Multiple Linear Regression Viewpoints

A Publication sponsored by the American Educational Research Association's Special Interest Group on Multiple Linear Regression: The General Linear Model

MLRV

Volume 25 • Number 2 • Fall 1999

Table of Contents

Multiple Linear Regression Viewpoints (ISSN 0195-7171) is published by the AERA Special Interest Group on Multiple Linear Regression (SIG/MLR) through Southern Illinois University at Carbondale. Subscription and SIG membership information can be obtained from Susan Tracz, MLR:GLM/SIG Executive Secretary, Department of Educational Research, Administration & Foundations, California State University-Fresno, 5005 N. Maple Ave., Fresno, CA 93740-8025. *MLRV* abstracts appear in CIJE, the ERIC system, and microform copies are available from University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106. *MLRV* is listed in the *EBSCO Librarians Handbook*. Second-class postage paid at Fresno, CA. POSTMASTER: Send address changes to:

Susan Tracz, MLR:GLM/SIG Executive Secretary,
Department of Educational Research, Administration & Foundations,
California State University, Fresno
5005 N. Maple Ave.
Fresno, CA 93740-8025

Using Simulated Annealing for Selection in Multiple Regression Analysis

Zvi Drezner
George A. Marcoulides
California State University - Fullerton

This note presents the simulated annealing heuristic search procedure as an alternative variable selection method for use in multiple regression analysis. The procedure performs better than traditionally used model selection techniques.

Drezner, Marcoulides and Salhi (1999) recently illustrated the heuristic Tabu search procedure as an alternative variable selection method for use in multiple regression analysis. The Tabu search procedure was compared to traditionally used regression analysis procedures (e.g., maximum R^2 and stepwise selection). The results of the study indicated the superiority of the Tabu search procedure over other model selection procedures in multiple regression analysis and comparability to the all-possible regression that may require prohibitive computer time. Using simulated data sets, Tabu search found the optimal solutions for all test problems examined without any computational difficulty.

The purpose of this note is to present the simulated annealing search procedure, which is a different heuristic search technique, for model selection in multiple regression analysis. To examine the capabilities of the simulated annealing search procedure, the same simulated data sets used by Drezner et al. (1999) were analyzed.

Simulated Annealing for Model Selection

Consider a multiple linear regression model with n observations and k independent variables. The most commonly used criterion to help in choosing between alternative equations in multiple regression is the R^2 (adjusted or unadjusted), the F-ratio based on R^2, along with the statistical significance of the F-ratio (Schumacker, 1994). Obviously, this criterion could easily be replaced by any other selection from those available in the literature. Based upon the selected criterion, the objective is to find the subset of independent variables that yields the lowest significance level among all possible subsets. For example, with 26 independent variables 2^{26} = 67,108,864 possible subsets must be calculated along with their significance levels. As such, it should be obvious that a very large number of equations need to be examined even when the number of independent variables is relatively small.

Simulated annealing (SA) is ideally suited for solving all types of large-scale optimization problems (Kirkpatrick, Gelat, & Vecchi, 1983). The process simulates the annealing of metals by starting with a high temperature and cooling the metal off. The process of simulated annealing has been successfully used for the solution of numerous optimization problems in the field of operations research (see Salhi, 1998 for a review and detailed description of the method).

The general simulated annealing (SA) approach is described below. Following the general description, we present the particular parameters used to solve the multiple regression model selection issue examined in this note. A FORTRAN coded computer program for model selection in multiple regression is available upon request from the authors.

The General SA Approach

1. A starting solution is selected.
2. A starting temperature T_0 is selected.

   (T_i is the temperature in iteration i.)
3. The following iterations are repeated N times.
4. At iteration i:

   a. A perturbation of the current selected set is randomly generated.

   b. The difference between the values of the objective function of the current set and the perturbed set, Δf , is calculated.

   c. If the perturbation results in a better objective function, it is accepted and the set of selected variables updated.

   d. If the perturbation results in a worse objective function, the quantity $\delta = \Delta f / T_i$ is calculated.

   e. The perturbed set is accepted with a probability of $e^{-\delta}$. Otherwise, the selected set remains unchanged and the perturbation ignored.

   f. The temperature T_i is changed to T_{i+1}.

Specific Parameters needed for Multiple Regression

1. The empty set was selected as a starting solution (i.e., no independent variables).
2. The starting temperature was set to $T_0 = 1$ This means that if the perturbation doubles the significance level, it is accepted 37% of the time.
3. A perturbation of the current selected set is created by randomly selecting an independent variable. If the variable is in the current set, it is moved out, and if it is not in the current set it is put in.
4. The number of iterations was set to N=10,000.
5. Since our objective function is a significance level, which varies a lot among problems, we replaced the change in the objective function Δf with the relative change in the objective function $\Delta f / f$ where f is the value of the objective function of the current set.
6. The last selected set was selected as the solution. One may keep the best solution encountered throughout the iterations as the solution.
7. The success of the simulated annealing procedure depends on the selection of the starting temperature T_0, the way the temperature is lowered, and the number of iterations. We kept the temperature constant for blocks of 100 iterations each. When a block of 100 iteration is completed, the temperature is multiplied by the value 0.95. One hundred blocks of 100 iterations each were executed for a total of 10,000 iterations. This lead to a final temperature of 0.006. At the end of the procedure, a deterioration in the significance level by a factor of 1.05 is accepted with probability of only 0.0002.

Computational Results

The simulated annealing procedure was tested on the simulated data sets examined by Drezner et al (1999). The data sets used had 50 observations and k variables ranging in $17 \leq k \leq 26$. The data for the smallest problem with $k = 17$ independent variables and $n = 50$ are presented in Table 1 (the remaining data sets are available upon request from the authors). Using this data set, the optimal subset of independent variables includes #2, #6, #12, and #17. It is important to note that the proposed simulated annealing procedure found this optimal solution. In contrast, stepwise regression produced the set #1, #4, #5, #7, #12, #13, #17 (when the entry selection level was set to 0.15), the set #1, #5, #12, #17 (when the entry selection level was set to 0.05), whereas maximum R^2 found the set #2, #5, #7, #1, #13, #17. Interestingly, variable #6, which is in the optimal group, was never identified by any of the other procedures, and variable #5, which is not in the optimal group, was included by the other procedures. Table 2 presents the results of the comparison between the SA procedure and the maximum R^2 and stepwise procedures. As can be seen in Table 2, the SA procedure found the best subset for all the data sets examined. In contrast, the other procedures were not very systematic in selecting the optimal solution. It is important to note that the results obtained using the simulated annealing procedure were identical to those obtained by Drezner et al. (1999) using their proposed Tabu search procedure which is a different local search procedure.

References

Drezner Z., Marcoulides, G.A., & Salhi, S. (1999). Tabu search model selection in multiple regression analysis. *Communications in Statistics - Computation and Simulation*, *28*(2), 349-367.

Kirkpatrick S., Gelat, C.D., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, *220*, 671-680.

Salhi S. (1998). Heuristic search methods, in G.A. Marcoulides (Ed.) *Modern Methods for Business Research*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

Schumacker, R.E. (1994). A comparison of the Mallows *Cp* and principal component regression criteria for best model selection in multiple regression. *Multiple Regression Viewpoints*, *21*(1), 12-22.

Table 1. Data for the 17 Variable Problem

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	y
10	12	11	22	25	29	33	34	32	26	28	33	23	21	19	25	31	182
22	22	22	16	20	16	15	17	12	17	12	14	12	15	15	15	20	129
35	35	33	36	38	28	28	25	22	26	24	17	17	14	15	18	27	188
41	42	30	31	23	19	19	21	28	33	36	27	27	21	17	20	21	170
12	13	14	16	18	27	32	25	19	22	18	27	19	19	21	28	33	167
31	36	34	24	27	20	26	19	17	22	27	20	16	20	22	23	27	149
0	1	11	7	15	21	19	15	20	18	13	23	25	30	24	23	28	125
27	24	20	22	17	24	29	33	35	36	27	21	18	18	27	23	16	130
10	19	15	25	24	30	21	24	28	23	28	25	24	25	18	13	19	135
29	21	26	31	32	23	17	14	18	27	19	21	18	23	18	26	31	159
4	14	13	11	16	24	21	19	27	22	27	31	25	18	20	18	14	151
21	15	15	18	20	28	29	29	21	21	29	28	23	18	20	27	20	141
46	32	25	20	21	23	26	19	13	23	28	31	23	30	27	33	28	181
10	16	21	14	20	17	20	18	19	15	20	25	20	14	10	20	22	132
37	36	36	32	29	23	25	27	19	25	19	23	27	19	24	20	25	180
41	38	40	35	25	19	19	18	25	17	26	28	29	27	19	20	25	164
13	18	16	25	18	20	16	17	23	19	14	21	24	25	22	15	21	155
27	28	26	27	33	33	33	25	19	24	21	15	17	26	31	28	24	156
39	29	25	19	28	23	27	25	28	26	29	27	29	29	34	37	35	195
37	36	26	22	30	30	30	31	30	35	34	24	18	17	19	14	17	176
4	13	18	19	27	20	17	12	21	19	18	25	24	29	21	14	22	149
10	8	11	11	16	22	21	23	30	30	21	26	25	30	23	21	18	158
14	10	19	15	23	26	20	27	30	33	36	31	27	28	23	18	22	143
31	27	22	23	23	21	16	15	19	22	29	30	30	24	25	31	25	184
47	41	40	34	32	28	32	25	19	15	19	21	20	24	29	23	30	206
34	38	35	33	30	28	22	18	21	16	24	19	20	15	24	31	31	186
34	34	34	38	35	34	25	20	17	17	15	11	21	16	25	31	23	193
21	20	18	16	24	28	31	25	29	21	20	24	18	22	27	20	25	146
22	16	16	15	24	27	27	32	28	28	25	23	24	25	28	20	27	157
14	24	27	21	21	19	21	21	21	28	29	23	16	16	19	20	26	134
12	21	17	14	20	17	12	13	12	9	6	8	15	21	19	15	24	98
3	5	14	12	19	19	13	13	22	27	31	33	27	32	25	22	29	143
24	27	21	22	29	20	18	16	16	13	20	19	27	23	18	19	22	156
47	40	39	35	27	20	26	28	29	31	24	17	13	16	22	17	26	186
39	30	23	19	18	23	26	30	21	18	27	31	26	24	21	23	20	151
5	16	24	24	21	23	26	32	26	21	21	16	14	21	23	24	22	146
20	22	18	14	20	21	26	28	32	28	33	34	27	30	25	28	29	194
13	10	13	16	24	19	27	22	30	34	27	18	21	17	20	25	24	146
28	32	35	35	35	29	30	29	34	33	27	21	17	24	24	28	28	192
32	37	34	31	22	16	18	23	23	16	22	25	30	35	34	32	25	172
49	49	35	37	37	27	27	27	33	33	32	33	23	19	25	22	26	206
34	36	29	30	26	27	32	25	19	19	24	29	32	32	27	33	24	182
35	28	24	18	18	20	22	16	11	14	20	17	12	14	17	15	24	155
10	10	11	12	9	10	7	12	16	16	12	9	11	16	24	26	21	90
19	18	24	24	31	32	22	28	31	29	30	28	23	27	20	19	15	152
8	17	21	26	18	13	20	28	30	31	24	24	18	26	21	17	25	140
1	8	14	24	30	33	33	26	32	36	31	25	22	21	27	32	28	151
15	21	22	16	22	27	29	22	20	27	22	26	30	31	25	21	17	165
48	45	35	29	26	21	26	24	27	24	25	18	23	21	26	22	23	173
3	2	5	13	15	20	24	20	19	21	20	14	11	17	20	16	13	98

Table 2. Comparison of Simulated Annealing to Other Regression Procedures

Number of Variables	Variables in Optimal Solution	SA Procedure	Stepwise Procedure		Max R^2 Procedure	
			Include	Exclude	Include	Exclude
17	2, 6, 12, 17	*identical*	1,4,5,7, 13	2,6	5,7,13	6
18	1, 6, 12, 13, 16, 17	*identical*	7	—	7	—
19	2, 6, 12, 13, 17	*identical*	1	—	8,10,15	—
20	1, 5, 7, 12, 13, 16, 17	*identical*	4	—	4,8,15	
21	1, 6, 12, 13, 17, 18	*identical*	*identical*		*identical*	
22	1, 4, 6, 7, 9, 13, 17, 18, 22	*identical*	*identical*		*identical*	
23	1, 2, 3, 6, 12, 13, 17, 22	*identical*	4	—	4	—
24	1, 4, 5, 6, 9, 11, 12, 13, 16, 18, 22, 24	*identical*	*identical*		*identical*	
25	1, 3, 4, 6, 8, 10, 12, 13, 15, 17, 19, 22, 25	*identical*	5,18,23	3,4,8,10, 15,19	2,7	1
26	1, 3, 6, 12, 13, 16, 17, 18, 20, 22, 24, 26	*identical*	5	—	8	—

Note: The term ***'identical'*** indicates that the final set of variables selected by that procedure is the same as the optimal set. The columns headed by **'Include'** indicate that the given procedure includes variables not in the optimal solution, and those headed by **'Exclude'** indicate that the given procedure excludes variables which are members of the optimal set.

Stability of Self-Esteem: Demonstrating the Effects of Various Assumptions in Regression-Style Analyses

Lee M. Wolfle
Virginia Polytechnic Institute and State University

Every statistic requires some assumptions. This paper examines some of the assumptions in regression-style analyses of the stability of self-esteem, and inspects the consequences of some of the assumptions one makes with regard to measurement error and the distribution of variables.

Every statistic requires some assumptions. And to the extent the assumptions are not met, the statistics calculated will vary consequently. Robustness studies can estimate to some extent those consequences by systematically varying the assumptions against known but arbitrary parameters. In most situations, however, one is faced with sample data with unknown parameters, with assumptions met with varying degrees of accuracy, and with unknown inferential consequences.

Background

One of the most frequently studied constructs in psychology is self-esteem. It has most often been measured with some form of Rosenberg's (1965) instrument, in which respondents rate themselves on a Likert-type scale from strongly disagree to strongly agree in response to items such as "I feel I am a person of worth, on an equal plane with others."

Just this question was asked in the National Longitudinal Study (NLS) of the High School Class of 1972 (Riccobono, Henderson, Burkheimer, Place & Levinsohn, 1981), which was designed to provide data on the development of educational, vocational, and personal aspects of the lives of adolescents as they made the transition from high school to the adult world. Included among the many items of data collected for over 20,000 respondents were four self-esteem items, including the one quoted above. The analysis reported here was restricted by listwise deletion of missing data to 3,511 U.S. white males who answered these items completely in 1972 during their senior year of high school and seven years later in a 1979 follow-up study, along with information provided about their postsecondary educational attainment.

The apparently simple question to be addressed here is the stability of self-esteem for U.S. white males in the seven years following high school. Do those who exhibit high (or low) self-esteem in 1972 continue to do so seven years later?

One approach toward addressing such a question would be to simply regress the 1979 responses on their 1972 counterparts and thus estimate the average rate of change of self-esteem in metric or standardized form (more complete information is conveyed, of course, by reporting both). The closer the estimated coefficient comes to unity, the greater the agreement between 1972 and 1979 responses.

However, such a regression approach requires certain assumptions that may or may not be met in varying degrees. Indeed, any analytic approach toward estimating the stability of self-esteem will involve certain assumptions, and it is the purpose of this short paper to briefly touch upon these as they affect the motivating question about the stability of self-esteem. The purpose is not to find a definitive answer, but rather to inspect the consequences of the assumptions one adopts in seeking an answer.

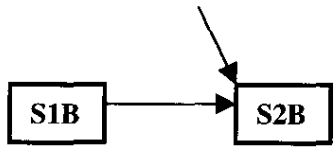
Regression Examples

Take the regression of 1979 responses to "I feel I am a person of worth..." on the 1972 responses to the same question. For this sample, the resulting regression equation was estimated to be:

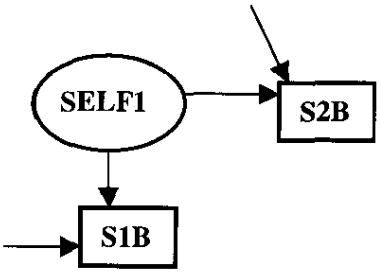
$$S2B = a + .150\ S1B + e$$
$$(\hat{\beta} = .160;\ R^2 = .026) \quad (.016) \qquad (1)$$

in which S2B stands for self-esteem measured at time 2 on item B (of four self-esteem items included in the NLS), and S1B stands for the same item at time 1; "*a*" is an intercept or constant; .150 is the estimated metric regression coefficient and its standard error is shown below it in parentheses; "*e*" is the error of prediction or residual; $\hat{\beta} = .160$ is the standardized regression coefficient, and $R^2 = .026$ is the coefficient of determination.

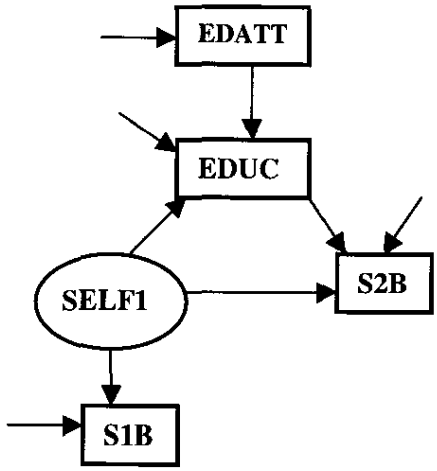
The same equation can be shown diagrammatically in Figure 1-A, in which the arrow from S1B to S2B indicates that S1B is thought to be a cause of, or to cause changes in, S2B; and the short disconnected arrow represents all sources of variation in S2B not explained by, and not correlated with, S1B.



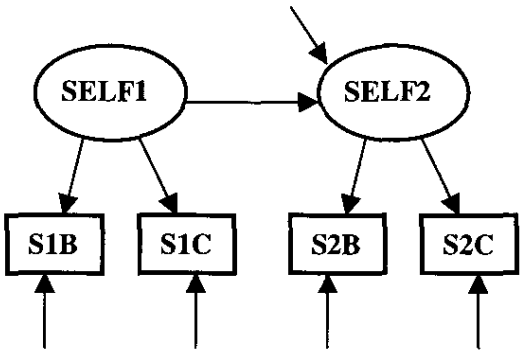
Model 1-A



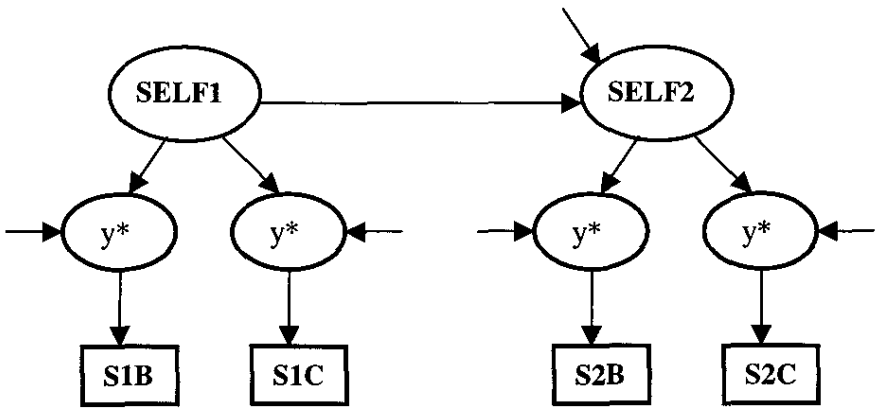
Model 1-B



Model 1-C



Model 1-D



Model 1-E

Figure 1. Models of the Stability of Self-Esteem

These results indicate that a one-unit change in self-esteem in 1972 may be expected to produce an average change of .150 units of self-esteem reported in 1979, and that less than 3 percent of the variation in self-esteem in 1979 is explained by self-esteem in 1972. The remaining 97 percent of the variation in self-esteem in 1979 is due to all other unmeasured (in this equation) sources of variation, including random changes, measurement error in self-esteem, deviations from linearity, and all other unspecified causes of self-esteem in 1979 uncorrelated with self-esteem in 1972.

The statement that a one-unit change in self-esteem in 1972 may be expected to produce an average change of .150 units of self-esteem reported in 1979, however, is based on an implicit assumption about the accuracy with which self-esteem was measured. As pointed out, for example, some time ago (Costner, 1969; Werts, Rock, Linn, & Jöreskog, 1976; Wolfle, 1979) and more recently (Rigdon, 1994), the regression approach assumes, in this instance, that self-esteem was measured perfectly! More specifically, it assumes that S1B was measured with reliability equal to 1.0. But as Schumacker and Lomax (1996, *p.* 38) have pointed out, the effect of unreliable variables on statistics can sometimes have dramatic effects.

Adding a measurement component to the analysis could be shown diagrammatically in Figure 1-B, in which Self1 is considered to be a so-called latent, unmeasured variable, thought to be a cause of S1B, the manifestly measured variable. In the previous equation, the coefficient thought to relate Self1 to S1B would be 1.0, and the residual of S1B (or, more accurately, the residual variance) would be zero, indicating no error of measurement. In this case, then, Figures 1-A and 1-B would be identical.

What if we relax the assumption of perfect measurement of the independent variable? What if we assume the reliability of the measurement of self-esteem in 1972 was less than unity? We can do this. Let us assume the reliability was 0.70. (This is not completely arbitrary, and is approximated from a confirmatory factor analysis with these data of the four self-esteem items included in the NLS.)

Assuming that the reliability of S1B was r_{XX} = .70 would imply that the error variance for S1B would be (1 - r_{XX})(variance of S1B) = (1 - .70)(.315) = .0945 (see Jöreskog & Sörbom, 1993, *p.* 37). The model thus implied in Figure 1-B was estimated with LISREL 8.30 (Jöreskog & Sörbom, 1993) using maximum likelihood estimates derived from the variances and covariances for the data described above.

The resulting (structural portion in LISREL terminology) equation was estimated to be:

$$S2B = a + .215\ Self1 + e$$
$$(\hat{\beta} = .192;\ R^2 = .037)\quad (.022) \qquad (2)$$

In this case, assuming S2B was regressed on S1B corrected for measurement error (i.e., Self1), one would expect that a one-unit change in self-esteem (Self1) in 1972 would be expected to produce a change of .215 units of self-esteem in 1979 (S2B) with an R-square of .037.

That errors in the independent variable reduce the coefficient in a bivariate regression is well known (e.g., Walker & Lev, 1953, *p.* 305). As seen here, the uncorrected regression coefficient of .150 underestimates the corrected (by measurement error) estimate of .215 by 30%. The converse is not true — errors in the dependent variable have no effect on ordinary least squares regression estimates, since such errors are absorbed as ordinary disturbances of prediction (Goldberger, 1964, *p.* 284).

The extension of this example to the case of two or more explanatory variables introduces unknowns into the system of equations that involve varying degrees of measurement error and multicollinearity among the predictors (Namboodiri, Carter & Blalock, 1975, *pp.* 541*ff*). In the present case, for example, one might expect that additional years of formal postsecondary education from 1972 to 1979 (see, e.g., Pascarella & Terenzini, 1991, *pp.* 162*ff*) might partially mediate (e.g., Baron & Kenny, 1986) the effect of self-esteem in 1972 on self-esteem in 1979, or equivalently said, that a part of the causal relationship of self-esteem in 1972 and 1979 occurs indirectly (e.g., Duncan, 1975) through the intervening accumulation of additional years of education.

Yet the effects of measurement error on such estimates is not necessarily predictable *a priori*. In this instance, consider the model shown in Figure 1-C, in which S2B is seen to be caused by Self1 and Educ, a measure of additional years of formal postsecondary education. Initially, the variable Educ is thought to be a latent variable measured without error, as is Self1. That is, initially Educ and EDATT are thought to be equivalent in which the error of EDATT is zero; similarly, Self1 and S1B are thought to be equivalent. The resulting set of two equations was estimated to be:

$$Educ = a + .204\ Self1 + e$$
$$(R^2 = .007)\qquad (.040)$$
$$S2B = a + .141\ Self1 + .044\ Educ + e \qquad (3)$$
$$(R^2 = .038)\quad (.016)\qquad (.006)$$

The difference in the estimated stability of .150 of self-esteem in Equation (1) and that of .141 in Equation (3) represents the indirect effect of 1972 self-esteem through educational attainment, namely (.204)(.044) = .009. In standard form, the same set of equations would be:

$$\begin{aligned} \text{Educ} &= .204\ \text{Self1} \\ \text{S2B} &= .151\ \text{Self1} + .113\ \text{Educ} \end{aligned} \qquad (4)$$

If we no longer assume perfect measurement, the results are not as tractable. If we assume, as before, that self-esteem in 1972 was measured with reliability of .70, and that educational attainment was measured with reliability of .85, then the model shown in Figure 1-C can be re-estimated by specifying that the errors of S1B and EDATT, respectively, are not zero, but rather (1 - .70)(.315) = .0945 for S1B and (1 - .85)(1.818) = .2727 for EDATT. The resulting set of two equations was estimated to be:

$$\begin{aligned} \text{Educ} &= a + .295\ \text{Self1} + e \\ (R^2 = .012) &\qquad (.058) \\ \text{S2B} &= a + .203\ \text{Self1} + .049\ \text{Educ} + e \\ (R^2 = .051)\ &(.023) \qquad (.008) \end{aligned} \qquad (5)$$

and in standard form:

$$\begin{aligned} \text{Educ} &= .111\ \text{Self1} \\ \text{S2B} &= .180\ \text{Self1} + .117\ \text{Educ} \end{aligned} \qquad (6)$$

These results are not easily related to those previously reported, except by their relative magnitudes, due to variations in reliabilities and multicollinearity (except to note its near absence in this instance), and have led many researchers to assume, for example, that "all instrumental variables are measured without error" (Wonnacott & Wonnacott, 1970, p. 371). Another approach (not pursued here) would be to construct overidentified models that could allow the estimation of, and correction for, random and systematic measurement error in variables (e.g., Wolfle, 1982).

Returning to the consideration of the stability of self-esteem without an intervening variable, with more information than just the single covariance of responses to the self-esteem items in 1972 and 1979 (and an assumed estimate of reliability of self-esteem in 1972 imposed on the model) one could estimate both the reliabilities of multiple self-esteem items and the stability of the latent estimates of self-esteem in 1972 and 1979. This can be accomplished by taking advantage of responses to multiple self-esteem items at the two time periods. Another such stem item was "I am able to do things as well as most other people." If we incorporate that item into the analysis, we could represent it appropriately as shown in Figure 1-D.

Figure 1-D specifies that a latent self-esteem variable, Self1, is the cause of two manifest items in the 1972 survey, S1B and S1C; similarly, Self2 is seen to be the cause of two identically worded items in the 1979 survey, S2B and S2C. In order to establish a metric for the latent variables, the slopes relating Self1 to S1B and Self2 to S2B were set to unity. The other two measurement parameter slopes were free to be estimated, as was the parameter relating Self1 to Self2. There were also two variances of latent variables to be estimated, as well as four error variances for the four manifest variables. With ten variances and covariances among the four manifest variables, this model is actually overidentified with one degree of freedom. Variations of this model have appeared before, some frequently with standardized variables, as early as some of Wright's (1934) work, and early work in the literature of path analysis in sociology (Siegel & Hodge, 1968). In standardized form, with minor restrictions, this is, of course, Spearman's (1907) correction for attenuation. It is also the model introduced by Costner (1969, Figure 4) that came to be known as the walking dog model, because of the visual appearance of the diagram.

This model (shown in Figure 1-D), unlike that of the model implied by Figure 1-A, makes no a priori assumption of error-free measurement. The estimate of the stability of self-esteem is thus adjusted for measurement error. The resulting structural equation was estimated to be:

$$\begin{aligned} \text{Self2} &= a + .291\ \text{Self1} + e \\ (\hat{\beta} = .255;\ R^2 = .065)\ &(.029) \end{aligned} \qquad (7)$$

This result was obtained with LISREL 8.30 (Jöreskog & Sörbom, 1993) using maximum-likelihood estimates from the covariance matrix, resulting in a model that exhibited a likelihood-ratio chi-square of 10.54 with 1 degree of freedom and a root mean square error of approximation (RMSEA; Steigler & Lind, 1980) of .052, which is numerically less than the cutoff value close to .06 recommended by Hu and Bentler (1999).

In substantive terms, then, having relaxed the constricting assumption of error-free measurement of self-esteem, one would expect that a one-unit change in self-esteem in 1972 (Self1) would be expected to produce a change of .291 units of self-esteem in 1979 (Self2), or .255 standard deviations, with an R-square of .065.

This latest estimate, however, is itself not free of assumptions of some kind. In particular, by

estimating the associations by the method of maximum likelihood, one assumes that the manifest variables are distributed multivariate normally, and as we shall see, this is an unrealistic assumption to make with regard to the measurement of self-esteem. For example, the NLS respondents were asked to respond to "I feel I am a person of worth, on an equal plane with others," on a 4-point scale, to which they could respond (1) agree strongly, (2) agree, (3) disagree, or (4) disagree strongly. (For this analysis, the items were reverse coded so that higher scores indicated higher self-esteem.) But these variables are highly skewed and kurtotic; for example, the estimate of skewness was -.377 (z = -4.66) for S2B and the estimate of kurtosis was -1.237 (z = -14.96), as calculated by PRELIS 2.30. Indeed, for these subjects no one agreed strongly with this statement! While non-normal bivariate distributions can occur with normal marginals (Kowalski, 1973), it may be said with near certainty that the non-normal univariate distributions seen here insure non-normal multivariate distributions.

Possible Solutions

A new (actually, a renewed) feature in PRELIS 2.30 (Jöreskog, Sörbom, du Toit & du Toit, 1999, pp. 162ff) provides one possible solution to this violation of the assumption of non-normality, namely to normalize the variables before analysis. The idea would be to substitute normal scores as a continuous variable rather than ordinal scores, but it is doubtful that this tact would work in general and simply does not work in the present case since the ordinal and the so-called normalized variable are correlationally equivalent.

A more useful approach would be to treat self-esteem scores as if they were ordinal and censored measures of latent, continuous normal distributions. Diagrammatically, this may be shown in Figure 1-E, in which the y^* variables represent unmeasured estimates of continuous normal variables thought to be caused by latent self-esteem factors, and thought in turn to underlie the ordinally, manifestly measured self-esteem scores (Muthén, 1984; Jöreskog, 1990, 1994).

Estimates of the moment matrix of the associations among the four y^* variables may be obtained with the use of polychoric correlations among the four variables, a procedure that is available in PRELIS 2.30 (Jöreskog & Sörbom, 1996). We are assuming in this case that there exist normally distributed, continuous y^* variables that underlie the ordinal self-esteem variables; furthermore, the associations among these four y^* variables can be estimated, but their metric is unknown, hence their associations are measured in correlational terms, specifically with polychoric correlations, where an underlying bivariate normal distribution is assumed for each pair. The assumption of bivariate normality among the y^* variables may be tested with a chi-square goodness-of-fit test implemented in PRELIS; for these variables even this assumptions is questionable since all such bivariate tests should be rejected at the .01 level.

In order to estimate the structure implied by Figure 1-E for these variables, a general fit function called "asymptotically distribution free" by Browne (1982, 1984) or "weighted least squares" by Jöreskog and Sörbom (1996) was employed. This procedure requires a weight matrix for the polychoric correlations obtained from PRELIS 2.30. As a practical matter, this weight matrix increases rapidly in size as the number of variables increase. In this instance, we have
$k = 4$ variables with $p = k(k + 1)/2 = 10$ unique moments, and the weight matrix is of the order p x p = 100 with $p(p + 1)/2 = 55$ unique elements. As a further practical matter, to estimate moments of the fourth order with reasonable precision requires very large samples (Jöreskog & Sörbom, 1996, *p.* 28).

With these caveats in mind, estimates for the model implied by Figure 1-E were obtained with PRELIS 2.30 and LISREL 8.30. The resulting structural equation in standardized form was estimated to be:

$$\text{Self2} = a + .292\ \text{Self1} + e$$
$$(R^2 = .085) \qquad (.031) \qquad (8)$$

The full model exhibited a likelihood-ratio chi-square of 8.58 with 1 degree of freedom and an RMSEA of .046. These results can be compared to the standardized estimate shown in Equation (7), and we see that the estimated standardized stability of self-esteem is now estimated to be .292 rather than .255. Which is to say, by specifying a model with self-esteem at two points in time, with two fallible indicators each, all distributed bivariate normally, but measured with censored ordinal variables, then the standardized estimate of stability is .292 with an R-square of .085.

Conclusions and Implications

That last sentence may be difficult to read with all of its clauses, but it represents most of the travails that got us to this point. We began by estimating the stability of self-esteem by assuming it was measured perfectly. That assumption was relaxed a bit by imposing a degree of measurement error on self-esteem as measured in 1972. With more than one

predictor, the estimated stability of self-esteem changes, and changes yet again depending on assumptions of measurement error imposed on the model. If multiple indicators of self-esteem are brought to the analysis, one no longer must assume lack of measurement error, or impose arbitrarily estimated levels of measurement error, but can estimate a model incorporating estimates of both the measurement properties and stability of self-esteem. But those estimates were purchased at the cost of distributional assumptions among the manifest variables. If the manifest variables cannot be assumed to be normally distributed, perhaps their underlying distributions can be, but the data-collection costs to obtain such estimates can be high.

Thus, the estimate of the stability of self-esteem for a sample of high school males depends on the assumptions one is willing to make about the variables involved. In a simulation study, one would start with known parameters and examine on the average how estimates deviated as a consequence of the effects of varying assumptions. In the present case, however, the parameters are unknown and we simply do not know if the correct (standardized) measure of stability is .160, .192, .151, .180, .255, .292, or some other value. It depends on the assumptions one makes, and is willing to defend. I don't think this is a reason to quit in frustration as I perceive some critics of regression-style structural modeling would have us do. Rather, I think it merely requires that one acknowledges the assumptions involved in any statistical application, and addresses them appropriately in the design of the research.

References

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*. Cambridge: Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.

Costner, H.L. (1969). Theory, deduction, and rules of correspondence. *American Journal of Sociology, 75*, 245-263.

Duncan, O.D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Goldberger, A.S. (1964). *Econometric theory*. New York: John Wiley & Sons, Inc.

Hu, L-T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Jöreskog, K.G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity, 24*, 387-404.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381-389.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International, Inc.

Jöreskog, K.G., & Sörbom, D. (1996). *PRELIS 2: User's reference guide* (3rd ed.). Chicago: Scientific Software International, Inc.

Jöreskog, K.G., Sörbom, D., duToit, S., & duToit, M. (1999). *LISREL 8: New statistical features*. Chicago: Scientific Software International, Inc.

Kowalski, C.J. (1973). Non-normal bivariate distributions with normal marginals. *American Statistician, 27*, 103-106.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Namboodiri, N.K., Carter, L.F., & Blalock, H.M., Jr. (1975). *Applied multivariate analysis and experimental designs*. New York: McGraw-Hill Book Company.

Pascarella, E.T., & Terenzini, P.T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass Publishers.

Riccobono, J., Henderson, L.B., Burkheimer, G.J., Place, C., & Levinsohn, J.R. (1981). *National longitudinal study: Base year (1972) through fourth follow-up (1979) data file user's manual*. Washington, D.C.: National Center for Education Statistics.

Rigdon, E.E. (1994). Demonstrating the effects of unmodeled random measurement error. *Structural Equation Modeling, 1*, 375-380.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Schumacker, R.E., & Lomax, R.G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Siegel, P.M., & Hodge, R.M. (1968). A causal approach to the study of measurement error. In H.M. Blalock, Jr. & A.B. Blalock (Eds.),

Methodology in social research. New York: McGraw-Hill Book Company.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18,* 161-169.

Steigler, J.H., & Lind, J.C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA, May.

Walker, H.M., & Lev, J. (1953). *Statistical inference.* New York: Holt, Rinehart and Winston.

Werts, C.E., Rock, D.A., Linn, R.L., & Jöreskog, K.G. (1976). Comparison of correlations, variances, covariances, and regression weights with or without measurement error. *Psychological Bulletin, 83,* 1007-1013.

Wolfle, L.M. (1979). Unmeasured variables in path analysis. *Multiple Linear Regression Viewpoints, 9*(1), 20-56.

Wolfle, L.M. (1982). Causal models with unmeasured variables: An introduction to LISREL. *Multiple Linear Regression Viewpoints, 11*(2), 9-54.

Wonnacott, R.J., & Wonnacott, T.H. (1970). *Econometrics.* New York: John Wiley & Sons.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics, 5,* 161-215.

School Identification and Dropping Out of School

Noreen Yazejian
University of North Carolina at Chapel Hill

The purpose of this study was to explore the relationship between school identification and school dropout using data from the National Eduation Longitudinal Study of 1988 (NELS:88). School identification was conceptualized as having components of belonging (social support and participation) and value (the extent to which students value school). In addition to school identification factors, eighth grade measures of achievement, retention history, parental expectations, and demographic characteristics, which have all been shown in previous research to relate to dropout, were included in the analysis. Logistic regression analyses indicated that the set of three school identification factors made a significant contribution in predicting dropout, above and beyond the contribution made by the academic, demographic, and parental expectation variables.

One of the major concerns facing educators today is high school dropouts. Although an enormous amount of theorizing and research has been directed to the problem of school dropout, relatively few large-scale studies have examined the issue as a complex event that is embedded within a larger social context. In addition, relatively few researchers have explored the problem using longitudinal data.

The extent to which students identify with school has emerged as an important correlate of problem behaviors in school, including perhaps dropping out of school. Also described by the terms "affiliation," "involvement," "attachment," "commitment," and "bonding," or by the negative terms "alienation," "withdrawal," and "disengagement," school identification has been defined as having two components: a sense of belonging in school and valuing school-related goals and outcomes (Finn, 1989). While both of these aspects of school identification have been linked empirically to various school outcomes, including, for example, grades, educational aspirations, and dropping out, few studies have examined school belonging and valuing together as the single construct of school identification. In addition, few studies have examined the relationship between school identification and dropping out using quantitative data and methods with large national samples.

The purpose of this study was to explore the relationship between school identification and school dropout using data from the National Education Longitudinal Study of 1988 (NELS:88). School identification was conceptualized as having components of belonging (social support and participation) and value (the extent to which students value school-related goals and outcomes). Students' school dropout status was regressed onto school identification measured at the eighth grade to explore whether identification predicts dropout and thus might be useful as a way of understanding why students drop out of school. In addition, demographic information (ethnicity, gender, and urbanicity) was included in the analysis, and several correlates of school dropout were controlled, including SES, parental expectations, grade retention history, and academic achievement.

Viewed at a broad level, the theoretical framework for this study can be described as a social-ecological systems model of development, in which the individual is seen as embedded within social settings (e.g., family, school, peer groups) that can influence one another (Bronfenbrenner, 1976; Bronfenbrenner & Morris, 1998). In addition, within this systems model, there are assumed to be interactions between the individual and the social settings over time (Sameroff, 1987). Thus, dropping out of school is viewed within this framework as an interplay between the child and the school, family, and peer group settings over time. This study looked at the child within the school setting while considering the influences of peers within that setting and also the influences of the family.

Within the broad social-ecological systems paradigm, this study was framed by current research on motivation in education. Specifically, Finn (1989) has advocated a participation-identification model for understanding school dropout as a developmental process. Within this model, students' identification with school leads them to participate in school and classroom activities and, in a cyclical fashion, this participation lead them to identify even more with school. Through a cycle of non-participation and nonidentification, students may become alienated and eventually drop out of school (Finn, 1989). Goodenow has demonstrated a

positive relationship between students' sense of belonging in school and subsequent motivation and school success (Goodenow, 1993a, 1993b; Goodenow & Grady, 1993). Voelkl (1996) recently has worked to develop a rating scale to measure the degree to which students identify with or disidentify from school. This scale, the Identification with School Questionnaire, includes items measuring belonging in the school setting and valuing school-relevant outcomes. Taken as a whole, the research of Finn (e.g., Finn, 1989), Goodenow and colleagues (e.g., Goodenow, 1993a, 1993b), and Voelkl (e.g., Voelkl, 1995, 1996) provides evidence that students' identification with school may be an important factor in understanding school outcomes and especially dropping out of school.

Literature Review

At least two literatures are pertinent to the current study: literature on school identification and literature on school dropouts. Literature in these two areas is review separately below.

School Identification

School identification has been defined as having two components: (a) a sense of belonging in school and (b) valuing school-relevant goals and outcomes (Finn, 1989; Voelkl, 1996). Voelkl constructed a measure of school identification that included items measuring belonging and valuing and administered the scale to 3,539 eighth graders in 163 schools in Tennessee. Confirmatory factor analyses of responses to the rating scale revealed that a one-factor solution, reflecting school identification, was essentially equivalent, in terms of fit indices, to a two-factor solution, reflecting belonging and valuing separately. However, the comparison of the one- and two-factor models was subjective. Because the two models were not nested, an empirical comparison could not be made. Voelkl concluded that a single factor may provide the best representation of the school identification construct. However, given the lack of an empirical comparison, more research is needed to determine whether a one- or two-factor representation of school identification is most appropriate.

The first component of school identification, belonging, has long been viewed as a fundamental need. In the 1960s, Maslow identified belonging—to be accepted and respected—as a basic human need that must be met before higher goals, including the educational objectives of knowledge and understanding, can be achieved (Maslow, 1968). More recently, Ryan and Powelson (1991) identified relatedness, or the "emotional and personal bonds between individuals" (*p.* 53), as one of three basic needs necessary for motivation, with the other two being the needs for autonomy and competence.

Using labels such as "school membership" (e.g., Newmann, Wehlage, & Lamborn, 1992; Wehlage, Rutter, Smith, Lesko, & Fernandez, 1989) and "belongingness" (Weiner, 1990), other researchers more recently have examined belonging as a potentially important factor in school learning. Goodenow (1993a, 1993b) has been most prominent in delineating the concept of school belonging through the development of a measure of adolescents' perceived belonging or, as she has termed the construct, "psychological sense of school membership" (1993a). Analyses of scores on the scale have revealed dimensions of teacher support, peer support, and participation in school life (Goodenow, 1993b; Hagborg, 1994). This empirical work partially supports the conceptualization of belonging advanced by Wehlage et al. (1989), who identified attachment to adults and peers and involvement in activities as conditions of social bonding. The involvement/participation component of school belonging pertains to participation in school life generally and therefore includes both extracurricular and in-class participation. Classroom participation is often referred to as classroom engagement and has been defined as ". . . active involvement, commitment, and concentrated attention, in contrast to superficial participation, apathy, or lack of interest" (Newmann et al., 1992, *p.* 11).

The second component of school identification, valuing school, is represented by students' assessment of the general importance of schooling and the utility of school for future success (Voelkl, 1996). Based on the literature on school identification and its components of belonging and valuing, school identification can be viewed as having four elements: teacher support, peer support, and participation in school life (as indicators of school belonging) and valuing school-related goals and outcomes.

High School Dropout

Other variables have been shown to relate to dropping out of school, and any analysis of school dropout should consider these variables. The work of Grant and Sleeter (1986) and Fernandez, Paulsen, and Hirano-Nakanishi (1989) emphasized the importance of separating analyses of school dropout by both ethnicity and gender. SES has been a consistent predictor of school dropout in the literature, often accounting for most of the differences in dropout rates between ethnic groups (e.g., Rumberger, 1995). The accumulated evidence on the association between academic variables—

including achievement, aspirations, and grade retention—and school completion has also been very consistent. Students who drop out of school are more likely to have lower grades and achievement test scores (e.g., Kaufman, McMillen, & Sweet, 1996; Rumberger, 1995), to report lower educational aspirations (e.g., Ensminger & Slusarcick, 1992; Rumberger, Ghatak, Poulos, Ritter, & Dornbusch, 1990), and to have been retained in grade (e.g., Grissom & Shepard, 1989; Kaufman et al., 1996) than their graduating peers. Finally, the role of parents in their children's educations must be considered. Although equivocal with respect to the type of participation that is most helpful to children, the literature on parent involvement generally supports its positive effects on student outcomes, including staying in school. One of the more consistent findings in this literature is that parental expectations, as a proxy for parental involvement, is positively associated with staying in school (e.g., Astone & McLanahan, 1991; Ensminger & Slusarcick, 1992; Rumberger, 1995). Based on this body of research on the correlates of dropping out, the present study was designed to include demographic information (gender, ethnicity, SES, and urbanicity), academic variables (achievement, aspirations, and grade retention), and parental expectations (as a measure of parental involvement) to examine whether measures of school identification predict dropping out above and beyond these correlates.

Method

Data and Samples

The data for this study came from the National Education Longitudinal Study of 1988 (NELS:88), which is a major longitudinal study sponsored by the National Center for Education Statistics, U.S. Department of Education. The base year survey was conducted in the spring term of the 1987-88 school year, with followups conducted in 1990, 1992, and 1994. The 8th through 12th grade longitudinal panel sample includes 16,489 students; this sample served as the basis for the present study. Due to insufficient sample sizes, Asian and Native American students were not included in the analyses. Dropping these students resulted in a sample size of 15,303. The 15,303 cases in the data set were randomly divided into two samples. Sample 1 (n=5,107) was used for exploratory factor analyses. Sample 2 (n=10,196) was used to confirm the scales developed with Sample 1 and to perform the logistic regression.

Variables

The dependent variable in this study was the dichotomous variable school dropout (a student is classified as either having dropped out or not having dropped out). For the purposes of this study, the dropout category consisted of any student who at any time during the first three waves of the NELS:88 data collection period (from March 1989 on) dropped out of school. This categorization included students who dropped out but later re-enrolled (termed "stopouts" in NELS:88 parlance) as well as students who were enrolled in or had already completed an alternative program (e.g., General Educational Development or other equivalency programs). The rationale for incorporating this more inclusive definition stems from the purposes of the study. I was interested in exploring how students' identification with school may be related to school dropout. Therefore I was interested in any occurrence of leaving school to explore whether the construct of identification may be of value in explaining school dropout. For the purposes of this study, I was not interested in whether students eventually become "completers," either through re-enrollment or finishing the requirements of an alternative program. I was interested in the phenomenon of school leaving and therefore used a broad definition of school dropout. The NELS:88 data set includes a variable indicating whether the student has ever dropped out. For the present study, this variable was coded as 1=never dropped out, 0=has dropped out.

The main independent variable of interest in the study was school identification. Based on the literature, this broad construct consisted of two components: (a) belonging, made up of social support (peer and teacher relationships) and participation (classroom and extracurricular participation) and (b) valuing school (the extent to which students value school and school-related outcomes). In the final analysis, factor scores were used to represent school identification. Using Sample 1 (n=5,107), I performed an exploratory factor analysis on the set of school identification items. Three school identification factors were found: (a) teacher supportiveness, (b) classroom participation, and (c) valuing school. These three factors were confirmed with Sample 2 (n=10,196), and the equality of the factor structure and loadings was confirmed across gender/ethnic, urbanicity, and SES subgroups. After conducting the confirmatory analysis, I used factor scores to represent school identification in the regression analysis, which was performed using Sample 2.

Demographic information also was included in the analysis to examine main effects for

demographic characteristics as well as two-way interactions among the demographic characteristics and school identification in predicting school dropout. The interactions were included to determine whether school identification was operating differently in predicting dropout for males versus females and for the different ethnic groups. On the NELS:88 questionnaire, students indicated their race as one of the following: 1=Asian/Pacific Islander, 2=Hispanic, 3=Black, not Hispanic, 4=White, not Hispanic, or 5=American Indian/Alaskan Native. As indicated previously, because of small sample sizes, only three of these groups were included in the analysis: Hispanic, Black, and White students. These categories were dummy coded into two variables: 1=White, 0=other and 1=Black, 0=other. (The Hispanic category did not have a dummy variable of its own but was the reference category.) Students' gender was dummy coded as 1=female, 0=male. The geographic location of the students' eighth-grade school was originally coded in the NELS:88 data set as 1=urban, 2=suburban, or 3=rural. These categories were dummy coded into two variables: 1=suburban, 0=other and 1=rural, 0=other, with urban being the reference category. Finally, students' socioeconomic status (SES) was included as a control variable in the analysis. A continuous SES variable is included in NELS:88 and was constructed using data from the base year parent questionnaire or the base year student questionnaire. Student SES was estimated from father's and mother's educational levels, father's and mother's occupations, and family income. The SES variable is a *z*-score (i.e., it has a mean of 0 and a standard deviation of 1).

Three additional variables known to be correlates of dropping out were included in the analysis. First, students' retention history was included. Research has shown that students who have been held back a grade or more are at an increased risk of dropping out of school (e.g., Roderick, 1994). Therefore, a dichotomous variable (0=held back once or more, 1=never held back) was included in the analysis. The NELS:88 data set includes both a student report and a parent report of grade retention. These two reports were compared, and 524 cases of inconsistent reports were found. However, these 524 cases were almost evenly divided between parents indicating a grade retention/students reporting no retention and parents reporting no retention/students indicating a retention. Therefore, it was arbitrarily decided to use the parent report as the primary data source, and to use the student report if the parent report was missing. Second, students' academic potential, consisting of their achievement and aspirations, was included as a control variable in the analysis. Using Sample 1, five achievement items—self-reported grades and achievement test scores in the areas of reading comprehension, mathematics, science, and history-citizenship-geography—and one aspirations item were analyzed with exploratory factor analysis techniques. A single academic potential factor was found consisting of the five achievement items. Using Sample 2 (n=10,196), this factor was confirmed, and the equality of the factor structure and loadings was confirmed across gender/ethnic, urbanicity, and SES subgroups. A factor score was then used as a measure of academic potential in the final analysis. Finally, parental involvement was included as a predictor variable in the analysis. Inconsistent findings have been reported in the literature about which types of parental involvement relate to student outcomes. However, one variable that has been consistently shown to relate positively to student success in school is parental expectations for their children. Therefore, a measure of parental expectations, from the parent data file, was included as a measure of parental involvement. This variable asked parents how far in school they expected their child to go, from 1=less than a high school diploma to 12=Ph.D. or M.D. level.

Models and Statistical Techniques

Exploratory and confirmatory factor analyses of the school identification variables were conducted to create school identification scales from individual NELS:88 items. Three school identification factors were found: (a) teacher supportiveness, (b) classroom participation, and (c) valuing school. In addition to the three school identification factors, eighth grade measures of academic potential (factor score), retention history, parental expectations, and demographic characteristics, which have all been shown in previous research to relate to dropout, were included in the analysis. Students' dropout status as of the 12th grade was regressed onto these 8th grade measures along with the three school identification factors using logistic regression.

Results

Descriptive statistics for the variables analyzed in the logistic regression are provided in Table 1. After deletion of cases with missing values, data from 8,291 students were available for the logistic regression analysis. This indicated that approximately 19% of the sample was not available for the analysis because of missing data. Analyses indicated that the data were not missing at random. A model predicting missing data from a set of demographic characteristics and dropout status revealed that the odds of having missing data were

Table 1. Descriptive Statistics for Variables Analyzed in the Logistic Regression (n=8,291)

Variable	*M*	*SD*	%
Academic Potential	0.07527	0.94430	
Parental Expectations	8.67351	2.76045	
SES	-0.00745	0.76770	
Classroom Participation	.00859	0.81481	
Teacher Support	-0.00512	0.85014	
Valuing School	-0.00434	0.79145	
Not Retained			83.7
Female			50.7
Black			11.6
White			78.9
Rural			32.6
Suburban			44.8
Dropout			15.2

greater for Black students versus Hispanics, for Hispanics versus Whites, for males, for students from lower SES backgrounds, for urban students, and for students who dropped out. The largest predictor of missing data in this model was dropout status—for students who dropped out, the odds of having missing data and thus not being available for analysis in the logistic regression were 1.8 times greater than for nondropouts.

Logistic regression uses maximum likelihood estimation, which produces parameter estimates that make the observed data most likely, i.e., these estimates maximize the likelihood of observing the data that were actually observed. The predictor variables for the logistic regression analysis were entered sequentially into the model as three blocks. First, variables that have been shown in the literature to be correlates of dropping out were entered. These variables included academic potential, retention, parental expectations, SES, gender, ethnicity, and urbanicity. Next, the three school identification factors were entered to examine whether these variables made a significant contribution to the prediction of school dropout, above and beyond that of the other variables. Finally, interaction terms of gender and ethnicity with the school identification factors were entered to explore whether there were differences among these groups in how the variables predict dropping out.

Model fit was initially assessed with χ^2 statistics, which measure the agreement of observed and modeled values. A test of the model with the academic, demographic, and parental expectations variables against a constant-only model was statistically reliable, $\chi^2_{(9)} = 1441.86$, $p < .0001$. This indicates that these predictors, as a set, reliably distinguished between dropouts and nondropouts. A test of this model against the model with the addition of the three school identification factors was also statistically reliable, $\chi^2_{(3)} = 84.94$, $p < .0001$. This χ^2 tested the null hypothesis that the coefficients for the school identification variables were zero. Comparing this model against a model that also included two-way interaction terms between (a) gender and the school identification factors and (b) ethnicity and the school identification factors returned a $\chi^2_{(9)}$ of 9.667, which was not significant ($p < .3781$). Therefore, the interaction terms were dropped, and the model with the set of academic, demographic, and parental expectations variables and the set of school identification factors was retained for interpretation.

In addition to the χ^2 statistics reported above, another measure of model adequacy involves examining a classification table, which is presented as Table 2. Prediction success for this model was mixed: 97% of nondropouts were correctly predicted, but only 23% of dropouts were correctly predicted, for an overall success rate of 86%. Hosmer and Lemeshow (1989) noted that this may be typical of many classification tables seen in applications of logistic regression. As they stated, "Classification is sensitive to the relative sizes of the two component groups and will always favor classification into the larger group, a fact that is also independent of the fit of the model" (p. 147). Thus they recommend using classification only as a supplement to more rigorous fit indices, such as the χ^2 reported above.

When assessing a model, it is also useful to examine how well the model fits each case and how much influence each case has on parameter

estimates. Four diagnostic measures were used to assess the adequacy of the model: residuals, leverage, Cook's D, and DfBetas. Residuals for each case are differences between the observed probability of an event and the predicted probability based on the model. Leverage values are used to detect cases that have a large impact on predicted values. They are somewhat analogous to leverage values in least-squares regression, but in logistic regression, leverage values depend on both the dependent variable scores and the design matrix. Cook's D is another measure of the influence of a case; it indicates the effect of deleting a case on residuals. DfBetas can be calculated for each coefficient and represent the change in logistic regression coefficients when a case is deleted from the model. Plots of residuals, leverage, Cook's D, and DfBetas were visually inspected as an assessment of model adequacy. Several cases had residuals with absolute values greater than 3, which indicates that the model does not fit well for some cases. The leverage, Cook's D, and DfBeta plots revealed that many cases exerted influence on the model, with a small number that appeared to be extreme, but given the large sample size, this might be expected. Given the overall χ^2 statistics and the logistic regression diagnostics, the model appeared to provide a satisfactory fit to the data.

The parameter estimates for this model are presented in Table 3. The Wald statistic tests the hypothesis that the coefficient for a particular variable is 0 and thus indicates which variables reliably predicted dropout status. The asterisks indicate that all of the variables in the model with the exceptions of ethnicity (White versus Hispanic) and valuing school reliably contributed to the prediction of dropping out. The column in the table labeled "R" contains the partial correlations between the dependent variable and each of the independent variables. The small values for R in the table indicate that the variables each had a small partial contribution to the model. The final column of the table contains the antilog of the coefficient estimate, which is the factor by which the odds change when the particular independent variable increases by one unit. (The odds of an event occurring are defined as the probability of the event occurring divided by the probability that it will not occur.) For example, holding all other variables constant, when the gender variable changes from 0 (male) to 1 (female), the odds of staying in school are decreased by a factor of .6558 (the odds of dropping out are greater for females than for males). Stated another way, the odds in favor of dropping out are 1.52 times greater for females than for males. For the school identification factors, the Wald statistic indicated that the valuing school factor was not significant in the prediction of school dropout. The other two factors, teacher support and classroom participation, each had significant coefficients. Holding all other variables constant, a change of one unit in each of these variables improves the odds in favor of staying in school by a factor of about 1.3 (1.28 and 1.29 for classroom participation and teacher support, respectively) (in other words, the odds of staying in school increased 30% with a one unit increase in each of these variables).

Table 2. Classification Table for Observed Versus Predicted Dropout.

	Predicted		
Observed	Dropped Out	Has Not Dropped Out	% Correct
Dropped Out	295	965	23.41%
Has Not Dropped Out	206	6825	97.08%
		Overall	85.88%

Discussion

The two school identification factors of teacher support and classroom participation were significant in predicting high school dropout. The coefficient for the valuing school factor was not significant. The teacher support and classroom participation factors each had about the same effects on the odds of dropping out. For a one unit increase in each of these variables, the odds of staying in school increased by a factor of about 1.3, net of the effects of the other variables in the model.

Previous research has found that valuing school is important in students' decisions to drop out. For example, Pittman (1991) found that 10th-grade students' perceptions of how useful math, English, and trade/business courses were for their futures had a significant effect on dropping out. In the present study, students' perceptions of valuing school were measured in the eighth grade. It may be that the eighth grade was too early to ask students about the utility of subjects for their futures. Such questions may have been too abstract at this grade level, which might explain the lack of relationship between valuing school and dropping out in this study. In addition, it may be that there are important differences in perceptions of valuing school in a general sense and the specific valuing of subject

Table 3. Logistic Regression Analysis of Dropout Status as a Function of Demographic, Academic, Parental Expectation, and School Identification Variables (n=8,291)

Variable	B	S.E.	Wald[a]	R[b]	Exp(B)[c]
Academic Potential	.6263	.0504	154.2086*	.1645	1.8706
Retention	1.1670	.0781	223.1009*	.1983	3.2122
Parental Expectations	.0978	.0124	61.8121*	.1031	1.1027
SES	.5361	.0555	93.2957*	.1274	1.7093
Race (Black)	-.2801	.1292	4.7024*	-.0219	.7557
Race (White)	-.0686	.1138	.3627	.0000	.9337
Gender (Female)	-.4219	.0719	34.4669*	-.0760	.6558
Rural	.2178	.0950	5.2614*	.0241	1.2434
Suburban	.2414	.0915	6.9525*	.0297	1.2730
ClassroomParticipation	.2443	.0446	30.0670*	.0707	1.2767
Teacher Support	.2539	.0506	25.1931*	.0642	1.2890
Valuing School	-.0699	.0527	1.7569	.0000	.9325
Constant	.5146	.1756	8.5875*		

Note: [a]The Wald statistic tests the null hypothesis that a coefficient is 0. [b]The R statistic is the partial correlation between the dependent variable and each independent variable. [c]Exp(B) is the change in odds associated with a one unit increase in each independent variable. * $p < .05$.

matter courses for the future. A student may strongly believe that education is important in a global sense, but may still not see the value of math, for example, for his or her future. Future research should explore the nature of valuing school and its association with dropping out.

The predictor with the largest effect on dropout status in the model was previous retention. For students who had not been retained, the odds of staying in school were 3.2 times greater than for students who had been retained. In addition, a one unit increase in the academic factor, consisting of achievement test scores and self-reported grades, increased the odds of staying in school by a factor of 1.87. Parental expectations also predicted dropping out. For a one unit increase in parental expectations, the odds of staying in school increased by a factor of 1.10, a small but statistically significant effect.

Previous literature has suggested that males drop out at higher rates than females, but that was not the case in the present model. Holding all other variables constant, the odds of dropping out were 1.52 more for females than for males. The difference between this model and previous literature in terms of gender and dropping out may stem from the manner in which dropouts were defined for this study. For the purposes of this study, a student was considered a dropout if s/he left school at any point before the spring term of the 12th grade during the NELS:88 study period, even if that student eventually returned to school or completed an alternative program (e.g., General Educational Development or other equivalency program). With this broad definition of dropouts, it may be hypothesized that because of pregnancies, females left school at higher rates than males. If females had later returned to school or completed alternative

programs, then they would not have been considered dropouts under other definitions. There is some support for this hypothesis in an analysis of NELS:88 data that did not count alternative program completers as dropouts. This study reported no significant difference in the dropout rates of males and females (Kaufman et al., 1996). It would appear then that the male-female discrepancy in dropout rates may no longer exist, and that females drop out at the same rate as males. Because the goal of the present study was to predict high school dropout from eighth grade measures, pregnancy was not included in the model, and therefore its effect on school leaving was not explored. However, future research should explore the consequences for students, perhaps especially females, of leaving school for a period of time to later return (stopping out). Research has examined the issue of stopouts from college, especially community college, but little research has looked at the long-term consequences for stopouts at the high school level.

Previous data have shown that non-Asian minority students consistently drop out at higher rates than White students. However, research has revealed that some of these differences in dropout rates become insignificant when family background variables, especially SES, are controlled (e.g., Rumberger, 1995; Velez, 1989). The results of the present study confirmed these earlier findings. After controlling for the other variables in the model, the odds of staying in school were the same for White and Hispanic students. However, the odds of staying in school were .756 less for Black students than for White and Hispanic students (in other words, the odds in favor of dropping out were 1.32 times greater for Black students than for other students). Thus, controlling for SES and the other demographic and academic variables in the model equalized the odds for Hispanic and White students, but Black students were still more likely to drop out. Rumberger (1995) found that the odds of dropping out for Black students remained higher than other students until a wide range of family background, parental involvement, academic, student attitude, and student behavior variables were controlled. It seems clear from this study and previous research that a gap remains between the educational outcomes of Black students and other students. What is not clear from this study are the reasons why this might be the case. Interactions between ethnicity and school identification were included in the analysis to explore whether differences in levels of identification might explain the disparity in dropout rates, but these interactions were not significant. Thus school identification factors were not operating differently across ethnic groups in the prediction of school dropout. Future work is called for in this area.

The results of the logistic regression must be considered in light of a relatively high percentage of missing data (19% of the sample) that was not available for the analysis. An examination of the missing data revealed that they were not missing at random and thus there might have been important differences between the cases excluded from the analysis and those included. In Sample 2, the sample used for the logistic regression, the percentage of students who dropped out was 17.7%. However, for the cases included in the analysis, the percentage who dropped out was 15.2%, while for the cases excluded because of missing data the percentage was 28.7%. It would seem that missing data, in and of itself, should have been included in the model as a predictor of dropping out. It could be argued that nonresponse on a questionnaire about educational experiences, administered at school during the school day, could be an indicator of disengagement from school. The majority of the cases excluded from the logistic regression analysis were missing data on the school identification factors, which means that these students failed to respond to one or more items in the set of nine teacher relationship, valuing school, and classroom participation items that were analyzed in the factor analysis. Because the students excluded from the analysis due to missing data were more likely to drop out, a complete picture of the dropout phenomenon cannot be pieced together from this study. Missing data unfortunately are an inherent part of research on dropouts. It is difficult to collect data on individuals who leave the school setting, and from this study, it appears that it is difficult to collect information from them even before they actually leave (i.e., in the eighth grade).

The results of the study imply that one strategy for reducing dropout rates would be to improve students' perceptions of teacher supportiveness and encourage classroom participation. Evidence suggests that these two elements, teacher support and classroom engagement, are reciprocally related in that they represent an exchange of commitments between teachers and students (Wehlage et al., 1989). Other researchers have explored the notion of teacher supportiveness as a sense of caring (e.g., Noblit, 1993; Noddings, 1988). The present study suggests that caring may serve as a protective factor against dropping out of school. In this study, teacher support and classroom participation had a positive correlation of about .25. While this is not a large amount of shared variance, the results suggest that adult behaviors that communicate respect and caring for students will increase students' classroom

engagement, and that both of these factors may help keep students from dropping out. Caring makes a difference.

Educators need to explore ways of creating a more caring culture in classrooms and schools. Writers have cautioned, however, that culture is not easy to create, and that caring is more than a set of activities or a particular program (Meier, 1993; Noblit, Rogers, & McCadden, 1995). Noblit et al. (1995) suggested that teachers "consider how helping, talking, and touching can be used to construct a caring culture" (p. 684). Newberg (1995) and Meier (1993) both championed the cluster model, in which students stay with the same small clusters of teachers for several consecutive years, perhaps even across school boundaries (e.g., from middle school to high school). The cluster model promotes a collective responsibility for student learning and offers continuity and the context for caring.

This study makes several important contributions to existing research. First, within the field of education, the research literature addressing the causes of school dropout generally can be described as either psychological, focusing on the individual, or sociological, looking at broader group, school, or societal forces. The present study represented an intersection between these two models in that it looked at student perceptions of the social context and thus offered a way of understanding the process by which broader social forces might influence individual student behaviors, in this case, dropping out of school. Second, much of the research on school dropout has focused on individual factors that are difficult if not impossible to change, including, for example, students' SES or family structure. This study controlled for some of these variables, but focused more specifically on school identification, a school social context variable that may be more amenable to change through school programs and policies that promote identification among students. Finally, this study furthered understanding of the construct of school identification. Finn (1989) delineated the concept as having the two components of belonging and valuing. Voelkl (1996) concluded that a single factor, rather than two separate factors of belonging and valuing, provided the best representation of the construct. The current study provided evidence concerning the structure of school identification, in addition to exploring whether the concept offers a valuable way of understanding school dropout.

This study also pointed to the need for further research on the causes of dropping out. While the model tested in this study showed that eighth grade measures of school identification were statistically significant in predicting dropout, the effects were quite modest. Clearly, other factors, not included in the model, are influencing students' decisions to leave school early. The adverse personal, social, and economic consequences of dropping out of school dictate that researchers continue to explore the factors that influence dropout behavior. Research should also focus on the subgroups that continue to drop out at higher rates than other students, including Black and lower-SES students. SES is consistently a strong predictor of dropping out, but many of the mechanisms that lead to higher rates of dropping out among this group are not known. It may be that there is an interaction between SES and school identification, and that this interaction helps explain dropout behavior. While this study focused on individual students and their perceptions of their social contexts, future research should also explore community-, school-, and individual-level factors that contribute to the decision to drop out.

A goal of this study was to contribute to the understanding of high school dropout. The construct that was hypothesized to help explain students' decisions to drop out, school identification, was indeed predictive of dropping out. However, the contribution that school identification made in the prediction model was modest. In addition, the original construct was proposed as consisting of aspects of belonging and valuing school, but the valuing school component was not significant in predicting dropout. While this study partially achieved its goal, it raised many more research questions about the components of school identification, the nature of this construct for different students, and the many complex factors that contribute to individual students' decisions to leave school early. As we have seen from this study, school identification components helped explain high school dropout, but much work remains to be done to advance our understanding of this complex event so that we can work to reduce the rates of dropping out in our schools.

Correspondence concerning this article should be addressed to:
Noreen Yazejian
Frank Porter Graham Child Development Center
University of North Carolina at Chapel Hill
Suite 300 NationsBank Plaza, CB# 8040
Chapel Hill, NC 27599-8040.
E-mail: yazejian@mail.fpg.unc.edu

References

Astone, N. M., & McLanahan, S. S. (1991). Family structure, parental practices and high school completion. *American Sociological Review, 56,* 309-320.

Bronfenbrenner, U. (1976). The experimental ecology of education. *Teachers College Record, 78,* 157-204.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmenal processes. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993-1028). New York: Wiley.

Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65,* 95-113.

Fernandez, R. M., Paulsen, R., & Hirano-Nakanishi, M. (1989). Dropping out among Hispanic youth. *Social Science Research, 18,* 21-52.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59,* 117-142.

Goodenow, C. (1993a). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30,* 19-90.

Goodenow, C. (1993b). Classroom belonging among early adolescent students: Relationship to motivation and achievement. *Journal of Early Adolescence, 13,* 21-43.

Goodenow, C., & Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *Journal of Experimental Education, 62,* 60-71.

Grant, C. A., & Sleeter, C. E. (1986). Race, class, and gender in education research: An argument for integrative analysis. *Review of Educational Research, 56,* 195-211.

Grissom, J. B., & Shepard, L. A. (1989). Repeating and dropping out of school. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp.34-63). London: Falmer Press.

Hagborg, W. J. (1994). An exploration of school membership among middle- and high-school students. *Journal of Psychoeducational Assessment, 12,* 312-323.

Hosmer, D.W., Jr., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Kaufman, P., McMillen, M. M., & Sweet, D. (1996). *A comparison of high school dropout rates in 1982 and 1992* (NCES 96-893). Washington, DC: U. S. Department of Education.

Maslow, A.H. (1968). *Toward a psychology of being* (2nd ed.). New York: Van Nostrand Reinhold.

Meier, D. (1993). Transforming schools into powerful communities. *Teachers College Record, 94,* 654-658.

Newberg. N. A. (1995). Clusters: Organizational patterns for caring. *Phi Delta Kappan, 76,* 713-717.

Newmann, F. M., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagmenet. In F. M. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11-39). New York: Teachers College Press.

Noblit, G. W. (1993). Power and caring. *American Educational Research Journal, 30,* 23-38.

Noblit, G. W., Rogers, D. L., & McCadden, B. M. (1995). In the meantime: The possibilities of caring. *Phi Delta Kappan, 76,* 680-685.

Noddings, N. (1988). An ethic of caring and its implications for instructional arrangements. *American Journal of Education, 96,* 215-230.

Pittman, R. B. (1991). Social factors, enrollment in vocational/technical courses, and high school dropout rates. *Journal of Educational Research, 84,* 288-295.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal, 32,* 583-625.

Rumberger, R. W., Ghatak, R., Poulos, G., Ritter, P. L., & Dornbusch, S. M. (1990). Family influences on dropout behavior in one California high school. *Sociology of Education, 63,* 283-299.

Ryan, R. M., & Powelson, C. L. (1991). Autonomy and relatedness as fundamental to motivation and education. *Journal of Experimental Education, 60,* 49-66.

Sameroff, A. J. (1987). The social context of development. In N. Eisenberg (Ed.), *Contemporary topics in developmental psychology* (pp. 273-291). New York: Wiley.

Velez, W. (1989). High school attrition among Hispanic and non-Hispanic White youths. *Sociology of Education, 62,* 119-133.

Voelkl, K. E. (1995). School warmth, student participation, and achievement. *Journal of Experimental Education, 63,* 127-138.

Voelkl, K. E. (1996). Measuring students' identification with school. *Educational and Psychological Measurement, 56,* 760-770.

Wehlage, G. G., Rutter, R. A., Smith, G. A., Lesko, N., & Fernandez, R. R. (1989). *Reducing the risk: Schools as communities of support.* London: Falmer Press.

Weiner, B. (1990). The history of motivation research in education. *Journal of Educational Psychology, 82,* 616-622.

Logit Regression: Best Model Selection

Randall E. Schumacker, University of North Texas
Cynthia Anderson, University of North Texas
James Ashby, Richardson ISD, Texas

The typical method of analyzing categorical variables is to use the chi-square statistic. However, with more than two categorical variables, simultaneous examination of main and interaction effects is not feasible. The logit regression technique permits analysis of categorical variables, the modeling of main and interaction effects, control of Type I error, and distribution freer assumptions. This study investigated parsimonious model fit related to the selection of the best set of categorical predictor variables. Findings indicated that the various variable selection criteria (L^2, z, log-odds ratio, R^2_L, model variance, and ΔC^2) provided different results. Order of variable entry also produced significantly different results. The use of a Tabu search procedure and ΔC^2 criteria is recommended to determine the best set of categorical independent predictor variables in logit regression.

Logit regression is a special case of log linear regression where both the dependent and independent variables are categorical in nature (Klienbaum, 1992). It offers distinct advantages over the chi-square method for analysis of categorical variables. Some of these advantages are: (1) control of Type I error rates, (2) modeling of interaction effects, and (3) distribution freer assumptions. The main objective of this study was to investigate the selection of the best set of categorical predictor variables in the presence of main and interaction effects. In logit modeling, natural log odds of the frequencies are computed which allow different models and different model parameters to be compared given the additive nature of the L^2 component for each model.

Logit regression is affected by sample size, outliers and inadequate expected frequencies in categorical cells (Demaris, 1992). This often occurs with too many categorical variables and small sample sizes, hence inadequate cell sizes. In has been understood that cell size should not have fewer than $n = 5$ (Hinkle, McLaughlin, & Austin, 1998; Kennedy, 1992). Another rule of thumb indicated that total sample size should be at least 4 to 5 times the number of cells in the model (Feinberg, 1981). Marasculio and Busk (1987) suggested that low expectancy in cells, possibly due to rare events, should be sampled until adequately filled, and if outliers are suspected, residuals be examined. Collapsing categories is also a reasonable option.

A theoretical logit regression model is generally postulated (null model or base model). A common practice is then to create one or more hierarchical models where each new model contains parameters of the previous model, plus a hypothesized new parameter. The theoretical model can be tested beginning with a null model and adding parameters, or with a saturated model deleting parameters. The best model is selected based on the likelihood ratio statistic, L^2. If the likelihood ratio statistic is significant, then the observed frequencies do not fit the expected frequencies, or in other terms, the data doesn't fit the theoretical model (hypothesized logit regression equation). Several logit regression models may "fit" equally well. In this case, the non-significant likelihood ratio statistics' for the competing models are subtracted yielding a L^2 difference test of model fit analogous to the change in R^2 in regression analysis. If the model change is not significant, then the most parsimonious model is typically chosen. Identification of significant variable parameters in the model is assessed by partitioning the L^2 into its additive components relative to the specified model. Post-hoc procedures generally evaluate fit of the data to individual cells based on standardized residuals or variance accounted for in the model.

Various criteria can be used to determine the predictors to include in a logit regression model:

1. Pearson chi-square or likelihood-ratio χ^2
2. z-test of parameters in model
3. log-odds ratio
4. Predictive efficiency (R^2 type measure)
5. ΔC^2 (difference between -2logL values for null and model)

The traditional Pearson chi-square and the likelihood-ratio chi-square with $(I\text{-}1)(J\text{-}1)$ degrees of freedom are similar because, as sample size becomes larger, the sampling distributions of both statistics become asymptotically chi-squared. The likelihood-ratio chi-square is computed as:

$$L^2 = 2\Sigma\Sigma n_{ij} \log[n_{ij}/ m_{ij}];$$

where n_{ij} = observed cell frequency, m_{ij} = expected cell frequency (Demaris, 1992, *p.* 4)..

The parameter estimates calculated using maximum likelihood estimation possess asymptotic properties. As sample size increases, the parameter

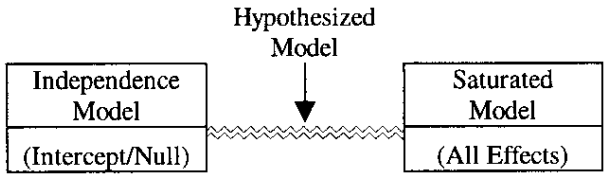
estimates become unbiased and consistent with population parameters. The sampling distribution also approaches normality with variance lower than other unbiased estimation procedures (least squares, etc.). Therefore, given larger samples, the test of a parameters'significance (independent categorical predictor variable) is a z-test calculated by:

$$Z = \hat{\beta}^1 / SE(\hat{\beta}^1)$$

Parameter estimates in logit models can also be readily interpreted as a log-odds ratio. This is calculated as e^{β} for a single parameter, or $e^{(\beta 1 - \beta 2)}$ for differences between two parameters. This is useful when examining contrasts between levels of two independent categorical predictor variables. The log-odds ratio will always agree with the expected cell frequencies.

Predictive efficacy refers to whether a model generates accurate predictions of group membership on the dependent variable. It is possible to have an excellent fit between the logit model and the data without having predictive efficacy. Recall, if $L^2 = 0$, a saturated model exists which perfectly fits the data, yet predictive efficacy (classification) can be far from perfect. In ordinary least squares regression, a saturated model would yield $R^2 = 1$. In SPSS, a classification coefficient (*c*) is calculated to indicate the amount of variance in the dependent variable accounted for by a set of predictors in the model. Given a 50-50 sample split, the base percent would be 50% for the independence model (intercept only model), thus *c* for each variable should be interpreted as a percent that contains this base percentage.

The R^2_L - type measure for logistic regression is not meant as a variance accounted for interpretation, as traditionally noted in least squares regression, because it under estimates the proportion of variance explained in the underlying continuous variables (an assumption made about categorical variables). Basically, a loss of power results when data are reduced from interval to ordinal to nominal. Instead, the R^2_L - type measure is an approximation (lower bound) for assessing predictive efficacy ranging from zero (0) [independence model] to one (1) [saturated model]. This can be depicted as:



The R^2_L-type measure (Hosmer & Lemeshow, 1989) is calculated as: $R^2_L = (SS_T - SS_E)/ SS_T$, where $SS_T = -2\log L_0$ and $SS_E = -2\log L_1$.

The ΔC^2 value provides a way to examine alternative logit models. The L^2 from one model is simply subtracted from the L^2 of the second model. This is similar to testing a full versus restricted modl in multiple regression. The calculation is simply: $L^2(2|1) = L^2_2 - L^2_1$ with the degrees of freedom equal to the difference in the degrees of freedom of the two models. In terms of the log values it is

$$C^2 = -2\log L_0 - (-2\log L_1).$$

If C^2 is non-significant, then additional independent categorical predictors in Model 2 are not needed. This type of test is only appropriate for the likelihood-ratio chi-square and not the Pearson chi-square because adding additional independent categorical predictor variables will never result in a poorer fit of the model to the data (similar to adding terms to a regression model that will never yield a lower unadjusted R^2). This property doesn't hold for the Pearson chi-square.

Logit Models

The logit model contains a categorical dependent variable and a set of categorical independent predictor variables. If a non-significant likelihood-ratio chi-square (L^2) value is computed, then a given model fits the observed data, which is what we desire. On one extreme of the logit model continuum is the *saturated model* or model with perfect fit, yielding a $L^2 = 0$ and *df*=0. The saturated model has as many parameter estimates as degrees of freedom, so it always perfectly reproduces the cell frequencies. For example, a model with all variable main effects and all interaction effects would lead to a saturated model. The *independence model*, in contrast, sets all parameter estimates to zero, resulting in the null model or intercept only model. Consequently, we have a model continuum ranging from the saturated model (all parameters estimated) to the independence model (no parameters estimated). A *hypothesized model* should fall somewhere between these two end-points and reflect a model with fewer parameter estimates than degrees of freedom, so that the degrees of freedom equals the total number of cells minus the number of parameters to be estimated in the model. A model containing only main effects would be an example.

The problem for a researcher becomes one of finding the best set of independent categorical predictor variables. However, what criteria should a researcher use to determine data-to-model fit? Oftentimes, main effects and/or interaction effects are included in a model to predict a dependent variable. For the purpose of this study, two examples are given which focus on the prediction of high-school dropout percent given a set of independent categorical predictor variables. Can we

Table 1. Logit Regression Models

	Loglinear Model	Model Designation
1	$\lambda_i + \lambda_j + \lambda_k + \lambda^R_{ij}$	[RL, D]
2	$\lambda_i + \lambda_j + \lambda_k + \lambda^R_{ij} + \lambda^R_{ik}$	[RL, RD]
3	$\lambda_i + \lambda_j + \lambda_k + \lambda^R_{ij} + \lambda^R_{ik} + \lambda^L_{jk}$	[RL, RD, LD]
4	$\lambda_i + \lambda_j + \lambda_k + \lambda^R_{ij} + \lambda^R_{ik} + \lambda^L_{jk} + \lambda^R_{ijk}$	[RLD]

predict dropout/non-dropout status based on a set of independent categorical predictor variables? Given this research question, we were concerned with the predictive efficacy of the logit model.

Study One

Method and Data

The National Education Longitudinal study of 1988 (NELS) data base was used for data analysis. Subjects were 391 twelfth grade students selected randomly from the NELS data base. Dropout status was treated as the categorical dependent variable. Grade repeat status and locus of control were designated as categorical independent variables. The main and interaction effects research questions were: (1) Do drop-out rates differ significantly between students who repeat a grade versus not repeat a grade?; (2) Do drop-out rates differ significantly between students who have high versus low locus of control?; and (3) Do drop-out rates differ significantly given an interaction between grade repeat status and locus of control? This basic study analysis was gleamed from a previous presentation by Anderson (1995).

The null model and alternative models are specified in Table 1. Model 1 is a null model which hypothesized that drop-out rates (D) are the same regardless of grade repeat status (R) and locus of control (L). Model 2 hypothesized a main effect for grade repeat status (R). Model 3 hypothesized a main effect for locus of control (L). Model 4 hypothesized an interaction between grade repeat status (R) and locus of control (L) in predicting drop-out rates (D).

Results

The calculation of L^2 is affected by the order of entry of independent categorical variables, consequently Table 2 indicates grade repeat status entered first (Method A) compared to locus of control entered first (Method B). Method A indicated a non-significant main effects L^2 value for grade repeat status and locus of control. Method B however indicated a non-significant main effects L^2 value for grade repeat status only. No interaction was indicated. A subsequent approach was to use the additive properties of the likelihood ratio statistic to assess the specific contribution of each parameter in the model specified by calculating the L^2 difference. Table 3 indicates the component L^2 values which are the difference between two modeled L^2 values. Model 2 (grade repeat status main effect) is statistically significant accounting for 93% of the total modeled L^2. Locus of control main effect and interaction effects are not significant.

The variance accounted for approach is yet another way to assess how much of the Null Model L^2 (48.58) is attributed to a hypothesized logit model, in this case Model 2 L^2 (45.13) in Table 2. It follows that 45.13 divided by 48.58 equals 93% of the Total L^2. Obviously, the other modeled L^2 values account for the remaining percent of the Total L^2. SPSS does compute and list a *c* value which indicates the percent classification.

Several post-hoc procedures have been suggested including standardized residuals (Hinkle et al., 1988), scheffe-type contrasts (Marascuilo & Busk, 1987), log-odds ratio of parameter estimates (Kennedy, 1992), and variance accounted for indicated above. A further investigation of this technique and analysis is presented in a second study to clarify best model selection strategies given multiple categorical independent predictor variables in logit regression models.

Study Two

Method and Data

There were 29,124 students enrolled in grades 7-12 in Richardson ISD. Of these students, 754 were dropouts (2.6%) and 28,370 were non-dropouts (97.4%). To facilitate the analysis, a random sample of 754 students was taken from the non-dropout students. The dependent variable was dropout status (dropout, non-dropout). The categorical independent predictor variables were: gender (male, female); ethnicity (asian, black, hispanic, white); grade (7,8,9,10,11,12); retained in grade (not retained, retained 1+ times); parent (natural, step/in-law); suspensions from school (none, 1, 2+); economic disadvantaged (no, yes); and number of courses missed (none, 1-5, 6+).

The research question of interest was in predicting dropout/non-dropout status from several independent categorical predictor variables. Consequently, predictive efficacy or classification status was the focus of the study. Basically, What set of independent predictors provides the best classification of dropout/non-dropout?

Table 2. Logit Regression Model Fit.

Method A	Residual			Component			
Model	L^2	*df*	*p*		L^2	*df*	*p*
Null Model – (1) [RL, D]	48.58	3	.0001				
Main Effects – (2) Grade Repeat [RL, RD]	3.45	2	.1779	$L^2_{(1-2)}$	45.13	1	.0001
Main Effects – (3) Locus of Control [RL, RD, LD]	0.24	1	.6242	$L^2_{(2-3)}$	3.21	1	.0421
Interaction – (4) Grade Repeat x Locus of Control [RLD]	0	0	---	$L^2_{(3-4)}$	0.24	1	.6242
Method B							
Null Model – (1) [RL, D]	48.58	3	.0001				
Main Effects – (2) Locus of Control [LR, LD]	41.85	2	.0001	$L^2_{(1-2)}$	6.73	1	.0071
Main Effects – (3) Grade Repeat [LR, LD, RD]	0.24	1	.6242	$L^2_{(2-3)}$	41.61	1	.0001
Interaction – (4) Grade Repeat x Locus of Control [RLD]	0	0	---	$L^2_{(3-4)}$	0.24	1	.6242

Results

A preliminary univariate analysis of each categorical independent predictor with the dependent variable dropout status is in Table 3. It is apparent that gender differences are not significant in determining dropout/non-dropout status. Similarly, economic disadvantaged doesn't yield a high L^2 or χ^2 relative to the other predictor variables. The slight difference in L^2 and χ^2 values is due to sample size, as noted before these values will be more similar as sample size increases because the sampling distributions are asymptotically chi-squared. If one were to interpret these individual results, the number of course failures would best predict dropout/non-dropout status, followed by number of times retained in grade, number of suspensions, grade level, et cetera. Variable entry order, however, does affect results (see Appendix).

Table 4 indicates the main effects for the eight predictor variables and several criteria which are used to judge the significance of categorical independent variable entry in the logit model equation. A comparison of the hypothesized logit models with single predictors to the intercept model (independence model) is given by ΔC^2. A continuation of this table to include all 2-way interactions, 3-way interactions, 4-way interactions, et cetera would be required to determine the best set of predictor variables using the ΔC^2 criteria . Subsequently, one could compare the predictive efficacy of each logit model equation provided by the Δc value which indicates the percent above and beyond the *c* value for the intercept model. Calculation of the total number of logit model equations, i.e., 256, (2^m) is beyond the scope of this paper.

Table 3. Univariate L^2 and χ^2 on Dropout Status

Categorical Variable	L^2	χ^2	*df*	*p*
Gender	2.736	2.735	1	.09800
Ethnicity	52.859	52.481	3	.00001
Grade	99.508	97.137	5	.00001
Retained	139.872	133.628	1	.00001
Parent	34.084	32.189	1	.00010
Suspend	112.089	108.991	2	.00001
Economic	6.602	6.590	1	.01000
Course Failure	324.900	306.150	2	.00001

Note: L^2 and χ^2 are asymptotically chi-squared and become similar as sample size increases.

All Possible Subsets

The logit main effects and interaction effects model in study two would contain 256 equations in a saturated model. This is calculated by 2^m, where m=8 (Freund & Littell, 1991, *p.* 107). This does not take into account the fact that the order of entry for the categorical independent predictor variables would change the results. Many of the criteria for determining the best set of predictors have inherent problems. For example, the individual univariate Pearson chi-square or likelihood-ratio chi-square tests don't reflect interaction effects; the z-test of parameters in the logit model would change based on the order of entry in the equation and number of variables in the equation; the log-odds ratio because it is the exponentiation of the parameter estimate would also differ depending upon the order of entry and number of variables in the logit model equation; and the predictive efficacy (classification percent) is not necessarily a function of the significance of the parameters in the logit model. Consequently, the ΔC^2 (difference between -2logL values for null and hypothesized models) appears to be the most useful. A problem still remains in that SPSS and SAS do not provide a test for subsets of predictors nor do they generate all possible subset equations (Demaris, 1992, *p.* 68).

A new procedure, TABU (Drezner, Marcoulides, & Salhi, 1999), provides a solution to model selection in multiple regression which is directly applicable to logit modeling, and provides better results than a previously determined Mallows' *Cp* criteria (Schumacker, 1994). The Tabu program generates the *F*-ratio based on the L^2 and/or χ^2 value for all possible equations between the independence (null model) and the saturated model. Given a best model selection criteria of ΔC^2, one could easily pick the best set of categorical independent predictor variables. Subsequently, of the best ΔC^2 models, predictive efficacy could be compared (Δc).

Conclusions

In the first study, Table 2 indicated that a grade repeat status main effect was statistically significant accounting for 93% of the total modeled L^2. Locus of control main effect and interaction effects were not significant. [Please note that in Table 2, $L^2_{1\text{-}2}$, is the same as ΔC^2 in Table 4.] With only a few independent predictors one can easily hand calculate all of the possible subsets of equations. The entry order of independent predictor variables did have an impact on parameter estimates.

In the second study, Table 4 indicated that ethnicity, grade level, retained in grade, parent, suspensions, economic disadvantaged, and course failures main effects were statistically significant in the prediction of dropout/non-dropout status. Gender was not significant. A relative comparison of the ΔC^2 values for these main effects suggests that number of course failures followed by number of times retained in grade and number of suspensions would provide a possible best subset model. However, a researcher would not ultimately know the best subset model unless all possible subsets were calculated and compared on ΔC^2. The use of a Tabu search procedure to generate all possible subsets is therefore needed.

Educational Importance

The logit regression technique is **not** widely used in education even though it offers several advantages over the use of the chi-square statistic in analyzing categorical variables (Green, 1988). The type of variables used in these two studies are typical of the data recorded in school districts. A better understanding of this statistical technique, its applications, and interpretation will hopefully increase awareness of its value to educational researchers (Tabachnick & Fidell, 1989; Stevens, 1992).

Direct Correspondence to:
Randall E. Schumacker
College of Education
Matthews Hall 304
P.O. Box 311377
University of North Texas
Denton TX 76203-1337
E-Mail: rschumacker@unt.edu

Table 4. Logit Models: Main Effects Only

Main Effects	-2logL$_i$	ΔC^2	*df*	*p*	R^2_L	*c*	Δc
Intercept	2090.532					50%	
Gender	2087.796	2.736	1	.09800	.002	52%	2%
Ethnicity	2037.673	52.859	3	.00001	.034	59%	9%
Grade	1991.024	99.508	5	.00001	.064	61%	11%
Retained	1950.660	139.872	1	.00001	.089	62%	12%
Parent	2056.448	34.084	1	.00010	.022	54%	4%
Suspend	1978.443	112.089	2	.00001	.072	62%	12%
Economic	2083.930	6.602	1	.01000	.004	53%	3%
Course Failure	1765.632	324.900	2	.00001	.194	68%	18%

References

Anderson, C. (January, 1995). *Practical Applications for Logit Modeling*. Paper presented at the annual Southwest Educational Research Association, Dallas: Texas.

Demaris, A. (1992). *Logit Modeling Practical Applications*. Newbury Park: Sage Publications.

Drezner, Z., Marcoulides, G.A., & Salhi, S. (1999). Tabu search model selection in multiple regression analysis. *Communications in Statistics - Computation and Simulation*, 28(2).

Feinberg, S. E. *(1981). The analysis of cross-classified categorical data. (2nd Ed).*Cambridge, MA: MIT Press.

Freund, R.J. & Littell, R.C. (1991). *SAS System for Regression* (2nd Ed.). Cary, NC: SAS Institute, Inc.

Green, J. A. (1988). Loglinear Analysis of Cross-classified Ordinal Data: Applications in Developmental Research. *Child Development, 59*, 1-25.

Hinkle, D. E., McLaughlin, G. W. & Austin, J. T. (1988). "Using Log-Linear Model in

Higher Education Research" in *New Directions for Institutional Research*, Terenzini, P.T., (ed.) : San Francisco: Jossey-Bass.

Hosmer, D.W. & Lemeshow, S. (1989). *Applied Logistic Regression*. NY: John Wiley.

Kennedy, J.J. (1992). *Analyzing Qualitative Data Log-Linear Analysis for Behavioral Research* (2nd Ed.). New York: Praeger Publishers.

Kleinbaum, D.G. (1992). *Logistic Regression: A self-learning text*. Springer-Verlag: New York.

Marasculio, L. A. & Busk, P. L. (1987). Loglinear Models: A way to study main effectsand interactions for multidimensional contingency tables with categorical data. *Journal of Counseling Psychology, 34*, 443-455.

Schumacker, R.E. (1994). A comparison of the Mallows *Cp* and principal component regression criteria for best model selection in multiple regression. *Multiple Regression Viewpoints, 21*(1), 12-22.

Stevens J. *(1992). Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tabachnick B. G. & Fidell, L. S. (1989). *Using Multivariate Statistics*. New York: Harper Collins Publishers.

APPENDIX
Variable Entry Order

FAILURE ENTERED FIRST

Null Model (Intercept Only): -2 Log Likelihood = 2090.5319

Hypothesized Model: -2 Log Likelihood = 1742.100

R^2: .206

Model Chi-Square (df=5): 348.432

Classification Overall: 71.15%

Variables in the Equation

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
FAILURES			128.0864	2	.0000	.2436	
FAILURES(1)	-2.0569	.1839	125.0690	1	.0000	-.2426	.1278
FAILURES(2)	-1.0140	.1612	39.5877	1	.0000	-.1341	.3628
RETAINED(1)	-.6551	.1681	15.1808	1	.0001	-.0794	.5194
SUSPEND			7.5456	2	.0230	.0412	
SUSPEND(1)	-.4718	.1730	7.4344	1	.0064	-.0510	.6239
SUSPEND(2)	-.2892	.2217	1.7012	1	.1921	.0000	.7489
Constant	1.9468	.1852	110.5476	1	.0000		

FAILURE ENTERED LAST

Variables in the Equation

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
RETAINED(1)	-.6510	.1648	15.6055	1	.0001	-.0807	.5215
SUSPEND			7.5251	2	.0232	.0411	
SUSPEND(1)	-.4708	.1729	7.4136	1	.0065	-.0509	.6245
SUSPEND(2)	-.2887	.2218	1.6940	1	.1931	.0000	.7493
FAILURES	1.0301	.0910	128.0361	1	.0000	.2455	2.8012
Constant	-.1093	.2522	.1878	1	.6648		

Note: *df* = 4, Model Chi-square = 348.417

Interpreting Regression Analysis Results: An Example

Carl J Huberty
J. Brady Allen
University of Georgia

An example of the application of multiple regression is presented in some detail. Predictor variable scores are based on the three parts of the Graduate Record Examination (GRE). Criterion variable scores are based on the performance of graduate students in an introductory statistical methods course. Even though the general predictive power of the GRE measures is assessed, the focus of the interpretation is on the prediction of the criterion for specified profiles of predictor measures.

The Graduate Record Examinations (GRE) are in widespread use across graduate-level universities in the United States. Typically, performances on the verbal test (GREV), the quantitative test (GREQ), and the analytical test (GREA) are used for admission purposes; sometimes score sums, such as GREV + GREQ, are utilized. Not only is the use widespread, so is the questioning of the use of GRE performance for admission purposes (Morrison & Morrison, 1995). The questioning position often espoused is usually based on the low relationship between performance on the GRE and performance in graduate school (as assessed by graduate grade point average). That is, what is being questioned is the predictive validity of GRE performance relative to performance in graduate school.

It might be argued that there are two related difficulties in assessing GRE performance predictive validity. One difficulty is that variability in the GRE scores is necessarily restricted because only those students with higher GRE scores are typically admitted to graduate school. The other difficulty is that the variability of *overall* performance in graduate school, as typically assessed by A-B-C grading, is quite restricted because of typical current grading practices (see, e.g., Cole, 1993).

But, how about using GRE scores in predicting the performances in a particular area of study in graduate school? Some restriction was considered by Kluever and Green (1992) in their prediction study for students in a college of education. An intent of the current study is to determine how useful performance on the GRE is in predicting performance in introductory graduate level statistical methods courses. The overriding purpose of this article is to illustrate how specific prediction information might be obtained; that is, information more specific than an overall index of relationship.

Analysis Units

A graduate statistical methods course taught at The University of Georgia in the College of Education might be titled Educational Statistics I. Topics covered in ESI include data description, correlation, and inference regarding a mean, proportion, and correlation.

The analysis units used in this study are students from a collection of six ESI classes. There was a total of 135 ESI students. Eleven of these students had not taken any of the GRE and, therefore, were not considered analysis units. [Class performance data for these 11 students were, however, used in calculating standard scores for the remaining 124 students.] Of the 124 ESI students, 48 were master level, 16 specialist level, and 60 were doctoral level. About 92 percent of the 124 ESI students were enrolled in Education graduate programs. Two textbooks were used with the ESI classes. For the first five classes, Moore and McCabe (1989, chaps. 1-8) was used; for the sixth class, Moore (1995, chaps. 1-7) was used.

Criterion Measure

Student performance was based on three types of assessment, four quizzes, a test, and an examination. The sequence of assessments used is Q_1 , Q_2 , T, Q_3 , Q_4 , E. The final examination covered material in the second half of each course. Three scores were obtained for each student in each class: (highest) sum of three 10-point quizzes, score on the 35-item test, and score on the 45-item examination. [All items on the quizzes, test, and examination were of the multiple-choice variety, focusing mostly on concepts. Typical score ranges were approximately 27-15 for the quiz sum, 33-15 for the test, and 35-15 for the examination.] Each of these three scores was transformed to z scores using the mean and standard deviation based on all six classes for ESI. [In "real life" the three scores are transformed using data on a current class plus the three most recent classes.] A composite of the three z scores,

$$Z = 0.5\, z_Q + 1.0 z_T + 1.5 z_E ,$$

served as the criterion variable score for this study. [The composite Z is the basis used in course grading.]

Table 1. Descriptors for GRE and Class Scores for ESI students

Variable	MIN	C_{25}	C_{50}	C_{75}	MAX
V	320	450	510	580	800
Q	350	350	560	620	800
A	250	500	550	610	760
V + Q	700	990	1045	1185	1510
V+Q+A	1070	1493	1605	1768	2160
Z	-8.58	-1.47	0.27	1.65	5.26

Note: V = GREV, Q = GREQ, and A + GREA.

Measurement characteristics of the quizzes, the midterms tests, and the examinations are judged to be acceptable. Specifically, content validity of the three types of scores is judged to be very respectable. Values of the Kuder-Richardson 20 index (of internal consistency) for the five midterm tests ranged from about .65 to about .85. It is to be expected that K-R_{20} values for the quizzes would be lower; recall that the sum of the three highest quiz scores was used for each student. The internal consistency of the scores on the final examination was somewhat higher than that for the midterm test scores. It is also assumed that a common scale of measurement is used across classes for the quizzes, for the tests, and for the examinations.

Predictor Measures

Three parts of the Graduate Record Examinations were utilized in this study to serve as bases of predictor measures; Verbal (GREV), Quantitative (GREQ), and Analytical (GREA). Thus, the three predictor variables considered are verbal aptitude, quantitative aptitude, and analytical aptitude. For ESI, the data matrix has 124 rows and four columns (three predictors and one criterion). Completion of the basic data matrix will now be briefly discussed.

Of the 124 ESI students, six had not taken the Analytical part of the GRE. Two ways of imputing these six scores were considered. One way was simply to use the mean GREA based on the remaining 118 students. The second imputation method used was to regress GREA on GREQ and GREV using the complete data on the 118 students. To determine the way of choice, two 4x4 correlation matrices were determined using the three GRE scores and the composite, *Z*; one matrix was based on the GREA mean and the other was based on regressed GREA scores. The benchmark correlation matrix is the "available case" matrix where all but six correlations are based on 124 students; the remaining six correlations are based on 118 students. The three correlation matrices were visually compared; for the purposes of this article, the regressed GREA value was used to replace the six missing scores. So, a full 124 x 4 data matrix was used in the analyses.

Results

Three sets of predictor measures were considered: (1) GREV (denoted V), GREQ (Q), GREA (A); (2) V + Q, A; and (3) V + Q + A. So then, the composite *Z* was regressed on V, Q, and A, on V + Q and A, and on V + Q + A. Table 1 shows descriptive information for each predictor and the criterion measured on the 124 ESI students.

The correlations among the five predictors and between each predictor and *Z* are reported in Table 2. [All three scatterplots (not reported herein) revealed reasonable linearity; normal probability-plots indicated no aberrations, as did a plot of *Z* versus .]

The predictability of performance in an introductory level statistical methods course as measured by *Z*) using the GRE scores as predictor scores may be *broadly* assessed via a multiple correlation coefficient value. The broad results for the three regression analyses are given in Table 3. The adjustment used to get R^2_{adj} is that proposed by M. Ezekiel in 1930 presented by Huberty (1994) wherein the *F*-test involving R^2_{adj} is also discussed.

Very often in prediction studies, the researcher is interested in determining a relative ordering of the predictor variables. That is, it may be of interest to determine the most and least important predictors. In a multiple regression context, we view the most important predictor as the one which when deleted from the total set of predictors will decrease the error mean square value (or, equivalently, the R^2_{adj} value)) the most -- focus is on overall predictive accuracy. This approach to assessing predictor importance is discussed by Huberty (1989) and Huberty and Petoskey (1999).

For the first prediction model that involved three predictors (V, Q, A), an ordering of the importance of the predictors may be obtained by deleting, in turn, each predictor. The *adjusted* two-predictor R^2 values were $R^2_{(Q)}$ = .250, $R^2_{(A)}$ = .370, and $R^2_{(V)}$ = .386, where $R^2_{(Q)}$ denotes the adjusted obtained by deleting Q. Thus, Q (i.e., GREQ) is judged to be the most important predictor, with V and A of about equal importance (or unimportance). For the second prediction model (using V + Q and A), it was found

Table 2. Correlations among GRE and Class Scores

Variable	Q	A	V+Q	V+Q+A	Z
V	.05[ns]	.35	.73	.66	.34
Q		.30	.72	.63	.53
A			.45	.80	.47
V + Q				.90	.60
V+Q+A					.64

Note: [ns] indicates not significant ($p > .05$). All other correlations were significant ($p < .001$).

Table 3. Overall Prediction Results for ESI data.

Predictor(s)	R^2	R^2_{adj}	F	df
V, Q, A	.436	.422	15.27	3.97, 120
V+Q, A	.412	.402	21.02	2.65, 121
V+Q+A	.409	.404	42.04	1.33, 122

Note: All models were significant ($p < .001$).

that the R^2_{adj} values obtained by deleting each of the predictors were $R^2_{(V+Q)}$ = .226 and $R^2_{(A)}$ = .354. Clearly, V + Q is more important than A when it comes to broad, overall prediction of *Z*.

More specific prediction information may be obtained by examining some particular predictor profiles. To do this, five clusters of profiles of ESI students were identified for the three-predictor model, one for the two-predictor model, and one for the one-predictor model (see Table 4). Our rationale for the cluster definitions is based on the various prediction models used at different universities; some use only predictors V and Q, some include A along with V and Q, and others use V + Q and/or V + Q + A. Also, we were interested in determining prediction quality for those who are generally high test scorers, low test scorers, and those who were high on some predictors and low on others. The question then becomes: How well can the composite *Z* score for ESI students be predicted for each the various profiles? The goodness of prediction was based on the magnitude of the standardized residual (see Montgomery & Peck, 1992, p. 68). If $|Z - \hat{Z}| < .80$, it was judged that we had a "good" prediction. [The composite *Z* scores typically ranged from about 5.00 to about -8.00. Other cut-offs may be more appropriate in other prediction situations.]

Table 5 summarizes how well *Z* scores can be predicted for students with each of the 12 profiles.

Table 4. Clusters of GRE Profiles for ESI students

Cluster	Size	Part	Centile	Score
1	8	Q ≥	70	650
		A ≥	60	590
		V ≥	50	480
2	17	A ≥	80	650
3	12	Q ≥	80	710
4	9	Q ≤	30	480
		A ≤	40	510
		V ≤	50	480
5	23	Q ≥	50	570
		V ≤	50	480
6	25	A ≥	60	
		V+Q ≥		1100
7	20	V+Q+A ≥		1800
		V+Q+A ≤		2000

Note: V = GREV, Q = GREQ, and A + GREA.

Some summary statements are given below:

1. Of the five ESI GRE score profiles for the three-predictor model, clusters 1 and 2 had over 80% small ($|Z - \hat{Z}| < .80$) prediction errors. That is, for students with high Q, A, and V scores or students with very high A scores, it was judged that the percent of good prediction of class performance was respectable.
2. For students performing poorly on all GRE parts (cluster 4), prediction was not considered very respectable (only 44% good prediction).
3. For those students who score above the median on GREQ and below the median on GREV (cluster 5), prediction was not very respectable.
4. For the two-predictor model (V+Q and A), respectable prediction resulted (80% small prediction errors) for students with "high" scores on both predictors (cluster 6).
5. Groups of students for whom respectable prediction resulted (clusters 1, 2, 6) were not dominated by high-performing (i.e., "A") students, except, possibly, for cluster 1 students with three "high" GRE scores.
6. Residuals for students in cluster 1 (with high GRE scores) indicated a dominance of over-prediction, while for cluster 6 (also with high GRE scores) there was a dominance of under-prediction; it was for these two clusters that residuals were judged respectable.

Conclusions

The results of this study would indicate that prediction of performance in an *introductory* graduate-level statistical methods course (ESI) using scores on the three GRE parts can be accomplished in a fairly successful manner. First of all, accounting for about 42% of the variability in overall course performance is judged to be fairly high, especially in relation to that found by Elmore, Lewis, and Bay (1993) and Goldberg and Alliger (1992) where *unadjusted* R^2 values ranged from .09 to .29. Secondly, accuracy of prediction in the current study was judged to be respectable for some subgroups of students. Success resulted in predicting overall course performance (as assessed by a composite Z) for subgroups/clusters of students with "high" GRE scores. Also, prediction of Z for students with "low" GRE scores was judged to be poor.

The above conclusions suggest to us that restricting the view of regression analysis results to looking at an adjusted R^2 value may very well result in placing unnecessary limitations on interpretation possibilities. For a real practical research situation it may very well be informative to learn about the predictive accuracy for analysis units with particular predictor variable score profiles.

References

Cole, W. (1993, January 6). By rewarding mediocrity we discourage excellence. *Chronicle*, pp. B1-B2.

Elmore, P. B., Lewis, E. L., & Bay, M. L. G. (1993, April). *Statistics achievement: A function of attitudes and related experiences.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Goldberg, E. L., & Alliger, G. M. (1992). Assessing the validity of the GRE for students in Psychology: A validity generalization approach. *Educational & Psychological Measurement, 52,* 1019-1027.

Huberty, C. J (1989). Problems with stepwise methods -- Better alternatives. In B. Thompson (Ed.), *Advances in social science methodology* (*pp.* 43-70). Greenwich, CT: JAI Press.

Huberty, C. J (1994). A note on interpreting an R^2 value. *Journal of Educational and Behavioral Statistics, 19,* 351-356.

Huberty, C. J, & Petoskey, M. D. (1999). Use of multiple correlation analysis and multiple regression analysis. *Journal of Vocational Educational Research, 24,* 15-43.

Kluever, R. C., & Green, K. E. (1992). Prediction of achievement of doctoral students in education. *Perceptual and Motor Skills, 74,* 419-423.

Montgomery, D. C., & Peck, E. A. (1992). Introduction to linear regression analysis. New York: Wiley.

Moore, D. S. (1995). *The basic practice of statistics.* New York: Freeman.

Moore, D. S., & McCabe, G. P. (1989). *Introduction to the practice of statistics.* New York: Freeman.

Morrison, T., & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the Graduate Record Examination with graduate grade point average representing the criterion of graduate success. *Educational & Psychological Measurement, 55,* 309-316.

Table 5. Results for Predicting Composite Z for Clusters of ESI students

Cluster[1]	Size	No.[2]	%	A	B	C-D	Pos.[3]	Neg.[4]
1	8	7	88	5	2	0	1	6
2	17	14	82	8	5	1	7	7
3	12	8	67	6	2	0	2	6
4	9	4	44	0	2	2	2	2
5	23	11	48	2	8	1	6	5
6	25	20	80	11	9	1	15	5
7	20	14	70	8	6	0	12	2

Note: [1]For Cluster definitions, see Table 4. [2]Number of students with a standardized residual magnitude < 0.8. [3] "Pos." indicates under-prediction. [4] "Neg." indicates over-prediction.