DOCUMENT RESUME

ED 381 561 TM 022 906

AUTHOR	Eller, Elizabeth Grace
TITLE	A Comparison of Analysis of Covariance and Blocking Designs in a Gender and Science Context.
PUB DATE	94
NOTE	33p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Nashville, TN, November 9-11, 1994).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Analysis of Covariance; Comparative Analysis; *Distance Education; Elementary Education; Elementary School Students; Pilot Projects; Pretests Posttests; *Research Design; *Science Instruction; Scores; *Sex Differences
IDENTIFIERS	*Blocking Paradigm; *Power (Statistics)

ABSTRACT

Data from a distance education project, Integrated Science 7 (IS7), were used to compare block designs with analysis of covariance (ANCOVA) for their ability to increase statistical power. The IS7 program enables students in grades six through eight to study science via satellite. A sample of 1,802 students from a pilot of the program yielded data for the study. The treatment by blocks designed were formed using the posttest aptitude scores as the dependent variable, the pretest scores as the concomitant or blocking variable, and the gender of the subjects of the independent variable. Two-block, 5-block, and 10-block designs were compared to 2 ANCOVA analyses. With the data used, the 10-block design appeared preferable to ANCOVA, but overall results suggest that there is no one optimal method. The use of ANCOVA versus blocking is dependent on several conditions, as discussed. Eleven tables present data from the analyses. Contains 49 references. (SLD)

A COMPARISON OF ANALYSIS OF COVARIANCE AND BLOCKING DESIGNS IN A GENDER AND SCIENCE CONTEXT

by

ELIZABETH GRACE ELLER, PH.D.

Paper presented at the annual meeting of the
Mid-South Educational Research Association
November 11, 1994

NASHVILLE, TENNESSEE

1994

INTRODUCTION

One task of educational researchers is to determine the most appropriate statistical analysis to apply to their data. A primary goal in experimental design is being able to select the most powerful statistical procedure. A lack of statistical power indicates that a study's examination of significant effects fails to show effects even when they exist (Type II error). The purpose of this study was to evaluate several statistical designs purported to increase power. Specifically, block designs were compared to analysis of covariance designs using data from a distance education project, Integrated Science 7 (IS7).

Influences on Statistical Power

Statistical power is influenced by several factors: "the size of the sample, the degree of variability in the dependent variable, the choice of research design and the method of statistical analysis, the significance level chosen by the researcher, and the magnitude of the treatment effect" (Porter & Raudenbush, 1987, p. 385). Often, the researcher has limited control over these factors. For example, once the population and dependent variables are selected, the researcher cannot always control the amount of variability. Additionally, the treatment effect is not controlled because it is not part of the design itself. Convention restricts the choice of the significance level to .05 or lower. The sample size is dependent upon such variables as cost, time restraints, the availability of subjects, and the number of trained observers. However, the researcher's choice of experimental design and statistical analysis may also influence power.

In determining what is meant by power, Benton (1990) indicated the following:

> The power of a statistical test is the probability, given the Ho (the null hypotheses) is false, of obtaining sample results that will lead to its rejection. . . In other words, a powerful test is one that has a high probability of claiming that a difference exists when it really does (p. 266).

He stated that the power of a test is dependent upon several factors including "(a) the size of true treatment effects, (b) the sample size, (c) the degree of error variance, and (d) the significance level" (p. 2). He explained that one of the first steps in planning an experiment should be the consideration of power. Benton further elucidated, "the power of a test is equal to one minus beta (where beta is the probability of a Type II error), and is determined by the four factors listed previously" (p. 2-3). Therefore, to increase the power of a test, the researcher must decrease the probability of a Type II error. "The smaller the Type II error (beta) the greater the power, and therefore, the greater the sensitivity of the test in detecting statistically significant difference" (Benton, 1990, p. 3).

To decrease the Type II error, the researcher can enlarge the number of observations and/or more precisely control the design of the experiment. Benton (1990) also stated, "The two most common procedures for increasing power are 1) to increase the size of the sample, and 2) to employ an experimental design that provides a more precise estimate of the treatment effects and a smaller error term" (p. 3).

A major source of error variance in behavioral science research can be contributed to individual differences among the subjects. These can be controlled partially by carefully selecting and assigning subjects who are similar in their characteristics (Benton, 1990). However, as in the IS7 pilot, the researchers cannot always carefully select subjects and must rely on statistical measures to reduce error variance. Methods such as blocking and ANCOVA are designs that can reduce error variance and improve estimates of treatment effects.

LITERATURE REVIEW

A popular experimental design in the social sciences and education involves the use of pretests and posttests (Hendrix, Carter, & Hintze, 1978). To add to the informational yield of the pretest and posttest experiment, the researcher can incorporate an additional independent or assigned variable. If properly structured, the use of such a variable can reduce the unexplained variance and increase the design efficiency. Further, as Kennedy and Bush (1985) explained, "building in assigned variables may even enable the experimenter to generalize his or her experimental findings across all levels of the assigned variable" (p. 349). The use of assigned variables is an attempt to reduce background noise or error variance. One specific strategy is the identification of a concomitant (or accompanying) variable that can be statistically correlated with the dependent variable. The concomitant variable is then used for subsequent blocking of observations (Kennedy & Bush, 1985; Lentner, Arnold, & Hinkelmann, 1989).

By blocking, one is classifying or grouping subjects by their scores on the concomitant variable. The primary advantage of blocking is to increase the design efficiency. Kennedy and Bush (1985) explain:

> Efficiency is improved when the investigator's blocking efforts result in greater homogeneity among measures within the levels of the blocking variable than would otherwise occur in a completely randomized arrangement. Because the variance among scores is smaller within factor-level combinations (cells), and because the estimate of population error variance is based upon within-cell variance, it follows that design efficiency can potentially be improved (p. 351).

In this study, the block designs employed were initially termed by Myers (1972) as "treatment-by-blocks." This design is a multifactor approach used by researchers who have a concomitant variable available that correlates with the dependent variable. The concomitant variable is used to improve the efficiency of the design and increase the chance of documenting treatment effects. Another advantage is that this system of analysis allows for the assessment of statistical interaction that cannot be done with one-factor or two-factor block designs (Kennedy & Bush, 1985).

The ANCOVA model offers another technique for reducing error variance and, thus, gaining statistical power. With ANCOVA, information is gathered from each subject on a concomitant variable. This variable, termed the covariable or covariate, is used to decrease error variance within a regression context (Kennedy & Bush, 1985). The dependent variable scores are regressed on covariable scores in the ANCOVA analysis. A one-way analysis of variance (ANOVA) is used on the resulting residual measures that represent the differences between scores expected in the least-square regression line and the actual dependent variable scores. Therefore, the regression model can account for a larger portion of each subject's dependent variable score (Kennedy & Bush, 1985). Kennedy and Bush (1985) stated the following:

> The salutary aspect of this consequence is that the great bulk of explained variability would constitute error variability in a standard one-way ANOVA. In ANCOVA, the explained variability is extracted and an analysis of variance is performed on the residual variability that is partitioned into two components: a) variability due to treatment group's differences, and b) variability which cannot be explained by the factor incorporated in the design of the study (error variability) (p. 396).

Therefore, ANCOVA is generally more powerful than an ANOVA.

Blair and Sawilowsky (1991) agreed that the power of an ANOVA test is increased by the introduction of a covariate and would assist in controlling extraneous variables. As Elashoff (1969) concurred, "the covariance procedure would reduce possible bias in treatment comparisons due to differences in the covariate *x* and increases precessions in the treatment comparisons by reducing variability in criterion scores 'due to' variability in the 'covariate,' *x*" (p. 384).

Review of ANCOVA versus Blocking

A review of the literature indicated conflicting results in the comparison of block designs to ANCOVA designs. Some of the literature supported the use of ANCOVA, while other literature supported block designs. The method to use in choosing between the two was also debated in the literature.

Cochran (1957) offered five advantages to using ANCOVA: its use can increase the precision in randomized experiments, remove effects of confounding variables in observational studies, add to the knowledge of the nature of treatment effects, fit regressions in multiple classifications, and assist in analyzing data when observations are missing. Greenberg (1953) and Gourlay (1953) were also among those who favored the use of ANCOVA. In similar studies, they compared ANCOVA to a matched block technique. Both recommended ANCOVA over the blocking. However, Greenberg added that when treatment groups are less than 10, blocking is preferred.

Keppel (1973) came to different conclusions in that he advocated the use of blocking over ANCOVA. Keppel offered the following advantages of blocking over ANCOVA: (1) the access to the block by treatment interaction, (2) the nonnecessity of variables being linear and (3) the ease in computation. In his discussion on the use of randomized block designs, post-hoc blocking, ANCOVA, and analysis of difference scores, he concluded the following:

> The analysis of covariance can be useful in increasing the precision of an experiment. The statistical model underlying its use is highly restrictive and thus not generally applicable. On almost every count, blocking is the method of choice and post-hoc blocking is a second-best technique to increase precision. The use of covariance should be questioned except in the simple clear cases, while the analysis of difference should generally be avoided (p. 516).

Feldt (1958) did a comparison of an ANCOVA, ANOVA of difference scores, and a blocking technique in which he termed a stratification of a factorial design. He found that the least effective procedure was the ANOVA of difference scores. Feldt indicated that the precision of the ANCOVA or the factorial approach depends upon the population correlation, p of the concomitant variable with the dependent variable. For p values $< .4$, the factorial approach is of equal or greater precision than the ANCOVA. For $p \geq .6$, ANCOVA is the more precise, and for $p < .2$, neither the ANCOVA nor the factorial design is more precise than a completely randomized design.

However, Maxwell, Delaney, and Dill (1984) argued that using the correlation between the dependent variable and concomitant variable when choosing between blocking and ANCOVA is incorrect. Based on a Monte Carlo study conducted by Maxwell et al., it is argued that two other factors should be considered. This includes (1) whether there is a linear relationship between the concomitant variable and dependent variable and (2) whether the scores are available on the concomitant variables for all subjects before subjects are assigned to treatments. If the scores are available on the concomitant variable before subject assignment, Maxwell et al. found blocking to have more power. If the relationship between the concomitant and dependent variables is linear, then ANCOVA is the recommended method. If the relationship is linear, Maxwell et al. recommended a two-way ANOVA or a generalized regression analysis.

Yet, Bonett (1982) argued in a previous article that "ANCOVA can be used when the concomitant variable is not linearly related to the dependent variable assuming the correct form of the regression equation is fitted" (p. 38). Further, Bonett pointed out that blocking is only more powerful than ANCOVA when the optimal number of blocks are used. To determine the optimal number of blocks, Bonett explained that the correlation of the dependent and concomitant variables must be known. However, Bonett explains that the ANCOVA does not require this information to be known. Bonett stated the following:

> The pooled within class regression coefficient is estimated directly from the sample size. To obtain maximum statistical power, the *magnitude* of the concomitant/dependent variable correlation must be known for the block design while the *form* of the relationship must be known for the ANCOVA (p, 37).

In another study, Wu (1993) made a comparison of power in an ANOVA, ANCOVA, and a two-block, four-block, and eight-block design. A main difference between this study and Wu's was he used a Monte Carlo method to obtain his data, and this study used data collected from participants in the IS7 program. By using simulated data, Wu was able to compare the designs under various levels of treatments, various number of subjects, and with different correlation coefficients. He found that when there is no correlation between the dependent variable and the concomitant variable, the one-way ANOVA is the more powerful. The block designs are more powerful when the correlation is low, and the ANCOVA is more powerful with high correlations. However, he found with moderate correlation, that the block design could be as powerful or more powerful than the ANCOVA when the number or treatments and number of subjects per treatment are large. Thus, he recommended a block design when the number of subjects and treatments are large, and an ANCOVA when they are not. Power increased in all five procedures "as the correlation coefficient, the number of treatments, and the number of subjects per treatment increased" (p. 27).

Problems with OVA Techniques

Researchers have examined the problems with using OVA techniques. The use of OVA methods requires that the independent variables be nominally scaled (Prosser, 1990; Thompson, 1988). The obvious problem with this is most variables are scaled higher than nominal (Prosser, 1990). Campbell (1989) found two other difficulties related to the use of OVA methods. These flaws involve a reduction of power against Type II errors and a distortion of the relationships among and distribution of the non-interval predictor variables. Further, Campbell presented three problems associated specifically with ANCOVA. She found that the assumption of reliable measurement of the control variables is often overlooked, that researchers too frequently regard ANCOVA as a "magical" technique for equalizing dissimilar groups, and that the critical homogeneity of regression assumption is regularly ignored.

Malgady and Colon-Malgady (1991) argued that little is gained from using ANCOVA and suggest it is better to do a simple comparison of gain scores. They disputed the implication that ANCOVA is more reliable than an analysis of gain scores. In a comparison of the two designs, they found little advantage in the use of ANCOVA. They indicated that ANCOVA suffers reliability problems in that the pre- and post-test reliabilities decrease as their intercorrelation increases. Malgady and Colon-Malgady made further suggestions:

> Rather than inviting further calamity, such as failure to satisfy the additional and often untenable assumptions of analysis of covariance, researchers might do just as well to perform analysis of variance on simple difference scores when their reliability is adequate. When it is not, analysis of covariance is not likely to help (p. 807).

However, the cautions of Malgady and Colon-Malgady (1991) were previously examined and refuted by Ware and McLean (1978), McLean (1979), and McLean (1989). Analysis of covariance and its use with different experimental designs was investigated by Ware and McLean. It was explained that ANCOVA is used correctly, to increase the accuracy of the design by reducing the unexplained, within-cell variance and incorrectly, to reduce differences among groups by adjusting dependent variable scores for the concomitant variable. The authors determined that when intact groups are used or covariates have low reliabilities, ANCOVA should be used with caution. However, the authors also warn against completely discarding the use of ANCOVA due to its limitations. An example is provided of a two-group experimental design in which groups were " 'equal' with respect to pretest means" (p. 18). When ANOVA was employed, no significant differences were found. Yet when an ANCOVA was employed, significant differences were found because the use of the covariate reduced the unexplained variation within the groups and increased the accuracy of the analysis (Ware & McLean, 1978). Further investigations indicated that ANCOVA will increase the precision of an analysis if there is a significant relationship between the covariate and dependent variable. However, ANCOVA will not adjust for pre-existing, among-group differences (McLean, 1979; McLean, 1989).

Assumption Violations of ANCOVA

The effect of assumption violations with ANCOVA has been discussed in the literature. Bennett (1983) found that if the sample sizes are equal, the power of ANCOVA would not be affected in the presence of heterogeneity of variance. Hamilton (1977) found that alpha levels can be maintained in the presence of heterogeneity of variance, but only when sample sizes are equal. However, Carver (1976) reported contrasting results. He found that a variation in power would be discovered, regardless of sample size, depending on the degree of heterogeneity. Further, Hollingsworth (1980) found that despite the size of the sample or the degree of the heterogeneity, heterogeneity of regression would affect the levels of alpha and power.

McLean (1979) also investigated assumptions associated with ANCOVA. He examined the importance of each assumption, illustrated methods for testing assumptions, and gave suggestions for alternative analyses when assumptions are not met. The assumptions explored by McLean are as follows:

(1) that the cases are assigned randomly to treatment conditions,
(2) that the covariate is independent of the treatment effect,
(3) that the covariate is measured without error (i.e., with perfect reliability),
(4) that the covariate is linearly related to the dependent variable,
(5) that the regression of the dependent variable on the covariate is the same for each group,
(6) that for each level of the covariate, the dependent variable is normally distributed, and

(7) that the variance of the dependent variable at the given value of the covariate is constant across treatment groups and is independent of the covariate (p. 3-4).

In examining the literature, McLean found that, in most cases, ANCOVA is robust to violations of the assumptions, homogeneity of regression and normal distribution. If the assumption that the covariate is related to the dependent variable is not met, the ANCOVA is still valid, but no more powerful than an ANOVA and may be less powerful due to the loss of the degree of freedom. The assumption of homogeneity of variance requires the same amount of consideration with an ANCOVA as it does an ANOVA design. The "most important" assumption to be met, according to McLean, is the independence of the covariate and treatment. McLean explained that the assumption of perfect reliability is not possible in the social sciences and becomes less important if the covariate and treatment become independent. Random assignment is an assumption that is built into the experimental design, and a violation of this will usually evidence itself with a violation in one or more of the other assumptions.

<u>Assumption Violations of Blocking</u>

A block design is an alternative when the homogeneity of variance assumption is violated. Kennedy and Bush (1985) found that if the number of subjects per cell is equal, an abundance of literature exists supporting the use of the block designs. It has been shown that the $\underline{F}$ tests in the block designs are robust against all but extreme violations of the assumption of homogeneity of variance (Kennedy & Bush, 1985).

Block designs can minimize the loss of information by accounting for the effects of nuisance factors that characterize the experimental material (Strange, 1990). The expectation of the block design is that the scores within the blocks are as homogeneous as possible, and scores in different blocks are as heterogeneous as possible. Lentner, Arnold, and Hinkelmann (1989) maintained that when these two expectations are met, a block design will yield "better" inference with respect to treatment effects than a design without blocking. However, if the scores in different blocks are not more heterogeneous than scores within blocks, the influence will not be as good as a non-block design.

Fawcett (1990) did a comparison of designs to examine the benefit of blocking versus the cost of blocking. He compared a Latin Square design, Graeco-Latin square design, and a completely randomized design to a randomized complete block design. The disadvantage of blocking included a loss in degrees of freedom and a stricter rejection of the null hypothesis. Nevertheless, the gain in the reduction of variability was substantial and Fawcett concluded that "the benefits of blocking more than compensated for the cost of blocking" (p. 205).

Optimal Level of Blocks

If a block design is to be employed, the question arises as to how many blocks are the optimal number. Kennedy and Bush (1985) maintained that this question must be answered in regard to the purposes of the research. If the researcher is using the blocking variable because of its intrinsic interest, then the subject matter will influence the number of levels of the blocking variable, but if the blocking variable is used only to reduce the noise of the nuisance factors, the levels of blocks would ideally be the number that maximized efficiency.

A discussion of the relationships that exist between the number of blocks and design efficiency is also presented by Kennedy and Bush (1985) in relation to the treatment-by-blocks design. They indicate that the stronger the relationship between the dependent variable and concomitant variable, the greater potential for design efficiency. Further, a strong relationship between these variables implies that increasing the number of blocks will reduce the average within cell variance (MS within). Reducing the MS within will decrease the noise and increase efficiency. However, for each number of levels of blocks added there will be a decrease in the degrees of freedom, within cells, by one degree of freedom. The effect of this on the design, as well as the effect of the sample size, must be taken into consideration. Kennedy and Bush indicated that increasing the sample size will contribute both to the efficiency of the design and the power of the F test. Therefore, the determination of the number of blocks is influenced by the following:

a) the correlation between the blocking variable and the dependent variable in the population, designated by Rho (ρ), b) the total number of sampling units (N) under

> study, and c) the number of levels associated with the treatment variable. Specific implications for determining the number of blocks in a treatment-by-blocks design are: a) the greater the magnitude of ρ, the greater the number of blocks, b) the greater the N, the greater the number of blocks, and c) the smaller the value of a, the greater the number of blocks (p. 372).

The definition of a is the number of levels that comprise the treatment variable. To apply this information, Feldt (1958, cited in Kennedy & Bush, 1985) created a table that enables the researcher to determine the optimal level of blocks based upon an integration of these relationships. This table is limited, however, to N's of 150 or less. The current study contained a sample size of 1,802.

PROCEDURES

Since the data used in the study were derived from a pilot of a distance education project, Integrated Science 7 (IS7), a brief history of IS7 is provided. Following this, the procedures used in the pilot of IS7 are defined. Next, the subjects in the study are defined followed by a description of the materials used. Finally, the procedure used to conduct the current study is presented.

Description of IS7

The IS7 program is broadcast from The University of Alabama's Center for Communication and Educational Technology. Dr. W. L. Rainey is the Project Director at the center. W. L. Rainey (personal communication, January 13, 1993) explained that The University of Alabama in partnership with the State Department of Education, Alabama Public Television, and corporate sponsors developed the program titled "Integrated Science 7." The ongoing goals of this program include having students (grades six through eight) study science via satellite, making the sciences understandable and enjoyable for all students, and presenting the sciences in a sequenced, well coordinated, and engaging series that draws from biology, chemistry, physics, and earth and space science. The University of Alabama is providing personnel and services to be the broadcast cite where classes originate and are designed.

Integrated Science 7 emphasizes direct, hands-on experience and practical applications so that students can relate the science to their everyday lives.

Procedure of IS7 Pilot

As explained by W. L. Rainey (personal communication, January 13, 1993), the 1991-92 IS7 pilot program was beamed by satellite or shown on public television from studios at The University of Alabama to participating schools in Alabama, Georgia, Mississippi, Florida, and Oklahoma. The broadcast was interactive in approximately half of the programs, meaning that selected students could converse with the IS7 instructor during IS7 class. The daily broadcast occurred, Monday through Friday for 30 minutes and was followed by 30 minutes of instruction in the classroom. The broadcast was conducted by a lead teacher and visiting scientists who typically introduced a science concept and demonstrated that concept on camera. The following half-hour was conducted by the cooperating teachers in participating schools. The cooperating teachers also presented IS7 curriculum material, recapped the broadcast, and guided students in forming, testing, and drawing conclusions to presented hypotheses. The cooperating teachers were provided with tutors based at The University of Alabama that were available by telephone following each broadcast. A computer bulletin board system was provided to each participating school so that the cooperating teachers could pose questions and contribute commentary on the IS7 pilot (Rainey, 1993).

Gender was a variable available in the IS7 pilot. Gender was of interest because of the demonstrated widespread differences among males and females in science achievement and attitudes, with males favored (Steinkamp, 1982). This was also demonstrated by a nationwide assessment conducted by The National Assessment of Educational Progress which revealed achievement differences in science favoring males, especially in the physical sciences for 9-, 13-, and 17-year old students (Crawley & Coe, 1990). It has been well documented that females are underrepresented in the science fields (Dix, 1987; Olstad, 1981; Reat, 1981).

Females, when compared to males, often avoid advanced science classes, are less motivated, and fail to see the usefulness of such classes (Khoury & Voss, 1985).

Due to the documented difficulties that students have in science (Nolen & Haladyna, 1990; O'Malley & Scanlon, 1990; Trefil, 1991), an obvious choice of investigation in the IS7 pilot was science aptitude. It has been shown that as students progress through school in science, achievement levels steadily decline, as do attitudes (Cannon & Simpson, 1985), and this is true across grade levels (Simpson & Oliver, 1985). For comparison purposes, the pre- and post-test aptitude scale was administered to all the seventh-graders participating in the IS7 pilot. This comparison was made due to the demonstrated positive relationships between attitudes and achievement in science (Gardner, 1975; Omerod & Duckworth, 1975; Ward, 1976). Research has indicated that science attitude scales can be used to predict science related behavior (Shrigley, 1990). Crawley and Coe (1990) found attitude to be one of the sole predictors of whether eighth grade students chose to enroll in an elective high school science course.

Subjects

The subjects for the study were participants in the IS7 pilot during the 1991-92 school year. W. L. Rainey (personal communication, January 13, 1993) indicated that a nationwide advertisement was posted in March of 1990 describing the IS7 program and informing schools how to participate. By May of 1991, school systems from Alabama, Georgia, Mississippi, Florida, and Oklahoma replied and paid fees to receive the IS7 broadcast and materials. Subjects for this study included the students in seventh-grade science classes whose schools participated in the IS7 pilot program. All IS7 pilot students were used as subjects; however, some data were unreadable or not returned and could not be used. The total population size was 2,414. From this, subjects were discarded who had not returned all four scores from the pre- and post-testing of the attitude and aptitude scales or who had not recorded their gender. This yielded a sample size of 1,802 for purposes of this study.

Materials

A science ability scale and science aptitude scale was used for the pre- and post-testing. The aptitude scale used was the instrument, Processes of Science (Yager, Blunt, & Ajam, 1990). The attitude scale used was the instrument, Attitudes, Preferences, and Understanding for Grades 4 through 12 (NEAP, 1980).

The pretest aptitude and attitude scales were mailed to participating schools from The University of Alabama in August of 1991 and were administered in September of 1991. The cooperating teachers administered the scales to students and returned the forms to the University in September of 1991. Cooperating teachers were mailed the posttest ability and attitude scales in March on 1992. The cooperating teachers administered these scales in April of 1992 and returned the materials to The University of Alabama in April and May of 1992 (W. L. Rainey, personal communication, January 13, 1993). The data from the pre- and post-testing were stored and analyzed using The University of Alabama's IBM 3090-400E mainframe. The software packages used included the CMS operating system with SAS version 6.07 (SAS, 1985).

Procedure of Current Study

The treatment-by-blocks designs were formed using the posttest aptitude scores as the dependent variable, the pretest aptitude scores as the concomitant or blocking variable, and the gender of the subjects as the independent variable. Three levels of blocks were employed: a two-block, five-block, and ten-block design. These same designs were replicated, but the pretest attitude scores were used as the concomitant variable. The results from these analyses were compared to each other and to similar ANCOVA designs.

Two ANCOVA analyses were also used and compared as to their effectiveness in increasing significance and power. As stated, the ANCOVA designs were also compared to the block designs. In the first ANCOVA analysis, the dependent variable was the posttest data from the aptitude scale, the independent variable was gender, and the covariate was the pretest data from the aptitude scale. The variables used in the second analysis were identical, except for the covariate, which was changed to the pretest scores from the attitude scale. Therefore, the ANCOVA designs made use of the same variables as did the block designs.

A comparison of the sensitivity of the designs used were made. Sensitivity was defined by the resulting *p* values and power values from the ANCOVA and treatment-by-blocks designs.

Data Analysis

A SAS program was written for each analysis and the *p* values were generated. The SAS code used for each analysis is presented in Appendix A. Power was determined using charts derived by Pearson and Hartley (1951, cited in Kirk, 1968) that are based on a procedure by Tang (1938, cited in Kirk, 1968) for calculating power. Power is calculated by entering a parameter ϕ in the charts.

Research Questions

To answer the research questions of the present study, an examination was done of the resulting values for *p* and power. The first research question (RQ) addressed whether there are differences in *p* values between the ANCOVA designs and the blocking designs. This comparison is represented by the following:

RQ_1: $p_1 = p_2; p_1 = p_3; p_1 = p_4;$

$p_5 = p_6; p_5 = p_7; p_5 = p_8$

The second research question was identical to the first except comparing power values instead of *p* values. This comparison is represented in the same manner, but substituting power values for *p* values.

RQ_2: $1-\beta_1 = 1-\beta_2; 1-\beta_1 = 1-\beta_3; 1-\beta_1 = 1-\beta_4;$

$1-\beta_5 = 1-\beta_6; 1-\beta_5 = 1-\beta_7; 1-\beta_5 = 1-\beta_8$

The third research question was that there is no difference in *p* values among the block designs, and the forth was that there is no difference in the power values. These are represented on the following page:

RQ_3: $p_2 = p_3 = p_4; p_6 = p_7 = p_8$

RQ_4: $1 - \beta_2 = 1 - \beta_3 = 1 - \beta_4; 1 - \beta_6 = 1 - \beta_7 = 1 - \beta_8$

The fifth and sixth research questions compared the *p* and power values, respectively, stating that there is no difference between the aptitude and attitude values in the ANCOVA design in terms of *p* and power. They are represented in the succeeding relationships:

RQ_5: $p_1 = p_5$

RQ_6: $1 - \beta_1 = 1 - \beta_5$

Research question seven asked whether a difference exists between the *p* values in each of the block designs using the aptitude scores as the blocking factor and the *p* values of each of the block designs using the attitude scores as the blocking factor. The last research question addressed a comparison of the power values of each of the block designs using aptitude to each of the power values of the block designs using attitude.

RQ_7: $p_2 = p_6; p_3 = p_7; p_4 = p_8$

RQ_8: $1 - \beta_2 = 1 - \beta_6; 1 - \beta_3 = 1 - \beta_7; 1 - \beta_4 = 1 - \beta_8$

RESULTS

The results from the ANCOVA designs are represented in Tables 2 and 3. The degrees of freedom (DF), sum of squares (SS), mean squares (MS), F values, and *p* values are presented. Table 2 gives the results from the ANCOVA using the pretest aptitude scores as the covariate, and Table 3 shows the results of the ANCOVA using the pretest attitude scores as the covariate.

Table 2

ANCOVA Summary Table Using Aptitude as Covariate

Source	DF	SS	MS	F Value	Pr > F
Model	2	27414.94	13707.47	757.67	.0001
Aptitude	1	27393.22	27393.22	1514.14	.0001
Gender	1	14.86	14.86	.82	.3649
Error	1799	32546.73	8.09		
Corrected Total	1801	59961.67			

In this analysis, the overall model's *p* value was less than .01; therefore, it was significant. The r value, which is the correlation between the covariate and the dependent variable, was .46. The variable, gender, was not significant ($p > .01$), indicating that the scores of the males and females did not differ significantly.

Table 3

ANCOVA Summary Table Using Attitude as Covariate

Source	DF	SS	MS	F Value	Pr > F
Model	2	22.10	11.05	.33	.7178
Aptitude	1	.38	.38	.01	.9150
Gender	1	21.63	21.63	.65	.4205
Error	1799	59939.57	33.32		
Corrected Total	1801	59961.67			

In this analysis, the overall model was not significant ($p > .01$). The r value was .002. The variable, gender, was not significant ($p > .01$), indicating that the scores of the males and females did not differ significantly.

Assumptions Tested

The assumption of independence of the covariate and treatment was tested for both ANCOVAs. In both designs, there was not a significant relationship ($p > .05$); this suggests that there was no significant relationship between gender and the covariates. The assumption of homogeneity of regression was also tested and neither ANCOVAs indicated significance ($p > .05$), implying that there were no interactions between the covariates and the variable gender. Thus, the essential assumptions for ANCOVA were met.

Tables 4, 5, and 6 are the results from the two-block, five-block, and ten-block designs, respectively, that used the pretest aptitude scores as the concomitant variable. Tables 7, 8, and 9 are the findings from the two-block, five-block, and ten-block designs, respectively, which used the pretest attitude scores as the concomitant variable. The blocking variable is represented by Blocks.

Table 4

Two-Block Design Summary Table Blocking on Aptitude

Source	DF	SS	MS	F Value	Pr > F
Model	2	20423.54	10211.77	4647.64	.0001
Blocks	1	20401.82	20401.82	928.29	.0001
Gender	1	1.01	1.01	.05	.8306
Error	1799	39538.13	21.98		
Corrected Total	1801	59961.67			

Table 5

Five-Block Design Summary Table Blocking on Aptitude

Source	DF	SS	MS	F Value	Pr > F
Model	5	26334.82	5266.96	281.31	.0001
Blocks	4	26313.10	6578.275	351.34	.0001
Gender	1	15.40	5.01	.82	.3645
Error	1796	33626.85	18.72		
Corrected Total	1801	59961.67			

Table 6

Ten-Block Design Summary Table Blocking on Aptitude

Source	DF	SS	MS	F Value	Pr > F
Model	10	27497.91	2749.79	151.70	.0001
Blocks	9	27476.19	3052.91	168.43	.0001
Gender	1	20.39	20.39	1.12	.2898
Error	1791	32463.76	18.13		
Corrected Total	1801	59961.67			

In the previous three block designs, the overall models were significant ($p < .01$). The concomitant variable was significant in all three designs ($p < .01$). The eta squared values were .34 for the two-block design, .44 for the five-block design, and .46 for the ten-block design. The variable, gender, was not significant in any of the designs, indicating that the scores of males and females did not differ.

Table 7

Two-Block Design Summary Table Blocking on Attitude

Source	DF	SS	MS	F Value	Pr > F
Model	2	22.36	11.18	.34	.7149
Blocks	1	.65	.65	.02	.8893
Gender	1	21.83	21.83	.66	.4184
Error	1799	59939.31	33.32		
Corrected Total	1801	59961.67			

Table 8

Five-Block Design Summary Table Blocking on Attitude

Source	DF	SS	MS	F Value	Pr > F
Model	5	253.15	50.63	1.52	.1794
Blocks	4	231.43	57.86	1.74	.1385
Gender	1	20.75	20.75	.62	.4296
Error	1796	59708.53	33.24		
Corrected Total	1801	59961.67			

Table 9

Ten-Block Design Summary Table Blocking on Attitude

Source	DF	SS	MS	F Value	Pr > F
Model	10	454.40	45.44	1.37	.1893
Blocks	9	432.68	48.08	1.45	.1626
Gender	1	21.26	21.26	.64	.4238
Error	1791	59507.27	33.23		
Corrected Total	1801	59961.67			

In the block designs using attitude, the overall models were not significant ($p > .01$), and the concomitant variables were not significant ($p > .01$). The eta squared values were .000009 for the two-block design, .0036 for the five-block design, and .0064 for the ten-block design. The variable, gender, was not significant in any of the designs, indicating that the scores of the males and females did not differ. The table on the following page summarizes the sizes of the population (n), means, and standard deviations (SD) of the variables used in the study.

Table 10

Descriptive Statistics for Study Variables

Variable	n	Mean	SD
Dependent (Post Aptitude)			
Male	830	20.2	5.99
Female	972	20.4	5.58
Covariate One (Pre Aptitude)			
Male	830	18.5	5.97
Female	972	18.6	5.61
Covariate Two (Pre Attitude)			
Male	830	2.6	.52
Female	972	2.6	.49

Comparison of Analysis Procedures

To examine the research questions, the *p* values and computed power values were entered into the Table 11.

Table 11

Comparison of Procedures

Analysis Method	Covariate or Blocking Variable: Aptitude		Attitude	
	p	Power	*p*	Power
ANCOVA	.3649	$\geq$ .9999	.4205	$\leq$ .32
Two-Block	.8306	$\geq$ .9999	.4184	$\leq$ .32
Five-Block	.3645	$\geq$ .9999	.4296	$\leq$ .40
Ten-Block	.2890	$\geq$ .9999	.4238	$\leq$ .40

To explore the first research question, the *p* value of the ANCOVA design using aptitude was compared to each of the block designs using aptitude, and the *p* value of the ANCOVA design using attitude was compared to each of the block designs using attitude. The ANCOVA design using aptitude had a slightly higher *p* value than the five-block design using aptitude, and a higher *p* value than the ten-block design using aptitude. However, the ANCOVA displayed a lower *p* value than the two-block design using aptitude. When comparing the *p* values of the ANCOVA design using attitude to the *p* values of the respective blocking designs, the *p* value differences were not noteworthy.

When considering the second research question, an examination of the power values was done. The power values of the designs using aptitude are all extremely large. This is due to the large size of the sample which inflated the values for power. The power values of the designs using attitude were also similar due to the extremely small effect sizes. This can be attributed to the small overall differences between the sexes.

To answer the third and fourth research questions, the *p* values and power values of each blocking design using the same blocking variable were compared to one another. The five-block and ten-block designs using aptitude had lower *p* values than the two-block design using aptitude. The ten-block design had a lower *p* value than the five-block design. Therefore, the larger the number of blocks, the lower the *p* value of the blocking designs using aptitude. The *p* values of the block designs using attitude did not differ by more than .01. As before, the power values did not differ for the designs using aptitude due to the large sample size, or for the designs using attitude due to the small differences in gender.

Research questions five and six compared the *p* and power differences of the ANCOVA design using aptitude as the covariate to the ANCOVA designs using attitude as the covariate. The *p* value of the ANCOVA design using aptitude was lower than the *p* value of the ANCOVA design using attitude. In addition, the power value of the ANCOVA design using aptitude was much greater than the power value of the ANCOVA design using attitude.

Answering the last two research questions involved a comparison of the block designs using aptitude as the concomitant variable to the block designs using attitude. In comparing the two-block designs, the design using aptitude indicated a higher *p* value than the two-block design using attitude. This was the only instance in which a design using aptitude had a higher *p* value than a design using attitude, and a power comparison of these two designs did not favor attitude. The power of the two-block design using aptitude was much greater than the power of the two-block design using attitude.

The five-block design using aptitude had a lower *p* value than the five-block design using attitude. The *p* value of the ten-block design using attitude was almost twice as large as the ten-block design using aptitude. Both the five-block and ten-block designs using aptitude indicated much higher power values than the five-block and ten-block designs using attitude.

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Comparison of Using Aptitude versus Attitude

There was evidence to indicate that the pretest scores of the aptitude scale were better covariate and concomitant variables than the pretest scores from the attitude scale with the Integrated Science 7 (IS7) data. The overall models of the designs using aptitude were significant which implies that the models were useful. However, the overall models of the designs using attitude were not significant, demonstrating that the models were not useful. The covariate was significant in all the designs using aptitude which indicated that the pretest aptitude scores were an effective predictor of the aptitude posttest scores with the ANCOVA and useful as a blocking factor for the blocking designs. The covariate was not significant in any of the designs using attitude, suggesting that the pretest attitude scores were not an effective predictor of the aptitude posttest scores with the ANCOVA and was not useful as a blocking factor with the blocking designs. The designs using the aptitude scores all indicated much greater power values than the designs using the attitude scores, and in all but one comparison (the two-block) the *p* values for the designs using aptitude were less than those using attitude. Lastly, the r value of the ANCOVA design using aptitude was much greater than the r value of the ANCOVA design using attitude.

Comparison of Designs Using Aptitude

Given that the designs using aptitude were superior, a comparison of these designs was considered. The power values of the four designs using aptitude were all greater than .9999; therefore, distinctions could not be made using power. However, comparisons of *p* values could be made.

Comparison of Block Designs Using Aptitude

A comparison of the *p* values of the block designs using aptitude was considered to determine the optimal number of blocks among the blocking designs. The *p* values for the five- and ten-block designs were less than the *p* value for the two-block design. In comparing the ten-block design to the five-block design, the ten-block design had the lower *p* value. Therefore, among the blocking designs considered, the optimal blocking number with the IS7 data appears to be 10.

In comparing the *p* value of the ANCOVA design using aptitude to the blocks designs using aptitude, the ANCOVA indicated a lower *p* value than did the two-block design. There was not a noteworthy difference between the *p* value of the ANCOVA and the *p* value of the five-block design in that the values only differed by .0004. However, the loss of degrees of freedom gave the five-block the advantage. The five-block design maintained a low *p* value despite losing more degrees of freedom than the ANCOVA.

The ten-block design indicated a lower *p* value than the ANCOVA, with the difference between the two *p* values being .0751. The lower *p* value was accomplished by the ten-block design even though nine degrees of freedom were lost. Therefore, with the IS7 data, it appears more advantageous to use the ten-block design in comparison to ANCOVA.

The analysis that appeared to be the most useful with the IS7 data was the ten-block design using the aptitude as the concomitant variable, but to generalize this to all data would be in error. Unlike Greenberg (1953) and Gourlay (1953) who indicated a preference towards ANCOVA or Keppel (1973) who revealed a bias towards blocking, there does not appear to be one optimal method. More evidence is available to suggest that the use of ANCOVA versus blocking is dependent upon several conditions (Maxwell, Delaney, & Dill, 1984; Wu, 1993). Cox (1957) and Feldt (1958) suggested that a determining factor should be the correlation between the dependent and concomitant variables. Wu (1993) found similar results suggesting that a block design is the better choice when the correlation coefficient is low, and an ANCOVA is favored the when the correlation is high. However, Wu also found ANOVA to be the preferred method when no correlation exists. Wu added that if a moderate correlation exists, block designs should be selected when there are large numbers of treatments and subjects per treatment. The present study supported Wu's findings, since the ten-block using aptitude was the preferred method and the this design contained large numbers of subjects per treatment and a moderate correlation. However, Cox stated that a block design is preferred if the correlation is less than .4; Feldt said neither design is preferred if the correlation is less than .2; and Wu said ANOVA is the better design if there is no correlation. The correlation between the covariate and the dependent variable was .02 when attitude was used as the covariate. None of the designs indicated an advantage, a result supportive of Feldt and Wu but not Cox.

Maxwell, Delaney, and Dill (1984) also offered guidance in choosing between ANCOVA and blocking. They found that if the scores on the concomitant variable are available before subject assignment to treatments, then the block design is more powerful. Maxwell et al. also found that if the relationship between the concomitant variable and dependent variable is linear, ANCOVA is the more powerful. If the relationship is not linear, Maxwell et al. suggested a two-way ANOVA or a generalized regression analysis. This coincides with McLean (1979), who stated that when the dependent and concomitant variables are not related, ANCOVA is no more powerful than ANOVA and may be less due the loss of the degrees of freedom.

Discussion of Assumption Violations

The two assumptions measured in the ANCOVA designs were homogeneity of regression and the independence of the independent variable and the covariable. Testing of these assumptions gave no reason to believe that they were not met in the present study ($p > .05$). However, since random sampling was not part of the design of the IS7 pilot, the assumption of random sampling was not met in the present study. The effect of this on the ANCOVA analyses is not known. With other assumptions, McLean (1979) found ANCOVA to be robust against most violations of the normality. He also stated that the assumption of homogeneity of variance "requires about the same amount of concern in ANCOVA as it does in the analysis of variance" (p. 7). Moreover, McLean pointed out that the assumption of perfect reliability is unattainable in the social sciences and is of less importance when the assumption of independence of treatment and covariate is met.

Concerning block designs, Kennedy and Bush (1985) reported that block designs appear to be robust against most violations of homogeneity of variance. The block design also does not require that the relationship between the concomitant variable and dependent variable be linear (Keppel, 1973) or that a high correlation exist between the dependent variable and concomitant variable (Cox, 1957; Feldt, 1958).

Suggestions for Further Research

Since the IS7 data meet the assumptions of homogeneity of regression and independence of covariate and independent variables, researchers could look at the effects when assumptions are not met to determine if blocking would be a better choice. The assumption of random sampling is not met in this study and could have impacted the results. Further research could examine data obtained through random sampling to determine its impact. The data also did not indicate a significant difference between the genders. A gender difference could be forced by a researcher, and then a comparison made between ANCOVA and blocking.

Arguments were presented that the ANCOVA using aptitude was the best choice for the IS7 data when compared to the other designs. However, none of the designs, including the ANCOVA, indicated a very high r value.

The present study indicated power values of greater than .99 for the models using aptitude as the concomitant or covariate. The inflation of the power values is due to the large sample size. Other research could examine ANCOVA and block designs using smaller sample sizes to obtain lower power values.

The present study indicated that the ten-block design using the pretest aptitude scores as the concomitant variable had the lowest *p* value. It would be of interest to determine if increasing the number of blocks by more than ten would have continued to produce lower values of *p*. Further research could examine increasing the number of blocks until each subject's score became its own block to assist in determining the optimal number of blocks.

The results of this study are limited by the IS7 pilot data but give insight into the uses of ANCOVA versus blocking designs. These results can be used to support past research and spur the interest of future researchers. As Wu (1993) stated in his work, "The greatest contribution of this study might not be the specific results reported, but the potential for examining other situations" (p. 29).

REFERENCES

Bennett, R. P. (1983, January). Comparison of two analysis techniques for the pretest/posttest control group experimental design. Paper presented at the meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 226 069)

Benton, R. L. (1990, January). The statistical power of planned comparisons. Paper presented at the meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Reproduction Service No. Ed 314 494)

Blair, R. C., & Sawilowsky, S. S. (1991, April). Confounding covariates in nonrandomized studies. Paper presented at the meeting of the Midwestern Association, Chicago, IL. (ERIC Document Reproduction Services No. ED 336 412)

Bonett, D. G. (1982). On post-hoc blocking. Educational and Psychological Measurement, 42, 35-38.

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

Cannon, R., & Simpson, R. (1985). Relationships among attitude, motivation, and achievement of ability grouped, seventh-grade, life science science students. Science Education, 69(2), 121-138.

Carver, D. W. (1976). The effects of nonhomogeneous regression slopes upon the power of the ANCOVA (Doctoral dissertation, The University of Texas at Austin, 1976). Dissertation Abstracts International, 12, 7707a.

Crawley, F. E., & Coe, A. S. (1990). Determinants of middle school students; intention to enroll in a high school science course: An application to the theory of reasoned action. Journal of Research in Science Teaching, 27(5), 461-476.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. Biometrics, 13, 261-281.

Cox, D. R. (1957). The use of a concomitant variable in selecting an experimental design. Biometrika, 44, 150-158.

Dix, L. C. (1987). Women: Their underrepresentation and career differentials in science and engineering proceedings of a Workshop. Washington D. C.: National Academy Press, Office of Scientific and Engineering Personnel. (ERIC Document Reproduction Service No. ED 285 752)

Elashoff, J. D. (1969). Analysis of Covariance: A delicate instrument. American Educational Research Journal, 6, 383-401.

Fawcett, R. F. (1990). The alignment method for displaying and analyzing treatments in blocking designs. The American Statistician, 44(3), 204-209.

Feldt, L. S. (1958). A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika, 23(4), 335-352.

Gardner, P. L. (1975). Attitudes to science: A review. Studies in Science Education, 2, 1-41.

Gourlay, N. (1953). Covariance analysis and its applications in psychological research. British Journal of Statistical Psychology, 6, 25-34.

Greenberg, B. G. (1953). The use of analysis of covariance and balancing in analytical surveys. American Journal of Public Health, 43, 692-699.

Hamilton, B. L. (1977). An empirical investigation of the effects of heterogeneous regression slopes in analysis of covariance. Educational and Psychological Measurement, 37, 587-601.

Hendrix, L. J., Carter, M. W., & Hintze, J. L. (1978). A comparison of five statistical methods for analyzing pretest-posttest designs. Journal of Experimental Education, 47, 96-102.

Hollingsworth, H. H. (1980). An analytical investigation of the effects of heterogeneous regression slopes in analysis of covariance. Educational and Psychological Measurements, 40, 611-618.

Kennedy, J. J., & Bush, A. J. (1985). An introduction to the design and the analysis of experiments in behavioral research. Lanham, MD: University Press of America, Inc.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Khoury, G., & Voss, V. (1985, April). Factors influencing high school students' science enrollments patterns: Academic abilities, parental influences, and attitudes toward science. Paper presented at the meeting of the National Association for Research in Science Teaching, French Lick Springs, IN. (ERIC Document Reproduction Service No. ED 254 408)

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

Lentner, M., Arnold, J. C., & Hinkelmann, K. (1989). The efficiency of blocking: How to use MS(blocks)/MS(error) correctly. The American Statistician, 43(2), 106-108.

Malgady, R. G., & Colon-Malgady, G. (1991). Comparing the reliability of difference scores and residuals in analysis of covariance. Educational and Psychological Measurement, 54(4), 803-807.

Maxwell, S. E., Delaney, H. D., & Dill, C. A. (1984). Another look at ANCOVA versus blocking. Psychological Bulletin, 95(1), 136-147.

McLean, J. E. (1979, November). The care and feeding of ANCOVA. Paper presented at the meeting of Mid-South EducationalResearch Association, Little Rock, AR. (ERIC Document Reproduction Service No. ED 191 867)

McLean, J. E. (1989, November). ANCOVA: A review, update, and extension. Paper presented at the meeting of Mid-South Educational Research Association, Little Rock, AR.

Myers, J. L. (1972). Fundamentals of experimental design (2nd ed.). Boston: Allyn and Bacon.

NEAP. (1980). Student preferences and understandings. Iowa City, IA: University of Iowa Science Education Center.

Nolen, S. B., & Haladyna, T. M. (1990). Motivation and studying in high school science. Journal of Research in Science Teaching, 27(2), 115-126.

Olstad, R. G. (1981, March). Inhibitors to achievement in science and mathematics by ethnic minorities. Project description at The University of Washington, Seattle, Washington. (ERIC Document Reproduction Service No. ED 223 404)

O'Malley, C. E., & Scanlon, E. (1990). Computer-supported collaborative learning: Problem solving and distance education. Computers Education, 15(1-3), 127-136.

Omerod, M. B., & Duckworth, D. (1975). Pupils' attitudes to science. London: National Foundation for Education Research.

Porter, A. C., & Raudenbush, S. W. (1987). Analysis of covariance: Its model and use in psychological research. Journal of Counseling Psychology, 34(4), 383-392.

Prosser, B. (1990, January). Beware the dangers of discarding variance. Paper Presented at the meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 314 496)

Reat, J. H. (1981, December). Girls and science technology (GASAT), Volume 1. Proceedings at the First International GASAT Conference, Eindhoven University of Technology, The Netherlands. (ERIC Reproduction Service No. ED 262 995)

SAS Institute, Inc. (1985). SAS user's guide: Statistics, version 6 edition, Cary, NC, Author.

Shrigley, R. L. (1990). Attitude and behaviors are correlates. Journal of Research in Science Teaching, 27(2), 97-113.

Simpson, R., & Oliver, S. (1985). Attitude toward science and achievement motivation profiles of males and females science students in grades six through ten. Science Education, 69(4), 511-526.

Steinkamp, M. (1982). Sex-related differences in attitude towards science: A quantitative synthesis. Paper presented at the Annual Meeting of the American Education Research Association, New York, NY. (ERIC Document Reproduction Service No. ED 220 285)

Strange, R. S. (1990). Introduction to experimental design for chemists. Journal of Chemical Education, 67, 113-115.

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

Trefil, J. (1991). Science Literacy: The enemy is us. Science, 251, 266-267

Ward, W. H., Jr. (1976). A test of the association of class size to students' attitudes toward science. Journal of Research in Science Teaching, 9, 147-156.

Ware, W. B., & McLean, J. E. (1978). A note on overlimiting the use of analysis of covariance. CEDR Quarterly, 11(4), 16-18, 21.

Wu, Y. (1993). To block or covary a concomitant variable: Which is more powerful. Unpublished doctoral dissertation, The University of Alabama, Tuscaloosa.

Yager, R. E., Blunt, S. M., & Ajam, M. (Eds.). (1990). Iowa assessment package for evaluation in five domains of science education. Iowa City, IA: University of Iowa Science Education Center.