DOCUMENT RESUME

ED 059 248 TM 001 062

AUTHOR Nichols, K. E.

TITLE A Research Project Using the Safran Student Interest Inventory (SSII): Discriminant Analysis of University Majors.

INSTITUTION Calgary Separate School Board (Alberta).

PUB DATE 31 May 71

NOTE 13p.; Paper presented at the Annual Convention of the CGCA, Toronto, Canada, May 71

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *College Bound Students; *College Majors; Counseling; Data Analysis; *Discriminant Analysis; Grade Prediction; Grouping (Instructional Purposes); *Predictive Validity; *Standardized Tests; Student Interests; Student Placement; Testing Programs; Test Results

IDENTIFIERS Differential Antitude Test; *Safran Student Interest Inventory

ABSTRACT

Safran Student Interest Inventory (SSII) data was gathered on 135 university students registered in five different faculties. A discriminant analysis of the data indicated that the SSII was a good test for separating students into faculties and therefore would make a good counselling instrument. Some results are also present using Differential Aptitude Test scores to predict first year grade point averages. (Author/DLG)

A research project using the Safran Student Interest Inventory (SSII): Discriminant Analysis of University Majors *

K. E. Nichols **
Research Psychologist
Calgary School Board

Introduction

The type of research where one wants to predict future position requires a longitudinal approach. Either one must begin when the students are still in high school and follow up on those that go to univers'ty or one can begin when the students are in university and then trace their test results back through cumulative records. This second method has one major advantage over the first in that it can be carried out at one time. It is not necessary to wait for the students to finish high school and then go to university. One problem with this approach, though, is the headache of incomplete records and the necessity to search large quantities of records to find complete data on enough subjects to make the study worthwhile.

Because of the desire to get quick results it was this second approach that was used here. The University of Calgary Registrar's Office cooperated in making available the names, faculties registered in, the majors, and the first year grade point averages of all the students from the Calgary schools who were, at that time, students there. Only those faculties with enough students from the Calgary system were used. This included Education, Arts and Sciences, Engineering, Business, and Fine Arts.

Sample

At the time these second year university students were in grade 10, the SSII was not universally given in the Calgary High Schools. This

* Presented at the annual CGCA Convention, Toronto, May 31, 1971
** Presently at the Laboratory of Personality and Group Analysis, U. of Illinois, Champaign

resulted in a high attrition rate for the initial sample of over 600 names. As data was also collected on the Differential Aptitude Test (DAT) at the same time, this even further limited the number of subjects with complete data. The use of the DAT as a predictive instrument was also analyzed in this project but will not be reported here because of space limitations. Table 1 shows the numbers in each group who had complete data on the SSII alone and also the numbers who had complete data on both tests.

Table 1

Group Sizes

	SSII	SSII + DAT
Arts and Sciences	54	37
Business	22	13
Education	25	17
Engineering	29	22
Fine Arts	5	(1) (not used)
Total	135	89

Instruments

SSII -- There have been 2 versions of the SSII at this time, a first version and a revised version, each with 7 scales. The one we are concerned with here is the first version and only 5 of the scales as that is all we had data on. The items in each scale are based on a forced choice answer, i.e., either one or the other. The scores are their total choices for each scale. The scale names used here are:

1. Economic
2. Technical
3. Humane
4. Artistic
5. Scientific

This test takes about 1 hour to complete and is scored by the student who takes it. This gives it a high efficiency coefficient as defined by Cattell (1968).

The SSII was published in final form in 1969 (Safran and Wright) along with a handbook containing some technical data. Because of its newness it has been the subject of practically no research.

DAT -- This test has 8 subtests, namely: 1. Verbal Reasoning; 2. Numerical Ability; 3. Abstract Reasoning; 4. Clerical Speed and Accuracy; 5. Mechanical Reasoning; 6. Space Relations; 7. Spelling; and 8. Grammar. The scores from this test were converted to percentiles which are not very satisfactory for further analysis as they are not based on equal interval scaling. This test takes about 4 hours to complete and may be machine scored.

Analysis of Data

The information obtained for each student was coded and punched on IBM data cards to make computer analysis possible. As there was more than one score for each subject and more than 1 faculty to be dealt with this problem was best handled by a technique of multivariate analysis. The predictor scores were continuous, normally distributed variables while the criteria were dichotomous dummy variables to indicate group membership.

Each faculty was considered as a variable. Students were given a one if they were in the particular faculty and a zero if they were not.

A discriminant analysis was carried out on the data. This was done three times, the first with the SSII only data, the second using the SSII as predictors but only using data complete on both the SSII and the DAT, and the third using the DAT scores as predictors. The second two were done to compare the usefulness of each test in prediction. This was best done by using the same subjects for each analysis.

A discriminant analysis finds the best weightings of the predictors that will maximally separate the groups in question. A significant overall value of the lambda or D^2 statistic says that the predictors separate at least one group from at least one other at the level used in the test of significance. This is not usually a very interesting piece of information as one is more often interested in which groups the predictors work on and which ones they don't work on.

The program used in this analysis was developed by the author and gives the following information:

1. means and standard deviations for each group on each variable
2. dispersion matrix
3. roots and weights of the discriminant functions
4. correlations of each variable with each function
5. centroid means for each group
6. univariate F-tests for each variable
7. multivariate F-tests of differences between means for each pair of groups
8. Classification matrix of how many were correctly classified into each group by the analysis
9. discriminant scores for each subject

The tables and description of the results will be discussed in the next section.

Results and Discussion

The first analysis to be described included all the subjects who had SSII scores. Tables 2 and 3 give the means and standard deviations for each group on each variable. It can be seen that there are considerable differences even in raw score means between the groups.

Table 2

Means for each group on each variable

	A & S	Bus	Ed	Eng	F. A.
Economic	13.7	17.1	17.6	10.9	13.0
Technical	11.9	17.6	7.9	19.7	3.8
Humane	15.6	10.3	16.7	10.4	19.6
Artistic	14.1	8.9	17.2	10.5	23.6
Scientific	18.6	19.9	15.6	21.4	14.4

Table 3

Standard Deviations for each group on each variable

	A & S	Bus	Ed	Eng	F. A.
Economic	6.4	6.5	5.8	5.0	5.5
Technical	8.2	6.8	7.2	5.7	2.6
Humane	6.2	5.3	5.3	5.1	2.1
Artistic	6.1	4.7	7.4	6.1	5.3
Scientific	5.1	5.1	4.3	3.8	2.6

The overall test of significance indicates that these predictors separate at least 2 of these groups at a very significant level. The Mahalanobis D^2 value is 111.17 which can be tested as a chi-square with 20 degrees of freedom and is significantly different from zero at greater than the .001 level. (Another test, the Wilkes lambda, equals .505 and is significant at the same level.)

Of the four roots extracted, it can be seen from Table 4 that only 2 are significant. These 2 account for 97.37% of the total variance accounted for; therefore only these two will be considered further.

Table 4

Roots and their significance

Root	% of variance	Chi-square	d.f.	Probability
1	79.90	67.69	8	.000
2	17.47	18.09	6	.007
3	2.58	2.83	4	.589
4	.06	.07	2	.967

Table 5

Weights for the two significant factors

Variable	Factor 1	Factor 2
1. Economic	.2278	.9636
2. Technical	-.1747	.1310
3. Humane	.6586	-.0457
4. Artistic	.6381	-.1698
5. Scientific	-.2770	-.1527

These weights multiplied by an individual's scores and then summed give his discriminant score on each factor. Correlating these new discriminant scores with each of the original variables indicates which variables are most closely aligned with the reference axis, or what the discriminant factor is measuring. That is, they describe the discriminant dimensions in terms of the original variables. The square of these correlation coefficients indicates the percentage of variance of that variable accounted for by that factor.

Table 6

Correlations of variables with factors

Variable	Factor 1	Factor 2
1. Economic	.247	.953
2. Technical	-.842	-.093
3. Humane	.709	-.159
4. Artistic	.738	-.281
5. Scientific	-.617	-.266

Table 6 indicates that Factor 1 accounts for over 50% of the variance of variables 2, 3, and 4, and only 6% of the variance of variable 1 and 38% of the variance of variable 5. Factor 2 consists mainly of variance from variable 1 and discriminates on this basis. Squaring and summing across rows indicates the proportion of each variable's variance that is accounted for in this analysis. For example, $.617^2 + .266^2 = .458$, which says only 45.8% of the variance of variable 5 is accounted for here (96% of variance of 1 is accounted for).

Table 7 gives the centroid values for each of the factors. These are found by multiplying the factor weights by the group means and summing the result.

Table 7

Group Centroids

	Group	Factor 1	Factor 2
1.	Arts & Sciences	15.14	8.87
2.	Business	7.80	13.74
3.	Education	20.34	11.95
4.	Engineering	6.67	7.50
5.	Fine Arts	26.27	5.92

These scores plotted in 2 dimensions on a graph indicate better what is happening for each factor.

Figure 1

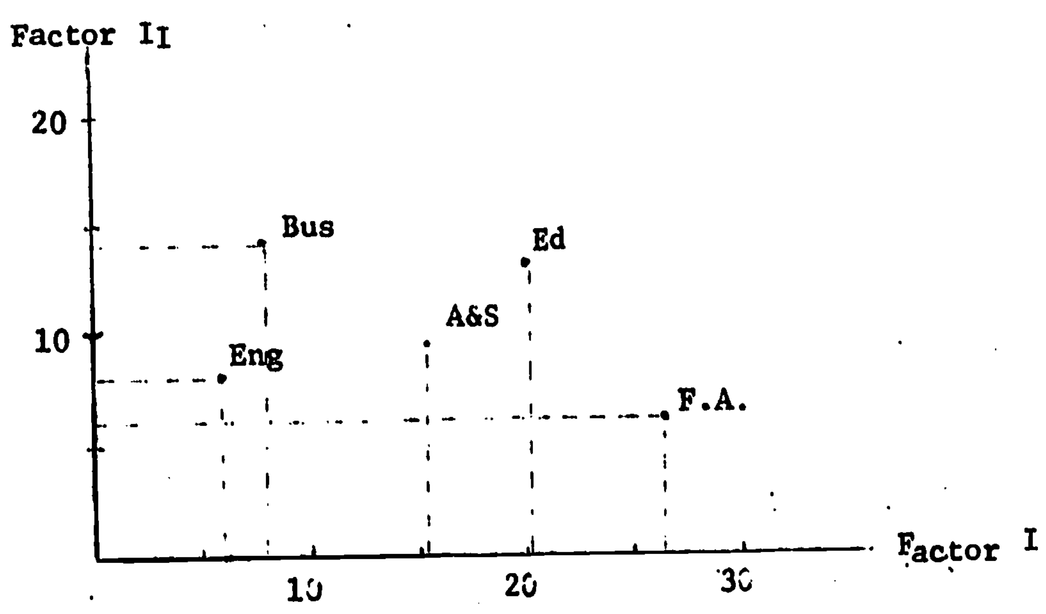

Figure I indicates that all but Business and Engineering are separated by some distance on Factor I, while only Business and Engineering, and Business and Fine Arts are separated on Factor II. Whether or not the distance separating them is significant is the next question. To answer this, each pair of groups has been compared on the basis of a multivariate test of the difference between means. Because each pair of groups are not independent from all the rest, this test suffers from the same problems as multiple student t-tests on groups which are not independent. A method for finding significance between groups in analysis of variance seems appropriate in the multivariate sense. This was proposed by Tukey and is called the Dunn method. In principle it says that if one wishes to test for significance at the α (say .05) and has tests dependent on one another then one must use a level that is $\frac{\alpha}{k}$ for all tests to be sure all will be significant at the specified level (k = number of groups to be compared). In this case if we want all 5 groups to be significantly different at the .05 level before we reject the null hypothesis, we must conduct each test at the $\frac{.05}{5}$ or .01 level.

Table 8 presents these tests and shows that 7 out of the 10 are significant at the .05 level. (We look for the .01 level.) Also this table indicates that only one pair of groups is really not discriminated, that is, Education from Fine Arts, and this would likely improve if the n for Fine Arts were a little larger. Another explanation is that there is overlap between the faculties depending on the students' major.

Table 8

Specific tests for each pair of groups

Pair	F-ratio	d.f num.	d.f den.	Probability
A&S - B	5.32	5	70	.00
A&S - Ed	2.92	5	73	.01
A&S - En	5.93	5	77	.00
A&S - FA	2.66	5	53	.03
B - Ed	7.29	5	41	.00
B - En	2.46	5	45	.04
B - FA	6.13	5	21	.00
Ed - EN	11.10	5	48	.00
Ed - FA	1.27	5	24	.31
En - FA	6.16	5	28	.00

Another statistic of interest is the univariate F-tests for each variable. This tells whether or not each individual variable by itself causes a significant difference between the groups. It is like 5 one-way analyses of variance with each of the predictor variables as the dependent variable in turn. This is useful to indicate whether or not any variable does not differ significantly in means for the groups. If one did not, then it likely would be adding little to the discriminant analysis.

Table 9 shows that all differ significantly. This can also be seen in the matrix of means. No one variable has a constant mean over the 5 groups.

Table 9

Univariate F-tests for each variable

d.f. between = 4, d.f. within = 130

Variable	F-ratio	Probability
Economic	5.41	.00
Technical	13.29	.00
Humane	8.91	.00
Artistic	10.12	.00
Scientific	6.48	.00

Table 10 indicates how many of the students were properly placed by using the weights of this analysis. It can be seen that there is considerable overlap between the groups. This can be partially explained by the fact that the majors in each area overlap. For example, one might major in Chemistry, in Education, or Arts and Science and also may take Chemical Engineering. Another explanation is that all students are not properly placed. This is only conjecture of course, but is worth considering as we know many students switch faculties after their first year.

Table 10

Classification Matrix

	A&S	Bus	Ed	Eng	FA	% correct
Arts & Sciences	15	8	11	11	9	28
Business	2	11	3	6	0	50
Education	3	4	12	1	5	48
Engineering	4	7	1	17	0	58
Fine Arts	0	0	1	0	4	80

It is not surprising that A&S and Education have the lowest percentage of correct placements as they have the greatest diversity of subject matter areas.

The other two analyses will be discussed briefly here just to show that one does not always get highly significant results by using discriminant analysis.

Table 11

Results of two discriminant analyses

Result	Using SSII	Using DAT
Overall test (lambda)	.521	.697
F-ratio	3.97	1.25
Probability	.00	.20
Number of specific tests significant	4/6	0/6
Number of univariate tests significant	5/5	2/8
Overall percentage of correct placements	54	41

This information serves to show two things. First, that all analyses don't give significant results and second, the SSII is a better discriminator than the DAT.

Conclusion

The above data supports the hypothesis that the SSII is a useful instrument for counselling college bound students. It is short, easy to administer, easy to score, and it discriminates well between faculties.

In this paper the DAT was used for a purpose for which it was not really meant. If one were to use the DAT or the SSII to predict grade point averages the DAT would prove far superior. Therefore, it too is a useful counselling tool in helping a student decide whether or not to go to college.

References

Cattell, R. B., and Warburton, F. Objective Personality and Motivation Tests. University of Illinois Press: Urbana, Illinois, 1968.

Safran, C., and Wright, E. N. Safran Student's Interest Inventory: Counsellor's Manual. Thomas Nelson and Sons: Don Mills, Ontario, 1969.