DOCUMENT RESUME

ED 422 801		HE 031 559
AUTHOR	Eno, Daniel; McLaughlin, Gerald W.; Brozovsky, Paul; Sheldon, Phyllis	
TITLE	Predicting Freshman Success Based on High School Record and Other Measures. AIR 1998 Annual Forum Paper.	
PUB DATE	1998-05-00	
NOTE	29p.; Paper presented at the Annual Forum of the Association for Institutional Research (38th, Minneapolis, MN, May 17-20, 1998)	
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	*Academic Persistence; College Attendance; *College Freshmen; *Dropout Research; Educational Research; High Schools; Higher Education; School Holding Power; Student Adjustment; Student Attrition; Undergraduate Students	
IDENTIFIERS	*AIR Forum; Virginia Polytechnic Inst and State Univ	

ABSTRACT

Most research on success in first-semester retention of college freshmen has looked either at performance in specific courses or at performance in a given course across all colleges. In the fall of 1997, Virginia Tech (Blacksburg) implemented a new academic eligibility policy, which defined success as earning an average grade of C or better. This report presents results from the second stage of a study that extended previous research by examining additional high school data, such as the number of courses in a specific area of study and grade point average in those courses, and results from the preceding year's freshman class. The study examined the value of detailed high school information, the effects of advising, and the relationship between actual performance and overall difficulty of students' schedules. The results indicate that performance prediction can be improved by using detailed high school information; that prediction of performance needs to be supported by operational processes; and fully understanding the outcome of the advising system depends on recording its results. (Contains 13 references.) (CH)

PREDICTING FRESHMAN SUCCESS BASED ON HIGH SCHOOL RECORD AND OTHER MEASURES

Daniel Eno, Research Associate
Gerald W. McLaughlin, Director
Paul Brozovsky, Senior Researcher
Phyllis Sheldon, Assistant Director
Institutional Research and Planning Analysis
Virginia Tech
Blacksburg, VA 24061

Prepared for

The Association for Institutional Research
1998 Forum
Minneapolis, MN

May 1998

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Eighth Annual Forum of the Association for Institutional Research held in Minneapolis, Minnesota, May 17-20, 1998.
This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of AIR Forum Papers.

Dolores Vura
Editor
AIR Forum Publications

PREDICTING FRESHMAN SUCCESS BASED ON HIGH SCHOOL RECORD AND OTHER MEASURES

Abstract

Supporting freshmen requires that the institution anticipate the difficulty students will have in their various courses. Initial research involved predictions based on SAT scores and overall high school performance. The current research expands the prediction equations to include high school courses taken and grades therein. The models used last year are validated and the role of schedule difficulty in students' success is investigated. The results indicate that the high school data are important. The equations used last year were still valid. The results did not support the hypothesis of a major "schedule difficulty" effect.

PREDICTING FRESHMAN SUCCESS BASED ON HIGH SCHOOL RECORD AND OTHER MEASURES

Accountability for Success

Perspectives on accountability are changing. Peter Ewell and Dennis Jones describe this change in their article, "Assessing and Reporting Student Progress: A response to the 'New Accountability,'" as "altering the focus of accountability for higher education from equitable access and efficient operation toward 'return on investment'" (1991). Along with many other higher-education institutions, Virginia Tech is concerned about this shift in focus (McLaughlin, Brozovsky, and McLaughlin, 1998). In college, one type of "return on investment" for students is success in a class. For fall 1997, Virginia Tech implemented a new academic eligibility policy that defines "success" as earning an average of 'C' or better in a class. This new policy generated considerable interest, internally, in the ability to anticipate the performance of students in their classes, specifically in the ability of new freshmen to make a "C" or better in their classes.

Previous Research

Most previous research on the topic of success prediction in specific courses has looked either at performance in specific sets of courses rather than in individual classes or at performance in a given course across colleges.

An example of the research on average grades is the work done by Williford (1996). Her work shows that freshman GPA is significantly related to educational aspirations, self-perceptions of ability, expectations for success, study habits, and willingness to seek academic-support resources. Boling (1996) found that participation in various activities seems to influence grades after considering the effect anticipated based on the standard academic measures of high school grades and test scores. Nobel and Sawyer (1997) also show that academic ability, as measured by the ACT and by high school grades have predictive validity sufficient to set admissions criteria for selection.

Some research has also looked at performance in specific courses across colleges. One earlier study by Noble and Sawyer (1987) focused on the ability to anticipate grades in a limited set of 18 courses across 233 colleges where American College Testing [ACT] data were available. Noble and Sawyer present, and chart, an excellent summary of research specifically on prediction of course grades from 1970 through 1985. These early studies used ACT scores, SAT scores, high-school grades, and standard subject-specific tests as predictors; courses were grouped into the subject areas of English, mathematics, social studies, and natural sciences, with a large proportion of the courses being in mathematics. The studies tended to be limited in scope in that none examined a full range of representative freshman courses.

Noble and Sawyer the studied the same four subject areas of English, Mathematics, social studies, and natural sciences. They based their analyses on student records from a variety of institutions that participate in ACT's Standard Research Service (SRS), using ACT subtest scores and self-reported high school grades as predictor variables, and grades for one or more specific freshman courses as criterion variables. Since more than one institution was involved, much work went into determining which courses at different institutions were comparable and into obtaining student-specific and course-specific grade information. A total of 576 specific courses from 233 institutions were selected for the final study. Research methodology was in two parts: the first part involved computing regression statistics for each specific course grade over the most recent year's data; the second part involved cross-validation of prediction equations where more than one year's data was available. Results indicated that a combination of the ACT scores and high school grades was a better predictor than either of these measures when used alone. As for accuracy, statistical analysis showed a median multiple correlation of about 0.5 in English, mathematics, and social studies, and about 0.59 in natural sciences. To eliminate much of the variability in predictive accuracy across course groups, courses, and subject areas, Noble and Sawyer recommend that local course-grade prediction equations be developed.

The work by Noble and Sawyer follows several years of investigating the relationship between ability as measured on an achievement test, high school performance, and performance in a course. Some studies in the area used performance in a specific coursed, such as mathematics, over a broad range of colleges (Bridgeman and Lewis, 1996; Wainer and Steinberg, 1992). Others looked at performance of students at a specific college in a specific course and built the sample size by looking at the results of students from multiple entering classes (Spencer, 1996). These studies generally find some relationship of grades to the various general measures, such as the SAT and high school grades.

Work by Bank, Biddle, and Slavings (1994) has shown that the grades students make in their initial courses seem to have a positive and significant effect on students' self-concepts. While they did not find that the same relationship held for student preferences, they did find that the initial grades were associated with some changes in the personal norms of the students. In short, the literature supports the likelihood that grades in individual courses can be anticipated and that the grades in initial courses will be associated with the feelings about self-worth.

Previous Research on Success at Virginia Tech

Virginia Tech's concern with the effect of first-semester success on retention has resulted in an effort to identify and use information that improves the advising process for new freshmen. The first step in the investigation looked at course-level performance for some 12,000 first-time freshmen from three years of entering students in a broad range of first semester courses (Beaghen, Brozovsky, & McLaughlin, 1996; Eno, Brozovsky, & McLaughlin, 1997). These students took some 80,000 classes. The dependent measure was making a grade of C or better and regression models were developed in 53 different classes. The independent measures were overall high school grades (GPA), high school rank (HSR), high school class size, SAT scores, major, gender, race, and entering year. Methodology involved creating separate equations for each of 53 classes. For each course, entering students were placed in a category depending on the predicted probability that they would get a grade of C or higher. Students who were predicted to have less than a 50% likelihood of a C or better were placed in the "x" category for the course. Students who were predicted to have a likelihood of 50% to 70% were placed in the "m" category. Students predicted to have more than a 70% likelihood of a C or better were placed in the "s" category. The Director of University Studies requested this dependent measure to obtain results that would be easy for students and advisors to use. Three models were developed for each course: the first to be used if all information was available, the second if part of the data was missing, and the last if no data at all was available. In the last case, prediction was based just on the average success rate for first semester freshmen for the course. Grades, rank, class size, and SAT's were ultimately used when available; major, gender, and race seemed to produce random effects and were dropped from further consideration.

This Research:

The following presents and discusses results for the second step in our research (Eno, Brozovsky, & McLaughlin, 1997). It extends the previous research in several ways: additional high school data, such as number of courses in a specific area of study and GPA in those courses, were available for modeling; and results from the preceding year's freshmen were available along with several measures of student performance. The research was done in three phases: In Phase I, models were built for use in predicting difficulty of courses for advising the Fall 1997 entering cohort using an augmented data set of number of courses and average grade in five broad discipline groups. Following this, Phase II of the research was undertaken to detect any effect from the use of the previous year's ratings to advise the Fall 1996 entering freshmen. The success rate for freshmen the preceding year was modeled and the model was used to estimate the success rate of the current group. Actual success (having a C or better) was compared to anticipated success (predicted to have a 'S' or 'M' in the course).

Phase III sought evidence that a relation does exist between overall schedule difficulty and student success. As part of this last phase, an “overall ability” index was created for each student by listing the total number of freshman courses that each student was predicted to find extremely difficult. Multiple tests were used to accumulate evidence.

Phase 1: Value of Detailed High School Information

This first phase of the study determined if the addition of detailed high school information improves ability to anticipate performance in specific first-semester classes for first-time freshmen. A student was considered a first-time, entering freshman if all of the following five conditions were met:

1. The student's academic level was not greater than freshman.
2. The student was not a transfer student.
3. The student was first enrolled at Virginia Tech in the fall semester of a given year, or the summer immediately preceding that fall semester.
4. The student enrolled for at least 12 credit hours of course work during his or her first fall semester of enrollment (this corresponds to "full-time" enrollment).
5. The student had completed high school within 9 months of enrolling at Virginia Tech.

Students were included if they met these criteria for the fall semesters of 1994, 1995, and 1996.

Independent measures included SAT scores, high school ranks, declared major areas of study, overall high school grade point averages, and experiences in specific high school classes. (Note: SAT scores for students entering before Fall 1996 were "re-centered" to make them comparable to current SAT scores.) The five subject categories of interest for this study were English, History and Social Sciences, Foreign Languages, Mathematics, and Laboratory Sciences. Measures were grade point averages by subject category and the number of courses taken in each subject category.

As stated above, the goal of this phase of the project was to build models to predict the difficulty that incoming freshman students would have in various courses. These models were built based on the success of such students during the fall semesters of 1994, 1995, and 1996. For the purposes of this study, students were considered successful in a course if they earned a grade of C or higher. Students were considered unsuccessful in a course if they earned a grade of C- or lower. This corresponds to the need for a GPA of 2.0 for continuing students. It should be noted that success in a course was based on students' first reported grades, regardless of whether or not they chose to invoke the [i]"freshman rule" for that course.

During the 1995-1996 academic year, models were built to predict difficulty based on three methods: multiple linear regression, logistic regression with binomial response, and logistic regression with ordinal response. Comparison showed the predictive abilities of the three modeling methods to be similar. Therefore, multiple linear regression was preferred because of its ease of computation and interpretation.

---

[i] The "freshman rule" at Virginia Tech allows a student to eliminate 6 credit hours of courses from the computation of his or her GPA.

For the current year's project, only multiple linear regression modeling was used. Models were built for each course in which at least 100 first-time freshman students enrolled during the fall semesters of 1994, 1995, and 1996 (combined). This size cut-off was selected for two reasons: 1) efforts would be directed toward obtaining results which incoming students are likely to find relevant; and 2) this size provided sufficient data to perform meaningful statistical analyses. There were 79 such courses.

For each such course, multiple linear regression models were built, first using a forward variable selection routine, and second using a backward variable selection routine. A variable was required to have a p-value at least as small as 0.05 to enter a model (in forward selection) or to stay in the model (in backward selection). Since the response variable was a dichotomous categorical variable, the p-values calculated were only approximate. In most cases, the two variable-selection procedures yielded models that were in close agreement. For cases in which the procedures yielded different sets of predictors, backward selection was applied to the set of all predictors identified by either of the first two selection procedures in order to obtain a final model.

The procedure described in the previous paragraph was used to build models based both on the full data set, consisting of the variables obtained from the student census files augmented with the data obtained from the high-school transcript extract files, and on the reduced data set consisting only of variables obtained from the student census files. This was done in order to compare the predictive ability of models based on the entire data set with that of models based on the reduced data set, to determine if the effort required to obtain the high-school transcript data is warranted.

Of the 79 courses which had at least 100 first-time freshman students enrolled during the fall semesters of 1994, 1995, and 1996 (combined enrollment), 74 had statistically significant variables. These models were used to predict, for each student taking a course, the probability that the student would successfully complete that course. This predicted probability was converted to a categorical variable, for use as an advising tool in the future. In particular, if a student was predicted to have less than a 50% probability of successfully completing a course, the course was assigned a difficulty rating of "Extremely Difficult" (denoted "X") for that student. If the predicted probability of success was at least 50% but less than 70%, the course was assigned a difficulty rating of "Moderately Difficult" (denoted "M") for that student. Finally, if the predicted probability of success in the course was at least 70%, the course was assigned a difficulty rating of "standard" (denoted "S") for the student. This criterion was developed as a way to convey sufficient information in a usable fashion.

Our primary interest in this phase of the study was building models to classify courses into difficulty categories for individual students. The existence of statistically significant p-values for the variables in our models was, therefore, of secondary importance. For example, in some courses, models could have been built in which there were statistically significant variables for predicting the probability of student success, yet the predicted difficulty rating fell into a single category for the entire sample of students used to build the model. There was no value in reporting an equation for these models. In 55 of the 79 courses considered, we had both a significant model and meaningful differentiation among difficulty ratings.

As noted above, the modeling procedure was performed both with the entire variable set and with the reduced set consisting only of the variables available from the student census files. The results of these two procedures were compared. The full variable set out-performed the reduced variable set in 43 of the 55 courses for which significant prediction and difficulty classification were possible. In five of the courses, the two data sets led to the same model (and therefore the same classification), and, in seven of the courses, the reduced variable set led to better predictive ability. When comparing these models, it was tempting to use the statistical significance of including the detailed high school experiences in the equations to see if these details contribute significantly to the expected class scores. This method does not, however, test to see if there was any variation in the usefulness of the equations rating classes for three levels of difficulty.

For each course and for each student in that course, the probability of success for a particular student was predicted based on a model in which the coefficients were computed without that student in the data set. In this way, each student's predicted success was, in fact, a prediction, rather than a fitted value. Similar procedures are common for comparing the predictive ability of linear regression models for measures such as the PRESS statistic.

Contingency tables were then created for each course with variables "Difficulty Rating" (X, M, or S) on the horizontal axis and "Actual Success" (0 or 1) on the vertical axis. Goodman-Kruskal's "gamma" provided a measure of association between these two variables. This statistic takes values between -1 and +1, with positive values indicating a positive association and negative values indicating a negative association. Gamma represents the degree to which the co-occurrence of two ordinal measures results in concordant pairs of observations, rather than discordant pairs.

A concordant pair is where an observation that is higher on one measure than another observation is also higher on the other variable than the other observation. We found that in 43 of the 55 courses in which prediction was possible, the full variable set led to a model with a more positive value of gamma than did the reduced data set without the detailed high school experiences.

The results described above have two consequences. First, these models can be used to predict the difficulty rating of various courses for freshman students entering Virginia Tech in the fall semester of 1997. These predictions will be used for advising purposes for students entering the University Studies program, and possibly for other students as well. Second, since the augmented data set (including high-school transcript data) tends to out-perform the reduced data set (using the student census file data only), the high-school transcript data will be used.

Building final models to predict difficulty for the fall 1997 entering class required taking into account the fact that not all students had complete data sets. If the 1997 class is like previous ones, as many as 30% of the incoming students will not have their high school rank information and/or their high school GPA available on the student census file. In addition, about 5% of the students will not have the area-specific high school transcript data available. Hence, it seemed practical to build several models for each course, based on different data sets.

Four models were built for each course in which some prediction seemed possible. The first model was based on all of the variables considered in the study (SAT scores, high school rank and GPA, and area-specific high school transcript variables). The second model uses just SAT scores and area-specific high school transcript variables. The third model was based only on SAT scores. Finally, the fourth model was an "intercept only" model. This final model uses the overall proportion of freshman students who were successful in a given course to determine its predicted difficulty. Hence, this model can always be used, even for students with all variables missing. The use of the independent measures in the equations is shown in Table 1.

Table 1.
Frequency of Measures in Three Regression Models

Measure	Full Model	W/o HSR/GPA	Only SAT
SAT Math	23	26	20
SAT Verbal	27	31	27
HS Rank	13	X	X
HS GPA	32	X	X
GPA English	12	19	X
GPA Hist/Soc Sci	13	29	X
GPA Languages	12	12	X
GPA Mathematics	19	23	X
GPA Lab Science	27	28	X
# English	3	4	X
# Hist/Soc Sci	7	5	X
# Languages	3	2	X
# Mathematics	2	3	X
# Lab Science	3	3	X

A value of "X" means that the measure was not available for use in the model.

As noted above, the forward and backward variable selection routines generally yielded similar models. To build the final models, only backward selection was used. This method seems to yield models that are as good as those obtained from the combined forward/backward selection procedure. The decision to use a single variable selection routine will make updating the prediction equations in the future much less tedious and more consistent from year to year than would be possible if a single routine were not prescribed. However, some thought will always need to be given before the resulting models are actually applied. For example, each equation should be checked to be sure it yields significant classification among difficulty groups (i.e., not all students are classified into the same group, and the association between predicted success and actual success is positive). Failure to do this checking could result in misleading predictions.

Two final points should be made regarding the application of the models. First, the study showed that a student's major ("University Studies" vs. "not University Studies") was rarely a significant predictor of success. In the cases in which this was a significant predictor, the direction of the effect was not consistently either positive or negative. Since there is little reason to expect declared major to influence success, this variable was eliminated from the selection process for final models. For the same reasons, the student's race and gender - considered during last year's project - were omitted this year. Second, during the final model building process, extra care was taken to ensure that all students' high school transcript data was comparable in content. In particular, only courses taken during high school grades 9, 10, and 11 were used. Also, students' high school transcript data was required to have courses listed for each of these grades. The above requirements were intended to counteract possible differences in high school transcript reporting practices. For example, many students would not have 12th grade transcripts on the databases prior to the summer orientation.

Phase 2: Effect of Advising

The second phase of our study involved trying to detect any effect that advising might have had on the University Studies students entering in fall 1996. This was the first group of students to be advised based on predicted difficulty they might have in various courses.

Our analyses during this phase of the study were carried out by comparing the fall 1995 class of entering freshmen in the University Studies program with the fall 1996 class. A word of caution is in order concerning interpreting the results. In order to test for differences in these groups resulting from the advising process, the authors made the assumption that the two groups were essentially "equivalent", and that the courses taught were equivalent as well. Although no systematic differences were apparent, it is impossible to determine if this assumption was correct. Perhaps these two groups of students had systematic differences due to unknown factors, or perhaps changes in syllabi of some courses between the years affected results. In fact, many differences could exist between the two groups and their experiences at Virginia Tech. However, in a study such as this, it is not feasible to perform a ***designed*** experiment, where unknown effects could be reduced, randomized, or eliminated. Hence, the following analyses are presented with the knowledge that the results can not be entirely convincing. However, if similar results are seen over several years, the credibility of the conclusions may increase.

The first analysis performed in this phase of the study was the validation of the models that predict course difficulty. In particular, it was important to confirm that the models built during the 1995 - 1996 academic year project were capable of predicting success for the fall 1996 class of entering freshman students. These models were built based on the 1993, 1994, and 1995 entering classes of freshman students in the University Studies program. Such validation helps assure that prediction is possible when using models based on past students' performance to predict future students' success. For this analysis, we built a contingency table for the entire fall 1996 entering freshman class, with actual success on the vertical axis and predicted difficulty on the horizontal axis (Table 2). For this analysis, the predicted difficulty ratings were classified into only two groups, namely "X" and "M or S". When a log-linear model was fit to the table, the association between predicted difficulty and actual success was extremely significant ($p < 0.0001$). In courses in which a student was predicted to have extreme difficulty, only 40% of the students achieved a success rating of "1" (successful) for the course. On the other hand, in courses in which a student was predicted to have a moderate or standard difficulty rating, 77% of the students successfully completed the course. This result has verified that prediction is possible across years.

Table 2.

Predicted and Actual Difficulty in Courses for Fall 1996

	Predicted Performance	
Actual performance	**'M' or 'S'**	**'X'**
"C" or better	5492 (77.4%)	206 (40.3%)
Below "C"	1601 (22.6%)	305 (59.7%)

The second question of interest concerned the course-taking pattern of the Fall 1996 class of entering freshman University Studies majors, as compared with that of the Fall 1995 class. In particular, it was of interest to determine if the advising process based on predicted success caused students to take fewer extremely difficult ("X") courses than they would have taken without the advising. For this analysis, a contingency table was built with "year" (1995 or 1996) on the vertical axis and "number of X courses taken" (0, 1, 2, or 3) on the horizontal axis (Table 2.). When a log-linear model was fit to this table, there did seem to be a significant association between "year" and "number of X courses taken" ($p = .008$). However, it is not entirely clear how to interpret the association. As shown in the table, the 1996 class had a higher percentage of students taking no "X" courses than did the 1995 class. The 1996 class also had a higher percentage of students taking 2 "X" courses than did the 1995 class. However, the 1996 class had lower percentages of students taking 1 or 3 "X" courses than did the 1995 class. More comparisons (in future years) will be necessary to further clarify the effect of the advising on course taking patterns.

Table 3.
'X' Courses Taken by Year

	0 'X'	1 'X'	2 'X's	3 'X's
1995	432 (82.1%)	85 (16.2%)	6 (1.1%)	3 (0.6%)
1996	593 (84.8%)	84 (12.0%)	22 (3.1%)	0(0.0%)

The final two analyses performed in this phase of the study were concerned with assessing overall performance differences between the two groups of students. It was thought that a student might have a better chance of successfully completing a difficult course if he or she had prior knowledge that the course would be difficult (perhaps the prior knowledge would cause the student to put more effort into the course). For each difficulty category (X, M, and S), a contingency table was created, relating "year" (1995 or 1996) to "success in course" (0 or 1). In none of the three difficulty categories was there a statistically significant association between year and success (all p-values were greater than 0.05; two of the three p-values were greater than 0.20). Thus no improvement in success rates was detected in 1996 as compared with 1995.

The final analysis of this phase, also designed to detect overall performance differences between the 1995 and 1996 classes of students, was to test for a difference in average QCA earned during the two fall semesters. The QCA for each student in the fall 1995 incoming class of freshman students and the fall 1996 class was calculated based on grades reported in the corresponding first-semester grade tapes. Note that these QCA's were not adjusted for the "freshman rule." The Wilcoxon Rank Sum test was performed on the resulting data. The result was that there was not a statistically significant difference in average QCA between the Fall 1995 class and the Fall 1996 class of University Studies majors ($p = 0.35$). A similar result held for students in majors other than University Studies.

Phase 3: Relationship between Actual Performance and Overall Difficulty

The third phase of the study involved looking for relationships between students' performance in various courses and the overall difficulty of their schedules. It is difficult to perform analyses to detect such relationships, because many factors interact between the variables of interest. For instance, it might be hypothesized that a student who takes four "X" courses in a single semester would have less likelihood of success than a student who takes one "X" course. It is likely that such an effect could be detected with a simple analysis. However, it is important to realize that a student who has a strong background (from high-school) is less likely to be given a predicted difficulty rating of "X" in any given course than is a student with a weaker background. The student with the weaker background is thus more likely to take a schedule of courses that contains many "X" courses. Care must be taken to try to separate the students' overall ability from their schedule difficulty.

One specific question addressed in this phase of the project was: "Is a student's success in a given course affected by the total number of 'X' courses in his or her schedule?" In order to address this question, log-linear modeling was used to test for association between "Actual Success" (0 or 1) and "Number of 'X' courses taken" (0, 1, 2, ...). For the reasons mentioned above, it was not appropriate to build a single contingency table and proceed with the analysis without first making an effort to control for the extraneous interactions which were present.

An effort was made to group the student population into "overall ability" categories. Predicting every course's difficulty rating for every student having full data did this. It did not seem advisable to introduce the further complication of making predictions based on several different models for each course. The number of courses in which a student was predicted to have extreme difficulty was used as a measure of the student's overall ability. Each student could thus be classified into one of the following four "overall ability" groups:

1. Group 0 - not predicted to have extreme difficulty in any course,
2. Group 1 - predicted to have extreme difficulty in 1 to 5 courses,
3. Group 2 - predicted to have extreme difficulty in 6 to 10 courses, or
4. Group 3 - predicted to have extreme difficulty in 11 or more courses.

An effort was also made to separate the overall difficulty of each course from the interactions of interest. This was done by calculating the overall percentage of students taking the course who were successful, and then using this percentage to arbitrarily group the courses into three categories of "course difficulty." The "easiest" third of the courses were given a "course difficulty" rating of 2, the middle third of the courses were given a rating of 1, and the "hardest" third of the courses were given a rating of 0.

For each course and for each student who took that course, a "predicted difficulty" rating was calculated. Contingency tables were then built relating "actual success" (0 or 1) with "number of X courses taken" (0, 1, 2, ...). This was done separately for each level of "overall ability" (1, 2, and 3), each level of "predicted difficulty" (X, M, and S), and each level of "course difficulty" (0, 1, and 2). Note that no tables were made for students in "overall ability" Group 0, since it was not possible for such students to take any "X" courses. The procedures described above were designed to reduce external factors that might influence the association between success in courses and schedule difficulty. Of course, it is uncertain whether such factors were totally eliminated.

The results of the above analysis were not very conclusive. Many tables were generated as a result of the classification process, of varying size and with varying numbers of subjects. Of the 16 tables with moderately large numbers of subjects (over 300), six displayed significant positive associations between "success" and "schedule difficulty". Four of the significant results were from tables with "overall ability" level 1. A possible explanation for this is that, for a student with a fairly strong background to be predicted to have extreme difficulty in a course, that course is likely to have a high overall difficulty rating (low overall probability of success). Of course, in an attempt to control for this, separate analyses were run based on "course difficulty," but it is possible that grouping of the courses into three categories did not adequately separate out the most difficult courses.

As noted above, the purpose of the described analyses was to determine if taking a difficult schedule of courses decreases the probability of a student being successful in a given course. It seems reasonable to expect such a relationship to exist. The analyses provide some evidence that it does exist, at least among some groups of students, but do not provide conclusive support for the hypothesis.

The above process required creation of an "overall ability" index for each student involved in this phase of the study. This led to an unanticipated result. If the total number of courses in which a student is predicted to have extreme difficulty is left as a "raw" variable (i.e., not classified into Groups 1 - 4), and a chart is made showing how many students attain each level of this variable, a distinctive pattern is easily discernible. This result is not of central interest to the current project, but is interesting in its own right. Perhaps a result such as this could be useful for admissions activities, or other purposes.

Use of the results

As noted earlier, this was the second year in which prediction of grades in first-term classes was done. This year, predictions were again provided for the advisors in University Studies. Requests came in asking that the equations be run for all of the first-time freshmen. The various colleges used these forms with various levels of enthusiasm and intensity. Colleges such as Engineering, where the first term is quite structured, did not see a need for the process. Colleges where there were many more options were much more interested in the process. Advisors received a single sheet for each student where for each class the sheet showed the "S," "M," or "X" as discussed above. The important element was that the grade was not predicted for the student. The student was told how difficult students with similar characteristics have found the class in the past.

The value of the advising and the validity of the process will be reviewed again within the context of the new academic eligibility policy.

Conclusions

The primary goal of this project was the creation of prediction equations to be used for building an advising tool for incoming first-time freshman students. This goal was accomplished, and the results will be used to aid the Fall 1997 class of entering students in course selection. As the university continues to use and improve the methodology to anticipate grades and to advise students, the following are some of the "lessons learned" to date:

- The prediction of performance in classes can be improved by using detailed high school information, but there will need to be improvement of our database containing high school data.
- Once the equations are developed, the actual prediction of performance needs to be supported by an operational production process.
- To fully understand the outcome of the advising system, the results of its use must continue to be recorded.

- The results of using the system will most likely change when the new academic eligibility rules are instituted.
- Similar equations for transfer students will not be needed: there are too few such students for accurate modeling and the gap in time between their high school performance and their transfer tends to be irregular thus making the high-school data a poor predictor.
- Continuing to consider such demographic measures as gender, race, and major is important, but no systematic relevance is anticipated based on previous findings.

In terms of the previous research, this research shows that using the categories of high school academic activity used by Noble and Sawyer (1987) gives statistically valuable information when looking at grades. It shows that the work on predicting grades in individual courses can be extended to a range of courses and not just focused on math, English, and chemistry.

Several next steps also seem appropriate based on the previous research. The work done by Noble and Sawyer (1997) suggests the possibility of using graphs that show a probability band for the various courses. In this case one might use a measure of high school achievement from courses and grades for one axis and a measure of test performance from the SAT tests to form the other axis for the graph. A second step suggested by their research is to compute and analyze the three indices they discuss in their work. The accuracy rate, the success rate, and the failure rate for the various courses might prove to be helpful in the various advising processes with students and parents. These indices might also be helpful in differentiating the ability to predict grades in various courses at different levels of precision.

In general, the ability to anticipate grades in courses is a very feasible statistical exercise. The use of three years of data provides an adequate sample size and the use of high school grades gives provides adequate predictive validity. Moving from achieving statistical significance to producing the best practical significance in helping students schedule and sequence their courses provides the greater challenge and one which will require continued work if our institutions are to meet the challenge of accountability set forth by Ewell and Jones.

References

Bank, B. J., Biddle, B. J., & Slavings, R. L. (1994, October). First-semester grades, thought modes, and undergraduate persistence. Contemporary Educational Psychology, v. 19, pp. 416-429.

Beaghen, M., Brozovsky, P., & McLaughlin, G. (1996, October). Explaining grades in individual courses. Presentation at Southern Association for Institutional Research, Mobile, AL.

Boling, B. (1996). The Relationship between Participation in Extracurricular Organizations and the Academic Performance and Retention of College Freshman. Virginia Tech dissertation: Blacksburg, VA.

Bridgeman, B., & Lewis, C. (1996, Fall). Gender differences in college mathematics grades and SAT-M scores: a reanalysis of Wainer and Steinberg. Journal of Educational Measurement, v. 33, pp. 257-270

Eno, D., Brozovsky, P., & McLaughlin, G. (1997, March). Predicting freshman success based on high school record and other measures. Virginia Tech IRPA Report, Vol. 96-97, No. 51. Available netscape: http://www.vt.edu:10021/admin/ir/irpa/irpa9651.html.

Eno, D., McLaughlin, G., & Brozovsky, P. (1997, October). Predicting freshman success in individual courses. SAIR/SCUP Conference, Hot Springs Arkansas. Available netscape: http://www.vt.edu:10021/admin/ir/irpa/irpa9751.html.

Ewell, P. T., & Jones, D. P. (1991, July). Assessing and Reporting Student Progress: A Response to the "New Accountability." National Center for Higher Education Management Systems: SHEEO, Denver CO.

McLaughlin, G. W., Brozovsky, P. V., & McLaughlin, J. S. (1998, February). Changing perspectives on student retention: a role for institutional research. Research in Higher Education, v. 39, n. 1. Human Sciences Press, Inc.: New York.

Noble, J. P. , & Sawyer, R. (1987). Predicting Grades in Specific College Freshman Courses from ACT Test Scores and Self-Reported High School Grades. ACT Research Report Series, 87-20: Iowa City, IA.

Noble, J., & Sawyer, R. (1997). Alternative Methods for Validating Admissions and Course Placement Criteria. AIR Professional File, Number 63, Spring; Association for Institutional Research for Management Research, Policy Analysis, and Planning: Thirty-Fifth Annual AIR Forum, Boston.

Spencer, H. E. (1996, December). Mathematical SAT test scores and college chemistry grades. Journal of Chemical Education, v. 73, pp. 1150-1153.

Wainer, H., & Steinberg, L. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: a bi-directional validity study. Harvard Educational Review, v. 62, pp. 323-336.

Williford, L. E. (1996, October). The freshman year: how do personal factors influence academic success and persistence? 1996 SAIR Annual Report: Charting the Course in a Changing Environment. Southern Association for Institutional Research and Southern Region of the Society for College and University Planning Conference: Mobile, AL.

NOTICE

REPRODUCTION BASIS