ED 306 299 TM 013 179

AUTHOR	Eason, Sandra H.; Daniel, Larry G.
TITLE	Trends and Methodological Practices in Several Cohorts of Dissertations.
PUB DATE	Jan 89
NOTE	33p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 26, 1989).
PUB TYPE	Information Analyses (070) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Analysis of Variance; Cohort Analysis; *Construct Validity; *Doctoral Dissertations; Educational Research; Higher Education; Literature Reviews; Methods Research; Multivariate Analysis; Reliability; *Research Methodology; Sample Size; Sampling; *Statistical Analysis; Surveys; *Trend Analysis
IDENTIFIERS	External Validity; Internal Validity

ABSTRACT

A framework for understanding methodological practices from the perspectives of internal validity, external validity, statistical control validity, and construct validity is presented. One hundred doctoral dissertations completed between 1980 and 1988 at a single urban public university were analyzed for various methodological practices and types of statistical techniques used. Techniques were further coded into categories of univariate or multivariate designs and subcategories within these categories. In all, 201 techniques were discovered across papers, which were generated by the Departments of Educational Leadership and Foundation, Curriculum and Instruction, and Special Education of the university. Demographics, size, and other information about sampling techniques were also assessed. Randomization and convenience were the most typical sample selection techniques. Means of establishing reliability of the dependent variables and whether surveying was involved for each study were also determined. Additional features concerning the discussion of results in the dissertations were coded as well. The analyses indicate that the instance of use of specific statistical techniques has changed little during the last 9 years. Three data tables are included. (TJH)

TRENDS AND METHODOLOGICAL PRACTICES

IN SEVERAL COHORTS OF DISSERTATIONS

Sandra H. Eason — Larry G. Daniel

University of New Orleans 70148

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, January 26, 1989.

ABSTRACT

The quality of methodological practices reflects on the quality of research findings. A framework for understanding methodological practices from the perspectives of internal validity, external validity, statistical control validity, and construct validity is presented. One hundred dissertations completed during the 1980's were analyzed for various methodological practices and types of statistical techniques. The analyses indicate that the instance of use of specific statistical techniques has changed little during the last nine years.

TRENDS AND METHODOLOGICAL PRACTICES

IN SEVERAL COHORTS OF DISSERTATIONS

Generally speaking, the quality of research is only as good as the quality of the methodological practices employed by the researcher. Failure to use a representative sample, for instance, may cause serious questions about what may appear to be theoretically and perhaps even statistically noteworthy results. Similarly, reliance upon instruments which have not been appropriately validated as measures of variables under consideration may confound or perhaps even invalidate a study's findings. Likewise, use of statistical techniques which do not honor the true relationships among the variables under study may cause the researcher to draw inaccurate conclusions about causality or correlation among variables. Considering the vast array of factors falling under the umbrella of "methodology," it is imperative that researchers take caution in teaching and practicing appropriate methodological techniques.

The purpose of the present study was to investigate the types of statistical techniques and various methodological trends employed in doctoral dissertations over a nine year period at an urban public university. A general theoretical framework based on four types of methodological validity is presented, and concerns relevant to each type of validity are discussed. One hundred doctoral dissertations by education majors were reviewed. Several variables were noted in the review. The primary variable investigated was the type of research technique employed by the researcher. Techniques were coded into categories of univariate or multivariate designs. Within each of these broad categories, the techniques were further coded into subgroups with similar characteristics.

A Framework for Understanding Methodological Validity

Cook and Campbell (1979) and Mitchell (1985) have addressed the issue of quality of research methodology from the perspective of a given study's validity. In the most general sense, study validity may be conceptualized as being either "internal" or "external." Cook and Campbell (1979, p. 37) made the following distinction between these two types of validity:

> Internal validity refers to the approximate validity with which we infer that a relationship between two variables is causal [or correlational] or that the absence of a relationship implies the absence of cause [or association]. External validity refers to the approximate validity with which we can infer that the presumed causal [or correlational] relationship can be generalized to and across alternate measures of the . . .[variables] and across different types of persons, settings, and times.

Two additional types of validity are also addressed in the literature (Cook & Campbell, 1976, 1979; Mitchell, 1985)--construct validity (the degree to which a measure of a construct adequately measures the construct) and statistical conclusion validity (the relative stability of statistical results resulting from minimization of random error variance and appropriate use of statistical tests). Although other types of validity could also be considered, these four broad categories subsume most of the major issues pertaining to the quality of research methodology. Each of these four varieties of validity will merit further discussion.

Internal Validity

Internal validity is concerned with issues relative to relationships between or among the variables under consideration. Issues of causality are usually involved when considering validity threats of this variety. The researcher may conclude that one variable causes another based upon the results of a given statistical test while, in actuality, there may be a third intervening variable which is actually the cause of the statistical differences among cases.

Internal validity threats may include history, maturation of subjects during an intervention, and mortality among members of the sample. Other internal validity threats are caused by subjects' attitudes about being involved in the study. Members of the experimental group in a given study, for instance, may perform better than those in the control group as the result of feeling special about the extra attention given to them during their participation in the study (the so-called "Hawthorne effect"). On the other hand, if placed in the control group, participants may work extra hard to outdo the performance of the members of the control group (the so-called "John Henry" effect). However, more often than not, internal validity threats are related to between-group differences among the groups included in the sample.

Frequently the researcher will not (and possibly cannot) take all of the steps necessary to ensure the equivalence at the outset of the study of members in both the control and experimental groups with respect to the dependent variable. When subjects are not randomly assigned to conditions, the possibility of between-group differences increases, and threats to the internal validity of the study are likely to occur. Problems with implementing true random assignment are common in educational experiments. In many instances, educational researchers must use intact groups (e.g., established classrooms within schools), and therefore may face the problem of non-equivalence of groups.

In an attempt to correct for non-equivalence of groups, and thereby ensure the internal validity of a study, many researchers will rely upon various statistical controls such as covariate adjustments of posttest scores. Covariate adjustments may be appropriate in experimental studies when true random assignment is used. However, as previously noted, educational experiments must frequently rely upon the use of intact, convenient groups of subjects. Statistical control methods (e.g., analysis of covariance) assume homogeneity of regression across all treatment groups; that is, in adjusting dependent variables, group membership is completely ignored. ANCOVA is <u>not</u> robust to the violation of this assumption.

In cases in which treatment groups have inherent differences, homogeneity of regression assumptions cannot be met, and, as a result, covariate adjustments can seriously distort results (Elashoff, 1969; Thompson, 1988). For example, in a study to determine the effects of alternate compensatory educational programs on the achievement of students, Campbell and Erlebacher (1975) showed how use of a covariate pretest score could artifically make compensatory education programs appear to be harmful to students. The use of the covariate assumed that the relationship between the independent variable (type of program) and the dependent variable (posttest score) was the same for both the experimental and control groups. In actuality, the experimental group (those offered the compensatory program) was comprised of those students deemed eligible for a compensatory program based upon certain entrance criteria. The control group consisted of students whose previous achievement had precluded them from being eligible for the compensatory program. When posttest scores were adjusted ignoring group membership, low achieving students in the experimental group were evaluated as if they learned at the same rate as higher achieving control group students, and thus the compensatory programs appeared to have a negative effect upon the achievement of the experimental group students.

The results of Campbell and Erlebacher's study illustrate well that in most educational experimentation, an ounce of random assignment is far superior to a pound of covariate cure. Although researchers should strive to maintain the internal validity of experimental studies, they must be careful not to employ statistical controls which may distort true relationships among variables. When covariate adjustments are used, researchers should routinely test for homogeneity of regression, to assure relative equivalence of groups with respect to covariate regression equation adjustments (Thompson, 1986).

External Validity

External validity addresses issues relative to the generalizability of results across times, settings, and persons. Threats to external validity are often the fault of poor sampling procedures. Educational research usually involves the use of parametric techniques which are designed to produce results generalizable to a larger population of interest. In such cases, it is desirable to select subjects on the basis of their representativeness of the larger population. When intact groups or samples of convenience are used, results may not be generalizable since it may be difficult to determine what target population the sample actually represents (Cook & Campbell, 1979).

In a review of 126 correlational studies in three organizational behavior journals, Mitchell (1985) reported that only 17 percent of the studies used true random samples, and that in most cases it was "unclear whether the sample [was] representative of anything--even the organization from which it was drawn" (p. 202). Researchers would do well to give demographic descriptions of the people included in samples in cases in which random samples are not used. Much information could at least serve as an informal indicator of the type of individuals or groups to which the sample could be compared (Mitchell, 1985). In addition, when rsponse rates are low, researchers should routinely compare respondents to nonrespondents to determine how closely the responding sample represents the population of interest. Eason and Thompson (1988) report a study illustrat.ng these tests of representativeness.

Statistical Conclusion Validity

A given research study can be said to have statistical conclusion validity if the statistical tests employed in the study are free of systematic bias and if measures of the study's variables are proven to be reliable (Cook & Campbell, 1979). If measures are unreliable, or if statistical tests are inappropriately used, the study's statistical conclusion validity will be at risk. Lack of statistical conclusion validity or "instability" of results is "concerned with drawing false conclusions about population covariation from unstable sample data" (Cook & Campbell, 1979, p. 37).

Three closely related threats to statistical conclusion validity are low statistical power, misinterpretation of statistical significance testing, and underinterpretation of estimates of effect size. Most inferential statistical tests involve testing of null hypotheses, i.e., hypotheses predicting no relationship or no differences across groups. For any given statistical test, a researcher must determine the probability level for rejecting a null hypothesis based on sample results when the null is actually true in the larger population of interest (i.e., the probability of making a Type I error). This level of probability is known as alpha or *p* critical. Generally, alpha levels are kept rather small (.05 or below), so as to minimize the possibility of making a Type I error. However, for a fixed sample size, the smaller the alpha level, the greater the possibility that the researcher will make a Type II error (failure to reject a false null hypothesis).

Most researchers who use statistical significance tests rarely if ever test for the possibility of making Type II errors. Of course, when results are statistically significant a Type II error is impossible. But when results are not statistically significant, this failure to evaluate power is particularly disturbing when one considers that statistical significance is largely an artifact of sample size (Carver, 1978; Thompson, 1987b), i.e., as sample size increases, the likelihood of obtaining statistical significance increases. Hence, when sample size is small, the probability of not rejecting a false null hypothesis (making a Type II error) is increased. Consequently, as an adjunct to statistical significance testing, Cook and Campbell (1979) recommend that researchers more frequently conduct power analyses (Cohen, 1970) as a protection against Type II errors.

Another common problem caused by heavy reliance upon statistical significance testing has to do with misinterpretation of results. Statistical significance is a test of sampling error. The basic question the researcher asks when performing a test of statistical significance is: If the sample I am using represents a population in which the null is exactly true, how likely is this result? (Carver, 1978). Many researchers, however, feel that a statistically significant result is always an important result. This misperception is fostered by the frequent use of the term "significant" in place of "statistically significant" in scholarly writing (e.g., Tuckman, 1988, Chapter 11).

As a result of this common misperception, statistically significant results are often regarded as noteworthy even when the actual effect size for the variables of interest is negligible. Thompson (1987b) recognized the importance of interpreting effect size estimates when performing any statistical test, as these measures are a true indicator of the practical importance of the statistical results. Cook and Campbell (1979) further emphasize the advantage of "magnitude estimates" over statistical significance tests as the magnitude estimates are much less dependent on sample size.

Further threats to statistical conclusion validity are possible when the researcher fails to use the statistical technique which is most appropriate for interpreting the data at hand. Errors are often made simply because educational researchers are unaware of the considerable variety of statistical techniques available to them. Advances in computer hardware and software have made even the most difficult and advanced techniques available to computer users. Among these more advanced methods are various multivariate techniques (e.g., discriminant analysis, MANOVA, factor analysis, canonical correlaion), which prior to the widespread use of computers were impractical even to the most seasoned statistician due to the mathematical complexities involved in calculating results (McMillan & Schumaher, 1984; Thompson, 1986). More recently, however, with the advances brought about by computer technology, researchers are able to employ many of the techniques that were not previously feasible.

That multivariate methods are now readily available to educational researchers is most fortunate since multivariate methods tend to reflect appropriately the full network of the relationships which exist among behavioral variables (Fish, 1988). As Thompson (1986, pp. 8-9) has noted,

> The fundmental reason why multivariate statistics are almost always vital is that these methods usually best honor the reality about which the researcher is attempting to generalize. This is usually a reality in which the researcher cares about multiple outcomes, in which most outcomes have multiple causes, and in which most causes have multiple effects.

Use of numerous univariate statistical tests when fewer multivariate techniques could be used is a threat to statistical conclusion vaildity for yet another reason, namely the inflation ot the researcher's experimentwise Type I error rate (Fish, 1988; Thompson, 1986, 1988). As previously stated, when testing a null hypothesis the researcher must determine a level of probability for making a Type I error. This level of probability, known as alpha, is the probability of making a

Type I error for any one test. However, the testwise error rate is not necessarily equivalent to the error rate for the entire study (Ryan, 1959). In actuality, the experimentwise error rate (the possibility of making a Type I error in the study as a whole) is a function of the degree of intercorrelation among the variables being studied and the number of statistical tests performed using a single sample (Thompson, 1986).

Considering the im- -ance of honoring the reality of behavioral phenomena, the danger of using multiple univariate tests with data from a single sample, and the current availability of computer packages which simplify the mathematical computations of the various multivariate techniques, it would follow that educational researchers should be using multivariate techniques with reasonable frequency. However, evidence exists suggesting that more traditional but less sophisticated techniques still dominate most educational research methodology. Goodwin and Goodwin (1985) tabulated the statistical methods used in articles appearing in the American Educational Research Journal over a five year period. Only 17 percent of the articles over the period reported use of multivariate techniques. This figure represented little change over data compiled by Willson (1980), who studied statistical methods used in AERJ articles in the 10 year period immediately preceding the time frame of the Goodwin and Goodwin study. A similar review of techniques covering the years 1978 through 1987 (Elmore & Woehlke, 1988), yielded even more alarming results, with multivariate techniques accounting for only about 10 percent of the techniques used in articles in three different educational research journals.

Two additional problems related to the selection of statistical methods involve use of stepwise analytic techniques and reduction of internally-scaled predictor variables to nominal categories in order to employ chi-square or analysis of variance techniques. The use of stepwise analytic techniques is problematic for several reasons, as explained by Thompson (1988). These problems include the lack of sensitivity of stepwise techniques to sampling error and lack of consideration of these selection techniques to the degree of intercorrelation among the variables in the predictor set.

The reduction of interval data to nominal categories in order to employ chi-square tests or analyses of variance may be even a more serious problem. These data conversions actually throw away valuable information which the researcher has gone to a great deal of trouble to collect, and, in so doing, reduce the amount of true variance in the predictor variable(s) (Kerlinger, 1986; Thompson, 1988).

Even when an appropriate statistical method is used, the researcher should be concerned with the stability of the statistical results in relation to the population (Mitchell, 1985). Stability of statistical estimators is particularly at risk when sample size is small (Frank, Massey, & Morrison, 1965). To address this problem, researchers and statisticians have developed a number of procedures for assessing the stability of statistical estimators. "Invariance" procedures, for instance, involve random splitting of an original sample into two roughly equivalent subgroups, one for deriving an estimator, and the other for cross validating it. More sophisticated techniques such as the "U-method" (Mantel, 1967) and the "jackknife statistic" (Gray & Schucany, 1972), use averages of weighted composites of the estimator derived by splitting the original sample into a number of small, equivalent subsets and running the statistical procedure numerous times with alternate subsamples omitted from the analysis at each repetition. Daniel (1989) illustrates such techniques.

A final threat to statistical conclusion validity involves the degree to which measures of variables under consideration are deemed to be reliable or stable. Depending upon the type of measure employed, reliability can be assessed in a number of ways. In some cases, particularly those involving subjective judgments, interrater reliability is most appropriate. In the majority of cases in education, however, either test-retest or internal consistency (e.g., split half, alpha) reliability is used. Measures with low reliability cannot be regarded as consistent and accurate measures "because unreliability inflates standard error of estimates and these standard errors play a crucial role in inferring differences between . . the means of different treatment groups" (Cook & Campbell, 1979, p. 43).

It is extremely important that researchers routinely cite previous studies which have established reliability data relative to instruments they employ, or, in studies involving development of new instruments, that researchers conduct field tests to determine whether the instruments are reliable prior to the instruments being substantively applied in research studies. In fact, it is always advisable that any researcher compute relability data for any test given to any sample. Even in cases in which an established instrument is used, it should be remembered that reliability is always a function of a given data set, and not a function of test items alone.

Construct Validity

Construct validity is concerned with whether the measures utilized in a given study adequately measure what they are supposed to measure. A study weak in construct validity would be subject to "confounding" of results (Cook & Campbell, 1979), i.e., what one investigator would regard as a relationship between variables A and B, another investigator might regard as a relationship between constructs A and C, B and D, or C and D. Confounding of results is often related to what Fiske (1982) has termed "method variance." The content of test items, the written or oral directions, the personality of the examiner, and characteristics of the items themselves (e.g., response bias, social desirability of response) may be considered various aspects of method variance (Mitchell, 1985). Failure to control these factors may result in distortion of construct validity.

Procedure

One hundred education dissertations completed between the years of 1980-1988 comprised the sample for the present study. All dissertations completed during this period were included for analysis regardless of the studies' research designs. The dissertations were analyzed with a focus on the following variables: (a) year of completion, (b) department, (c) sample characteristics, (d) reliability of dependent variable(s), (e) survey research characteristics, if applicable, (f) result descriptions, and (g) unit of analysis. Since the purpose of the present study was to observe possible trends, results of the codings were not intended as a comparison of strengths and weaknesses. Judgement on design or analyses was not the focus.

The studies were reviewed by two judges. Interrater agreement was established at 86 percent. A coding instrument was created to guide the analysis. Categories one and two allowed the gathering of basic information about each study. Within the section on "unit of analysis" each technique employed by the dissertations was coded. Since some studies employed multiple techniques, total coding for the variable exceeded the number of total dissertations.

Results

The 100 dissertations completed between the years of 1980 and 1988 included a total of 201 techniques. The years in which more dissertations were completed were 1981 and 1982 with 26 percent and 1986 and 1987 with 31 percent.

A second variable, department, was coded to designate three departments within the College of Education that grant Ph.D. and Ed.D. degrees. The percentages of dissertations produced were the departments of Educational Leadership and Foundation (42%), Curriculum and Instruction (35%), and Special Education (23%). The size of the departments was not weighted in comparison to the number of dissertations since that factor was irrelevant to the present analysis.

Sample characteristics, the third variable of interest, included demographic information such as the grade or level of the subjects, size of the samples, and sample selection techniques. These findings are presented in Table 1. The subject group that accounted for the largest percentage of the studies was elementary and/or secondary school students (36 percent) followed by administrators and teachers (16 percent). The category designated "Others" included counselors, parents, married couples, high-risk infants, pharmacists, social workers, business education graduates, and meta-analysis studies of research.

Sample sizes ranged from three, a single subject design, to 6155 subjects. For ease of presentation, sample size was arbitrarily divided into five groups. The most populous category, 175 subjects or less, accounted for 56 percent of the total studies with 86 percent of the dissertations having under 500 subjects.

Studies were coded for sample selection and assignment. Two techniques, randomization and convenience, accounted for 94.9 percent of the total sample selection. Studies which used populations instead of samples and studies not reporting their method of selection constituted the remainder. Numerous randomization techniques were coded, including random selection, random assignment, combinations of random selection and random assignment, and stratified random selection. Selection was coded "convenience" when subjects were not randomly selected. Studies that utilized a random selection technique accounted for 52 percent of the dissertations as compared to studies that selected subjects by convenience, accounting for 42.9 percent of the dissertations. Moreover, considering all dissertations, more investigators used convenience samples in the years of 1981-1982 and 1986-1987. Interestingly, these years coincide with the most productive years for completed dissertations.

The third variable of interest was how the various authors established reliability of the dependent variable(s). Reliability refers to consistency of measurement (Cook & Campbell, 1976). The higher the reliability, the more consistently the instrument measures the research questions; thus, the less measurement error present in the analysis. The dissertations analyzed in the present study employed established instruments, self-made instruments, and combinations of both. Established instruments were used in 83.3 percent of the studies while self-made instruments were used in 28.1 percent of the dissertations. Categories coded for reliability of the dependent variable are presented in Table 2.

Survey research was conducted in 61 percent of the dissertations. The most frequently used data collection methods were mailed questionnaires (41 percent) and questionnaires administered personally (37.7 percent). Interviews were used the least (3.3 percent). Questionnaires personally administered by the investigator provided the highest response rate. The Likert scale was the most frequently used response format with a range from 4 to 15 steps comprising the scale. Five-point scales were the most dominant and accounted for 42 percent of the dissertations. Scales with 9 to 15 points comprised six percent of the total. A wide range of scale points maximizes variance and thus increases reliability (Thompson, 1981). Twenty-four percent of the studies did not report the range of points for their Likert scales.

Additional features involving the results discussion of the dissertations were coded and are presented in Table 2. The paradigm of statistical significance was used in almost all of the studies (98%). Forty-one percent of the studies reported effect size estimates. These estimates provide a measure of the amount of variance explained by a given variable. As previously noted, Thompson (1987b) suggests that the failure to use effect size estimates is in part due to the influence of the significance testing paradigm. However, not all studies reporting effect sizes focused interpretation on these results. When effect sizes were reported for multivariate analyses, Wilks' lambda was the dominant technique.

Also indicated in Table 2 are additional codings of result descriptions. The use of post hoc analyses was coded. The only post hoc coding included analyses performed on univariate statistics. Scheffe was the primary post hoc technique utilized. A priori constrasts were coded in twelve percent of the dissertations. However, it is interesting that some of the studies reporting a priori contrasts still conducted omnibus tests.

The final variable coded for the dissertations was "statistical analysis." Of interest within the variable were the categories of univariate and multivariate statistical analyses. A breakdown of all statistical techniques coded is presented in Table 3. Ninety univariate analyses and 91 multivariate analyses were coded. The sole univariate analyses that had increased in use over the time period was ANCOVA. The number of dissertations utilizing ANCOVA ranged from one for 1980-1984 to ten for 1985-1988. The instance of use of other univariate analyses remained relatively stable over the time period. Also, studies using ANOVA and ANCOVA and the respective multivariate techniques were coded for the test of homogeneity of variance. Out of 61 studies employing one of these analyses, 20 percent tested for the homogeneity of variance. Homogeneity of variance tests the assumption that the variances of all cells of a design are equal. Such tests are considered to be an important process in "OVA" statistical designs (Shavelson, 1981).

Descriptive studies provided a third coding for the unit of analysis variable. Twenty studies devoted large sections of their results to descriptive analyses and six of the dissertations were almost completely descriptive. Seven descriptive studies were coded for 1980-1984 and 13 for 1985-1988.

Trends in multivariate techniques for the years 1980-1984 and 1985-1988 indicated both increases and decreases in use. MANOVA was the only multivariate technique that increased in use across the time period. Between 1980-1984 only five dissertations used MANOVA, although this had increased to 14 in 1985-1988. However, three other multivariate techniques experienced a decrease in application: discriminant analysis from 15 to 8, factor analysis from 17 to 9, and canonical correlation from 14 to 4.

In addition to the frequency of univariate analyses, the number of tests performed within each analysis was recorded. The focus for the coding was on the possible inflation of experimentwise error rate. As previously noted, experimentwise error rate refers to error in a study where several similar analyses have been conducted based on data from the same subjects. Each of the analyses test for statistical significance by setting a low alpha level, the probability of making a Type I error. Additional analyses, each having an alpha of .05, raise the probability of making a Type I error to greater than 5 percent. Multivariate methods are suggested as the type of analysis to avoid inflation of experimentwise error rates (Thompson, 1986).

Univariate analyses susceptible to experimentwise error rate include t-tests, chi square, ANOVA, and ANCOVA. These combined analyses were utilized a total of 57 times in the dissertations. Most susceptible to experimentwise error rate were three dissertations which analyzed ten or more t-tests each, three dissertations which analyzed more than 20 chi squares each, six dissertations which analyzed more than 10 ANOVA's each, and one dissertation which analyzed nine ANCOVA's. Out of the nine ANCOVA studies, one study tested for the homogeneity of regression. The homogeneity of regression concerns the requirements of a common slope for all groups.

Two additional univariate analyses coded were correlational and multiple regression. The coding of correlational techniques yielded Pearson product moment (93.3%) as the most frequently used correlation statsitic. Out of 16 multiple regression codings, five dissertations indicated using cross validation procedures. Regression's most frequently used method for entering variables was stepwise (67%) followed by direct (27%).

Multivariate techniques coded for additional information were factor analysis and canonical correlation. Of interest in factor analysis were the techniques used for rotation and factor extraction. The coding indicated varimax as the only type of rotation utilized and principal component as the sole factor extraction method.

Discussion

The data reported here provided an overview of the statistical techniques and trends being used in dissertations over the last nine years. The various analyses failed to show any difference between the primary variables of interest. Neither univariate nor multivariate analyses dominanted the studies. Designs experiencing increases were ANCOVA and MANOVA. Designs decreasing in usage were discriminant analysis, factor analysis, and canonical correlation analysis.

Half of the studies used randomization techniques of some kind. For these studies internal validity, external validity, statistical conclusion validity, and construct validity were stronger than in studies selecting subjects by convenience. That is, subjects who agree to participate may differ in characteristics from subjects who did not agree. However, researchers in education are frequently not able to inplement "textbook" studies. One method to increase validity that is available to researchers is comparison of respondents to nonrespondents. The technique checks for between group differences. For further discussion of methodological problems see Thompson (1988).

Reliability of the dependent variable, important for statistical control validity, was not reported for all studies. Of the several methods for establishing reliability, literature citations of reliability studies by others was the most dominant. However, even in published articles only about half the authors report reliability (Willson, 1980).

In conclusion, dissertations completed in the 1980's appear similar in their methodologies to the methodologies identified in the analyses of journal articles by Willson (1980) and Mitchell (1985). There appears to be no increase in the use of research techniques regardless of the increase in availability of computer hardware and software. In some studies, the lack of internal validity, external validity, statistical conclusion validity, and construct validity continue to constitute design control problems. Thus, there remains room for progress.

REFERENCES

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

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Cohen, J. (1970). Statistical power analysis for the behavior sciences. New York: Academic Press.

Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 223-326). Skokie, IL: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.

Eason, S. H., & Thompson, B. (1988, October). The effect of persuasion on the perception of program evaluation information. Paper presented at the annual meeting of the American Evaluation Association, New Orleans.

Daniel, L. G. (1989, January). Use of the jackknife statistics to establish the external validity of discriminant analysis results. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston.

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

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17 (9), 19-20.

Fish, L. J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Fiske, D. W. (1982). Convergent-discriminant validation in measurements and research strategies. In D. Brinberg & L. H. Kidder (Eds.), Forms of validity in research (pp. 77-92). San Francisco: Jossey-Bass.

Frank, R. E., Massey, W. F., & Morrison, D. G. (1965). Bias in multiple discriminant analysis. Journal of Marketing Research, 2, 250-258.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14, (2), 5-11.

Gray, H. L., & Schucany, W. R. (1972). The generalized jackknife statistic. New York: Marcel Dekker, Inc.

Kerlinger, F. N. (1986). Foundations of behavioral research

(3rd ed.). New York: Holt, Rinehart, and Winston.

McMillan, J. H., & Schumacher, S. (1984). Research in education: A conceptual approach. Boston: Little, Brown.

Mitchell, T. R. (1985). An evaluation of the validity of correlational research conducted in organizations. Academy of Management Review, 10, 192-205.

Shavelson, R. J. (1981). Statistical reasoning for the behavioral sciences. Boston: Allyn and Bacon.

Thompson, B. (1981, November). Factor stability as a function of item variance. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas. (ERIC Document Reproduction Service No. ED 205 553)

Thompson, B. (1986, November). Two reasons why multivariate method are usually vital: An understandable reminder with concrete examples. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN.

Thompson, B. (1987a, January). Peer review of doctoral dissertations as a quality control mechanism: Some methods and examples. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas. (ERIC Document Reproduction Service No. ED 282 499)

Thompson, B. (1987b, April). The use (and misuse) of statistical significance testing: Some recommendations for improved editorial policy and practice. Paper presented at the annual meeting of the American Educational Research Association, Washington D.C. (ERIC Document Reproduction Service No. ED 287 868)

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

Tuckman, B. (1988). Conducting educational research (3rd ed.). San Diego, CA: Harcourt, Brace, Jovanovich.

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

Table 1
Sample Characteristics

Catagory	Percentage
Grade or level of subject:	
Administrators	9.0
Teachers	12.0
Adminstrators & teachers	16.0
Teachers & students	7.0
College students	7.0
Elementary/secondary students	36.0
Others	13.0
Size of sample:	
3 - 175	56.0
176 - 325	20.0
326 - 500	10.0
501 - 675	7.0
676 - 6155	7.0
Sample selection and assignment:	
Random assignment	22.4
Random selection	14.3
Random assign. & random select.	3.1
Stratified random	11.2
Stratified	3.1
Convenience	42.9
Population	2.0
Not given	2.0
Not applicable	1.0

Table 2
Overall Results

Instruments:
Established 83% Self-made 28%

Reliability:
Alpha 33% Test retest 42% Split half 46%
Literature citations 59% Interrater Interobserver 26%

Response Format:
Likert 50% Scale most frequently used - 5 points

Survey Research:
Response Rate:
Range 23% - 100% Not given in three studies
Data collection method:

Questionnaires administered personally	38%
Mailed questionnaires	41%
Delivered but not administered	18%
Interviews	3%
Comparison of respondents to nonrespondents	1%

Results:

Post hoc analysis of univariate statistics	19%
A priori analysis	12%
Statistical significance	98%
Effect size	41%
Balanced design	13%
Pilot study	27%

Table 3
Statistical Analyses

	Year									
	80	81	82	83	84	85	86	87	88	Tot.
Dissertations per yr	9	15	11	8	8	9	15	16	9	100
Univariate analyses:										
t tests	1	1	2	-	-	-	3	3	1	11
Chi square	-	-	1	-	-	2	2	-	-	5
Correlation	1	3	2	2	5	3	4	4	4	28
Multiple Reg.	-	2	1	1	-	-	1	-	-	5
ANOVA	3	2	2	4	4	2	3	6	4	30
ANCOVA	-	-	1	-	-	1	2	5	2	11
Multivariate analyses:										
Discriminant ana.	1	9	3	1	1	3	1	3	1	23
MANOVA	-	2	1	2	-	2	1	10	1	19
MANCOVA	-	-	1	-	-	-	-	-	-	1
Factor analysis	4	7	5	-	1	4	-	5	-	26
Canonical cor.	1	7	3	1	2	1	1	2	-	18
Latent trait	-	1	1	-	-	-	-	-	-	2
LISREAL	-	-	-	-	-	-	1	1	-	2
Descriptive analysis	3	-	1	2	1	2	5	2	4	20