DOCUMENT RESUME

ED 159 223 TM 007 706

AUTHOR Olejnik, Stephen F.; Porter, Andrew C.
TITLE Estimating Treatment Effects and Precision for Quasi-Experiments Assuming Differential Group and Individual Growth Patterns.
PUB DATE [Mar 78]
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; *Analysis of Covariance; Analysis of Variance; Comparative Statistics; Hypothesis Testing; *Mathematical Models; *Reliability; *Research Design; Sampling; Standard Error of Measurement; Test Reliability; Tests of Significance
IDENTIFIERS Fan Spread Hypothesis; *Quasi Experimental Research; *Sample Size

ABSTRACT

The statistical properties of two methods of estimating gain scores for groups in quasi-experiments are compared: (1) gains in scores standardized separately for each group; and (2) analysis of covariance with estimated true pretest scores. The fan spread hypothesis is assumed for groups but not necessarily assumed for members of the groups. Sample standard deviations provide a biased estimate of treatment effects when the first procedure is used with small samples and poorly correlated measures. Under a nonlinear model of within-group growth, the true score analysis of covariance underadjusts for initial group differences. Under the assumption of linear within-group growth and with large samples, both methods estimate the desired effect with equal precision. When samples are small, neither of the two methods are appropriate for the nonlinear model of within-group growth. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED159223

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

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

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephen F. Olejnik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ESTIMATING TREATMENT EFFECTS AND PRECISION FOR QUASI-EXPERIMENTS ASSUMING DIFFERENTIAL GROUP AND INDIVIDUAL GROWTH PATTERNS

Stephen F. Olejnik

Institute for Research on Teaching
Michigan State University

and

Andrew C. Porter

Institute for Research on Teaching
Michigan State University

Paper presented at the 1978 AERA meetings in Toronto, Canada.

TM007 706

ESTIMATING TREATMENT EFFECTS AND PRECISION FOR QUASI-EXPERIMENTS ASSUMING DIFFERENTIAL GROUP AND INDIVIDUAL GROWTH PATTERNS

Stephen Olejnik

Institute for Research on Teaching
Michigan State University

and

Andrew C. Porter*

Institute for Research on Teaching
Michigan State University

Abstract

The appropriateness of standardized gain scores and analysis of covariance adjusted for errors of measurement were considered for quasi-experiments conforming to the fan spread hypothesis. Previous confusion in this area was resolved by considering a linear and a non-linear model of within group growth. For the linear model both procedures estimated the desired effect with equal precision when samples were large. With small samples analysis of covariance was appropriate but standardized gain scores estimated the wrong effect with spurious power. For the non-linear model only standardized gain scores were appropriate with large samples. Neither procedure was appropriate for the non-linear model with small samples.

ERIC

ESTIMATING TREATMENT EFFECTS AND PRECISION FOR QUASI-EXPERIMENTS ASSUMING DIFFERENTIAL GROUP AND INDIVIDUAL GROWTH PATTERNS

Selecting an appropriate analysis strategy for a study based on a quasi-experimental research design has been a topic of considerable controversy. The nonequivalent control group design (Campbell and Stanley, 1963) in particular has received a great deal of attention. Recently, the discussion has focused on the issue of academic rate of growth and its implications for traditional analyses procedures. Specifying the appropriate analytic model is dependent on how individuals change over time. Some authorities (Campbell and Boruch, 1975) have suggested that the initial difference between the comparison groups on a pretest achievement measure implies that the comparison groups are growing academically at different rates. Initial achievement differences have been found frequently in quasi-experimental studies like the evaluation of compensatory education programs. This differential growth rate problem has been labeled the fan spread hypothesis (Campbell and Erlebacher, 1970). The theory suggests that without a treatment, the difference between the comparison group means would increase over time and there would be a proportional increase in the within group variability. This relationship between the increasing mean difference and the within group variance can be represented algebraically as the following:

$$\frac{\mu_{xpt} - \mu_{xct}}{\sigma_t} = K$$

where:

μ_{xpt}, μ_{xct} are the population means on measure (X) for the program and control groups, respectively, at time t;

σ_t is the pooled within-group standard deviation of the outcome measure at time t; and

K is a constant.

ERIC

The use of traditional analyses techniques under this conceptualization has been challenged as inappropriate on the basis that they underadjust for group differences (Campbell and Boruch 1975; Kenny, 1976). Not everyone agrees however with this assessment and evidence supporting the use of traditional analyses procedures have been presented (Porter, A. C. and Chibucos, T., 1974; Bryk, A.S. and Weisberg, H., 1977).

The confusion concerning this issue of an appropriate analytic strategy can be attributed to a large extent to conflicting assumptions made implicitly on the nature of individual growth. The previous discussions of the fan spread hypothesis have concentrated on the differential growth rates between comparison groups. The growth rates for these groups have been assumed to be linear over time. The growth rates of individuals within these groups however have been ignored. For the average group growth to be linear, individuals may grow linearly or non-linearly. A linear model of within group growth implies that individuals are changing at a constant rate (see figure 1). The correlation

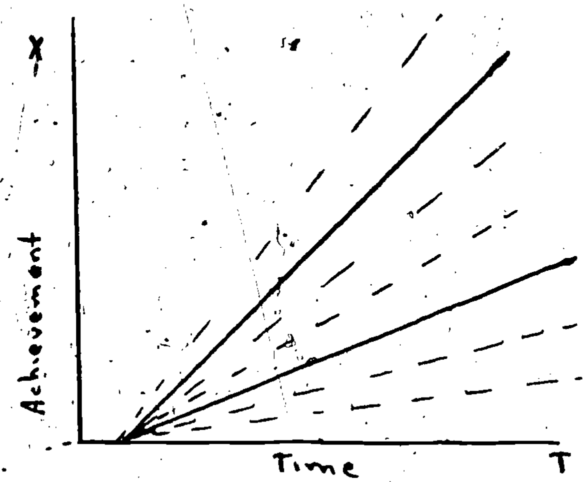

Figure 1. The fan spread hypothesis with the linear model of within group growth. The solid lines represent average group growth while the dashed lines represent individual growth.

between a pretest and a posttest should therefore equal unity except for measurement errors. The non-linear model of within group growth suggests that an individual may grow academically at varying rates across time. That is, academic growth may occur in spurts (see figure 2). The correlation between a pretest

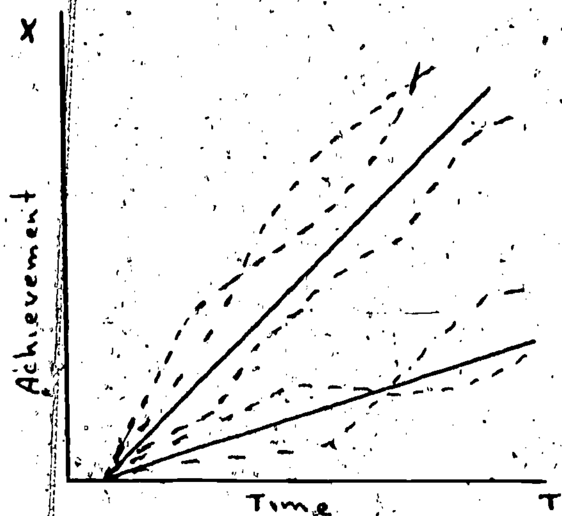

Figure 2. The fan spread hypothesis with a non-linear model of within group growth.

and a posttest measure under this model would not equal unity even with a perfectly reliable instrument. Selecting the appropriate analysis strategy depends on which of these models of within group growth is appropriate.

Purpose

The purpose of the present study was to consider two competing analytic strategies in light of the two models of within group growth suggested above. The two procedures considered were the following: 1) standardized gain scores or more appropriately gains in standard scores and 2) single covariable analysis of covariance with estimated true scores. The first technique involved the use of the analysis of variance model with the dependent variable created by taking the difference between a pretest and a posttest after standardizing each measure. Standardization is achieved by dividing each measure by the pooled within group standard deviation at each point in time. The second technique uses the estimated true score of the pretest rather than the observed pretest score as the covariate in the analysis of covariance model (Porter, 1967) This procedure corrects for the errors of measurement found in the pretest data. The appropriateness of each technique under the two models of within group growth was based on two criteria: 1) the effect estimated by each procedure and 2) the precision with which each effect was estimated. The two analyses procedures studied have been considered previously in the literature on the basis of the first criterion (Kenny, D., 1975; Bryk and Weisberg, 1977). Because the nature of within group growth assumed was not explicitly stated, the results and recommendations have been contradictory. The second criterion proposed in the present study has not been considered in the previous studies examining the two competing analytic strategies. Precision provides a basis on which an analysis technique may be selected in situations where the same desired effect is estimated by two or more procedures.

<u>Estimation under the fan spread hypothesis</u>

The fan spread model of growth suggests that concomitant with an increase in mean difference between comparison groups is a proportional increase in within group variability. Furthermore this relationship between the mean difference and pooled standard deviation remains constant across time. Algebraically this relationship can be presented as the following:

$$\frac{\mu_{xp} - \mu_{xc}}{\sigma_x} = \frac{\mu_{yp} - \mu_{yc}}{\sigma_y}$$

The terms are as defined previously. From this definition of the fan spread hypothesis, the appropriate adjustment for differences on a posttest (Y), given no treatment effect, is the product of the pretest difference and the ratio of posttest to pretest standard deviations:

$$\alpha = \mu_{yp} - \mu_{yc} - \frac{\sigma_y}{\sigma_x}(\mu_{xp} - \mu_{xc}).$$

An analytic strategy having this adjustment factor provides an unbiased estimate of group differences in situations conforming to the fan spread model of growth. Since the definition of the fan spread hypothesis does ~~not~~ include a reference to the nature of within group growth patterns, the above adjustment is appropriate for both the linear and non-linear models of within group growth.

Estimation with gains in standard scores

Standardizing the scores at each point in time adjusts for the increasing within group variability suggested by the fan spread model of growth. The effect estimated by computing the gains in standard scores can be written as the following:

$$\alpha_{GSS} = \mu_{yp} - \mu_{yc} - \frac{\sigma_y}{\sigma_x}(\mu_{xp} - \mu_{xc}).$$

This estimate is identical to the adjustment strategy suggested above based on the definition of the fan spread model of growth. Thus gains in standard scores provides an unbiased estimate of a treatment effect in situations conforming to the fan spread model. A problem arises however when standardization is achieved by using the pooled within group standard deviation calculated on the sample data. That is, the expected value of the ratio of pooled within group standard deviations, s_y/s_x does not equal the desired ratio of the population standard deviations. Table 1 presents the expected value of the ratio of two non-independent

Table 1

The expected value of the ratio of two non-independent standard deviations assuming equal population standard deviations.

n	$\rho = .9$	$\rho = .7$	$\rho = .4$
20	1.0055	1.0146	1.0238
40	1.0026	1.0068	1.0112
60	1.0017	1.0044	1.0073
80	1.0012	1.0033	1.0054
100	1.0010	1.0026	1.0043

sample standard deviations for various sample sizes and selected relationships between the two measures when the population standard deviations are equal. If the population standard deviations are unequal, as suggested by the fan spread hypothesis, the descrepencies are even greater. Thus the use of the sample standard deviations to standardize scores provides a biased estimate of a treatment effect in studies involving small samples and poorly correlated measures. For large samples and highly correlated measures however the procedure is appropriate. Furthermore, since the technique is not affected by the relationship among the individuals within comparison groups, the gains in standard scores is appropriate for both the linear and non-linear models of individual growth.

Estimation with true score analysis of covariance

The second analytic strategy considered for situations conforming to the fan spread model was the true score analysis of covariance procedure. The procedure assumes knowledge of the appropriate population reliability coefficient in order to calculate the true scores of the pretest data. The effect estimated by using the true scores as a covariate in the analysis of covariance model can be written as the following:

$$\alpha_{ACTS} = \mu_{yp} - \mu_{yc} - \frac{\rho_{xy}}{\rho_{xx}} \frac{\sigma_y}{\sigma_x} (\mu_{xp} - \mu_{xc}).$$

Where ρ_{xy} and ρ_{xx} are the population correlation coefficient and reliability coefficient respectively. The other terms are as defined previously. The above estimate is identical to the adjustment strategy suggested by the fan spread definition except for the ratio of ρ_{xy}/ρ_{xx}. This ratio provides the correction for errors of measurement. Therefore if the true relationship between the two measures is perfect, as proposed by the linear model of within group growth, the ratio of the correlation to the reliability of the covariate will also equal unity. Thus for the linear model of within group growth the analysis of covariance model with estimated true scores provides an appropriate adjustment for the fan spread situation.

Under the non-linear model of within group growth, the true relationship between the pretest and posttest does not equal unity, even after correcting for errors of measurement. Given this model of within group growth, true score analysis of covariance underadjusts for the initial group differences.

Precision

The above discussion on estimation showed that both strategies estimate the desired effect for the linear model of withi group growth. In this situation the precision with which each technique estimates the effect is of great interest. It provides a basis on which a researcher could select between the competing procedures. Precision is defined in terms of the variability of the effect estimated. The greater the variance of the estimate the lower the precision of the test. An index of precision is therefore found in the standard error of the simple contrast of interest.

Both gains in standard scores and true score analysis of covariance can be conceptualized as an index of response having the following form:

$$W = Y - KX$$

Where:

W is the adjusted variable;

X, Y are the pretest and posttest measures respectively;

and K is the adjustment coefficient.

The two procedures differ in how the adjustment coefficient is determined. The gains in standard scores approach defines the adjustment coefficient as the ratio of the pooled within group standard deviations of posttest to pretest, S_y / S_x. The adjustment coefficient for the true score analysis of covariance on the other hand is defined as the ratio of the pooled within group regression slope of posttest on pretest to the population test-retest reliability coefficient $\left(\frac{b_{y.x}}{\rho_{xx}}\right)$ Given this model, the contrast of interest is the simple difference between the group means on the adjusted variable W. The index of precision by which the two strategies ~~can~~ be compared is the following:

$$\sqrt{Var(\overline{W}_p - \overline{W}_c)}$$

The variance of the contrast can be written as the following:

$$\mathrm{Var}(\overline{W}_p - \overline{W}_c) = 2\left[\mathrm{Var}(\overline{Y}) + (E(K)^2 + \mathrm{Var}(K))\mathrm{Var}(\overline{X}) - 2E(K)\mathrm{Cov}(\overline{Y},\overline{X}) + (\mu_{x_p} - \mu_{x_c})\mathrm{Var}(K)\right]$$

(Keesling and Wiley, 1977)

In calculating the adjusted variable, both gains in standard scores and true score analysis of covariance require that the adjustment coefficient (K) be estimated on the sample. This estimate is likely to change from sample to sample. The adjustment coefficient is therefore a random variable which must be taken into consideration by including the Var(K) as a factor in the variance of the contrast. Since both strategies being considered have the same form for the variance of the effect estimated, differences in precision must result from differences in the expected value and variance of the adjustment coefficients.

The expected values for the ratio of two non-independent standard deviations were presented in table 1. The adjustment coefficient for true score analysis of covariance assumes that the population test-retest reliability coefficient is known. The expected value for this adjustment factor is therefore equal to $\frac{1}{\rho_{xx}} E(b_{Y \cdot X})$. Since the sample regression slope is an unbiased estimator of the population regression slope the following is true: $E(b_{Y \cdot X}) = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$. For the linear model of within group growth the expected value of the true score adjustment coefficient therefore equals unity since $\rho_{XY} = \rho_{XX}$ and equal pretest and posttest variance have been assumed.

The variance of the adjustment coefficient for the gains in standard scores procedure was derived from the density function for the ratio of two non-independent standard deviations (Olejnik, 1977)

Table 2 presents the variance for the adjustment coefficient given varying sample sizes and selected relationships between the

Table 2

The variance of the ratio of two non-independent standard deviations assuming equal population standard deviations.

n	$\rho = .9$	$\rho = .7$	$\rho = .4$
20	.01125	.03049	.05066
40	.00515	.01388	.02296
60	.00334	.00899	.01485
80	.00246	.00663	.01096
100	.00196	.00527	.00870

two measures. The variance of the true score adjustment coefficient assuming knowledge of the population test-retest reliability coefficient equals: $\frac{1}{\rho_{xx}^2} \mathrm{Var}(b_{y \cdot x})$. The variance of the sample regression slope is equal to $\frac{1-\rho_{xy}^2}{n-3} \cdot \frac{\sigma_y^2}{\sigma_x^2}$. Table 3 presents for varying sample sizes and selected relationships between the two measures, the variance of this adjustment coefficient. As with the previous tables the variance of the pretest and posttest have

Table 3

The variance of the adjustment coefficient suggested by the true score analysis of covariance assuming equal pretest and posttest variance

n	$\rho = .9$	$\rho = .7$	$\rho = .4$
20	.01379	.06122	.30883
40	.00634	.02813	.14189
60	.00412	.01826	.09211
80	.00305	.01360	.06818
100	.00242	.01073	.05412

been assumed to be equal. Finally to facilitate the comparison between the two procedures in terms of precision, table 4 compares the sum of the expected value squared and the variance of the adjustment coefficient for both strategies.

Table 4

The sum of the expected value squared and variance $[E(K)^2+Var(K)]$, for the gain in standard scores and true score analysis of covariance assuming equal pretest and posttest variance

	$\rho = .9$		$\rho = .7$		$\rho = .4$	
	Gains in Standard Scores	True Score Analysis of Covariance	Gains in Standard Scores	True Score Analysis of Covariance	Gains in Standard Scores	True Score Analysis of Covariance
n	$K = s_y/s_x$	$K = \frac{b_{y.x}}{\rho_{xx}}$	$K = s_y/s_x$	$K = \frac{b_{y.x}}{\rho_{xx}}$	$K = s_y/s_x$	$K = \frac{b_{y.x}}{\rho_{xx}}$
20	1.02228	1.01380	1.05489	1.06122	1.09883	1.30883
40	1.01036	1.00634	1.02753	1.02813	1.04549	1.14189
60	1.00674	1.00411	1.01781	1.01826	1.02949	1.09211
80	1.00486	1.00305	1.01324	1.01352	1.02179	1.06818
100	1.00396	1.00242	1.01048	1.01073	1.01732	1.05412

Discussion

A comparison of the expected value and variance associated with the adjustment coefficients for the two competing analytic strategies indicates almost no differences when the relationship between the pretest and posttest is high, $\rho_{xy} = .9$ or $\rho_{xy} = .7$. This is true regardless of the sample size. When the relationship between the measures is low, $\rho_{xy} = .4$, some differences become apparent. In situations where the relationship is low but the sample size is large the two procedures appear to have equal precision. As the sample size decreases, greater precision is associated with the gains in standard scores approach.

The above conclusions on precision have been based on the standard error presented in this paper. There is, however, an important difference between the calculations provided here and the standard error generally associated with the gains in standard scores procedure. In actual practice gains in standard scores are determined in a two-stage process; first the adjusted variable is determined; and second it is used as the dependent variable in the analysis of variance model. This procedure assumes that the adjustment coefficient determined in step one is a constant across replications. This standard error therefore differs from the correct standard error presented earlier by eliminating all factors involving the variability of the adjustment coefficient. This reduced form of the standard error produces spurious precision which leads to a liberal test of the hypothesis under investigation.

The degree to which the gains in standard scores procedure as generally determined, is too liberal, is dependent on the variability of the adjustment coefficient. Table 2 provides some information on this question. When the relationship between the pretest measure and the posttest measure is high and the sample size is large, the variability of the adjustment coefficient is essentially zero. As discussed earlier these are the only conditions under which the procedure estimates the correct effect. Thus under those conditions the gains in standard scores procedure, as they are generally calculated, is likely to be appropriate. As the sample becomes small and the relationship weakens, the probability of error increases as does the bias in estimating the effect.

ERIC

These observations were based on the situation when the variance of the measures are equal across time. In situations conforming to the fan spread model, the variability associated with the adjustment coefficient increases and with it the inappropriateness of the gains in standard scores technique. Thus, unless the sample is large and the relationship between measures is high the use of gains in standard scores as generally calculated should be avoided. On the other hand, the estimated true score analysis of covariance technique does take the variability of the adjustment coefficient into consideration in determining its standard error. The analysis of covariance is not a two-stage process but rather the adjustment coefficient and the estimation are computed in a single analysis. The discussion presented here, has assumed that the population test-retest reliability coefficient was known. If it must be estimated on the data set or an independent sample, the variability of the reliability coefficient must be taken into consideration in calculating the appropriate standard error. Substituting a sample estimate of the test-retest reliability coefficient in the true score analysis of covariance approach can also result in spuriously high precision. Given that the population reliability coefficient is known, the true score analysis of covariance provides an appropriate estimate of a treatment effect under the linear model of within group growth and greater percision than the gains in standard scores technique.

Summary

The present study has resolved the problem of conflicting recommendations found in the literature concerning the gains in standard scores and analysis of covariance adjusted for errors of measurement as appropriate analysis strategies in quasi-experiments conforming to the fan spread hypothesis. The confusion resulted from conflicting implicit assumptions on the nature of individual growth patterns. By clearly stating the nature of within group academic growth, the appropriateness of the two competing analytic strategies were evaluated. Under the linear model of within group growth with large samples both gains in standard scores and true score analysis of covariance estimate the desired effect with approximately equal precision. With small samples however, the true score analysis of covariance technique estimates the desired effect with the appropriate standard error but the effect estimated by the gains in standard scores approach is not the desired one and the estimate is made with spurious power. For the non-linear model of within group growth true score analysis of covariance does not estimate the desired effect. Gains in standard scores does estimate the desired effect with the appropriate standard error only when the samples are large. When samples are small neither of the two strategies considered are appropriate for the non-linear model of within group growth.

References

Bryk, A. S., & Weisberg, H. "Use of thenon-equivalent control group design when subjects are growing". Psychological Bulletin, 1977, 84, 950-962.

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Campbell, D.T. & Boruch, R.F. "Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects." In C. A. Bennett (Ed.), Evaluation and experiment. New York: Academic Press, 1975.

Campbell, D. T. & Erlebacher, A. "How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful." In J. Hellmuth (Ed.), Compensatory education: A national debate (Vol. 3) The disadvantaged child. New York: Brunner/Mazel, 1970.

Keesling, J.W. & Wiley, D. "Measurement error and the analysis of quasi-experimental data, version IV." Unpublished manuscript, 1976.

Kenny D. "A quasi-experimental approach to assessing treatment effects in the non-equivalent control group design." Psychological Bulletin, 1975, 82, 345-362.

Olejnik, S. "Data analysis strategies for quasi-experimental studies where differential group and individual growth rates are assumed." Unpublished Ph.D. dissertation East Lansing: Michigan State University, 1977.

Porter, A. C. & Chibucos, T. R. "Selecting an analysis strategy." In G. Borich (Ed.), Evaluating Educational Programs and Products. Educational Technology Press, 1974

Porter, A. C. "The effects of using fallible variables in the analysis of covariance." Unpublished Ph.D. dissertation, Madison: University of Wisconsin, 1967.