 to 75; the median percent loss was 63. As a result, it was not possible to develop prediction models for those time periods. This was due to a number of causes: Many students failed to respond to the Enrolled Freshman Survey during the fall and/or spring terms. In addition, many of the questionnaires received did not contain complete data; hence, varying amounts of data were missing for many of the predictors. Finally, several schools identified no students dropping out at certain periods.

Results for the final logistic regression models are shown in Tables 5 through 9. Each table provides the predictor variables for each institution, along with their associated regression estimates, for a given time period. First-term persistence was modelled in two ways: initially, only pre-enrollment variables were used to predict persistence (shown in Table 5); subsequently, fall Enrolled Freshman Survey and enrollment variables were also included (shown in Table 6). As a result of including the additional variables, there was a certain amount of data loss (as shown in Table 4); therefore models could not be developed for all schools. The same approach was attempted in modelling spring persistence; however, the extent of data loss was such that a second model could be developed for none of the five schools.

Although relatively few individual predictor variables were present in more than one model, most of the significant predictors could be grouped in terms of the categories described in the Data section. The most frequently occurring categories at all time periods were the following: goal commitment (e.g., academic goal at this institution, number of credits, expect to complete freshman year, certainty of career choice), institutional commitment (e.g., I like this college, satisfaction with academic reputation of college, satisfaction with availability of major), and academic fit/integration (e.g., number of credits dropped, number of credits earned, GPA, availability of courses wanted, use of academic advising). High school preparation/background (e.g., number of math courses taken, ACT Mathematics score, high school GPA, high school per pupil expenditure) was also significant for most time periods. Plans to work while in school was important in predicting first-term persistence at two schools, but was not a significant predictor for later time periods. Another issue that appeared significant only to first-term persistence was state residency classification, which is related to tuition. Both plans to work and nonresidence status might be interpreted as indicators of financial stress. Social fit indicators (e.g., use of college-sponsored or off-campus activities and programs) were significant predictors in a few instances. Finally, a few variables related to personality were significant (like school, enjoy being with people socially).

The discrepancy variables examine the points in Tinto's model at which expectations meet actual experiences, and goal and institutional commitments are re-evaluated. In several of the models, discrepancies between responses on Entering Freshman Survey items and corresponding items on the fall or spring Enrolled Freshman Survey were statistically significant predictors. For these items, a higher value on the later survey than on the earlier survey created negative values on the discrepancy variables. Conversely, if the later response was lower than the initial response, the discrepancy was positive.

Some of the relationships between survey response discrepancies and persistence were not intuitively obvious. For example, as shown in Table 9, the discrepancy between students' initial satisfaction and later satisfaction with employment opportunities was positively related to persistence. This suggests that if students' satisfaction with their employment opportunities decreased over time, they were more likely to persist. If their satisfaction increased, they were more likely to drop out. A possible explanation for this is that students who find that they have better employment opportunities than they expected will be likely to drop out of school to work, or fail as a result of working too many hours while in school.

In Table 7, the discrepancy between initial and later concern about having to drop classes due to poor grades was a statistically significant predictor for Institution 5. Students who became more concerned appeared more likely to persist, while students who became less concerned were more likely to not return the second term. It is possible that the students who indicated increased concern had a realistic sense of the challenges they faced, and consequently worked harder to perform well. On the other hand, students who expressed less concern may have become apathetic about school or unrealistic about expectations and consequently dropped out.

The results support Tinto's assertion that models of student persistence should be institution-specific. For example, Table 9 shows that in predicting second year re-enrollment, measures of academic fit were important at three of the five institutions, social fit indicators were significant at two institutions, and goal commitment indicators were important at two of the five institutions. This, as well as the fact that no specific variable appeared in more than one model (in Table 9), indicates that characteristics of institutions uniquely interact with characteristics of their particular student populations.

The results are consistent with previous research using Tinto's model. Other studies have found that goal and institutional commitment are key predictors of student persistence (e.g., Hackman & Dysinger, 1970; Cope & Hannah, 1975; Noble, 1988; Webb, 1989). Furthermore, other research has found academic fit to be a more salient predictor of persistence than social fit (e.g., Munro, 1981; Halpin, 1990).

Practical Utility of Logistic Regression Results

Using the predicted probabilities of success from a logistic regression model, prediction accuracy can be examined by constructing a simple decision table such as the one shown below:

	Predicted outcome	
Actual outcome	**Below critical value**	**Above critical value**
Persist	A	B
Drop out	C	D

It is possible to identify a critical point on the scale of obtained index values; a student whose index value is at or above the critical point would be predicted to persist, and a student whose index value is below the critical point would be expected to drop out. Observations are categorized into one of the four possible outcomes. B is the number of "true positives", that is, the number of students who were expected to persist and actually did; C is the number of "true negatives—the students who were predicted to drop out and actually did. A represents the false negatives and D represents the false positives; these are the groups for which incorrect decisions were made. B + C represents the number of students for whom correct decisions were made; when presented as a proportion of the total group this is also referred to as the accuracy rate.

Retention Programs: Identification and Intervention with High-Risk Students

The information provided by logistic regression models can be used to assist in identifying high-risk students and designing intervention strategies to address their needs. The proportion high-risk, probability of retention, and accuracy rate values can be used, with the Index scale, to set critical points for identifying high-risk students. The logistic regression model provides the variables most strongly associated with persistence for each institution. High-risk students' performance or responses on these variables can be used in identifying areas where interventions can be focused.

Before establishing a retention program for identifying and intervening with high-risk students, an institution must consider several factors not addressed by statistical analysis: students attend and drop out of college for many reasons. According to Tinto (1987), a student dropout should be considered a failure only if both the student and the institution fail to meet their goals. If the student's intent is to attend school for a year, and then transfer, this does not necessarily make him a "failure" at the school. A goal of some institutions may be to minimize student transfer; others may encourage it (e.g., two-year colleges). Further, many variables related to student persistence may not be under the control of institutions (e.g., race, gender, health status). Consequently, it may not be feasible or cost-effective to attempt intervention with some students. Each institution must determine, before implementing a retention program, what its goals are with regard to student persistence, the types of dropouts with which it wishes to intervene, and the resources to be made available to the program.

An Example

The logistic regression results for Institution 2 across all four persistence time periods are shown in Table 10. For each time period, the significant ($p \leq .05$) predictor variables are identified, with their corresponding regression weights. As shown in the table, the significant predictors were not constant over time; in fact, the student's goal in attending the institution was the only recurring variable.

The regression weights describe the direction and strength of the relationship between each variable and student persistence. For example, the number of hours a student planned to work was negatively related to fall term persistence at this school; as the number of hours a student planned to work increased, probability of dropping out increased. For fall term persistence based on pre-enrollment and fall survey data, the discrepancy between expected completion of the freshman year at entry and at mid-fall term was positively associated with persistence. Increased expectations of completion were associated with higher probabilities of staying in school. Per-pupil expenditure for the high school attended was negatively associated with spring term persistence; students who attended wealthier high schools were more likely to drop out. Further, students who were more satisfied with job opportunities during the mid-spring term than at the beginning of the year were less likely to persist through the beginning of their sophomore year.

In summary, institutional and goal commitment variables initially tended to be significant predictor variables for students at this institution. For later time periods, however, academic and social student/institutional fit variables predominated, though goal and institutional commitment variables were still present.

Figures 1 through 4 illustrate the prediction accuracy of these models using three indicators of prediction accuracy. Each figure corresponds to one of the four time periods for the institution whose prediction models are summarized in Table 10. The horizontal axis represents the Index scale, computed from the relevant logistic regression model. The vertical axis represents a probability, or proportion, associated with each of the three indicators shown in the Figure. The **probability of retention** curve indicates the probability of persisting in college for a student with a given Index value. As the Index value increases, the probability of retention increases to its maximum value. The probability of retention is always 50% for an Index value of 0.

The **proportion high risk** and **accuracy rate** curves can be used in setting a critical point on the Index scale. For each point on the Index scale, the curves illustrate the expected results of using that Index value as a cutoff point for identifying high-risk students. The proportion high-risk curve indicates the proportion of students who would be identified as high-risk for a given cutoff on the Index scale. This curve increases as the cutoff value increases; at the highest Index value all students would be flagged as high-risk. The accuracy rate curve indicates the proportion of students correctly classified for any given cutoff point.

From Table 4, it can be seen that the actual dropout rate during the first term at Institution 2 was about 10% of the sample. Hypothetically, this institution could choose to intervene with only 5% of its students. The costs of intervention would need to be weighed against the benefits to determine the best target percentage, however. Figure 1 can be used to determine the expected results of such a decision. Drawing a line across from .05 on the vertical axis to the proportion high-risk curve and then down to the Index scale, the resulting Index cutoff value would be 1. This would mean that a student with an Index value at or below 1 would be identified as high-risk. The probability of retention curve shows that a student with an Index value at or below 1 would have about a 74% or lower probability of staying in school. Correspondingly, the accuracy rate associated with an Index value of 1 is about .9, meaning that about 90% of the students would be correctly identified. Note that the accuracy rate begins to decrease at this point, as the Index is increased; if the Index cutoff were set at 2, the accuracy rate would be .8. The greatest prediction accuracy would be achieved when the Index value associated with the maximum accuracy rate is used.

Similar decisions can be made about the other time periods for Institution 2, as shown in Figures 2, 3 and 4. For first term persistence (using pre-enrollment and first term data), as shown in Figure 2, an Index cutoff value of 1 would identify somewhat less than 5% of the students as high-risk, with a probability of about 75% of staying in school, and an accuracy rate of .98. For spring term completion and fall re-enrollment, slightly lower Index cutoff values might be identified, due to the relatively gradual decline in the accuracy rate curve. A cutoff value of .5 might be used to maximize the accuracy rate while still targeting a relatively small proportion of high-risk students. Approximately 2.5% of the students would be identified as high-risk in the spring term, and 8% would be identified as high-risk for fall re-enrollment persistence.

Targeting Correlates for Developing Intervention Strategies

Tables 11 through 14 list selected students from each time period that were flagged as high-risk using the Index values suggested in the previous section. The Index value for each student is given, along with his or her responses or performance on the significant predictor variables.

The results shown in Table 11 highlight several areas for potential intervention during the fall term. In developing interventions, schools may want to look at the intentions of high-risk students with regard to educational goals. For some students who plan to take only one or two courses then leave, it may not be worthwhile to intervene. For example, of the 41 high-risk students (only a subset of the 41 are shown) at Institution 2, four attended the college for self-improvement, to take a few courses, or with the intent of transferring. Depending on the institution's definition of dropout, these students might or might not be targeted for intervention.

Of the 41 students identified as high-risk at Institution 2 during the fall term, 18 (44%) had no definite purpose in mind for attending the school. An investigation of the career and educational counseling provided by the school, the extent to which information about these services is made available to students, and the form in which the information is given to students (e.g., orientation, advising, written materials) might provide additional information for future revision or modification. It is possible that, with additional education and career information and guidance, these students might have persisted. Further support for this investigation is found by examining the mathematics course work and the student's goal in attending the school. Twenty-five of the 41 students had taken 1 or fewer mathematics courses in high school. Further, of the 14 students who intended to obtain a Bachelor's degree, 11 had taken no mathematics course work in high school. Nearly all of these 11 students were enrolled full-time and were planning to work 20 to 29 hours weekly. In fact, for the total group, all but 4 were enrolled full-time, and all but 7 were planning to work at least 20 hours each week.

The results for fall persistence using pre-enrollment and fall survey data, as shown in Table 12, were less clear for the purpose of identifying possible student interventions. Only two of the 27 identified students (of which a sample is shown) were planning to transfer, six had no definite plans, and three were taking courses for self-improvement. All others were planning to complete an Associate or Bachelor's degree. Of the 27 students, 16 changed their expectation of completing their freshman year from "yes" to "no" between entry and mid-fall term. Four students indicated at midterm that they intended to complete the year, but did not persist. Using this information as a preliminary indicator, other variables such as midterm grades and satisfaction with the institution could be examined to further identify factors related to these students' dropout behavior.

For spring term persistence, a key variable related to student persistence was the difference between credit hours enrolled and credit hours earned in the fall term. As shown in Table 13, the students identified as high-risk typically lost over 9 credit hours in the fall term, and their GPAs from the courses for which they did earn credit tended to be less than 1.00 (less than a "D"). Only two students had no definite purpose in attending; all others were planning to achieve certification (1) or to complete an Associate (2) or Bachelor's (8) degree. Investigating the hours the students were working, the credit hours enrolled, the college courses in which they were enrolled, and their high school course work would provide further information about why the students were failing. Additional guidance/advising about appropriate college course work and academic support might reduce the potential for student failure.

The list of high-risk students for re-enrollment in the sophomore year is provided in Table 14. Two findings are clear: over 50% of these students were not degree-seeking students, and these students' responses about liking the college tended to be negative. Eight of these students were planning to transfer, and 6 had no definite plans in attending the school. The undecided students might benefit from educational and career counseling, as noted in discussions of earlier time periods. In addition, examining the students' responses regarding their satisfaction with specific aspects of the college would assist in identifying potential areas needing improvement and intervention.

In conclusion, the results from this particular institution illustrate a few important points. It is apparent that later dropouts are more likely to be academic failures, as seen in the spring term persistence data, or to have entered the institution with shorter-term goals, as seen in the sophomore re-enrollment data. The earlier dropouts, of which there are generally more, may be the more difficult students to identify as at-risk. In addition, they may offer more opportunities for successful interventions.

Discussion

There were several problems that may have hindered the interpretability of this study. Future studies of this type should take care to address these problems.

Because of the fact that there were multiple data collection points throughout the year, there was, inevitably, a certain amount of data loss. Each time a survey was administered, some portion of the original sample failed to respond. This amount varied, depending on the administration methods that each school employed. For example, although this was strongly discouraged, some schools administered the Enrolled Freshman Survey by mail. In other cases, schools administered the surveys in freshman classes, but simply did not have any particular classes in which most freshmen enroll; this was a more severe problem in the spring term.

Another source of data loss resulted from the fact that questionnaire respondents often leave some items blank, perhaps due to oversight, fatigue/boredom, or uncertainty about how to answer an item.

In multiple linear or logistic regression procedure, such as those used in this study, a case-wise deletion process is typically used to eliminate missing data. It is possible to use some method of estimating values for missing data in incomplete cases, which would alleviate this problem. It should be noted, however, that the most straightforward methods of value estimation may result in biased regression parameters; more appropriate methods would involve iterative techniques that prove to be expensive and time consuming (Anderson, Basilevsky, and Hum, 1983).

An unexpected statistical difficulty was the relatively low proportions of students who withdrew. The attrition rates were, in most cases, significantly lower than were expected. At some institutions, the intercept-only model gave a high probability of persistence, and there was very little to gain by adding predictor variables to the model. In such cases, there were greater opportunities for chance factors to influence results. It is possible that the first-term attrition rates in this study were artificially low because students who withdrew very early in the term were not reflected on the initial enrollment files. One ofthe institutions in the study was able to confirm that the enrollment files were only accessible after the first few weeks of class had passed, thereby eliminating any record of students who had already withdrawn. In addition, it was noted that percent of data loss due to nonresponse or incomplete surveys tended to disproportionately reduce the number of dropouts. That is, students who were likely to drop out were also less likely to complete the surveys. In future persistence studies, some of the persistence periods could be combined in order to increase the number of dropouts in each period.

Finally, there were many predictor variables included in the study. Because of the large number of variables included in the study, it was unlikely that a particular set of variables would consistently emerge as significant predictors across all institutions. The exploratory nature of the research made it desirable to include variables addressing all constructs that past research had shown to be related to retention. Furthermore, the study specifically addressed the question of how dropouts at different periods during the freshman year differ; consequently, it was important to collect data at several different points in time. In the follow-up study currently underway, factor analyses have been performed with the survey variables with the goal of stabilizing results by replacing individual variables with composite variables.

References

The American College Testing Program (1987). The College Planning Search Book. Iowa City, Iowa: Author.

Anderson, A. B., Basilevsky, A., & Hum, D. (1983). Missing data. In A. Rossi, J. D. Wright, & A. B. Anderson (Eds.). Handbook of Survey Research (pp. 415-494). Orlando: Academic Press, Inc.

Bean, J. P. (March, 1986). Assessing and reducing attrition. In D. Hossler (Ed.). Managing College Enrollments: Number 53. New Directions for Higher Education (pp. 47-61). San Francisco: Jossey-Bass, Inc.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression Diagnostics, New York: John Wiley and Sons.

Blanc, R. A., DeBuhr, L. E., & Martin, D. C. (1983). Breaking the attrition cycle: The effects of supplemental instruction on undergraduate performance and attrition. Journal of Higher Education, 54, 80-90.

Cope, R. G. (1978). Why students stay, why they leave. In L. Noel (Ed.). Reproducing the Dropout Rate: Number 3. New Direction for Student Services (pp. 1-11). San Francisco: Jossey-Bass, Inc.

Hackman, J. R. & Dysinger, W. S. (1970). Research notes: Commitment to college as a factor in student attrition. Sociology of Education, 43, 311-324.

Halpin, R. L. (1990). An application of the Tinto model to the analysis of freshman persistence in a community college. Community College Review, 17(4), 22-32.

Lenning, O. T., Beal, P. E., & Sauer, K. (1980). Retention and Attrition: Evidence for Action and Research. Boulder, Colorado: National Center for Higher Education Management Systems.

Munro, B. H. (1981). Dropouts from higher education: Path analysis of a national sample. American Educational Research Journal, 18(2), 133-141.

Noble, J. (1988). Using Pre-enrollment and Survey Measures to Test Tinto's Model of Attrition. A paper presented at the Annual Forum of the Association for Institutional Research in Phoenix, Arizona.

Pallett, B. H. (1984). The Use of ACT Pre-enrollment Measures to Test Tinto's Theory of Attrition. Unpublished doctoral dissertation, Kansas State University, Manhattan.

Pascarella, E. T. & Terenzini, P. T. (1983). Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: A path analytic validation of Tinto's model. Journal of Educational Psychology, 75(2), 215-226.

Tatsuoka, M. M. (1971). Multivariate Analysis: Techniques for Educational and Psychological Research, New York: John Wiley & Sons.

Terenzini, P. T. & Pascarella, E. T. (1977). Voluntary freshman attrition and patterns of social and academic integration in a university: A test of a conceptual model. Research in Higher Education, 6, 25-43.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. Review of Educational Research, 45, 89-125.

Tinto, V. (1987). Leaving College: Rethinking the Causes and Cures of Student Attrition. Chicago: The University of Chicago Press.

Webb, M (1989). A theoretical model of community college student degree persistence. Community College Review, 16(4), 42-49.

Table 1. Description of Participating Institutions.

School	Region	Type	Control	Admissions policy	Enrollment range	Number of students studied	Sampling	Calendar type
1	Midwest	4 yr.	Public	Selective	over 15,000	2100	Random classes	S
2	Mtn/Plns	4 yr.	Public	Liberal	5-15,000	1400	Other	S
3	East	2 yr.	Public	Open	1-5,000	600	Random	S
4	Midwest	4 yr.	Private	Selective	1-5,000	450	Whole	S
5	West	4 yr.	Public	Selective	1-5,000	1400	Other	Q

Note: The "other" category of sampling refers to representative samples of students enrolled in specific classes.

Table 2. Background and Educational Characteristics of Students by School (in percent).

Characteristics	School 1	School 2	School 3	School 4	School 5
Males	50	40	39	37	32
Age = 17-18 Age = 22 or older	84 2	83 1	88 1	93 <1	87 0
White Black Asian	90 4 2	96 1 <1	91 2 <1	97 <1 <1	84 3 6
Single	98	99	99	99	100
High school GPA of 2.50 or less	7	15	33	3	1
Father's ed level-HS diploma or less Mother's ed level-HS diploma or less	35 37	40 40	43 39	26 25	15 15

Table 3. Sample Size and Number of Dropouts for Each Retention Period, by Institution

	Institution 1		Institution 2		Institution 3		Institution 4		Institution 5	
Retention period	N	Dropouts	N	Dropouts	N	Dropouts	N	Dropouts	N	Dropouts
Fall persistence	1508	25	994	95	478	73	439	22	333	6
Spring re-enrollment	1483	76	899	167	405	9	417	84	327	28
Spring persistence	1407	23	732	4	396	9	333	3	299	3
Sophomore re-enrollment	1384	158	879	151	387	15	330	48	296	36

Table 4. Sample Size and Number of Dropouts for Each Prediction Model, by Institution

Retention period	Institution 1		Institution 2		Institution 3		Institution 4		Institution 5	
	N	Dropouts	N	Dropouts	N	Dropouts	N	Dropouts	N	Dropouts
Fall persistence (using pre-enrollment data only)	1293	13	903	84	417	62	901	79	328	5
Fall persistence (using pre-enrollment and fall enrollment data)	937	7	751	42	244	22	236	5	--	0
Spring re-enrollment	940	42	--	0	222	7	226	29	155	14
Spring persistence	902	16	781	19	396	9	--	0	--	0
Sophomore re-enrollment	412	20	335	52	307	11	--	*	146	17

*For institution 4, sophomore re-enrollment information was not available.

Table 5. Logistic Regression Models for Predicting First-term Retention, by Institution. (Pre-enrollment Data)

Institution	Predictor variable	Regression weight
1	- Credit hours enrolled	0.51
	- High school extracurricular activities	0.38
	- Credit hours x extracurricular activities	0.02
2	- Number of math courses taken in high school	0.41
	- Goal in attending this school	0.14
	- Number of hours plan to work	-0.41
	- Full-time/part-time enrollment	2.26
3	- Enjoy school	1.96
	- Number of hours plan to work	-0.30
	- Goal at this school x credit hours enrolled	0.02
	- Goal at this school x satisfaction with academic reputation	0.09
	- Goal at this school x enjoy school	-0.25
4	- Credit hours enrolled	0.28
	- ACT Mathematics score	0.09
	- High school athletic accomplishments	0.13
5	- High school GPA	1.52
	- Importance of beauty of campus/buildings	1.96

Table 6. Logistic Regression Models for Predicting First-term Retention, by Institution (Pre-enrollment and Fall Enrolled Data)

Institution	Predictor variable	Regression weight
1	- Residency classification	-3.18
	- Importance of entrance requirements	-1.46
	*- Does college offer all courses students want (mid-fall)	-1.97
2	*- Expect to complete freshman year (at entry)	-4.33
	- Discrepancy: expect to complete freshman year (entry minus mid-fall term)	2.48
	- College located in home state	1.80
3	- Satisfaction with availability of a particular major (mid-fall)	.47
	- Discrepancy: expected grades minus actual fall grades	-.84
	- Discrepancy: expect to complete freshman year (entry minus mid-fall term)	1.54
4	- Use of college-sponsored social activities	3.05
	- Discrepancy: find a stimulating intellectual atmosphere (entry minus mid-fall term)	-1.10

*Survey item was reverse coded; higher values assigned to negative responses

Table 7. Logistic Regression Models for Predicting Second-term Re-enrollment, by Institution

Institution	Predictor variable	Regression weight
1	- Importance of facilities for the handicapped (mid-fall)	-.82
	- I like attending this college (mid-fall)	.64
	*- *Expect to complete freshman year (mid-fall)*	-2.06
	- Fall GPA	1.02
3	- Satisfaction with recognition of prior credit earned (at entry)	1.74
	- Certainty of career choice (mid-fall)	-4.35
	- Number of credits dropped in fall	-.59
4	- Importance of recognition of prior credit earned (at entry)	.92
	- Use of academic advising (mid-fall)	-1.35
	- Use of credit-by-exam programs (mid-fall)	1.73
	- Number of hours spent studying (mid-fall)	-.68
	- Number of credits earned in fall	-.21
5	- Discrepancy: concern about having to drop classes due to poor grades (entry minus mid-fall term)	-1.11
	- Library services/facilities will be sufficient (at entry)	1.02
	- Enjoy being with people socially (at entry)	1.01
	- Importance of recognition of prior credit earned (mid-fall)	1.39
	*- Does college offer all courses student wants (mid-fall)	-1.74
	- Number of credits dropped in fall	-.54

*Survey item was reverse coded; higher values assigned to negative responses.

Table 8. Logistic Regression Models for Predicting Second-term Retention, by Institution

Institution	Predictor variable	Regression weight
1	- Satisfaction with the college (mid-fall)	.81
	*- Expect to complete freshman year (mid-fall)	-1.18
	- Number of classes missed per week (mid-fall)	-1.24
2	- Per pupil expenditure for high school attended	-.13
	- Credit hours dropped in fall term	-.24
	- Fall GPA	1.14
3	- Number of credits dropped in fall term	-.71

*Survey item was reverse coded; higher values assigned to negative responses.

Table 9. Logistic Regression Models for Predicting Sophomore Re-enrollment, by Institution

Institution	Predictor variable	Regression weight
1	- Discrepancy: satisfaction with academic reputation (entry minus mid-spring term) - Use of cultural programs (mid-spring) - Spring GPA	.92 .96 .66
2	- Discrepancy: satisfaction with opportunities for employment (entry minus mid-spring term) - Discrepancy: expected grades minus self-reported grades mid-spring term - Goal in attending this school spring term - I like attending this school (mid spring)	.52 -.66 .24 .95
3	- Discrepancy: highest education level expected (entry minus mid-spring term)	-1.06
4	- Discrepancy: importance of entrance requirements (entry minus mid-spring term) - Concern about having to drop out because of poor grades (mid-fall) - Need help with study skills (SPS)	-1.48 -1.43 -1.56
5	- Opportunities to participate in off-campus cultural/recreational activities (at entry) - Availability of courses you want at times you can take them (mid-fall)	1.10 -1.09

Table 10. Persistence Models for a Particular Institution, by Persistence Period

Persistence period / predictor variables	Significant predictor variable	Regression weight
Fall term / pre-enrollment variables only	Number of mathematics taken in high school	.41
	Goal in attending this school	.14
	Number of hours plan to work	-.41
	Full-time/part-time enrollment	2.26
Fall term / pre-enrollment and fall Enrolled Survey variables	Expect to complete freshman year (at entry)	-4.33
	Discrepancy: expect to complete freshman year (entry minus mid-fall term)	2.48
	College located in home state	1.80
Spring term / all fall variables (including credit hours earned and GPA)	Per pupil expenditure for high school attended	-.13
	Credit hours dropped in fall term	-.24
	Fall GPA	1.14
Re-enrollment fall term sophomore year / all variables (spring Enrolled Survey variables)	Discrepancy: satisfaction with opportunities for employment (entry minus mid-spring term)	.52
	Discrepancy: expected grades minus self-reported grades mid-spring term	-.66
	Goal in attending this school spring term	.24
	I like attending this school	.95

Table 11. Selected Variables Related to First Term Persistence for Selected Students (Pre-enrollment Variables Only)

Student	Index	Probability of persistence	Number of mathematics courses taken in high school	Goal in attending this school	Number of hours plan to work	Full/part time enrollment
1	-2.0	.12	3	No plans	40 or more	Part
2	-1.0	.27	1	Bachelor's degree	20-29	Part
3	-0.6	.35	2	Bachelor's degree	20-29	Part
4	-0.1	.48	0	No plans	20-29	Full
5	0.0	.50	0	Bachelor's degree	40 or more	Full
6	0.3	.57	1	No plans	20-29	Full
7	0.3	.57	0	No plans	10-19	Full
8	0.4	.60	0	Bachelor's degree	30-39	Full
9	0.5	.62	0	Associate degree	20-29	Full
10	0.6	.65	3	Few courses	40 or more	Full
11	0.7	.67	0	Transfer	20-29	Full
12	0.7	.67	2	No plans	20-29	Full
13	0.8	.69	0	Bachelor's degree	20-29	Full
14	0.8	.69	1	Self-improvement	10-19	Full
15	0.8	.69	1	Vocational-technical	20-29	Full
16	1.0	.73	2	Associate degree	30-39	Full
17	1.0	.73	0	Associate degree	10-19	Full

Table 12. Selected Variables Related to Fall Term Persistence for Selected Students (Pre-enrollment and Enrolled Survey Variables)

Student	Index	Probability of persistence	Expect to complete freshmen year (at entry)	Expect to complete freshman year (mid-fall term)	Discrepancy (entry - mid-fall)	College in home state
1	-2.9	.05	Undecided	No	-1	Yes
2	-2.9	.05	Yes	No	-2	No
3	-1.0	.27	Yes	No	-2	Yes
4	-0.4	.40	Undecided	Undecided	0	Yes
5	0.2	.55	No	Yes	2	Yes
6	0.3	.57	Undecided	Yes	1	No

Table 13. Selected Variables Related to Spring Term Persistence for Selected Students (All Fall Variables)

Student	Index	Probability of persistence	HS per-pupil expenditure	Number of credits dropped in fall term	Fall GPA
1	-2.1	.11	$3600-$3799.99	16	0.0
2	-1.8	.44	$3200-$3399.99	16	0.0
3	-1.7	.15	$3400-$3599.99	15	0.0
4	-1.6	.17	$2600-$2799.99	17	0.0
5	-1.2	.23	$6000 or more	9	0.7
6	-0.5	.38	$4400-$4599.99	12	1.0
7	-0.4	.40	$4600-$4799.99	9	0.6
8	-0.3	.43	$3000-$3199.99	12	0.37
9	0.0	.50	$6000 or more	6	1.10
10	0.1	.52	$3600-$3799.99	10	0.66
11	0.1	.52	$2600-$2799.99	11	0.25
12	0.4	.60	$2600-$2799.99	11	0.50
13	0.5	.62	$3200-$3399.99	9	0.53

Table 14. Selected Variables Related to Re-enrollment, Sophomore Year for Selected Students (All Variables)

Student	Index	Probability of persistence	Opportunities for employment (SA--SD at entry)*	Opportunities for employment (SA--SD mid-spring term)*	Discrepancy (entry - mid-spring)	Expected grades	Spring midterm self-reported grades	Grade discr. (entry - mid-spring)	Goal in attending this college (spring)	I like attending this college (SA--SD)*
1	-1.6	.17	Agree	Agree	0	A	B	1	Transfer	Strongly disagree
2	-1.4	.20	Agree	Agree	0	C	C	0	No plans	Disagree
3	-1.3	.21	Agree	Neutral	1	C	D	1	Self-improvement	Disagree
4	-1.2	.23	Neutral	Neutral	0	B	C	1	No plans	Neutral
5	-1.2	.23	Agree	Strongly agree	-1	B	C	1	Transfer	Disagree
6	-1.1	.25	Disagree	Agree	-2	B	B	0	Few courses	Neutral
7	-1.0	.27	Agree	Strongly agree	-1	B	B	0	No plans	Neutral
8	-0.9	.29	Agree	Agree	0	B	B	0	Transfer	Strongly disagree
9	-0.5	.38	Agree	Agree	0	C	C	0	No plans	Neutral
10	-0.4	.40	Agree	Agree	0	A	B	1	Bachelor's degree	Disagree
11	-0.2	.45	Neutral	Agree	-1	C	D	1	Transfer	Neutral
12	-0.1	.48	Agree	Neutral	1	B	C	1	Transfer	Disagree
13	0.0	.50	Neutral	Neutral	0	B	B	0	Transfer	Disagree
14	0.3	.57	Agree	Agree	0	C	C	0	Bachelor's degree	Disagree
15	0.3	.57	Agree	Neutral	1	B	C	1	No plans	Agree
16	0.4	.60	Agree	Agree	0	B	A	-1	Self-improvement	Neutral
17	0.5	.62	Strongly agree	Agree	1	C	B	-1	Bachelor's degree	Strongly disagree

* Scale ranges from strongly agree (SA) to strongly disagree (SD)