The Misuse of ANCOVA: The Academic and Political Implications of Type VI Errors In Studies of Achievement and Socioeconomic Status

Susan M. Tracz	Laura L. Nelson	Isadore Newman	Adrian Beltran
California State University, Fresno	University of Wisconsin, La Crosse	University of Akron	California State University, Fresno

This paper examines ANCOVA designs which use SES as the covariate for achievement and Type VI errors. Type VI errors are inconsistencies between the research question and the research methodology, and these errors are discussed in the context of general semantics. The consequences of a Type VI error in studies of achievement differences covariating for SES can be highly misleading. When research with a Type VI error concludes that there are no significant differences in achievement across groups when statistically controlling for SES, the tacit implications are that actual achievement is consistent across groups and that SES can be causally controlled or is somehow not influential. Neither is correct. Authors suggest conducting validity studies of adjusted outcome scores to insure accuracy in interpreting results.

The study of student achievement is a major focus of educational researchers and practitioners. With the recent passage of the No Student Left Behind legislation, the study of the factors related to increasing achievement scores has intensified. One salient variable which is correlated to achievement is SES (Attewell & Battle, 1999; Chapell & Overton, 2002; Gregory, 2000; O'Brien, Martinez-Pons & Kopala, 1999; Verna & Campbell, 1998), and researchers have used it as a covariate for achievement (Dillon & Schemo, 2004; Ferguson, 1981; Kaplan, 2002). The assumed reasoning is that if the variance attributable to SES is removed, the unique variance in achievement can be examined and explained. However, many errors in conceptualization and interpretation are possible with such designs. The purposes of this paper are:

1. To review the way in which ANCOVA is typically explained in textbooks commonly used in university statistics classes,
2. To provide a conceptual framework for discussions of numerical descriptions and the use of language,
3. To explain how Type VI errors, which are inconsistencies between the research question and the research methodology, can lead to inaccurate conclusions and/or interpretations of the data,
4. To provide descriptions of common errors in studies testing for group differences in achievement in which SES is used as the covariate,
5. To discuss the educational and political implications of such errors, and
6. To suggest that the adjusted scores in ANCOVA designs be correlated with other appropriate measures to determine their validity and correct interpretability.

ANCOVA Designs

Isolating and examining the unique variance in a dependent variable in studies of group differences can be undertaken by logical argument, by research design, and by statistical control (McNeil, Newman & Kelly, 1996). While argument is obviously the weakest method, and research design is the strongest, including all confounding variables in a design is not always possible. In such cases, the statistical control provided by analysis of covariance can be a viable alternative. However, there are stringent underlying assumptions which must be met for its appropriate use and interpretation, and these assumptions are frequently violated.

The ANCOVA is a statistical technique used to ascertain group differences on an adjusted dependent variable. This statistical analysis is similar to ANOVA in that it is a vehicle for determining group differences with the exception that instead of examining group means, adjusted group means are studied. In fact, each score is adjusted when the effects of the covariate are statistically removed from the dependent variable.

In his classic work, Pedhazur (1982) presents the mathematical logic of covariance with the example of achievement as the dependent variable, intelligence as the covariate, and a treatment as the independent variable. Recalling that when "a variable is residualized, the correlation between the predictor variable

and residuals is zero" (p. 496), if intelligence is used to predict achievement, the residuals are then zero correlated with intelligence. These residualized scores are then called the adjusted scores for achievement and are analyzed for group differences. Pedhazur summarized as follows:

$$Y_{ij} = \bar{Y} + T_j + b(X_{ij} - \bar{X}) + e_{ij}$$

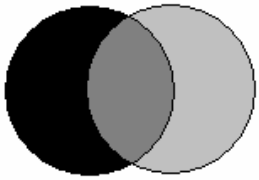
where Y_{ij} is the score of subject i on treatment j, $\bar{Y}$ is the grand mean of the dependent variable, T_j is the effect of the treatment, b is the regression coefficient, X_{ij} is the score on the covariate for subject i under treatment j, $\bar{X}$ is the grand mean for the covariate, and e_{ij} is the error term. This formula also can be

$$Y_{ij} - b(X_{ij} - \bar{X}) = \bar{Y} + T_j + e_{ij},$$

which shows that the adjusted score is equal to the grand mean plus a treatment effect plus error.

Variance in the dependent variable

Figures 1A



Overlap in the variance between the dependent variable and the covariate

1B

Variance in the adjusted dependent variable

1C

Finally, it is important to remember that the outcome or dependent variable in ANCOVA is an adjusted score. To reiterate this point visually, Figure 1A represents the total or 100% of the variance in the dependent variable, for example, achievement. Figure1B represents the overlap between the total variance in the dependent variable and the covariate, for example, achievement and SES. After the effects of the covariate have been statistically controlled or removed from the dependent variable (Figure 1C), the error variance is all that remains. This residualized or adjusted dependent variable is no longer the same as the original dependent variable.

Numbers and the Use of Language

To put the distinctions between the original dependent variable and an adjusted outcome into context, some discussion of general semantics may be useful. Polish mathematician Count Alfred Korzybski (1948) applied the perspective of mathematics to limitations and problems of language in order to help people use language more precisely and effectively, thereby avoiding common problems in communication. His germinal work, *Science and sanity: An introduction to non-Aristotelian systems and general semantics*, gave rise to a highly influential movement in language and communication studies, general semantics. Korzybski recognized a number of highly significant truths about human sign systems that apply to both language and numbers, of which three are relevant here. First, the word or the number is not the phenomenon that is labeled with that word or measured by that number. There is always more to the real world phenomenon than a word or number can capture. Korzybski and his followers summarize this with the aphorism, "The word is not the thing," which might also be paraphrased as "The number is not the thing." A further parallel might be "One's IQ score is not one's intelligence."

Second, the words or numbers used to describe the concrete material world in more complex ways can never represent 100% of what is described, summarized with the aphorism, "The map is not the territory." Anyone who has used a map but still gotten lost has had a practical experience of this truth. Nor can a battery of test scores hope to represent the concrete complexity of a child.

Finally, Korzybski pointed out that while signs in code systems allow communication with one another about the concrete material world in useful ways, abstraction above and beyond that world is inherent in their use. When researchers talk about or measure the immediate environment, they have moved one step away from that environment with those words or numbers. When they discuss the talk or

average those measurements, they have moved two steps away. When researchers analyze the discussion about the original talk, or manipulate the averages of the original measurements, they have distanced themselves yet again, and so on. As researchers measure achievement, calculate means, covary for still other measures, the original concept of achievement becomes so abstract as to be something else entirely.

That is, it is difficult to understand what is actually being measured by the adjusted score. It could be a meaningful concept, or it could be something else entirely, possibly error. For example, the original criterion may have been reading achievement, which had validity estimates and made logical sense from a nomological net. The achievement test may have made sense and had adequate validity in terms of its relationship or correlation to other achievement estimates, such as other tests and teacher evaluations. However, when the variance from socioeconomic status is removed from that achievement test, the residual or adjusted score may not have the validity support that the original, unadjusted, reading achievement score had. Therefore, it is possible that the adjusted score, as a criterion, may actually have less validity than the original unadjusted score.

Types of Covariates

To further complicate the situation, the word, "control," can be interpreted in multiple ways. In an attempt to clarify the use of the word, "control," in connection with ANCOVA, Ferguson (1981) proposes that covariates fall into two categories, intrinsic and extrinsic. Intrinsic covariates are attributes which are internal to the subjects such as pretest scores or motivation and may be affected by some aspect in a study, while extrinsic covariates, such as teacher's years of service, are external to both the subjects and the study. SES is, of course, an extrinsic covariate, and as such, is not under the researcher's influence or ability to change. The technical terminology, "controlling for the covariate," however, implies the opposite to those unfamiliar with the language of statistical testing.

Type VI Errors

Given the preceding discussion of semantic meaning, research is still conceptualized as having multiple purposes including predicting; adding to the knowledge base; having a personal, social, institutional, and/or organizational impact; measuring change; understanding complex phenomena; testing new ideas; generating new ideas; informing constituencies; and examining the past (Newman, Ridenour, Newman, & DeMarco, 2003). In the cases of interest here where the effects of SES are removed from achievement, the researcher generally aspires to predicting outcomes or measuring change, both quantitative questions in nature. Such studies have appeal for educational researchers, evaluators and school district personnel, who often speculate whether significant differences exist in achievement between specific programs or ethnic groups but recognize that achievement is confounded with SES.

However, such a research design may be making a Type VI error (Newman, Fraas, Newman & Brown, 2002). Type VI errors occur when the research question does not match the research design. Type VI errors include: "practices that (a) fail to distinguish between statistical analysis and research design issues, (b) do not match the model used in structural equation modeling with the research question, (c) analyze a research question that involves practical significance with an analytical technique that fails to do so, (d) use methods designed to control for inflated Type I error rates that do not match the nature of the research question, and (e) employ multivariate data analysis techniques for research questions that require the application of univariate techniques" (p. 138).

The real question of interest in the ANCOVA design discussed here is whether there are significant differences in achievement, and while researchers may hope such differences exist or not, they really want to know about achievement, the phenomena, and not adjusted achievement, an abstract number. To emphasize this point, predicted scores for the dependent variable of GPA, for example, can be generated in ANOVA and ANCOVA, yet actual GPA and adjusted GPA would look quite different. Adjusted GPA is the residualized variable after the effects of a covariate such as SES has been removed, and it lacks the same meaning as GPA. Adjusted GPA is not in the same metric as GPA, or in other words, it does not have the same mean and standard deviation as GPA. It is inaccurate and misleading to draw conclusions about one variable (GPA) when the analysis has been conducted on a different variable (adjusted GPA).

Common ANCOVA Patterns

SES is so commonly used as a covariate for achievement that such designs have been reported by the popular press (Dillon & Schemo, 2004). In these kinds of designs, the outcome variable is some measure of achievement, the covariate is SES, and the grouping variable may be ethnicity, developmental levels (accelerated, normal, slow), treatment (treatment group, control), school ratings (high, medium, low in achievement), time (pretest, posttest), school type (public, charter, private), or community area (urban, suburban, rural) to name a few of the possibilities.

Typically, these studies begin by accurately describing their variables and analyses, and then slip into some common but unsupportable practices. First, they don't test for homogeneity of regression and so never look for interactions. Second, they fail to understand the implications of the fact that the adjusted scores or residuals will be zero correlated with SES, which overlaps greatly with achievement. Third, their research questions often seem to expect conceptually counterintuitive outcomes, such as students from the poorest performing schools should exhibit academic performances comparable to those students from the highest performing schools if the effects of SES are removed. Fourth, they slip into using achievement and adjusted achievement interchangeably, ignoring the fact that these represent vectors of different scores with different means, different standard deviations, and different patterns of variability. They may even neglect to provide tables of original and adjusted means, or label figures with adjusted means as simply "achievement." Finally, if group means for achievement show one pattern, but means for adjusted achievement appear comparable or even exhibit reversals, misleading conclusions can be made if adjusted means are presented but referred to as original achievement. For example, reading achievement means may show normal-developing students outperforming slow-developing student, but adjusted reading achievement means may show they are comparable. If the results discuss adjusted means but simply refer to them as reading achievement, an erroneous conclusion may be that the two groups are the same, when the truth is that the slow-developing group is still just that, slow-developing and poor in reading.

Determining the Validity of Adjusted Scores

One possible remedy for this problem of interpreting adjusted scores is to treat them as typical variables and to subject them to recognized and acceptable methodological practices. It is a common research practice to assess whether the outcome variable is a valid measure. To ascertain this, the outcome variable is correlated with other recognized, widely used measures of the same construct (Groth-Marnat, 1999; Huck & Cormier, 1996, Nunnally, 1978). Adjusted scores in ANCOVA designs should also undergo the same procedure. If they correlate with other similar measures, their usage is acceptable, and those adjusted scores can be interpreted in a meaningful way. If the adjusted scores do not correlate with other similar measures, such as achievement for example, they are merely residuals or random error, and as such, cannot be interpreted as the original variable, achievement or anything else.

Conclusion

Inconsistencies between the research question and the research methodology, indicative of a Type VI error, can be pervasive and subtle in their semantic expression. Although adjusted scores are used in the analysis cited in this example, conclusions and graphs often tend to refer simply to achievement or to shift in subtle linguist ways that imply that unadjusted achievement is the outcome. Such conclusions are not only inaccurate, but may lead to inappropriate recommendations.

The consequences of a Type VI error in studies of achievement differences covariating for SES can be highly misleading. When research with a Type VI error concludes that there are no significant differences in achievement across groups when statistically controlling for SES, the tacit implications are that actual achievement is consistent across groups and that SES can be causally controlled or somehow is not influential. Neither is correct.

The consequences of a Type VI error in studying achievement when covarying for variables such as SES can be highly misleading. Generally this happens when the researcher asks a question about achievement, conducts the analysis on adjusted achievement scores, but interprets the results in terms of plain, unadjusted achievement. Adjusted achievement is no longer achievement because meaningful, predictable, overlapping variance has been statistically removed. Adjusted achievement is an abstract, unknown construct. That is, what is meant by achievement after the effects of the covariate of SES is

statistically removed is unknown. However, achievement scores for children are increasingly being used for high stakes decisions. Researchers need to keep the two straight, and educators need to teach the difference between the two.

The semantic substitution of achievement for adjusted achievement has a far-reaching impact on subsequent discussions or recommendations. What researchers are testing is not what the research question hopes to ascertain, although they sound quite similar. Achievement is a complex interplay of shared variance, some of which can be influenced by teachers, educational systems or other factors and caused to improve, while some of this variance is attributable to factors such as SES which is external and not controllable in the context of educational research. Higher levels of abstraction are extremely valuable as they can reveal truths and trends that would never be perceived without the human ability to abstract through the use of words and numbers. The other edge of the sword is that higher and higher levels of abstraction may also distort those perceptions, misrepresent the concrete material world in significant ways, and mislead thinking as researchers attempt to understand whatever phenomenon is the focus of their attention.

It is also notable that children who have low SES are much more likely to be children of color. When conclusions that no significant differences in achievement between programs or among ethnic groups are reached, the recommendations that follow may find special programs to help students with specific needs or to create parity are unnecessary. Therefore, it is essential that researchers, policy makers and practitioners carefully distinguish between the manner in which achievement is defined and is validated in research questions and research methodologies. These definitions must coincide for conclusions and recommendations to be viable. Furthermore, if professionals lack clarity in differentiating between adjusted and unadjusted means in achievement, there is little reason to expect that the general public will understand this seemingly subtle distinction or will understand that adjusted achievement won't look anything like a child's ability to read.

References

Attewell, P., & Battle, J. (1999). Home computers and school performance. *Information Society, 15*(1), 1-10.

Chapell, M. S., & Overton, W. F. (2002). Development of logical reasoning and the school performance of African American adolescents in relation to socioecomonic status, ethnic identity, and self-esteem. *Journal of Black Psychology, 28*(4), 295-317.

Dillon, S., & Schemo, D. J. (2004, November 23). Charter schools fall short in public school matchup. *New York Times.* Retrieved June 27, 2005, from http://www.nytimes.com

Ferguson, G. A. (1981). *Statistical analysis in psychology and education (5th ed.).* New York: McGraw-Hill.

Groth-Marnat, G. (1999). *Handbook of psychological assessment (3rd ed.).* New York: John Wiley & Sons.

Gregory, S. T. (Ed.). (2000). *The academic achievement of minority students: Perspectives, practices, and prescriptions.* Lanham, MD: University Press of America.

Huck, S. W., & Cormier, W. H. (1996). *Reading statistics and research (2ne ed.).* New York: Harper Collins.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education, 77*(4), 189-215.

Korzybski, A. (1948). *Science and sanity: An introduction to non-Aristotelian systems and general semantic,*(3rd ed.). Lakeville, CN: International Non-Aristotelian Library, Institute of General Semantics.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Newman, I., Fraas, J., Newman, C., & Brown. (2002). Research practices that produce Type VI errors. *Journal of Research in Education, 12*(1), 138-145.

Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P., Jr. (2003). A typology of research purposes and its relationship to mixed methods. In A. Toshakkori & C. Teddie, Eds., *Handbook of Mixed Methods in Social & Behavioral Research* (pp. 167-188), Thousand Oaks, CA: Sage.

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.).* New York: McGraw-Hill.

O'Brien, V., Martinez-Pons, M., & Kopala, M. (1999). Mathematics self-efficacy, ethnic identity, gender, and career interests rerlated in mathematics and science. *Journal of Educational Research, 92*(4), 231-235.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart and Winston.

Verna, M. A., & Campbell, J. R. (1998). *The differential effects of family processes and SES on academic self-concepts and achievement of gifted Asian American and gifted Caucasian high school students.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 13-17. (ED 419 025).

Send correspondence to: Susan M. Tracz
California State University, Fresno
Email: susant@csufresno.edu