DOCUMENT RESUME

ED 384 654 TM 023 888

AUTHOR	Buser, Karen P.
TITLE	Dangers in Using ANCOVA To Evaluate Special Education Program Effects.
PUB DATE	18 Apr 95
NOTE	16p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Analysis of Covariance; Compensatory Education; Evaluation Methods; *Evaluation Problems; Intervention; Regression (Statistics); Reliability; *Research Methodology; *Special Education; *Statistical Bias
IDENTIFIERS	*Homogeneity of Regression

ABSTRACT

Analysis of covariance (ANCOVA) has been recommended as one vehicle with which to evaluate special education and other intervention impacts (M. J. Taylor and M. S. Innocenti, 1993). Common misinterpretations of this methodology for these purposes are explained. These misapplications of ANCOVA include: (1) ignoring the assumption of homogeneity of regression; (2) using ANCOVA even given a lack of random assignment of subjects to groups; and (3) lack of attention to reliability of covariate scores. It is noted that such misapplications of ANCOVA "can mistakenly make compensatory education look harmful" (D. T. Campbell and A. Erlebacher, 1975, p. 597). At the very least, researchers should exercise caution in using ANCOVA to consider differences between groups, looking carefully at conditions that may bias the data and give false information about treatments. Contains 31 references. (SLD)

ancova.wp 3/30/95

Dangers in Using ANCOVA to Evaluate Special Education Program Effects

Karen P. Buser

Texas A&M University 77843-4225

Paper presented at the annual meeting of the American Educational Research Association (session #6.64), San Francisco, April 18, 1995.

ABSTRACT

ANCOVA has been recommended as one vehicle with which to evaluate special education and other intervention impacts (Taylor & Innocenti, 1993). Common misinterpretations of this methodology for these purposes are explained. These misapplications of ANCOVA include (a) ignoring the assumption of homogeneity of regression, (b) using ANCOVA even given a lack of random assignment of subjects to groups, and (c) lack of attention to reliability of covariate scores. It is noted that such misapplications of ANCOVA "can mistakenly make compensatory education look harmful" (Campbell & Erlebacher, 1975, p. 597).

Education, by its very nature, often makes difficult the use of true experiments. When working with children, instructional programming, and intrinsic attributes, researchers often need to account for variation attributable to variables other than the independent variables of primary concern, especially when random assignment to treatments is unavailable to the researcher. Analysis of covariance (ANCOVA) (Huitema, 1980) has long been used as an external statistical control to explain extraneous variation in research designs. Using this procedure, the researcher hopes to increase precision, by crediting the covariates with a portion of the variation, thus reducing error variance.

The educational literature is replete with examples of measuring treatments on intact groups. ANCOVA analyses are occasionally employed to equalize preexisting differences among these intact groups. While some ANCOVA analyses have been done appropriately, others appear to be cases medicinal ANCOVA, applying analysis of covariance to studies of weak design for a "quick fix." For years, ANCOVA was described in such terms as, "the preferred statistical method," to compensate for inequalities between intact groups (Borg & Gall, 1983, p. 667). Unfortunately, ANCOVA is not good medicine in such cases, notwithstanding intermittent and incorrect protests (e.g., Taylor & Innocenti, 1993) to the contrary.

However, of late, several statistical messiahs have arisen preaching the dangers of inappropriate uses of ANCOVA analyses (cf. Campbell, 1989; Loftin & Madison, 1991; Thompson, 1988, 1992,

1994a). Misinterpretation and misapplication of ANCOVA data may result from ignoring the assumption of homogeneity of regression, lack of random assignment of subjects to groups, and lack of attention to reliability of covariate data. Novice researchers should heed warnings against committing these unpardonable ANCOVA sins.

The importance of homogeneity of regression

ANCOVA analysis employs a covariate to make predictions that "adjust" the dependent variable scores (using one regression equation computed ignoring group affiliation), followed by an ANOVA on the residualized dependent variable scores (sometimes called the "e" or error scores) (Thompson, 1992). This is appropriate only if the separate regression equation for the covariate(s) and the dependent variable are reasonably similar. Put differently, this single equation is only appropriate if the regression lines of the groups have parallel slopes (Loftin & Madison, 1991).

As Thompson (1992) notes, this

> ...assumption requires that the "b" weights applied to the covariate(s) be reasonably equal across each group. If this assumption is met, it is then reasonable to use a single average or "pooled" equation for all the subjects.... But when the assumption is *not* met, the covariate scores will be correlated with the group membership scores, and ANCOVA will require that the group membership scores also be residualized by the covariate. (p. ix, xi)

Similarly, Cliff (1987, p. 273) argued that, "It could be that the relationship between the dependent variable and the covariate is different under different treatments. Such occurrences tend to invalidate the interpretation of the simple partial correlations described above."

Testing for homogeneity of regression should be the first step taken when employing ANCOVA analysis. If the homogeneity of regression assumption is not met, the ANCOVA analysis may be employing a regression equation different from that of any of the separate groups, and mistakenly applying that single equation to all groups. Thus, this "line of best fit" would inaccurately adjustment the dependent variable scores of all the groups.

As Loftin and Madison (1991) state,

> This is exactly where most applications of ANCOVA fail, since researchers quite often have truly nonequivalent *k* groups for which the regression slopes indeed are different. (p. 141)

A pooled regression coefficient must be accurate for appropriate adjustment of dependent variable scores and thus means. Only then can a researcher decide if a difference exists between group means, after dependent variable scores are adjusted using the covariate(s).

The importance of random assignment to groups

When an experimental design calls for administering treatments to intact groups, the researcher may try to filter out preexisting differences by using analysis of covariance to adjust scores and

equalize groups. ANCOVA enthusiasts beware! This practice has been strongly questioned.

It would be wonderful if the ANCOVA "statistical correction" for pre-existing group differences could always be used. Some researchers incorrectly believe that ANCOVA has just such magic, and

> can "save" a shoddy experiment [with major, real, pre-existing group differences]. Some researchers overuse this method as in the instance of a person I once overheard asking of a researcher, "Where is your analysis of covariance?"--the understanding in his department was that it is <u>always</u> used in experimentation. (McGuigan, 1983, p. 231)

Unfortunately, there is no more magic in statistics than there is other aspects of life. If the groups are different (e.g., a compensatory intervention group and a group not eligible for the intervention) at the start of a study, ANCOVA <u>cannot</u> always be used to statistically adjust for these differences.

Thus, Lord (1969) concluded that there is *no* statistical procedure that can properly allow for uncontrolled, preexisting differences among groups (Hinkle, Wiersma & Jurs, 1994, p. 485). Others agree, noting that

> The situation is radically different (some say hopeless) when ANCOVA is used in quasi-experimental or in experimental research for the purpose of "equating" intact groups. (Pedhazur, 1982, p. 520)

> Ideally, ANCOVA should be used only to correct assignment error in cases where random assignment of individuals to groups was conducted. (Loftin & Madison, 1991 p. 145)

> Unfortunately, ANCOVA is not a panacea for equalizing dissimilar groups. (Campbell, 1989, p. 4)

> ANCOVA is only appropriate for use in conjunction with randomly assigned groups. (Thompson, 1991)

> ...The use of ANCOVA with nonexperimental designs to equate groups that were not randomly created at the time of the study represents a major misuse of the procedure. In most, if not all of the cases, ANCOVA is not statistically justified, since it depends critically on the assumption that individual subjects--not groups of subjects--have been assigned randomly to the treatment conditions. (Keppel & Zedeck, 1989, p. 483)

At the very least, researchers should exercise extreme caution when using ANCOVA to consider differences between groups, looking carefully for conditions that may bias the data and give false information about treatments. It must be remembered that, as Lord (1969) stated, "The answer depends on the means used" (p. 336). This is especially important in special education and other

compensatory interventions, because a misapplications of ANCOVA can actually make the compensatory intervention look like it is harming students who are actually being helped, as Campbell and Erlebacher (1975) so strongly emphasized!

The importance of reliability of data

The reliability of data associated with ANCOVA analyses is critical in making an accurate interpretation of results. However, as Campbell (1989) states, many researchers not only "do not report the measurement error of their variables", they "may inappropriately make statistical corrections using unreliable covariates that make random adjustments" (p. 4). This may occur because many researchers incorrectly believe that tests rather than scores are reliable, and such researchers therefore do not monitor the reliability of the scores they are actually using as covariates (Thompson, 1994b).

Few researchers even report the reliability of their covariates and other data (Meier & Davis, 1990; Willson, 1980), and even fewer seem to realize that reliability must be associated with scores, and not the tests themselves (Rowley, 1976; Sax, 1980). The importance of reliability is clearly stated by Thompson (1994a, p. 5),

> ...too few researchers act on the premise that score reliability establishes a ceiling for substantive effect sizes... The failure to consider score reliability in substantive research may exact a toll on the interpretations within research studies. We

may conduct studies that could not possibly yield noteworthy effect sizes. Or we may not accurately interpret our results if we do not consider the reliability of the scores we are actually analyzing.

Particular attention must be given to accounting for reliability when adjusting sampling error in an ANCOVA analysis. Otherwise, "...one will end up potentially adjusting sampling error with measurement error, and creating a mess" (Loftin & Madison, 1991, p. 145).

One journal's use of ANCOVA

Considering the historical propensity of researchers to ignore the need for homogeneity of regression, random assignment of subjects to groups, and reporting the reliability of their own data, attention was given to exploring one journal's reported use of ANCOVA analyses. Volumes 5 and 6 (1982-83) and 15 and 16 (1992-93) of the Learning Disability Quarterly were examined for treatment of ANCOVA data and trends during a ten-year period. Reports were categorized based on for reports of testing homogeneity of regression, use random assignment, and reporting covariate reliability.

A total of seven articles reported using ANCOVA analyses in these volumes (Bryan, Cosden, & Pearl, 1982; Bryan, Sonnefeld, & Grabowski, 1983; Harbor, 1983; Hutchinson, 1993; Pflaum, Pascarella, Auer, Augustyn, & Boswick, 1982; Schunk & Rice, 1992; Welch, 1992). Of the four studies in the earlier two volumes, two reported tests for homogeneity of regression, and two structured

their studies with random assignment of subjects to groups. Three of the four studies reported reliability of data associated with their measurements. Two of the studies did not meet the requirements for appropriate usage of ANCOVA in two of the three categories mentioned above.

In volumes 15 and 16 (1992-93), three studies used ANCOVA as part of their data analyses. Two of the three studies gave attention to homogeneity of regression, reliability of scores, and random assignment. The third study failed to report testing for homogeneity of regression, and did not employ random assignment of subjects to groups.

A data set of seven studies in four volumes does not allow a definitive judgment concerning reported use of ANCOVA analyses in Learning Disability Quarterly. However, it should be noted that in this journal, as in other journals, ANCOVA is used fairly infrequently (Elmore & Woehlke, 1988; Willson, 1980), perhaps because of the difficulties with ANCOVA noted earlier. However, there does appear to be a slight trend toward more complete attention being devoted to addressing the requirements for appropriate application of ANCOVA, as established by recent literature.

Conclusion

In the 1963 Handbook of Research on Teaching, Campbell and Stanley wrote an influential chapter on experimental and quasi-experimental design. Campbell and Stanley (1963, p. 193) suggested that "the use of this more precise analysis [e.g., ANCOVA] would

seem highly desirable." They also argued that "covariance analysis and blocking on 'subject variables' such as prior grades, test scores, parental occupation, etc., can be used, thus increasing the power of the significance test" (p. 196).

Campbell and Erlebacher (1975) subsequently issued what appeared to be a recant noting that the decision to blithely use ANCOVA statistical control when the homogeneity of regression assumption is not met "can mistakenly make compensatory education look harmful" (Campbell & Erlebacher, 1975, p. 597). In the context of this recant, it is ironic that the 1963 Handbook of Research on Teaching also included the admonition that ANCOVA "is never really satisfactory except as an adjunct to some appropriate randomization procedure" (Lumsdaine, 1963, p. 656).

In the wake of such powerful arguments by noted researchers in the field, it is difficult to ignore the call for an ANCOVA alert. While appropriate application of analysis of covariance may aid in the interpretation of results, "tragically misleading analyses" (Campbell & Erlebacher, 1975, p. 597) are always worth avoiding.

Authors of empirical literature and of statistical textbooks, and the professorate, should make continuing efforts to address both the advantages and disadvantages associated with the analysis of covariance methodology. Only when information is analyzed and reported accurately can we be certain of our conclusions regarding the impacts of educational interventions.

References

Borg, W. R., & Gall, M. D. (1983). Educational research. New York: Longman.

Bryan, T., Cosden, M., & Pearl, R. (1982). The effects of cooperative goal structures and cooperative models on LD and NLD students. Learning Disability Quarterly, 5, 415-419.

Bryan, J. H., Sonnefeld, L.J., Grabowski, B. (1983). The relationship between fear of failure and learning disabilities. Learning Disability Quarterly, 6, 217-221.

Campbell, D. T., & Erlebacher, A. (1975). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In M. Guttentag & E. L. Struening (Eds.), Handbook of evaluation research (Vol. 1, pp. 597-617). Beverly Hills: SAGE.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage, Handbook of research on teaching (pp. 171-246). Chicago: Rand McNally.

Campbell, K. T. (1989, November). Dangers in using analysis of covariance procedures. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 312 298)

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational

Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Harbor, J. R. (1983). The effects of illustrations on the reading performance of learning disabled and normal children. Learning Disability Quarterly, 6, 55-59.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin.

Huitema, B.E. (1980). The analysis of covariance and alternatives. New York: John Wiley & Sons.

Hutchinson, N. L. (1993). Effects of cognitive strategy instruction on algebra problem solving of adolescents with learning disabilities. Learning Disability Quarterly, 16, 34-50.

Keppel, G. & Zedeck, S. (1989). Data analysis for research designs. New York: W. H. Freeman and Company.

Loftin, L. B., & Madison, S. Q. (1991). The extreme dangers of covariance corrections. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 133-147). Greenwich, CT: JAI Press.

Lord, F. M. (1969). Statistical adjustment when comparing pre-existing groups. Psychological Bulletin, 72, 336-337.

Lumsdaine, A.A. (1963). Instruments and media of instruction. In N.L. Gage, Handbook of research on teaching (pp. 583-682). Chicago: Rand McNally.

McGuigan, F. J. (1983). Experimental psychology: Methods of research (4th ed.). Englewood Cliffs: Prentice-Hall.

Meier, S. T., & Davis, S.R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. Journal of Counseling Psychology, 37, 113-115.

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

Pflaum, S. W., Pascarella, E. T., Auer, C., Augustyn, L. & Boswick, M. (1982). Differential effects of four comprehension-facilitating conditions on LD and normal elementary-school readers. Learning Disability Quarterly, 5, 106-115.

Rowley, G.L. (1976). The reliability of observational measures. American Educational Research Journal, 13, 51-59.

Schunk, D. H., & Rice, J. M. (1992). Influence of reading-comprehension strategy information on children's achievement outcomes. Learning Disability Quarterly, 15, 51-63.

Sax, F. (1980). Principles of educational and psychological measurement and evaluation (2nd ed.). Belmont, CA: Wadsworth.

Taylor, M.J., & Innocenti, M.S. (1993). Why covariance? A rationale for using analysis of covariance procedures in randomization studies. Journal of Early Intervention, 17(4), 455-466.

Thompson, B. (1988, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)

Thompson, B. (1991). [Review of Data analysis for research

designs]. Educational and Psychological Measurement, 51, 500-510.

Thompson, B. (1992). Misuse of ANCOVA and related "statistical control" procedures. Reading Psychology, 13(1), iii-xviii.

Thompson, B. (1994a, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Thompson, B. (1994b). Guidelines for authors. Educational and Psychological Measurement, 54(4), 837-847.

Welch, M. (1992). The PLEASE strategy: A metacognitive learning strategy for improving the paragraph writing of students with mild learning disabilities. Learning Disability Quarterly, 15, 19-127.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.