

DOCUMENT RESUME

ED 122 874

JC 760 257

AUTHOR Capoor, Madan; Eagle, Norman
 TITLE Dropouts in Two-Year Colleges: Better Prediction with the Use of Moderator Subgroups.
 PUB DATE [Apr 76]
 NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage
 DESCRIPTORS Community Colleges; *Dropout Research; Dropouts; Junior Colleges; Measurement Techniques; *Multiple Regression Analysis; *Persistence; *Predictor Variables; *Research Methodology; Statistical Analysis

IDENTIFIERS *Moderator Variables

ABSTRACT

Failure to identify and account for the effect of moderator variables is an important reason for the low explanatory power of much educational research. Pre-existing subgroups such as sex, ethnicity, and curriculum offer an easily identifiable and theoretically meaningful source of moderator variables. Tests for intercept and slope differences in a multiple regression analysis offer a convenient and reliable way to test the significance of these moderator variables. Using this improved statistical methodology, this paper analyzes data on 32 cognitive and non-cognitive variables collected on the Bronx Community College freshman class at the time of their entry in fall 1972. The 32 variables include measures of personality, family background, attitudes, aspirations, and academic ability. At the end of one year, the 984 sampled students were classified as dropouts or persisters, and the variables were examined to determine their effectiveness as predictors of persistence. The categories of sex, ethnicity, and curriculum were selected to test for their moderating effect. Complete statistical data are included, as are separate tables listing significant predictor variables for the 12 curriculum sub-populations. Definitions of the predictor variables are appended. (Author/NHM)

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. Nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

ED122874

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Dropouts in Two-Year Colleges:
Better Prediction with the Use of
Moderator Subgroups

MADAN CAPOOR

Director of Institutional Research
Middlesex County College, Edison, New Jersey

NORMAN EAGLE

Director of Institutional Research
Bronx Community College, Bronx, New York

In a survey of all the articles appearing in The American Educational Research Journal (AERJ) from 1970 to 1974, Brown (1975) found seventy-eight citations of Multiple R or R^2 as an indication of the overall success of the multiple regression analyses used in the studies reported. Brown found that the mean R^2 for the seventy-eight cases was .259, and that in 87 percent of the cases the R^2 fell below .5. In fact, in 72 percent of the cases, the R^2 could not cross the .3 level. These results do not speak well of the explanatory power of educational research, particularly when it is remembered that the studies reported in the AERJ represent the best that is being done in the field of educational research.

One of the many reasons for such poor results found in educational research is the assumption made by many researchers about the homogeneity of the population being studied, and consequently their

failure to detect and account for any moderator variables, particularly in the form of certain embedded subgroups within their study sample. Such moderator variables can and do affect the relationship between the predictors and the criterion and a proper accounting for their moderating effects can enhance the explanatory power of a research analysis. While moderator variables can be profitably used in most of the statistical procedures employed in educational research, their discussion in this paper is confined to the context of a multiple regression approach.

Nissitz and Schoenfeldt (1974) used a method for detecting and utilizing moderator variables which does not appear to be very satisfactory. On the basis of a factor analysis of the biographical data on the subject of their study, they identified 21 male and 14 female subgroups. They conducted separate multiple regression analysis for each of the 35 subgroups, and then compared the overall results achieved through the subgroup analyses with those achieved through a single analysis of the total group. They found that the use of the separate subgroup analysis did not produce better results than those of the single total group analysis, and, hence, they concluded that "simple is best," and that "the total group least squares procedure consistently resulted in predictions equal to or better than did those of the other procedures." This is too rash a generalization to be based upon the findings of a single study, particularly when these findings might have resulted from a faulty methodology as explained below.

Any and every subgrouping variable cannot be assumed to be a moderator variable. The significance of the moderating effect of a categorical variable has to be established before that variable can be used as a moderator. Lissitz and Schoenfeldt apparently did not perform any tests of significance on the moderating effect of these subgroups. Instead, having established the 35 subgroups, they took their moderating power for granted. But, as their own findings reveal, the conduct of separate regression analyses for the 35 subgroups did not significantly increase the amount of criterion variance explained over that explained through the analysis of the total group, which indicates that these subgroups did not exercise a significant moderating influence on the criterion-predictor relationships in their study.

Moreover, the procedure of creating ad hoc subgroups on the basis of a factor analysis of a particular set of data has certain drawbacks. Firstly, these subgroups not being natural, may be difficult to interpret. Secondly, the created subgroups may partly be a function of the given data and may not prove stable unless they have been properly cross-validated.

A better and more convenient method is to identify the natural, pre-existing subgroups in a population and then test for the significance of their moderating effect. Moreover, even when the moderating power of the subgroups has been established, it is not

necessary to splinter the total sample into too many subgroups, because as Lissitz and Schoenfeldt rightly observe, by doing so one may be "trading an increase in homogeneity for a loss in precision of parameter estimation." (Page 72). A preferred procedure would be to use some of the important moderator variables to form the subgroups, while using interaction terms to account for the moderating effect of other categorical variables.

An alternative method, as suggested by Lissitz and Schoenfeldt and commonly practiced by educational researchers, which consists of using moderator variables as additional predictors in a single total group analysis, is also not adequate.

The effect of a moderating variable on the predictor-criterion relationship can occur in two ways. Firstly, the intercepts for the regression slope lines can be different for the different subgroups, which means that the predictor means in the different subgroups are not the same. For example, the mean score on the need for aggression for males may be higher than for females. On the other hand, the slopes of the regression lines in the different subgroups may not be parallel, which means that they intersect each other at some point. An example of this would be where the relationship between the need for aggression and dropping out is found to be positive for males, but negative for females (see Tables 9 and 10). In such a case, if one total group regression

analysis is conducted, the contrary relationships for males and females will cancel each other, and the need for aggression may fail to appear as a significant predictor of attrition and, thus, the explanatory power which could be contributed by the need for aggression would be lost.

The use of moderator variables as additional predictors can only take care of the intercept or the level differences, but such a use cannot account for the slope differences if there are any. Once the slope differences have been established, the only way to account for them is either by running separate subgroup analyses or by introducing interaction terms between the moderator variables and the different predictors in the regression equation.

To summarize, the proper procedure for taking care of any moderating effect of the subgrouping variables is to, first, test for the moderating effect through a test for slope differences. If the slope differences are found to be insignificant, the categorical variables may then be used as additional predictors. However, if the slope differences are found to be significant, their effect should be accounted for either by running separate subgroup analyses, if the resulting subgroups are not too small, or by utilizing interaction terms in the regression equation. This is the procedure followed in our own study, a description of which will make the details of the procedure more explicit.

On the basis of the past studies conducted locally and at other colleges, it was concluded that the lack of academic ability or preparedness was not enough to account adequately for the phenomenon of dropping out from college. Consequently, the focus was shifted on selected non-cognitive variables, and following an interaction model, it was hypothesized that students drop out because of a mismatch or conflict between their psychological and intellectual needs and attributes on the one hand, and the social, psychological and academic demands and characteristics of a college on the other hand (Spady, 1970). Holding the college environment constant by limiting the study to one college, data on 32 cognitive and non-cognitive variables was collected on a freshman class at the time of their entry in the Fall, 1972 to an urban community college. The 32 variables consisted of measures on personality, family background, attitudes, aspirations and academic ability. The details on these variables are given in Appendix A. In order to sample the ~~absentee~~ freshmen, a second administration of the research instrument was conducted for a sample of the freshmen who were absent at the time of the first administration. A multivariate analysis of variance on the continuous variables and chi square tests on the categorical variables revealed no significant differences on these variables between the two samples obtained through the two administrations of the research instruments. In view of this, the two samples were combined. At the end of one year the sampled students were classified as dropouts or persis-

tors depending upon their status in the college. The study is based upon the data for 984 subjects for whom information on all the research variables was available.

The categories of sex, ethnicity, and curriculum were selected to test for their moderating effect. This was done through the test for slope differences as recommended by Johnston (1972). The method consists of first conducting a multiple regression analysis on the total group and then separately for each of the categorical subgroups. The residual sums of squares obtained from the total group analysis are compared with the residual sums of squares pooled from the subgroup analyses. If the running of separate subgroup analyses result in a significant reduction in the residual sums of squares, as compared to the residual sums of squares obtained from the total group analysis, it indicates that the slopes of the regression lines are not the same in the different subgroups. In plain language, it means that the relationship between the criterion and some of the predictors is not the same among all the subgroups. In order to identify the predictors which have a significantly different relationship in different subgroups, interaction terms between the various categories of the subgrouping variables and the different predictors must be tested for significance in the total group regression analysis.

If, as indicated earlier, the tests for slope differences show no significance, it will be fruitless to either run separate subgroup

analyses or to introduce interaction terms in the total group equation. Introduction of the categorical variables as additional predictors in the total group equation should take care of the level differences, if any, among the subgroups. If this procedure is followed, one would not need any a priori omniscience, as Nevick (1974) demands, to choose between the total group or within group model of the least squares.

As Table I shows, slope differences for the categories of sex, ethnicity and curriculum were found to be significant in our study at the .05 or lower level. The figures in column three indicate that while the total group analysis could account for only 9.8 percent of the criterion variance, the subgrouping of the total group by the categories of sex resulted in an additional 2.8 percent of the explained criterion variance, subgrouping by ethnicity produced an additional 9.2 percent of the variance, while the use of curriculum subgroups increased the explained criterion variance by 15.1 percent, respectively.

Significant slope differences among the categories of sex, ethnicity and curriculum having been established there were two options to choose from: Either run 48 sex by ethnicity by curriculum subgroup analyses, or use subgrouping for one or two of the three categorical variables and use the remaining categorical variable(s) for producing interaction terms. As the first option would have

resulted in subgroups too small for any reliable regression analyses, the second alternative was chosen and it was decided to conduct 12 sex by curriculum analyses, and the slope differences due to ethnic categories were accounted for through the use of interaction terms. Tables 3 through 14 present the results of the 12 subgroup regression analyses, while Table 2 presents a summary of the overall results of the twelve analyses.

On an average, 45 percent of the criterion variance was explained through the twelve regression analyses as against the mean of 25.9 percent found by Brown in the studies reported in AERJ. In half of the 12 analyses, the explained variance was more than 50 percent, in two of these groups it went above 70 percent, while in the case of female students in Business transfer curriculums, it reached a peak of 78.3 percent. In a quarter of the analyzed subgroups, the explained variance fell between 25 percent and 50 percent, while in another quarter of the subgroups it fell below the 25 percent level.

These are impressive results. However, to our great regret and for reasons beyond our control, these results have not yet faced the crucial test of cross-validation. The data for cross-validation has been collected. As soon as the cross-validation is completed, the results will be reported through some research publication.

CONCLUSION:

Failure to identify and to account for the effect of moderator variables is an important reason for the low explanatory power of a large portion of educational research. Pre-existing subgroups such as those based on sex, ethnicity, and curriculum, offer an easily identifiable and theoretically meaningful source of moderator variables. For testing the significance of these moderator variables, tests for intercept and slope differences in a multiple regression approach offer a convenient and reliable way. It is hoped that efforts to identify and properly account for the moderator variables in a multiple variable study may offer better rewards in terms of improved level of prediction and explanation than the addition of more predictors.

TABLE I
RESULTS FOR THE TESTS OF SLOPE DIFFERENCES

Categorical Variance	No. of Subgroups	% of the Criterion Variance Explained	F-Ratio for Slope Differences	p
Sex	2	12.6%	1.79	< .05
Ethnicity	4	19.0%	1.91	< .01
Curriculum	6	24.9%	1.89	< .001
Total Group	1	9.8%		

TABLE 2
THE AMOUNT OF CRITERION VARIANCE EXPLAINED
IN THE TWELVE CURRICULUM SUBGROUPS

Subgroup	Multiple R	Amount of Criterion Variance Explained (R ²) in %	F-Ratio for Overall R
Lib. Arts Sc. Male	.716	51.3	6.95*
Lib. Arts Sc. Female	.424	18.0	6.85*
Lib. Arts Non-Sc. Male	.616	37.9	5.94*
Lib. Arts Non-Sc. Female	.393	15.4	4.29*
Business Transfer Male	.638	40.6	4.70*
Business Transfer Female	.885	78.3	10.83*
Engineering Sc. Male	.725	52.5	7.01*
Nursing Female	.764	58.3	6.30*
Business Career Male	.848	71.8	5.25*
Business Career Female	.420	17.6	6.51*
Technology Male	.675	45.6	9.33*
Technology Female	.724	52.4	7.70*

* p < .001

Table 3
Liberal Arts Males (science major)
 Predictors Contributing Significantly to Regression

No.	Predictor	B weight	F ratio
1.	Autonomy	-0.327	6.98*
2.	Exhibition	-0.310	6.29*
3.	Educational Goals	0.359	8.40**
4.	Grades Aimed At	0.385	9.04**
5.	Withdrawal Due to F.D.	-0.289	5.06*

* .05 level ** .01 level *** .001 level

Table 4
Liberal Arts Females (science major)
 Predictors Contributing Significantly to Regression

No.	Predictor	B weight	F ratio
1.	H.S. Grad. Date	0.220	7.09**
2.	Grades Aimed At	0.257	9.90**
3.	Fin. Hardship	0.229	7.90**
4.	Blacks x Curr. Satis.	-0.171	4.32*

* .05 level ** .01 level *** .001 level

Table 5
Liberal Arts Males (non-science major)
 Predictors Contributing Significantly to Regression

No.	Predictor	B weight	F ratio
1.	Dominance	-0.279	11.09**
2.	Impulsivity	0.183	4.60*
3.	Nurturance	0.184	5.17*
4.	Play	-0.359	17.76***
5.	Ability	0.219	7.37**
6.	Hisp. x Achievement	-0.814	14.11***
7.	Hisp. x Curr. Satis.	0.713	10.88**
8.	W. Cath. x Edu. Goals	0.283	6.59*
9.	W. Cath. x S.E.S.	-0.441	12.50***
10.	Blacks x Grades Aimed At	0.398	6.49*
11.	Blacks x Hours Employed	-0.370	5.90*

* .05 level ** .01 level *** .001 level

Table 9
Engineering Science Males
 Predictors Contributing Significantly to Regression

No.	Predictor	B weight	F ratio
1.	Aggression	0.322	6.35*
2.	Nurturance	0.290	5.01*
3.	Job vs Education	-0.315	7.80**
4.	W. Cath. x Hours Empl.	-0.324	7.07*
5.	Blacks x Curr. Satis.	-1.099	19.36***
6.	Blacks x Grades Aimed At	0.815	11.12**

* .05 level

** .01 level

*** .001 level

Table 10
Nursing Females
 Predictors Contributing Significantly to Regression

No.	Predictor	B weight	F ratio
1.	Aggression	-0.389	8.75**
2.	Social Recognition	0.814	29.79***
3.	Newspapers	0.282	5.08*
4.	Hispanics	-16.243	13.05***
5.	Hisp. x Hours Empl.	-0.599	11.42**
6.	W. Cath x H.S. Grad. Date	-0.878	4.23*
7.	W. Cath x Hours Empl.	-1.542	21.46***
8.	W. Cath x S.F.S.	-14.229	12.18**

* .05 level

** .01 level

*** .001 level

Table 11
Business Career Males
 Predictors Contributing Significantly to Regression

No.	Predictors	B weight	F ratio
1.	Achievement	0.463	12.33**
2.	Dominance	-0.333	8.11**
3.	Endurance	0.317	5.77*
4.	Play	0.456	16.99***
5.	H.S. Graduation Date	-0.654	16.24***
6.	Edu. Goals	0.282	6.31*
7.	Newspapers	0.308	9.35**
8.	Hours Employed	-0.368	8.83**
9.	S.E.S.	0.520	17.01***
10.	Ability	0.353	9.14**
11.	Blacks x Endurance	-1.110	8.97**
12.	Blacks x Order	1.319	20.85***
13.	Blacks x H.S. Grad. Date	0.607	7.37**
14.	Blacks x Edu. Goals	-0.555	4.27*
15.	W. Cath. x H.S. Grad. Date	1.746	23.02***
16.	W. Cath. x Curr. Satis.	-1.337	6.82*
17.	W. Cath. x Imp. of C.E.	-1.758	17.39***
18.	W. Cath. x Hours Emp.	0.929	9.60**

* .05 level

** .01 level

*** .001 level

Table 12
Business Career Females
 Predictors Contributing Significantly to Regression

No.	Predictor	B weight	F ratio
1.	Order	0.186	5.03*
2.	Chances of Grad.	0.252	9.96**
3.	W. Cath. x Achievement	-0.993	14.36***
4.	W. Cath. x Ability	0.935	12.76***

* .05 level

** .01 level

*** .001 level

REFERENCES

1. Brown, Daniel J., Down with the Lineal Model. American Educational Research Journal, 1975, 12, 491-505.
2. Johnston, J., Econometric Methods, New York: McGraw Hill, 1972.
3. Kerlinger, F. N., and Pedhazur, E. J., Multiple Regression in Behavioral Research, New York: Holt, Rinehart & Winston, 1973.
4. Lissitz, R. W., & Schoenfeldt, L. F., Moderator Subgroups for the Estimation of Educational Performance: A comparison of Prediction Models. American Educational Research Journal, 1976, 11, 63-75.
5. Novick, M. R., & Jackson, P. H., Further Cross-Validation Analysis of the Bayesian M-Group Regression Method. American Educational Research Journal, 1976, 11, 77-85.
6. Spady, W. G., Dropouts from Higher Education: An Interdisciplinary Review and Synthesis. Interchange, 1970, 1, 64-85.

Appendix A

Definition of the Criterion and the Predictor Variables

Criterion

Dropouts and persisters: Freshmen who entered B.C.C. in Fall, 1972 but did not register for any courses in Fall, 1973, and who left the college with no grade index or with an index of less than 2.0 are considered dropouts for this study. Freshmen who entered B.C.C. in Fall 1972 and who were registered at B.C.C. in Fall, 1973 are considered persisters for this study.

Predictors

Ethno-religious Groups: On the basis of the information obtained through a Biographical Inventory, the sampled students were categorized into six ethno-religious sub-groups as follows. Abbreviations used for these sub-groups are shown in parenthesis.

1. Black Americans (Blacks): All students identifying themselves with this ethnic category irrespective of their religious affiliation.

2. Puerto Ricans and Spanish Americans (Hispanics):
All students identifying themselves with this category irrespective of their religious affiliation.

3. White Catholics (White Cath.): All students identifying themselves as "White Americans" and indicating affiliation with Catholic religion.

4. White Protestants (White Prot.): All students identifying themselves as "White Americans" and indicating affiliation with Protestant Religion.

5. Jews (Jews):* All students identifying themselves as "White Americans" and indicating affiliation with Jewish Religion.

6. Others and Mixed (Mixed):* All students in the sample who are not covered by any of the five categories above.

Curriculum Groups: The information on students' curriculum group was obtained through the Biographical Inventory, and the liberal arts group was split into science and non-science groups on the basis of the pattern of courses taken by the students.

1. Liberal Arts, science major. (Lib. Arts. Sc.)
2. Liberal Arts, non-science major, or Performing Arts and Music (Lib. Arts, Non-Sc.)
3. Business Teaching or Administration (Business Transfer)
4. Engineering Science (Eng. Sc.)
5. Nursing (Nurs.)
6. Accounting, Retailing, Secretarial Studies or Data Processing (Bus. Career)
7. Electrical and Mechanical Engineering, or Plastics, Chemical and Medical Labs. Technology (Technologies)

Personality Needs¹

1. Achievement: Aspires to accomplish difficult tasks; maintains high standards and is willing to work toward distant goals; responds positively to competition; willing to put forth effort to attain excellence.

* As there were not enough subjects in these three categories they were not used in the analyses.

1. Measured through the Personality Research Form (PRF) by Douglas N. Jackson. The definitions of these needs are taken from the PRF manual.

III

2. Affiliation: Enjoys being with friends and people in general; accepts people readily; makes efforts to win friendships and maintain associations with people.
3. Aggression: Enjoys combat and argument; easily annoyed; sometimes willing to hurt people to get his way; may seek to "get even" with people whom he perceives as having harmed him.
4. Autonomy: Tries to break away from restraints, confinement, or restrictions of any kind; enjoys being unattached, free, not tied to people, places, or obligations; may be rebellious when faced with restraints.
5. Dominance: Attempts to control his environment, and to influence or direct other people; expresses opinions forcefully; enjoys the role of leader and may assume it spontaneously.
6. Endurance: Willing to work long hours; doesn't give up quickly on a problem; persevering, even in the face of great difficulty; patient and unrelenting in his work habits.
7. Exhibition: Wants to be the center of attention; enjoys having an audience; engages in behavior which wins the notice of others; may enjoy being dramatic or witty.
8. Harmavoidance: Does not enjoy exciting activities, especially if danger is involved; avoids risk of bodily harm; seeks to maximize personal safety.
9. Impulsivity: Tends to act on the "spur of the moment" and without deliberation; gives vent readily to feelings and wishes; speaks freely; may be volatile in emotional expression.
10. Nurturance: Gives sympathy and comfort; assists others whenever possible, interested in caring for children, the disabled, or the infirm; offers a "helping hand" to those in need; readily performs favors for others.
11. Order: Concerned with keeping personal effects and surroundings neat and organized; dislikes clutter, confusion, lack of organization; interested in developing methods for keeping materials methodically organized.
12. Play: Does many things "just for fun," spends a good deal of time participating in games, sports, social activities, and other amusements; enjoys jokes and funny stories; maintains a light-hearted, easy-going attitude toward life.

IV

13. Social Recognition: Desires to be held in high esteem by acquaintances; concerned about reputation and what other people think of him; works for the approval and recognition of others.
14. Understanding: Wants to understand many areas of knowledge; values synthesis of ideas, verifiable generalization, logical thought, particularly when directed at satisfying intellectual curiosity.

Other Predictors¹

15. Sex
16. Socio-economic status based on the education, income and occupation of parents (SES).
17. Ability Score: Based on the average of Nelson Denny reading score and high school average after standardization.
18. Date of high school graduation (H.S. Grad. Date).
19. Educational Goals (Edu. Goals)
20. Satisfaction with placement at B.C.C. (B.C.C. Satis.)
21. Satisfaction with curriculum placement (Curr. Satis.)
22. Student's perception of his chances of graduating from B.C.C. (Chances of Grad.)
23. Importance of college education for achieving success in life. (Imp. of C.E.)
24. What letter grades does the student work for. (Grades Aimed At)
25. Number of non-assigned books borrowed from any library during the last one year. (N.A. Books)
26. Frequency of newspaper readership during a week (Newspaper)
27. Preference between pursuing college education in order to get a good paying job after a few years, and leaving college to get a job and live comfortably now. (Job vs. Edu.)
28. Expected hours of employment per week during the current school year. (Hours Empl.)
29. Chance of student's withdrawal from college because of the difficulty in financing his college education (With F.D.)
30. The amount of financial hardship that student thinks will be caused to him or his family due to his attending college (Fin. Hard.)

1. Measured through a Biographical Inventory.

UNIVERSITY OF CALIF.
LOS ANGELES

JUN 11 1976

CLEARINGHOUSE FOR
JUNIOR COLLEGES