Journal of Educational Psychology
1990, Vol. 82, No. 4, 646–656


0022-0663/90/$00.75

# Causal Ordering of Academic Self-Concept and Academic Achievement: A Multiwave, Longitudinal Panel Analysis

Herbert W. Marsh
School of Education and Language Studies
University of Western Sydney, Macarthur, New South Wales, Australia

There is surprisingly little sound research on the causal ordering of academic self-concept and academic achievement in longitudinal panel studies, despite its theoretical and practical significance. Data collected in Grades 10, 11, 12, and one year after graduation from high school that were used in this study come from the large ($N$ = 1,456 students), nationally representative Youth in Transition study. It was found that reported grade averages in Grades 11 and 12 were significantly affected by academic self-concept measured the previous year, whereas prior reported grades had no effect on subsequent measures of academic self-concept. The results provide one of the few defensible demonstrations of prior academic self-concept influencing subsequent academic achievement, and the study appears to be methodologically stronger than previous research.



A positive self-concept is valued as a desirable outcome in many educational settings and is frequently posited as a mediating variable that facilitates the attainment of other desired outcomes such as academic achievement. A growing body of literature (e.g., Byrne, 1984; Hansford & Hattie, 1982; Marsh, 1986, 1987; Marsh, Byrne, & Shavelson, 1988; Shavelson & Bolus, 1982) indicates that academic self-concept is clearly differentiable from general self-concept and that academic self-concept is more highly correlated with academic achievement and other academic behaviors than is general self-concept. Marsh, Byrne, and Shavelson, for example, found that none of the general self-concept scales from three different instruments were significantly correlated with school grades in English, mathematics, or all school subjects, whereas academic self-concept scales were substantially correlated with achievement. This pattern of relations supports the construct validity of academic self-concept responses and the need for educational researchers to consider academic self-concept instead of relying on general self-concept scales.

Wylie (1978) suggested that students' perceptions of their academic ability are based largely on school performance, so that standardized ability test scores should add little to the prediction of self-concept beyond the contribution of school performance measures. Literature reviews (e.g., Hansford & Hattie, 1982; Wylie, 1979) have found school performance indicators to be more highly correlated with self-concept than are IQ or general academic achievement. However, I noted in Marsh (1987) (also see Davis, 1966), that school performance measures typically are normalized relative to other students within the school, whereas standardized tests are normalized in relation to a broader population. I suggested that high school students may use both frames of reference in forming their academic self-concepts. I also argued that school-based performance is more likely to be affected by effort and motivational influences than are standardized test scores, so that prior academic self-concept is more likely to affect subsequent school performance than to affect standardized test scores. For these reasons, I indicated the need to consider separately the effects of standardized tests scores and school performance in evaluating relations between academic self-concept and achievement.

## Causal Ordering of Academic Self-Concept and Academic Achievement

Perhaps the most vexing theoretical question in academic self-concept research involves determining the causal ordering of academic self-concept and academic achievement. This question is of practical importance because many self-concept enhancement programs are based on the assumption that an improvement in self-concept will lead to gains in academic achievement.

Byrne (1984) noted that much of the interest in the relation between self-concept and achievement stems from the belief that academic self-concept has motivational properties such that changes in academic self-concept will lead to changes in subsequent academic achievement. Calsyn and Kenny (1977) contrasted self-enhancement and skill development models of the relation between self-concept and achievement. According to the self-enhancement model, self-concept is a primary determinant of academic achievement. Support for this model would provide strong justification for the self-concept enhancement interventions that are explicit or im-

I would like to thank Raymond Debus, Rhonda Craven, and Rosalie Robinson for helpful comments on earlier drafts of this article. The data used in this manuscript were made available by the Inter-University Consortium for Political and Social Research and were originally collected by Jerald Bachman.

Correspondence concerning this article should be addressed to Herbert W. Marsh, School of Education and Language Studies, University of Western Sydney, Macarthur, P.O. Box 555, Campbelltown, New South Wales 2560, Australia.

plicit in many educational programs. In contrast, the skill development model posits that academic self-concept is primarily a consequence of academic achievement. According to this model, the best way to enhance academic self-concept is to develop stronger academic skills. Calsyn and Kenny (1977) argued that comparisons of the self-enhancement and skill development models have been severely hindered by conceptual and methodological limitations. Because self-concept and academic achievement are not readily amenable to experimental manipulations (but see Brookover & Erikson, 1975; Byrne, 1984; Marsh & Richards, 1988; Scheirer & Kraut, 1979), most research has relied on longitudinal data in which both self-concept and achievement are measured on at least two occasions (i.e., a two-wave, two-variable design).

## Cross-Lagged Correlation

Calsyn and Kenny (1977) proposed the method of cross-lagged panel correlation to establish which of the two variables, academic self-concept or achievement, was causally predominant. They found a reasonably consistent predominance of academic achievement over academic self-concept in a variety of comparisons, thus supporting the skill-enhancement model. Although their analytic technique represented an improvement over most previous research, it suffered from serious conceptual and methodological limitations. Conceptually, the question posed in cross-lagged correlation studies may be inappropriate for examining relations between self-concept and achievement. Theory (e.g., Marsh, Byrne & Shavelson, 1988; Shavelson, Hubner, & Stanton, 1976) and common sense dictate that academic self-concept must be determined at least in part by prior academic achievement. The critical issue is whether the relation is reciprocal such that prior academic self-concept also has a causal influence on subsequent academic achievement. In cross-lagged correlation, however, neither the causal predominance of one variable over the other nor the lack of any causal predominance can be used to test whether a reciprocal relation exists. Hence, cross-lagged correlation does not adequately test the nature of the relation between academic self-concept and achievement.

In a methodological critique of cross-lagged correlation that focused on the Calsyn and Kenny (1977) study, Rogosa (1979, 1980) demonstrated that the comparison of cross-lagged correlations was not a sound basis of causal inference. Furthermore, the problems of interpretation were exacerbated rather than facilitated by the use of multiple indicators of each construct and multiwave data based on three or more waves of data. Rogosa (1980) concluded that "no justification was found for the use of CLC [cross-lagged correlation]" (p. 257) and that "cross-lagged correlation is not a useful procedure for the analysis of longitudinal panel data" (p. 245). Rogosa (1979) also discussed the limitations in the traditional use of path analysis that is based on multiple regression for longitudinal data. The most important problems are due to fallible measurement and the implicit assumption in multiple regression that all variables are measured without error. As in cross-lagged correlation, the inclusion of multiwave data and multiple indicators of each latent construct is likely to exacerbate limitations in the traditional path analysis approach. In contrast to cross-lagged correlation and traditional multiple regression, structural equation models (SEMs) have distinct advantages for the analysis of multiwave longitudinal data. The application of this technique is facilitated by the inclusion of multiple indicators of each latent construct and by multiwave data. Nevertheless, Rogosa (1979) cautioned that "causal attribution is not an automatic process; useful causal conclusions are the product of careful thought, high-quality data, and sound data analysis" (p. 301).

## Methodologically More Adequate Studies: Minimum Design and Analytic Requirements

In her classic review of the academic self-concept research, Byrne (1984) examined studies purporting to test causal predominance between self-concept and academic achievement. Such studies, she noted, must satisfy three prerequisites: (a) A statistical relationship must be established, (b) a clearly established time precedence must be established in longitudinal studies, and (c) a causal model must be tested. At the time of her review, Byrne found only two studies satisfying her prerequisites: SEM studies by her (subsequently published as Byrne, 1986) and by Shavelson and Bolus (1982). She claimed that no conclusions about the causal ordering of self-concept and achievement were warranted from existing research.

I found six studies that tested causal models of the relation between self-concept and academic achievement using longitudinal panel data. Three are less relevant either because they did not consider both self-concept and school grades in more than one wave (Felson, 1984; Maruyama, Rubin, & Kingsbury, 1981) or because they used traditional path analysis instead of SEMs (Marsh, 1987). In the three more relevant studies (Byrne, 1986; Newman, 1984; Shavelson & Bolus, 1982), both academic self-concept and academic achievement were measured on at least two occasions and SEMs were used to test the causal models.

Byrne (1986) found no effect of prior achievement on subsequent self-concept or of prior self-concept on subsequent achievement. Both academic achievement and academic self-concept were inferred from multiple indicators, and the sample size was large. However, she considered data from only two occasions, collected in the same academic year. Byrne questioned the appropriateness of combining school grades and academic achievement into a single construct. One of the two indicators of academic self-concept was a subscale of the Coopersmith instrument (see Marsh & Richards, 1988) that was apparently weak. Furthermore, her models contained a Heywood case (a negative residual variance term), and this problem may reflect a misspecified model.

Shavelson and Bolus (1982) found that prior academic self-concept affected subsequent performance, whereas the effects of prior achievement on subsequent academic self-concept were not statistically significant. These results suggest the predominance of academic self-concept over academic achievement. Shavelson and Bolus, however, cautioned that the size and nature of their study (99 7th-grade students from a single school and an interval from Time 1 to Time 2 spanning only 4 months) dictated caution in generalizing from the results. Academic self-concept was inferred from multiple indicators, but academic achievement was based on a single indicator. Although Shavelson and Bolus did not explore this potential limitation, its implications may not be too serious because school grades are likely to be reasonably reliable.

Newman (1984) considered math achievement tests and math self-concept collected in Grades 2, 5, and 10. For the interval from Grade 2 to Grade 5 and from Grade 5 to Grade 10, prior achievement had a significant effect on subsequent math self-concept, but prior math self-concept had no effect on subsequent math achievement. The sample size was dubiously small by SEM standards (*N*s were 84 to 143 for different correlations when pairwise deletion was used to construct the correlation matrix, and $N = 75$ when casewise deletion for missing data was used). The most serious problem, however, was that academic self-concept was inferred on the basis of responses to a single self-response item. Newman examined this problem with a sensitivity analysis: The untestable reliability of each single-item self-concept factor was fixed at different plausible values, and other parameters were estimated for the various reliability values. The sensitivity analysis was an important addition that "made the best of a bad situation" (Newman, 1984, p. 868). The problem with the sensitivity analysis was that the analysis was conducted on a reduced model, in which all paths leading from prior self-concept to subsequent achievement had already been eliminated, and thus the sensitivity analysis provided no tests of the conclusion that self-concept does not affect subsequent behavior. In a reanalysis of the Newman data, I (Marsh, 1988) conducted a sensitivity analysis, using the same range of values for the reliability of the single-item self-concept factors that were considered by Newman, for the full model in which paths leading from self-concept to achievement were retained. Depending on the a priori reliabilities, self-concept sometimes affected subsequent achievement, whereas achievement sometimes had no effect on subsequent self-concept. On the basis of this sensitivity analysis, I argued that the data were not strong enough to justify either the conclusion that prior self-concept affects subsequent self-concept or Newman's conclusion that prior self-concept has no effect on subsequent achievement.

Despite apparent methodological limitations in these three studies, it is interesting to note how the findings vary depending on how academic achievement was inferred. Shavelson and Bolus (1982) inferred academic achievement from school grades and found the causal predominance of academic self-concept over school grades. Newman (1984) inferred academic achievement from standardized test scores and argued for the predominance of academic achievement over academic self-concept. Byrne (1986) inferred academic achievement from a combined construct based on both school grades and standardized test scores and found no support for the causal predominance of either construct. Although interpretations should be made cautiously, this pattern is consistent with Marsh's (1987) suggestion that prior academic self-concept is more likely to affect subsequent achievement if achievement is inferred from school grades that are responsive to motivational influences rather than from standardized test scores.

In summary, it may be useful to provide an overview of important design features in this area of research. Ideally, studies will achieve the following: (a) measure academic self-concept and academic achievement (school performance, standardized test scores, or preferably both) at least twice (i.e., a two-wave study) and preferably more frequently; (b) infer all latent constructs on the basis of multiple indicators; (c) consider a sufficiently large and diverse sample to justify the use of SEMs and the generality of the findings; and (d) fit the data to a variety of SEMs that incorporate measurement error and test for likely residual covariation among measured variables. If both test scores and school grades are collected in the same study, then they should be considered as separate constructs unless there is empirical support for combining them to form a single construct. If any of the latent constructs are measured with a single measured variable, an a priori estimate of reliability should be used and the sensitivity analysis should be conducted on the full model to determine the generality of the conclusions. On the basis of these criteria, none of the previous studies reviewed here is fully adequate. The purpose of the present investigation is to further examine the causal ordering of academic self-concept and academic achievement on the basis of a study that is apparently methodologically stronger than any previous study considered here.

## Method

### Sample and Procedure

Data came from the large, nationally representative Youth in Transition study of all 10th grade boys in public high schools in 1966 (Bachman, 1970; Bachman & O'Malley, 1977, 1986; also see Marsh, 1987). A two-stage sampling scheme was used in which a random sample of 87 public high schools was selected, and then approximately 25 students were randomly selected from each school. Data in the present investigation came from the commercially available longitudinal data file that comprises information from Wave 1 (early 10th grade; $N = 2{,}213$), Wave 2 (late 11th grade; $N = 1{,}886$), Wave 3 (late 12th grade; $N = 1{,}799$), and Wave 4 (one year after normal high school graduation; $N = 1{,}620$). For present purposes, data were considered from the 1,456 students who had complete data for the measures considered in the present investigation at Times 1, 2, and 3.

As a result of the two-stage cluster sampling scheme used in original data collection, standard errors based on the assumption of simple random sampling are biased. To compensate for this bias, Bachman and O'Malley (1986) suggested that an *N* of 1,000 be used for purposes of testing statistical significance in an analysis of 1,487 of the 2,113 students in the original sample. Because the sample size considered here is similar, an *N* of 1,000 was also used for testing statistical significance in the present investigation. It should be emphasized that this decision in no way affects actual parameter estimates.

### Variables and Their Measurement

All variables considered here are from the commercially available longitudinal data file from the Youth in Transition study (Bachman, 1975). Relations among the variables considered here are presented in Table 1 and are the basis of subsequent analyses. Three latent constructs were inferred from these variables.

Academic ability was measured only at Time 1. It was inferred from scores on four standardized tests: IQ (Ammons & Ammons, 1962), vocabulary (U.S. Department of Labor, 1962), reading comprehension (Gates, 1958), and mathematical reasoning (Guide to the Use of the General Aptitude Test, 1962). For the 1,487 students considered here, there were no missing values for any of these variables. All scores are standardized ($M = 0$, $SD = 1$) for present purposes.

Reported average grades were measured at Times 1, 2, and 3. At Times 1 and 2, this variable was inferred on the basis of a single self-report item, collected as part of an individually administered personal interview conducted by a trained interviewer. At Time 3 this variable was based on a single self-report item collected as part of a self-administered questionnaire. At Times 1 and 2, students were asked "What is the average grade you got in your classes last year?" At Time 3 students were asked "What is the average grade you have been getting in your classes this year?" Reported average grades were recorded into one of 13 categories: A+, A, A−, B+, B, B−, C+, C, C−, D+, D, D−, and F (or E). For the 1,487 students considered here, there were no missing values for any of these variables. For present purposes, reported grade averages at Time 1 were standardized ($M = 0$, $SD = 1$), and grades at Times 2 and 3 were standardized in relation to the mean and standard deviation of Time 1 grades.

Academic self-concept was measured at Times 1, 2, and 4. It was inferred from responses to three self-rating items: school ability (how does respondent rate in school ability compared to others?) Times 1 and 2; intelligence (how intelligent does respondent feel he is?), Times 1, 2, and 4; and reading (how good a reader is the respondent?), Times 1, 2, and 4. Students responded along a 6-point response scale that varied from *far below average* (1) to *far above average* (6). For the 1,487 students considered here, there were no missing values for any of the measures collected at Time 1 or Time 2. At Time 4 there were 1,329 and 1,290 responses to ratings of intelligence and reading. For purposes of these Time 4 variables only, correlations were computed using pairwise deletion for missing data.

## *Tests of Initial a Priori Model*

The initial a priori model (see Figure 1) was based in large part on the temporal ordering of the data collection, in that Time 1 variables preceded Time 2 variables, Time 2 variables preceded Time 3 variables, and the Time 3 variable preceded the Time 4 variable. At Time 1 there were three constructs: academic ability, reported average grade, and academic self-concept. Academic ability was posited to

Table 1
*Correlations, Means, and Standard Deviations for Variables in This Study*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ability at Time 1 | | | | | | | | | | | | | | | |
| 1. IQ-1 | — | | | | | | | | | | | | | | |
| 2. VOCB-1 | .643 | — | | | | | | | | | | | | | |
| 3. MATH-1 | .409 | .586 | — | | | | | | | | | | | | |
| 4. ENGL-1 | .620 | .667 | .539 | — | | | | | | | | | | | |
| Grades at Time 1 | | | | | | | | | | | | | | | |
| 5. GRADE-1 | .360 | .456 | .472 | .401 | — | | | | | | | | | | |
| Self-concept at Time 1 | | | | | | | | | | | | | | | |
| 6. SCHABL-1 | .346 | .396 | .334 | .321 | .447 | — | | | | | | | | | |
| 7. SRINTL-1 | .347 | .404 | .347 | .295 | .478 | .554 | — | | | | | | | | |
| 8. SREAD-1 | .363 | .406 | .233 | .313 | .278 | .337 | .433 | — | | | | | | | |
| Grades at Time 2 | | | | | | | | | | | | | | | |
| 9. GRADE-2 | .301 | .416 | .418 | .355 | .659 | .440 | .443 | .244 | — | | | | | | |
| Self-concept at Time 2 | | | | | | | | | | | | | | | |
| 10. SCHABL-2 | .367 | .402 | .342 | .319 | .416 | .502 | .488 | .304 | .471 | — | | | | | |
| 11. SRINTL-2 | .376 | .449 | .382 | .334 | .455 | .475 | .555 | .372 | .475 | .600 | — | | | | |
| 12. SREAD-2 | .379 | .453 | .287 | .336 | .315 | .332 | .395 | .660 | .279 | .386 | .485 | — | | | |
| Grades at Time 3 | | | | | | | | | | | | | | | |
| 13. GRADE-3 | .301 | .395 | .382 | .337 | .579 | .367 | .379 | .225 | .640 | .407 | .447 | .261 | — | | |
| Self-concept at Time 4 | | | | | | | | | | | | | | | |
| 14. SRINTL-4 | .301 | .338 | .298 | .251 | .328 | .349 | .468 | .325 | .358 | .441 | .535 | .367 | .351 | — | |
| 15. SREAD-4 | .328 | .366 | .191 | .298 | .214 | .227 | .309 | .569 | .198 | .294 | .346 | .634 | .195 | .428 | — |
| *M* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 4.31 | 4.33 | 3.15 | −0.70 | 4.40 | 4.36 | 4.11 | −0.13 | 4.29 | 3.99 |
| *SD* | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.79 | 1.06 | 1.01 | 0.81 | 0.83 | 1.06 | 1.02 | 0.74 | 0.98 |

*Note.* IQ = Quick Test intelligence test score; VOCB = Vocabulary test score; MATH = arithmetic reasoning test score; ENGL = reading comprehension test score; GRADE = reported grade average; SCHABL = self-rating of school ability; SRINTL = self-rating of intelligence; SREAD = self-rating of reading skills. The number following each variable name refers to the time of data collection: 1 = 10th grade; 2 = 11th grade; 3 = 12th grade; 4 = one year after graduation from high school.

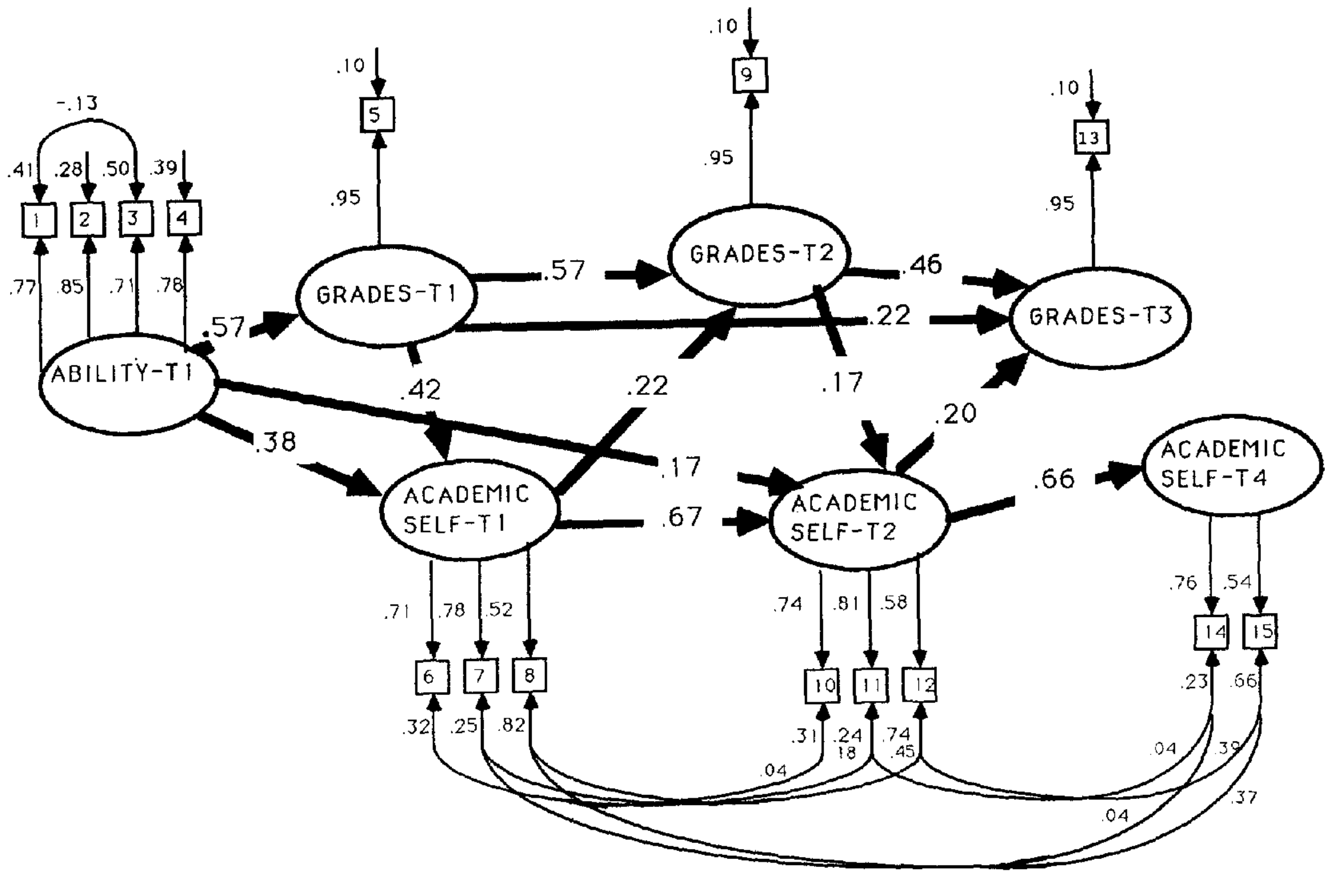


*Figure 1.* Standardized effects of prior ability, reported grade averages, and academic self-concept on subsequent grades and academic self-concept for Model 3. (SELF = Self-concept. The 15 boxes represent the 15 measured variables shown in Table 1. The ovals represent latent constructs inferred from the measured variables. The numbers following each latent construct represent the data wave in which it was collected: T1 = 10th grade, T2 = 11th grade, T3 = 12th grade, and T4 = one year after graduation. The straight, bold lines connecting the different latent constructs represent path coefficients: the direct effects of each construct on all subsequent constructs. Nonsignificant path coefficients are excluded for purposes of clarity, but they are presented in Table 3 under Model 3. The curved lines represent correlated residuals between measured variables.)

precede grades. Because students were asked to report their grades from the previous year, I posited that school grades preceded academic self-concept. Similarly, at Time 2 reported average grades were posited to precede academic self-concept. At Time 3 and at Time 4, only one construct was considered, so there was no need to posit a causal ordering within each wave. One should note, however, that the ordering of variables within a given wave has no influence on the overall goodness of fit of the model and almost no influence on the path coefficients relating variables from different waves. In this a priori model, correlated residuals relating the uniquenesses of the same indicator of academic self-concept administered at different points in time were also posited as shown in Figure 1. Such correlated residuals are usually found in longitudinal panel studies, and their existence is likely to inflate estimates of the stability of the underlying construct. These observations were substantiated by fitting a series of alternative models.

In preliminary analyses, three models were evaluated in terms of their ability to fit the data. Each of the three models was reasonable in that the iterative procedure converged to a proper solution; each of the constructs inferred from multiple indicators was well-defined; and the overall goodness-of-fit indices, particularly given the large sample sizes, was moderate to good (for a discussion of evaluating goodness of fit, see Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988). Model 1 (Table 2) did not include the correlated residuals that were hypothesized a priori. The fit of Model 1 is much poorer than the other two models, thus supporting the inclusion of the correlated residuals. Model 2 is the a priori model originally hypothesized and it fits the data very well. Inspection of the modification indices provided by LISREL, however, suggested that one additional correlated residual was required between two of the multiple indicators of academic ability (Model 3 in Table 2; also see Figure 1). The inclusion of this additional parameter made a small but statistically significant improvement in the goodness of fit. Model 3 provides an excellent fit and is the basis of subsequent analyses.

In SEMs latent constructs are automatically corrected for estimates of unreliability that are based on the design of the model, so long as

Table 2
*Goodness-of-Fit Indices for Alternative Models*

| Model | $\chi^2$ | *df* | TLI | BBI | Model description |
|---|---|---|---|---|---|
| 1 | 951 | 72 | .83 | .87 | No correlated uniquenesses |
| 2 | 246 | 65 | .96 | .97 | Model 1 with 7 a priori correlated uniquenesses |
| 3 | 198 | 64 | .97 | .97 | Model 2 with 1 a posteriori correlated uniqueness[a] |

*Note.* TLI = Tucker-Lewis Index; BBI = Bentler-Bonett Index. The goodness-of-fit indices are an index of the covariance among measured variables that is explained by the model; values greater than .90 are typically interpreted to indicate an adequate fit.
[a] See Figure 1 for illustration of Model 3.



there are multiple indicators of each construct. Whereas academic ability and academic self-concept were inferred from multiple indicators, reported average grades were based on only a single indicator. As noted in reviews of previous research, much of the value of the SEM approach is undermined when a construct is inferred on the basis of a single indicator. For present purposes, initial analyses were conducted in which I assumed that the reliability of this construct was .90 and that relations among the different grade estimates did not involve any correlated residuals. In subsequent sensitivity analyses, however, the implications of these a priori assumptions are explored further.

## Results

### *Causal Path Model*

Standardized parameter estimates for the final model are presented in Figure 1.[1] All nonsignificant path coefficients are excluded for purposes of clarity but are presented in Table 3. Of particular importance are the effects of latent constructs in one wave on latent constructs in subsequent waves. At Time 2 academic self-concept was influenced by academic ability and Time 1 academic self-concept but not by Time 1 reported average grades. At Time 2 reported grades were influenced by both Time 1 academic self-concept and Time 1 reported grades. Similarly, reported grades at Time 3 were significantly influenced by both Time 2 academic self-concept and Time 2 grades. Academic self-concept at Time 4 was significantly influenced by academic self-concept at Time 2 (there was no Time 3 academic self-concept measure) but not by Time 3 reported grades. Particularly because the results were replicated across two different intervals, the findings provide strong support for the effect of prior self-concept on subsequent reported grades. It is also interesting to note that in neither of the intervals did prior grades have a statistically significant direct effect on subsequent academic self-concept. In the language of earlier research, the effects of academic self-concept were "causally predominant" over those of reported grades, and these results provide strong support for the self-concept enhancement model of the relation between self-concept and achievement.

The correlation between any two constructs (see Table 3) can be divided into total effects (sometimes referred to as total causal effects) and noncausal relations. Noncausal relations are the components of the correlation that can be explained in terms of other variables that occur prior to or at the same point as the variable being considered. The total effects can be further divided into the direct effects shown in Figure 1 and indirect effects; direct effects are the unique influences of each construct, after controlling for all the other constructs, and the indirect effects are the effects of each influence that are mediated through intervening variables. Particularly for variables early in the causal chain, the total effects can be substantial, even when the direct effects are small or nonsignificant. In this study, for example, the total effects of each Time 1 and Time 2 construct on all subsequent constructs were large and statistically significant (see Model 3 in Table 2).

As noted previously, the path coefficients—the direct effects—relating constructs from different waves in Figure 1 varied little with the causal ordering of variables within each data wave. Model 3a (see Tables 2 and 3) differs from Model 3 only in that no causal ordering of constructs within each data wave is posited; constructs within the same wave are posited as merely being correlated. Because Model 3a is only a reparameterization of Model 3, the number of estimated parameters and goodness of fit are the same. Direct effects among variables within the same data wave were all zero, but the direct effects of variables among constructs from different data waves were nearly the same as in Model 3. However, because correlations among variables within the same data wave are assumed to represent noncausal relations, the total effects of each construct tended to be substantially smaller in Model 3a than in Model 3. Whereas Model 3 can be justified on the basis of theory, Model 3a corresponds more closely to models typically used to analyze longitudinal panel studies (e.g., Rogosa, 1979) and may be more parsimonious in terms of the number of a priori assumptions. For this reason, the primary conclusions of this study are based on the direct effects relating variables from different waves in which there are almost no differences between the two models.

### *Sensitivity Analysis*

The major apparent weakness of this investigation is that reported grades are inferred on the basis of a single indicator, thus undermining many of the advantages of the SEM approach that exist when there are multiple indicators of each construct. Two important limitations in analyses based on

[1] Whereas all models described here were originally tested in the covariance metric (see Table 1), the standardized parameter estimates presented in Figure 1 and Tables 2–4 were obtained in subsequent analysis of the correlation matrix to facilitate interpretations of the parameter estimates. It is important to emphasize, however, that chi-square values, goodness-of-fit indices considered here, and the t-test values associated with each estimated parameter were the same in analyses of the correlation and covariance matrices (see Jöreskog & Sörbom, 1988, p. 49).

Table 3
*Path Coefficients: Direct and Total Effects Among Latent Variables for Alternative Models*

| From:[a] | Ability | | | | | | Grades at Time 1 (G1) | | | | | Academic self-concept at Time 1 (SC1) | | | | Grades at Time 2 (G2) | | | Academic self-concept at Time 2 (SC2) | | Grades at Time 3 (G3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Effect / To:[b] | G1 | SC1 | G2 | SC2 | G3 | SC4 | SC1 | G2 | SC2 | G3 | SC4 | G2 | SC2 | G3 | SC4 | SC2 | G3 | SC4 | G3 | SC4 | SC4 |
| Model 3 | | | | | | | | | | | | | | | | | | | | | |
| Direct | .57 | .38 | .04 | .17 | .06 | .07 | .42 | .57 | −.05 | .22 | −.09 | .22 | .67 | −.10 | .13 | .17 | .46 | −.01 | .20 | .66 | .07 |
| Total | .57 | .62 | .51 | .65 | .48 | .54 | .42 | .66 | .34 | .55 | .22 | .22 | .71 | .14 | .60 | .17 | .48 | .14 | .20 | .67 | .07 |
| Model 3a | | | | | | | | | | | | | | | | | | | | | |
| Direct | .00 | .00 | .04 | .18 | .06 | .07 | .00 | .57 | .05 | .22 | −.09 | .22 | .70 | −.10 | .13 | .00 | .46 | −.01 | .20 | .66 | .07 |
| Total | .00 | .00 | .04 | .18 | .11 | .19 | .00 | .57 | .05 | .49 | −.04 | .22 | .71 | .14 | .60 | .00 | .46 | .02 | .20 | .67 | .07 |

*Note.* SC4 = self-concept at Time 4. Model 3 has a "full-forward" pattern of effects in that every latent variable is posited to affect all subsequent variables (see Figure 1). In Model 3a, latent constructs from the same data wave were posited to be correlated, whereas latent constructs in each wave were posited to affect latent constructs in all subsequent data waves. These two models represent reparameterizations of the same model in that the number of estimated parameters and goodness of fit is the same for both. All direct paths shown in Figure 1 are statistically significant, whereas the remaining direct effects presented here are not statistically significant. All total effects presented here that are greater than .1 are statistically significant ($p < .05$).
[a] Indicates variables from which path coefficients originate.
[b] Indicates resultant variables of path coefficients.

single-indicator constructs are particularly relevant to longitudinal panel studies. First, there is no empirical procedure to estimate reliability and to correct relations between latent constructs for unreliability. Jöreskog and Sörbom (1988) and many others argue that it is preferable to set the reliability of the single indicator to a plausible, a priori value, rather than to assume that the construct is perfectly reliable as is typically done. Newman (1984) extended this approach with his sensitivity analysis, testing his final model for a whole range of plausible values and determining how critical parameter estimates varied according to the posited reliability estimate. In Marsh (1988) I further refined the approach by showing that no parameters should be excluded from the model before conducting this analysis because previously nonsignificant parameter estimates may become significant when reliability estimates for the single-indicator constructs are varied. For purposes of this investigation, the results based on the a priori reliability estimate of .90 used in earlier analyses were compared with those based on reliability of 1.0, .95, .85, and .80.

The second limitation of analyses based on single-indicator constructs is the following. In longitudinal panel studies, correlations between the same construct measured on two different occasions are typically inflated by correlated uniquenesses in the multiple indicators used to infer each construct. Thus it is not unusual to find test–retest correlations that exceed the reliability estimates on each occasion so that the typical correction for unreliability would result in a test–retest correlation greater than 1.0. So long as there are parallel multiple indicators, the existence of correlated residuals is testable and controllable in SEMs, provided constructs are not inferred from a single indicator. To examine this possibility, the sensitivity analysis was extended to include the likely possibility of correlated residuals. For purposes of this investigation, residual covariances relating reported grades at Time 1, Time 2, and Time 3 were set at 0%, 12.5%, 25%, and 50% of the corresponding residual variances.[2]

The results of the sensitivity analysis are presented in Table 4. In general, parameter estimates varied systematically and logically with different residual variances (i.e., uniqueness or unreliability) and with residual covariances assigned to the single indicators of reported grades at Times 1, 2, and 3. Not surprisingly, the effect of prior grades on subsequent grades varied inversely with the assumed reliability and the sizes of correlated residuals. For present purposes the most important parameter estimates in the sensitivity analysis are those leading from prior self-concept to subsequent reported grade averages and from prior grades to subsequent self-concept.

With regard to the effect of prior self-concept on subsequent

[2] Technically, a priori estimates of reliability were varied by setting the uniqueness (in the diagonal of the Theta Delta LISREL design matrix, Jöreskog & Sörbom, 1988) at x% of the variance in the measured variable in which x is 1 − the a priori reliability estimate. The correlated residual was then fixed at 0%, 12.5%, 25%, or 50% of the square root of the product of the uniqueness terms. For variables standardized to have variances of 1.0, residual variances were fixed at 0, .05, .10, .15, and .20, and the corresponding correlated residuals were fixed at 0%, 12.5%, 25%, or 50% of these values.

Table 4
*Sensitivity Analysis: Direct Effects of Latent Variables for Alternative Versions of Model 3[a] With Different Uniquenesses and Correlated Uniqueness for the Single-Indicator Grade Variables*

| Correlated uniqueness[b] / From:[c] | Ability | | | | | | Grades at Time 1 (G1) | | | | | Academic self-concept at Time 1 (SC1) | | | | Grades at Time 2 (G2) | | | Academic self-concept at Time 2 (SC2) | | Grades at Time 3 (G3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| To:[d] | G1 | SC1 | G2 | SC2 | G3 | SC4 | SC1 | G2 | SC2 | G3 | SC4 | G2 | SC2 | G3 | SC4 | SC2 | G3 | SC4 | G3 | SC4 | SC4 |
| Uniqueness[e] = 0, reliability = 1.00 | | | | | | | | | | | | | | | | | | | | | |
| .00 | 54 | 42 | 07 | 17 | 06 | 06 | 38 | 47 | −03 | 22 | −08 | 25 | 67 | −08 | 12 | 14 | 38 | −01 | 22 | 66 | 06 |
| Uniqueness[e] = .05, reliability = .95 | | | | | | | | | | | | | | | | | | | | | |
| .00 | 56 | 40 | 06 | 17 | 06 | 07 | 40 | 51 | −04 | 22 | −08 | 24 | 67 | −09 | 12 | 16 | 42 | −01 | 21 | 66 | 06 |
| .00625 | 56 | 40 | 06 | 17 | 06 | 07 | 40 | 50 | −04 | 22 | −08 | 24 | 67 | −09 | 12 | 15 | 41 | −01 | 21 | 66 | 06 |
| .01250 | 56 | 40 | 06 | 17 | 06 | 07 | 40 | 49 | −03 | 22 | −08 | 25 | 68 | −09 | 12 | 15 | 40 | −01 | 22 | 66 | 06 |
| .02500 | 56 | 40 | 07 | 17 | 07 | 07 | 40 | 47 | −03 | 21 | −08 | 26 | 67 | −08 | 12 | 15 | 39 | −01 | 22 | 66 | 06 |
| Uniqueness[e] = .10, reliability = .90 | | | | | | | | | | | | | | | | | | | | | |
| .00000 | 57 | 38 | 04 | 17 | 06 | 07 | 42 | 57 | −05 | 22 | −09 | 22 | 67 | −10 | 13 | 17 | 46 | −01 | 20 | 66 | 07 |
| .01250 | 57 | 38 | 05 | 17 | 06 | 07 | 42 | 54 | −05 | 22 | −09 | 23 | 66 | −10 | 13 | 17 | 44 | −01 | 21 | 66 | 07 |
| .02500 | 57 | 38 | 06 | 17 | 06 | 07 | 42 | 52 | −04 | 21 | −09 | 24 | 66 | −09 | 13 | 17 | 42 | −01 | 22 | 66 | 06 |
| .05000 | 57 | 38 | 07 | 17 | 07 | 07 | 42 | 46 | −03 | 20 | −08 | 27 | 66 | −08 | 13 | 16 | 39 | −01 | 23 | 66 | 06 |
| Uniqueness[e] = .15, reliability = .85 | | | | | | | | | | | | | | | | | | | | | |
| .00000 | 59 | 36 | 03 | 17 | 05 | 07 | 45 | 63 | −08 | 22 | −11 | 20 | 66 | −12 | 13 | 20 | 51 | −01 | 19 | 65 | 08 |
| .01875 | 59 | 36 | 04 | 17 | 06 | 07 | 45 | 59 | −06 | 21 | −10 | 22 | 66 | −11 | 13 | 19 | 48 | −01 | 20 | 66 | 08 |
| .03750 | 56 | 36 | 05 | 17 | 07 | 07 | 45 | 55 | −05 | 20 | −10 | 24 | 66 | −10 | 13 | 18 | 45 | −01 | 21 | 66 | 07 |
| .07500 | 56 | 36 | 08 | 16 | 08 | 07 | 45 | 46 | −03 | 18 | −09 | 28 | 65 | −08 | 13 | 17 | 40 | −01 | 23 | 66 | 06 |
| Uniqueness[e] = .20, reliability = .80 | | | | | | | | | | | | | | | | | | | | | |
| .00000 | 61 | 33 | 00 | 17 | 05 | 07 | 48 | 71 | −12 | 19 | −13 | 16 | 66 | −13 | 14 | 24 | 59 | −01 | 18 | 65 | 10 |
| .02500 | 61 | 33 | 02 | 17 | 06 | 07 | 48 | 65 | −09 | 20 | −12 | 19 | 66 | −12 | 14 | 22 | 54 | −01 | 19 | 65 | 09 |
| .05000 | 61 | 33 | 04 | 17 | 07 | 07 | 48 | 58 | −06 | 19 | −11 | 23 | 65 | −11 | 14 | 20 | 49 | −01 | 21 | 61 | 08 |
| .10000 | 61 | 33 | 08 | 16 | 09 | 07 | 48 | 46 | −03 | 16 | −09 | 29 | 65 | −07 | 13 | 18 | 40 | −02 | 24 | 66 | 06 |

*Note.* SC4 = academic self-concept at Time 4. In the model shown in Figure 1, the reliability of the three grade variables was assumed to be .90 (i.e., uniqueness of .10) with no correlated uniquenesses among the three grade variables. For present purposes models were fit in which the reliability of the grade variable varied between 1.0 and .85 (uniquenesses of .00 to .15), and the correlated uniqueness was 0, .125, .25, or .50 of the uniqueness. Because these parameter estimates were fixed, the degrees of freedom and goodness of fit for all the models were necessarily the same.
[a] See Figure 1 for illustration. [b] For each model, the uniqueness covariance is set at 0% (i.e., no correlated uniqueness), 12.5%, 25%, or 50% of the uniqueness. Within a given model, the uniquenesses for the 3 grade variables and the 3 uniqueness covariances among the 3 grade variables were assumed to be equal. [c] The headings in this row that follow indicate variables from which path coefficients originate. [d] The headings in this row that follow indicate resultant variables of path coefficients. [e] Uniqueness = 1 − reliability in standardized form.

reported grades, the parameter estimates varied inversely with the assumed reliability of reported grades and directly with the size of the correlated residuals. All values for both intervals, however, were statistically significant for all the values presented in Table 4. The values presented in Figure 1 appear to be reasonably conservative in that they are among the lower of the values in Table 3. Because it is likely that at least some residual covariances exist between single indicators of reported grades on different occasions, the effect of prior academic self-concept on subsequent academic achievement is likely to be stronger than reported in Figure 1. With regard to the effect of prior reported grades on subsequent self-concept, all parameter estimates were negative but did not differ significantly from zero for any of the values considered in Table 4. The parameter estimates became more negative as unreliability increased and less negative as the size of correlated residuals increased. It is also important to note that the effect of correlated residuals is magnified by the unreliability in the single indicator. This, of course, reflects the way in which the sizes of correlated residuals were constructed. Nevertheless, correlated residuals must necessarily approach zero as the reliability approaches 1. These results, consistent with common sense, indicate that problems associated with the use of single-indicator constructs are likely to be particularly serious when the reliability of these single-item indicators is low.

Whereas the focus of the sensitivity analysis has been on parameter estimates of particular interest in this investigation, it is important to note that the statistical significance or nonsignificance was not dependent on the particular values of residual variances and covariances for any of the parameters summarized in Table 3. In summary, the results of the sensitivity analysis indicate that the findings of this study are apparently robust in relation to limitations associated with inferring reported grades from a single indicator.

## Summary and Implications

The causal ordering of academic self-concept and academic achievement is a particularly important issue for the study of self-concept in educational settings. The self-enhancement model, which is based on the assumption that prior academic self-concept affects subsequent academic achievement, is used implicitly to justify many educational programs designed to enhance self-concept. In contrast, the skill development model is based on the assumption that academic self-concept merely reflects academic skills so that the best way to enhance academic self-concept is to improve academic skills. In reality, both of these extreme positions are probably too simplistic in that relations between academic self-concept and academic achievement are likely to be reciprocal. Theory and common sense dictate that prior academic accomplishments must, at least in part, determine academic self-concept, though a growing body of research (e.g., Marsh, 1986; Marsh, Byrne, & Shavelson, 1988) demonstrates that other factors are important. Hence, it is important to determine the effect of prior academic self-concept on subsequent academic achievement even if prior academic achievement also affects subsequent academic self-concept. Given the important theoretical and practical implications of this issue, there is surprisingly little sound research. Furthermore, when evaluated in relation to desirable criteria, none of the previous studies considered here was fully adequate. Thus, the present investigation is important for at least two reasons. First, it is one of the few studies—along with, perhaps, Shavelson and Bolus's (1982) study—to provide defensible evidence for the effect of prior academic self-concept on subsequent academic achievement. Second, it is apparently methodologically stronger than previous research.

Despite the apparent strengths of this investigation, it is important to recognize its limitations. First, even if all the desirable characteristics of an ideal longitudinal panel study were present, causal inferences that are based on correlational data must be viewed cautiously (Freedman, 1987; Rogosa, 1987). Second, the fact that academic achievement in the present investigation was not inferred from multiple indicators undermined many of the advantages of the SEM approach. Even though the sensitivity analysis indicated that the conclusions were apparently robust in relation to this limitation, a better solution would have been to have multiple indicators of this construct. Third, the major criterion variable in this investigation, reported grade averages, was based on self-report data. Unlike most studies, however, Time 1 and Time 2 measures of reported grades were collected as part of a personal interview administered individually to each respondent by a trained interviewer, though the Time 3 measure was based on a traditional, self-administered questionnaire. Furthermore, to the extent that there were self-response biases in the reported grades, this effect would probably work against the finding that prior self-concept significantly affected subsequent achievement. Because the response biases would probably inflate the size of correlations between self-reported grades on the two occasions, the amount of unexplained variance that could be explained by academic self-concept would be reduced. Nevertheless, a stronger methodology would have involved basing this variable on actual school grades obtained from school records. Fourth, the results are based on a large, nationally representative sample of boys who attended high school in the late 1960s so that the generalizability of results to girls and to current high school students are not addressed.

A final purpose of this investigation is to stimulate further research on the important issue of the causal relation between academic self-concept and achievement. Additional research can provide tests of the generalizability of findings in this investigation and hopefully overcome many of its limitations. This additional research may take two different forms. First, the analytical techniques demonstrated here can usefully be applied to reanalyses of previous research. Second, there is a need for new research that more fully incorporates recent theoretical developments in self-concept research. In addition to the desirable methodological features emphasized here, there are a variety of important issues that should be pursued.

1. The intervals between each of the four data waves in this investigation were all approximately one year and each wave was collected in a different school year. Other research

has considered intervals that were considerably shorter (e.g., Shavelson & Bolus, 1982; Byrne, 1986) or considerably longer (e.g., Newman, 1984). Newman (1984) questioned the value of studies based on intervals of only 5 months. Results based on two waves collected within the same year in school and waves in which the interval is more than 2 or 3 years should, perhaps, be viewed cautiously. It would be useful to have data waves in the same school years and in different school years within the same study in order to evaluate the consistency of the findings. Because there is no clear understanding of the effect of the interval length, this is an issue that needs to be considered further.

2. Because there are so few methodologically adequate studies of the causal ordering of academic self-concept and academic achievement, there is little empirical basis for understanding how this relation varies developmentally. There is, however, evidence suggesting that the agreement between academic self-concept and academic achievement grows stronger with age at least through junior high school or early high school (e.g., Bloom, 1976; Hansford & Hattie, 1982; Newman, 1984; Nicholls, 1979). Hence, it is possible that the causal ordering of these variables varies with age.

3. Academic self-concept, because of the nature of available items, was treated as a relatively unidimensional construct in this investigation. Recent research (e.g., Marsh, Byrne, & Shavelson, 1988), however, clearly demonstrates that academic self-concept is a multidimensional construct and that there is a particularly clear separation between verbal and mathematical components of academic self-concept. Because there was no self-rating of mathematics ability in this investigation and because school grades were not obtained for different school subjects, this important distinction could not be pursued here but should be considered in future research.



## References


Ammons, R. B., & Ammons, C. H. (1962). The Quick Test (QT): Provisional manual. *Psychological Reports, 11*(Monograph Suppl. 7-VII), 1–7.

Bachman, J. G. (1970). *Young in transition. Vol. 2. The impact of family background and intelligence on tenth-grade boys.* Ann Arbor, MI: Institute for Social Research.

Bachman, J. G. (1975). *Youth in transition, data file documentation (Vol. 2).* Ann Arbor, MI: Institute for Social Research.

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology, 35,* 365–380.

Bachman, J. G., & O'Malley, P. M. (1986). Self-concepts, self-esteem, and educational experiences: The frogpond revisited (again). *Journal of Personality and Social Psychology, 50,* 33–46.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Bloom, B. S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Brookover, W. B., & Erikson, E. L. (1975). *Sociology of education.* Homewood, IL: Dorsey Press.

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research, 54,* 427–456.

Byrne, B. M. (1986). Self-concept/academic achievement relations: An investigation of dimensionality, stability, and causality. *Canadian Journal of Behavioural Science, 18,* 173–186.

Calsyn, R., & Kenny, D. (1977). Self-concept of ability and perceived evaluations by others: Cause or effect of academic achievement? *Journal of Educational Psychology, 69,* 136–145.

Davis, J. A. (1966). The campus as a frog pond: An application of theory of relative deprivation to career decisions for college men. *American Journal of Sociology, 72,* 17–31.

Felson, R. B. (1984). The effect of self-appraisals of ability on academic performance. *Journal of Personality and Social Psychology, 47,* 944–952.

Freedman, D. A. (1987). As others see us: A case study of path analysis. *Journal of Educational Statistics, 12,* 101–128.

Gates, A. I. (1958). *Gates Reading Survey—Form I.* New York: Columbia University Teachers College.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and academic/performance measures. *Review of Educational Research, 52,* 123–142.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications.* Chicago: SPSS.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23,* 129–149.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79,* 280–295.

Marsh, H. W. (1988). Causal effects of academic self-concept on academic achievement: A reanalysis of Newman (1984). *Journal of Experimental Education, 56,* 100–104.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80,* 366–380.

Marsh, H. W., & Richards, G. (1988). The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. *Australian Journal of Psychology, 40,* 281–298.

Maruyama, G., Rubin, R. A., & Kingsbury, G. G. (1981). Self-esteem and educational achievement: Independent constructs With a common cause? *Journal of Personality and Social Psychology, 40,* 962–975.

Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. *Journal of Educational Psychology, 76,* 857–873.

Nicholls, J. (1979). Development of perception of attainment and causal attributions for success and failure in reading. *Journal of Educational Psychology, 71,* 94–99.

Rogosa, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 263–302). San Diego, CA: Academic Press.

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88,* 245–258.

Rogosa, D. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12,* 185–195.

Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self-concept change. *Review of Educational Research, 49,* 131–150.

Shavelson, R. J., & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology, 74,* 3–17.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research, 46,* 407–441.

U.S. Department of Labor. (1962). *Guide to the use of the General Aptitude Test Battery.* Washington, DC: Author.

Wylie, R. C. (1979). *The self-concept.* (Vol. 2) Lincoln, NE: University of Nebraska Press.

Received August 17, 1989
Revision received May 9, 1990
Accepted June 14, 1990 ■