DOCUMENT RESUME

ED 174 634 TM 008 997

AUTHOR Holley, Charles D.; Dansereau, Donald F.
TITLE Some Considerations in Controlling for Transient Motivation in Cognitive Manipulation Studies.
PUB DATE Apr 79
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, Arpil 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Covariance; *Cognitive Measurement; *Control Groups; *Educational Research; Factor Analysis; Higher Education; Reading Comprehension; *Research Design; *Research Problems; Sampling; *Statistical Analysis; Statistical Bias; Verbal Ability
IDENTIFIERS Placebo Effect

ABSTRACT

Four types of control groups are commonly used in cognitive manipulation studies: (1) no-treatment; (2) practice with own methods; (3) practice and training with competing treatments; and (4) practice and training with irrelevant treatments. There are problems associated with the use of each group as a baseline for identifying the "true" treatment effect. A general, two-step control procedure was advocated in which the researcher first identifies the most appropriate control group for a given context and then supplements this experimental control with statistical control. In reference to the latter controls, procedures were presented for developing a task-specific index as a covariable to be employed with a suitable measure of verbal aptitude in a multiple analysis of covariance. The procedure was demonstrated on data from a quasi-experimental design in which it was shown that three different conclusions were possible, depending upon the control procedures employed. The converging operations obtained from the multivariate approach to the data analyses supported the results from the two-step control procedure. (Author/RD)

Some Considerations in Controlling for Transient Motivation in Cognitive Manipulation Studies

Charles D. Holley

Donald F. Dansereau

Texas Christian University

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

Abstract

The four types of control groups commonly employed in cognitive manipulation studies were identified and arguments were presented concerning the problematic nature of each group as a baseline for identifying the "true" treatment effect. A general, two-step control procedure was advocated in which the researcher first identifies the most appropriate control group for a given context and then supplements this experimental control with statistical control. In reference to the latter controls, procedures were presented for developing a task-specific index as a covariable to be employed with a suitable measure of verbal aptitude in a multiple analysis of covariance. The procedure was demonstrated on data from a quasi-experimental design in which it was shown that three different conclusions were possible depending upon the control procedures employed. It was argued that the converging operations obtained from the multivariate approach to the data analyses supported the results from the two-step control procedure.

Some Considerations in Controlling for Transient Motivation in Cognitive Manipulation Studies

The utilization of a placebo control has apparently become sine qua non in the field of educational research, as well as other areas involving human subjects. The rationale for this derives principally from the results of medical research in which it has been reliably demonstrated that controlling for placebo effects is a necessity (e.g., Beecher, 1959; Haas, Fink, & Hartfelder, 1959; Kennedy & Uphoff, 1939). These arguments have also been extended to other inquiries involving human subjects (e.g., Roethlisberger & Dickson--Hawthorne effect, 1939; Rosenthal, 1966, 1967, 1969). However, "across the board" generalization of the use of placebo controls as a panacea may be ill-advised. Such a caution was emphasized by Kaplan (1964):

> ...a conspicuously successful technique in some area of behavioral science is not only identified with "scientific method" but comes to be so mechanically applied that it undermines the very spirit of scientific inquiry. (p. 29)

Campbell and Stanley (1966) pointed out some of the inherent problems in identifying an appropriate placebo group and, in experiments on learning strategies, Dansereau, Collins, McDonald, Holley, Garland and Diekhoff (1979) and Holley, Dansereau, McDonald, Garland and Collins (in press) have argued that typical placebo controls (having students either "practice" their own learning strategies or participate in an alternate training program) can produce artifactually lowered mean performances on the dependent measure(s) thereby leading to an unnecessarily liberal interpretation of the actual effect of the treatment (X). For example, Collins, Dansereau, Holley, Garland and McDonald (in press) describe an experiment containing three treatment conditions and a "no-treatment" control group in which one of the treatment groups had lower mean scores on the dependent indices than the control group. Since the nature of this treatment condition was such that it was expected to facilitate performance, it could have been chosen as a placebo control in which case the interpretation of the actual effects of the other treatment would have been somewhat misleading. Similar results have been reported by Garland (1977) and Long (1976).

Additional evidence supporting the questionable technique of using placebo controls has been cited by Wilkins (1977):

> A similar consideration may be made of the placebo effect, which Archibald claimed is an area in which

> the self-fulfilling prophecy is evident. ...client prophecies and prophetic information delivered to clients were not found to have a reliable effect on actual therapeutic change (Borkovec, 1973; Emmelkamp, 1975; Rosen, 1976; Wilkins, 1973). In addition, the studies by Marcia, Rubin, and Efran (1969) and McGlynn and Williams (1970) show high-prophecy information to be associated with slightly less improvement than low- or no-prophecy information. In an academic setting, Beyer (1971) found the effects of the communication of prognostic information to students to be statistically nonsignificant. Beyer and Oetting (1970) reported a significant reversal effect: Students who were told that they would do quite well in school actually performed less well than students who did not receive that information. ...
>
> On the basis of the evidence that was omitted or erroneously reported by Archibald, the conclusion that self-fulfilling prophecies exist is unwarranted. ... (p. 56)

Consequently, it appears unreasonable to assume that the placebo problem, at least with cognitive manipulation studies (e.g., imagery and mnemonics training, depth of processing instruction, learning strategy training), can be resolved simply by incorporating a placebo group into the experimental design. One major purpose of the placebo group is to control for spuriously inflated "scores" by the experimental group(s) attributable to transient motivational aspects of the experimental situation (e.g., Rosenthal & Hawthorne effects); however, some of the previously cited empirical evidence suggests that the placebo group can produce a spurious effect in the opposite direction.

Four types of control groups are typically utilized in cognitive manipulation studies: no-treatment , practice with own methods, practice and training with competing treatments, and practice and training with irrelevant treatments. For convenience these control groups will hereafter be referred to as Groups A, B, C, and D, respectively. (Groups A and B implicitly require that the to-be-performed task is already in the student's repertoire of skills.) With each of these control groups it is possible, and in some cases highly probable, that the control subjects will be at a higher or lower motivational level than the treatment group at the time the dependent measures are administered. This differential motivation may be the result of factors attributable to both the control and treatment manipulations. (However, the arguments and procedures advanced in this paper are also generalizable to those situations in which differential motivation can be ascribed to a selection bias in the randomization procedure.) Without additional measurements it is not possible to isolate that portion of the total variance that results from these motivational factors.

Groups B and D may seriously violate the purpose of the placebo control by fostering negative motivational factors (e.g., frustration due to the non-meaningfulness of the manipulation). The importance of frustration effects in the experimental situation has been well documented in studies involving both infra-human (e.g., Amsel & Roussel, 1952; Mackintosh, 1974; Wagner, 1959) and human subjects (e.g., van der Keilen, 1978). Group C becomes suspect due to the possibility of differential experimenter presentations (i.e., the effect can be attributed to the experimenter rather than the treatment) and Group A is problematic due to the fact that this group generally spends less time in the experimental situation than the treatment group.

Given the foregoing arguments, the stated purpose of the placebo control may not be attainable with only experimental manipulation; the addition of statistical controls through covariates designed to assess the subjects' transient motivational states are necessary to provide a satisfactory index of the "true" treatment effect. A motivation index can be developed for members of both the treatment and control groups using such measures as self-reported motivation and task effort, functional time on task, distractability, and/or pretesting.

These multiple measures can be factor analyzed with the subsequent factor scores for the first, and possibly the second, factor(s) serving as the transient motivation index to be employed as a covariable (or covariables) in an analysis (or multiple analysis) of covariance (M/ANCOVA) in order to statistically control for these effects on the dependent measures. It is also suggested that, in academic tasks, an appropriate measure of ability be included as an additional covariate. (While self-reported prior content knowledge may represent a potentially useful covariate, results from our laboratory have indicated relatively low correlations between reported prior knowledge scores and the dependent indices of performance; see Note 1). In general, this use of M/ANCOVA is a subset of the Usage-1 and/or Usage-2 procedures detailed by Evans and Anastasio (1968).

The appropriateness of each of the measures of task specific motivation will be determined by constraints imposed by the particular experimental situation. For example, in the study to be reported in this paper, pretesting was employed as the principal measure of motivation since complete random assignment to groups was not feasible (non-equivalent control group, Campbell & Stanley, 1966). Such a design would potentially bias the utility of the self-report measures since the "pay-off" value for "honesty" in the reports may not be equivalent between groups (Carroll, 1972); this results in subjective reports being inappropriate covariates since these scores would potentially be confounded with treatment effects (Evans & Anastasio, 1968). (This assumption was supported by an analysis of the subjective reports; see Note 2.) While pretesting represents an impure measure of transient motivation (i.e., it also contains variance associated with task-specific ability, prior content knowledge, etc.), it is, nonetheless, preferrable in the present context.

Having advanced logical arguments to support the contention that each type of control group is inherently problematic, the objective of this paper is to demonstrate a two-step control procedure in which the experimenter first selects the control group presenting the fewest rival hypotheses to the interpretation of X and then controls for these rival interpretations through the use of M/ANCOVA. Application of this methodology (slightly altered for the non-equivalent control group condition) with Group A (no-treatment control) is demonstrated on data from a learning strategy experiment. Rationale for selecting this as the most appropriate group for this context was given by Holley, et al. (in press):

> ...Attempts at equating training time by having students practice their own or less effective, competing methods have led to suppression of mean performance in comparison to "untrained" students using their own techniques...It should also be emphasized that the college-age students participating in these experiments have had 12 to 14 years of experience and practice with their own study methods (in particular with naturally occurring prose) and can therefore be considered no-treatment controls in name only. (pp. 10-11)

Additionally, study protocols collected from untrained students indicate that these students employ a wide range of rather sophisticated study methods (Holley, Note 3); consequently, the term "ad-lib", rather than no-treatment, appears to be a less misleading label for Group A.

Method

Subjects

Data were collected on 99 undergraduate students at Texas Christian University in the fall semester 1978. Fifty-seven of these students were enrolled in a learning strategy course and were randomly divided into two treatment groups. Forty-two general psychology students served as the (non-equivalent) ad-lib control group.

Procedure

Experimental Sessions. Since the nature of the subject population prohibited complete random assignment to groups, pre and post measures were collected. During the pre assessment phase all students participated in three experimental sessions. Session 1 consisted of completing the Educational Set Scale (Siegel & Siegel, 1965) and the Delta Vocabulary Scale (Deignan, 1973). During session 2 the students spent one hour studying and taking notes on a 2,542-word passage extracted from an introductory college textbook; pre and post-study questionnaires were also administered during this session. During session 3, which occurred five days after session 2, students were allowed to review the passage and their notes for three minutes.

After this review period the students were sequentially administered a pretest questionnaire, an essay exam (14 minutes; students were required to summarize the article), short answer exam (13 minutes; 15 questions), multiple-choice exam (12 minutes; 36 questions), and cloze exam (10 minutes; 24 questions--students were required to fill in important concepts that had been deleted from paragraphs extracted from the article).

The post assessment phase consisted of two sessions identical to the previously described sessions 2 and 3 with the exception that the stimulus passage discussed a different content.

Specification of the training procedures for the two treatment groups is irrelevant for purposes of the present paper. Such specification has been given elsewhere (Dansereau, Holley, Collins, Brooks & Larsen, Note 4) and the training procedures employed were based on previous learning strategies which have been shown to impact on prose processing (e.g., Dansereau, 1978; Dansereau et al., 1979; Holley et al., in press).

Data Analyses. Since four types of exams were employed as pretests, these scores were subjected to a principal factor analysis prior to applying the covariance analysis. The purpose of this procedure was to reduce the number of covariates while retaining as much information as possible from the original data base. Kirk (1968) argued that little utilitarian value could be served by employing more than two covariates; since verbal aptitude will also be employed as a covariable, it is preferable to reduce all other potential covariables to one, perhaps two, "good" composite covariate(s). These composites are represented by the factor scores.

Since performance on the pretest is mediated by a number of nuisance variables (e.g., general verbal ability, task specific ability, motivation and effort, prior content knowledge), prior to applying the covariance analysis to the (post) dependent measures the scores on the pertest were adjusted for the influence of verbal ability (hereafter referred to as COV-1). Thus the task-specific covariate (hereafter referred to as COV-2), while "free" of the influence of verbal ability still remains confounded by the other nuisance variables; this issue will subsequently be elaborated upon in discussing the results.

Since four types of exams were utilized as (post) dependent indices of performance these tests were also subjected to a principal factor analysis in an attempt to reduce the number of dependent variables; factor scores were employed for subsequent analyses.

The final step in the analysis was to employ a one-way multiple analysis of covariance on the dependent measure(s) and to use Tukey's HSD test for post hoc comparisons of the adjusted means.

Results

Means and standard deviations are reported in Table 1. The results for verbal ability indicate a distinct bias in favor of the control group (A) compared to either treatment group-1(T1) or treatment group-2 (T2) whereas the scores for the latter two groups are reasonably comparable.

Insert Table 1 about here

The principal factor analysis of the pretest scores resulted in a one factor solution accounting for 65% of the common variance (see Table 2 for the zero-order correlation matrix and Table 3 for the factor loadings). Mean performances on the resultant factor scores indicated that Groups A and T2 were equivalent and had substantially higher scores than Group T1 (see Table 1 and Figure 1). Adjusting these scores for verbal ability results in T2 being superior to A and A being superior to T1; the latter difference is considerably less than the former difference (see Table 1 and Figure 1). (The assumption of equality of regression slopes was not violated in any of the M/ANCOVAs reported in this paper.) Thus, assuming that deviation from the grand mean of these adjusted scores represents task-specific variance, it appears that group T2's scores are inflated by these sources of error, that group A's scores are slightly depressed and that group T1's scores are substantially depressed. (Since the grand mean of these adjusted scores was zero the deviation scores [COV-2] and the adjusted scores are the same; this will not always be the case.)

Insert Tables 2 and 3 and Figure 1 about here

The principal factor analysis for the ndent exams (the post measures) resulted in a one factor solution accounting for 63% of the common variance (see Table 2 for the zero-order correlation matrix and Table 3 for the factor loadings). Mean performances on the resultant factor scores indicated that groups T2 and T1 were generally equivalent and performing at a higher level than Group A. (A one-way ANOVA of these scores was non-significant.)

Adjusting these scores for verbal ability (COV-1) presents a slightly different pattern of results between the treatment groups (T1 becomes minutely superior to T2) with the differences between these groups and the control group becoming much more pronounced (see Table 1 and Figure 1). An ANCOVA was significant, $F(2,95) = 3.97$, $p \leq .022$; a Tukey HSD pair-wise comparison of means indicated that each of the treatment groups was performing at a significantly higher level ($p \leq .05$) than the control group.

Consequently, had the analysis terminated at this point, the research conclusion would have been that both treatments were effective and that they were equivalent in their effectiveness. However, when the task-specific covariate (COV-2) is introduced into the analyses the adjustment in performance scores is sufficient to warrant a different conclusion. This adjustment indicated that T1 was performing considerably better than T2 and that T2 was performing considerably better than A (see Table 1 and Figure 1). An ANCOVA was significant, $F(2,94) = 4.21$, $p \leq .018$; a Tukey HSD pair-wise comparison of means indicated that T1 was performing significantly better than A ($p \leq .05$) and all other comparisons were non-significant. As previously stated, COV-2 represents an impure measure of motivation; consequently, it is important to point out exactly what these latter adjusted scores represent: These scores represent how the students would have performed if they were equivalent on both general verbal ability and those sources of variance represented by COV-2 (e.g., task-specific ability, motivation and effort, prior content knowledge). It still remains to be determined how much of the differences in these latter scores can be specifically ascribed to non-placebo factors. (With complete randomization to groups, the self-reports would have been incorporated into the development of COV-2 thereby eliminating the need for the following rationale.) Three lines of argument can be employed to support the accuracy of these adjusted scores in representing the true treatment effect:

1. Possibly the most important argument follows from Wilkins (1977) in which it was argued that previous experimental investigations into placebo factors had neither confirmed or denied their existence. However, given the within group correlations between self-reported effort/motivation and performance (reported in Note 2), it could be argued that these changes in correlation do in fact represent the operation of placebo factors. If this can be assumed, then the important research question is not the existence of placebo factors per se but how much variance in the results can be ascribed to placebo effects. (Confidence in the research results would be inversely

related to the magnitude of this source of error.) The average within group correlation for the self-reported measures and performance on the pretest (no placebo effects assumed to be in operation) was .36; the correlation between the self-reported measures and performance on the posttest for the control group (no placebo effects assumed to be in operation) was .36. Thus even if the treatment groups did "try harder" on the posttest only 13% of the variance in the performance scores can be explained on the basis of effort/motivation (self-reported). Consequently, reports of trying harder appear to have little impact on the results.

2. Both treatment groups out-performed the control group; since subjects in the two treatment groups were randomly assigned to conditions and were exposed to similar interactions with, and conditions imposed by, the experimenter, it may reasonably be assumed that the differences between T1 and T2 represent genuine treatment effects (also, see Note 5).

3. Four dependent (post) tests were administered (later linearly collapsed via factor analysis into a single measure); examining the patterns of group performances on each of these exams indicates that the treatment groups did not perform substantially better "across the board" (see Table 4) as might be expected if the results were ascrib-

able simply to differential effort/motivation.

Insert Table 4 about here

Thus, our research conclusion is that while treatment-2 led to an improved performance this improvement could have occurred by chance whereas treatment-1 appears to represent a genuinely effective treatment. (Since the argument might be advanced that prescores represent a sufficient covariate in and of themselves, such adjusted scores were calculated and the results are discussed in Note 6.)

Conclusion

This paper advanced arguments and data to support the assertion that none of the four types of control groups typically employed in cognitive manipulation studies is satisfactory in controlling for placebo factors. An alternative procedure utilizing both experimental and statistical controls was suggested and partially demonstrated.

The general strategy suggested was a two-step procedure in which the researcher first identifies the most appropriate control group for his/her particular investigation and then employs multiple covariates (verbal ability; specific task abilities and motivation) to adjust the dependent measure(s) for these sources of error. Further, that principal factor analysis can be employed to reduce the number of measured variables for each covariate to a single linear combination (i.e., factor scores). For example, measurements of self-reported motivation and task effort, functional time on task, distractability and/or pretesting can be factor analyzed with the resulting factor scores for Factor I serving as the motivation covariate. (Depending on the factor solution, it may be necessary to use Factor II as well; however, given the typical relationships among such motivational measures, it is doubtful that going beyond Factors I and II would be necessary.)

The example presented in this paper demonstrated that three different conclusions could have been drawn from the same set of data depending on the use of no covariates, verbal ability as a single covariate, or verbal ability and task-specific knowledge/skill and motivation as multiple covariates. These three research conclusions were, respectively:

1. Neither of the two treatments employed in this study was effective.

2. Both of the treatments employed in this study were effective.

3. Only one of the treatments employed in this study was effective.

Examination of the pretest scores and the verbal ability scores argue against conclusions 1 and 2. First, both of the treatment groups were lower in verbal ability than the control group. (The importance of verbal ability in such a research context has been well documented by other researchers--e.g., Cronbach & Snow, 1977; Snow, 1977.) This indicates a need to equate the groups for the influence of verbal ability on performance. Second, although the students in the two treatment groups were randomly assigned to conditions (mean verbal ability was reasonably equivalent between these two groups) their performances on the pretests indicated substantial non-equivalence of the groups. This result indicated a need to equate the groups for the influence of specific task ability and motivation on performance. It was argued that this source of non-equivalence (after adjusting for verbal ability) represented an index of specific task abilities and motivational differences, albeit confounded, between the groups.

While the control procedures advanced in this paper appear reasonable and present an attempt to "go beyond" the mechanical application of typical experimental controls for placebo effects, these procedures are not advocated as a panacea. In particular, the investigator must exercise caution in the use of statistical controls since their misuse can produce spuriously significant results (Evans

& Anastasio, 1968; also, see Lord, 1960). However, as in the experiment reported in the present paper, by employing a multivariate experimental approach the researcher is able to establish a sufficient number of converging operations on the data to be reasonably confident of the accuracy of the conclusions.

Although the present demonstration was based on a non-equivalent control group design, the general two-step control procedure is applicable for either the quasi- or fully experimental situation (see Campbell & Stanley, 1966). However, particularly with regard to quasi-experimental designs, it is apparent that systematic bias in verbal ability and task-specific variance existed in the groups utilized in the present study; these variables were also highly correlated with performance and were, therefore, employed as multiple covariables. The underlying assumption is that, in the present research context, these covariables brought under control all relevant sources of systematic bias between the groups. To the extent that this assumption is violated, the advocated procedure must be viewed as palliative and the concomitant implications for comparison of treatment effects must be noted (cf. Cronbach & Furby, 1970).

Reference Notes

1. Correlations between self-reported prior content knowledge and the performance indices were .02 for the pretest and .13 for the posttest. Additionally, these self-report measures are subject to pay-off value considerations discussed in Note 2.
2. Within group correlation between the pretest scores and the (pre) self-reported scores of motivation were .39, .22, and .48 for treatment group-1, treatment group-2, and the control group, respectively; the within group correlations between the posttest scores and the (post) self-reported scores of motivation were .11, .08, and .36 for the respective groups. Thus, it appears that self-report measures have potential utility when the subject perceives the self-report to be independent of any potential pay-off accruing to him/her; in the present data the control group was believed to correspond to this situation and their pre and post correlations remained reasonably high and consistent. On the other hand, the treatment groups could have (falsely) perceived it to be to their advantage to erroneously report positive motivation; the low and erratic pre and post correlations for these groups appear to support such an interpretation.

3. Holley, C. D. *An evaluation of intact and embedded outlines as schema cuing devices with non-narrative text*. Doctoral dissertation in progress, Texas Christian University, 1979.

4. Dansereau, D. F., Holley, C. D., Collins, K. W., Brooks, L., & Larson, D. *Further evaluation of a systematic learning strategy program*. (Tech. Rep. in progress). Institute for the Study of Cognitive Systems, Texas Christian University, Fort Worth, Texas, 1979.

5. Considered in isolation from Group A, Groups T1 and T2 could be viewed as representing the typical placebo control experiment. While arguments were presented in the introduction concerning the problematic nature of such a design (control group C), these arguments are not necessarily applicable in the present context since the only difference in the experimenter manipulations for the T1 and T2 conditions was in the sequencing of the treatment materials.

   Further, Campbell and Stanley's (1966) comments concerning such comparisons should be considered:

   > ...any placebo therapy which is plausible enough to look like help to the student is apt to be as good a therapy as is the treatment we are studying. ...experimental tests of the relative efficacy of

> two therapeutic procedures are much easier to evaluate than the absolute effectiveness of either. (p. 16)

Note the implicit "Catch-22" in the above quote (applicable to control groups C and D in the present paper): It is pointless to test between a "real" treatment and a placebo treatment because the placebo treatment will produce actual treatment effects. (Arguments were made in the present paper that these effects can be either positive or negative.)

Failure to examine the absolute effectiveness of the treatments represents the logical fallacy of *petitio principii*. Determination of this absolute effectiveness is normally imperative but can only be obtained with a "neutral" placebo condition. A closer approximation of the condition of neutrality obtains when the researcher combines experimental and statistical control.

6. An ANCOVA of the (post) dependent measure using only the pretest as a covariate resulted in non-significant results ($p \leq .13$). The use of multiple analysis of covariance with verbal ability and the pretest as multiple covariables led to a reduction in the within group error terms compared to the within group error terms for either of these measures employed as a single covariate.

Consequently, it appears that all of the influence of verbal ability is not accounted for by performance on the pretest.

References

Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. _Journal of Experimental Psychology_, 1952, _43_, 363-368.

Beecher, H. K. Generalization from pain of various types and diverse origins. _Science_, 1959, _130_, 267-268.

Beyer, H. N. Effect of students' knowledge of their predicted grade point averages on academic achievement. _Journal of Counseling Psychology_, 1971, _18_, 603-605.

Beyer, H. N., & Oetting, E. R. Effect of encouragement and reassurance on students' performance in remedial English. _Journal of Educational Research_, 1970, _64_, 59-60.

Borkovec, T. D. The role of expectancy and physiological feedback in fear research: A review with special reference to subject characteristics. _Behavior Therapy_, 1973, _4_, 491-505.

Campbell, D. T., & Stanley, J. _Experimental and quasi-experimental designs for research_. Chicago: Rand-McNally, 1966.

Carroll, J. B. Defining language comprehension: Some speculations. In J. B. Carroll & R. O. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington, D. C.: John Wiley & Sons, 1972.

Collins, K. W., Dansereau, D. F., Holley, C. D., Garland, J. C., & McDonald, B. A. Control of affective responses during academic tasks. Journal of Educational Psychology, 1979, in press.

Cronbach, L. J. & Furby, L. How we should measure "change"--or should we? Psychological Bulletin, 1970, 74(1), 68-80.

Cronbach, L. J. & Snow, R. E. Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington, 1977.

Dansereau, D. F. The development of a learning strategy curriculum. H. F. O'Neil, Jr., (Ed.), Learning Strategies. New York: Academic Press, 1978.

Dansereau, D. F., Collins, K. W., McDonald, B. A., Holley, C. D., Garland, J. C., Diekhoff, G., & Evans, S. H. Development and evaluation of an effective learning strategy program. Journal of Educational Psychology, 1979, 71(1), 64-73.

Deignan, G. D. The Delta reading vocabulary test. Air Force Human Resources Laboratory, Lowry Air Force Base, Colo., 1973.

Emmelkamp, P. M. G. Effects of expectancy on systematic desensitization and flooding. European Journal of Behavioural Analysis and Modification, 1975, 1, 1-11.

Evans, S. H., & Anastasio, E. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. Psychological Bulletin, 1968, 69(4), 225-234.

Garland, J. C. The development and assessment of an imagery based learning strategy program to improve the retention of prose material. Unpublished master's thesis, Texas Christian University, 1977.

Haas, H., Fink, H. & Hartfelder, G. Das placeboproblem. Fortschritte der Arzneimittleforschung, 1959, 1, 279-454. Translated in Psychopharmacology Service Center Bulletin, 1959, 2(8), 1-65.

Holley, C. D., Dansereau, D. F., McDonald, B. A., Garland, J. C., & Collins, K. W. Evaluation of a hierarchical mapping technique as an aid to prose processing. Contemporary Educational Psychology, 1979, in press.

Kaplan, A. The conduct of inquiry. San Francisco: Chandler, 1964.

Kennedy, J. L. & Uphoff, H. F. Experiments on the nature of extra-sensory perception, III. The recording error criticisms of extra-chance scores. Journal of Parapsychology, 1939, 3, 226-245.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, CA.: Brooks/Cole, 1968.

Long, G. L. The development and assessment of a cognitive process based learning strategy training program. Unpublished doctoral dissertation, Texas Christian University, 1976.

Lord, F. M. Large-sample covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 1960, 55, 307-321.

McGlynn, F. D., & Williams, C. W. Systematic desensitization of snake-avoidance under three conditions of suggestion. Journal of Behavior Therapy and Experimental Psychiatry, 1970, 1, 97-101.

Marcia, J. E., Rubin, B. M., & Efran, J. S. Systematic desensitization: Expectancy change or counter-conditioning? Journal of Abnormal Psychology, 1969, 74, 382-387.

Mackintosh, N. J. The psychology of animal learning. New York: Academic Press, 1974.

Roethlisberger, F. J., & Dickson, W. J. Management and the worker. Cambridge, Mass.: Harvard University Press, 1939.

Rosen, G. M. Subjects' initial therapeutic expectancies and subjects' awareness of therapeutic goals in systematic desensitization: A review. Behavior Therapy, 1976, 7, 14-27.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R. Covert communication in the psychology experiment. Psychological Bulletin, 1967, 67, 356-367.

Rosenthal, R. Interpersonal expectations: Effects of the experimenters hypothesis. In R. Rosenthal & R. L. Rosnow (Eds.), Artifact in behavioral research. New York: Academic Press, 1969.

Siegel, L. & Siegel, L. C. Educational set: A determinant of acquisition. Journal of Educational Psychology, 1965, 56, 1-12.

Snow, R. E. Individual differences, instructional theory, and instructional design. (Tech. Rep. 4). Stanford, CA.: Aptitude Research Project, School of Education, Stanford University, 1977.

van der Keilen, M. Critical note on use of the terms "failure" and "frustration" in defining experimental situations. Psychological Reports, 1978, 43, 1269-1270.

Wagner, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect". Journal of Experimental Psychology, 1959, 57, 130-136.

Wilkins, W. Expectancy of therapeutic gain: An empirical and conceptual critique. Journal of Consulting and Clinical Psychology, 1973, 40, 69-77.

Wilkins, W. Self-fulfilling prophecy: Is there a phenomenon to explain? Psychological Bulletin, 1977, 84(1), 55-56.

Acknowledgements

1. Requests for reprints may be sent to Charles D. Holley, Department of Psychology, Texas Christian University, Fort Worth, Texas 76129.
2. This research was supported, in part, by grant #MDA-903-76-C-0218 from the Defense Advanced Research Projects Agency, Washington, D. C.

Table 1

Means and Standard Deviations on Verbal Ability and Other Indices of Performance

Group		Verbal Ability[a]	Pretest unadjusted	Pretest adjusted[b]	Posttest unadjusted	Posttest adjusted[b]	Posttest adjusted[c]
Treatment-1 (n=28)	$\bar{x}$	56.98	-.31	-.15	.04	.20	.26
	sd	16.32	.87	.82	.88	.83	.68
Treatment-2 (n=29)	$\bar{x}$	58.85	.12	.21	.08	.18	.09
	sd	15.15	1.04	.81	.97	.67	.63
Control (n=42)	$\bar{x}$	66.35	.13	-.05	-.08	-.25	-.23
	sd	14.85	1.00	.80	1.09	.83	.78

[a]Reported in percentages.

[b]Adjusted for verbal ability.

[c]Adjusted for verbal ability and task-specific variance; see text.

Table 2

Correlation Tables for the Four

Exams--Pretest and Posttest

Exam	Pretest Essay	Pretest Short-Answer	Pretest Multiple-Choice	Posttest Essay	Posttest Short-Answer	Posttest Multiple-Choice
Essay						
Short Answer	.61			.47		
Multiple-Choice	.48	.64		.49	.43	
Concept-Cloze	.42	.62	.41	.48	.48	.67

Table 3

Factor Loadings for the Four

Exams--Pretest and Posttest[a]

Exam	Pretest	Posttest
Essay	.78	.76
Short Answer	.90	.74
Multiple-Choice	.79	.83
Concept-Cloze	.75	.84

[a]Each of the factor analyses resulted in a one-factor solution accounting for 65% and 63% of the common variance in the pretest and posttest data, respectively.

Table 4

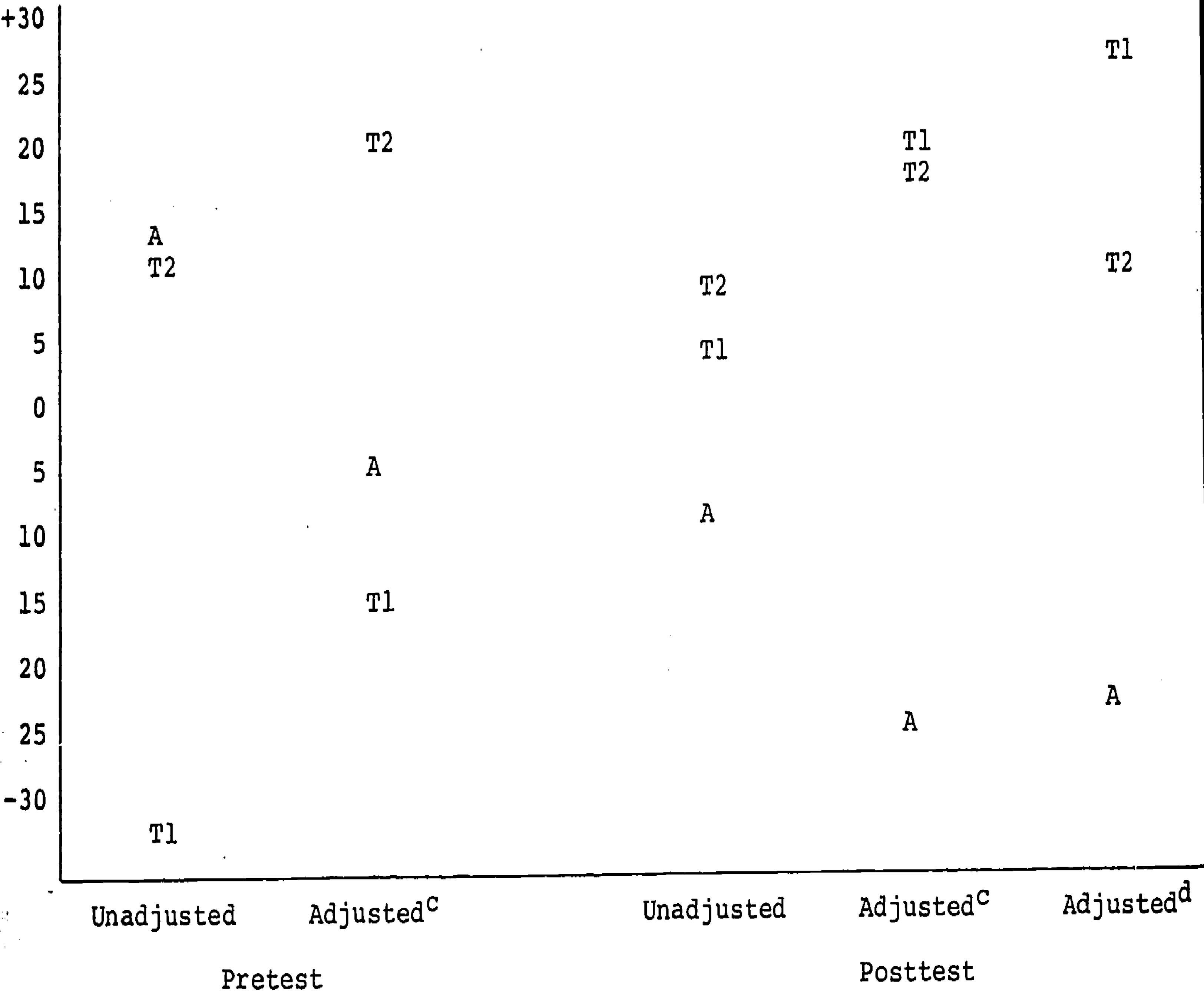
Means and Standard Deviations for Each of the Four Posttests after adjusting for COV-1 (Verbal Ability) and COV-2 (Task-specific Variance)[a]

Group		Essay	Short-Answer	Multiple-Choice	Concept Cloze
		Exam			
Treatment-1	$\overline{X}$	63.23	52.30	62.50	61.36
(n=28)	sd	24.05	15.12	9.13	15.51
Treatment-2	$\overline{X}$	59.35	46.52	63.39	58.31
(n=29)	sd	18.80	14.23	10.26	12.40
Control	$\overline{X}$	50.01	41.23	60.99	54.59
(n=42)	sd	23.94	16.80	10.01	14.74

[a]Unadjusted scores were expressed in percent of maximum score possible on each exam.

Figure Caption

Figure 1. Adjusted and Unadjusted Mean Performances of the Three Groups--Pretest and Posttest[a,b]



[a]See Table 1 for exact values.

[b]Symbols: T1=Treatment Group-1; T2=Treatment Group-2; A=Control Group

[c]Adjusted for verbal ability.

[d]Adjusted for verbal ability and task-specific variance.