DOCUMENT RESUME

ED 301 595   TM 012 582

AUTHOR   Thompson, Bruce
TITLE   Common Methodology Mistakes in Dissertations:
                Improving Dissertation Quality.
PUB DATE   Nov 88
NOTE   63p.; Paper presented at the Annual Meeting of the
                Mid-South Educational Research Association
                (Louisville, KY, November 8-11, 1988).
PUB TYPE   Reports - Research/Technical (143) --
                Speeches/Conference Papers (150)

EDRS PRICE   MF01/PC03 Plus Postage.
DESCRIPTORS   Chi Square; *Doctoral Dissertations; *Educational
                Research; Higher Education; *Misconceptions;
                Multivariate Analysis; Psychometrics; *Research
                Methodology; Research Problems; Scholarship;
                *Statistical Analysis; Statistical Significance
IDENTIFIERS   Covariance Structural Analysis; Stepwise
                Regression

ABSTRACT
                Dissertations are the cumulative, tangible "best
evidence" of interests of doctoral faculty and students in serious
and incisive scholarship. Thus, dissertations are thoroughly studied
by the program review teams periodically hired by boards of higher
education in most states. The present paper explores seven errors in
quantitative analysis in both published literature and dissertations.
They are associated with beliefs concerning statistical significance
testing, multivariate statistics, chi-square analysis, covariance
statistical corrections, stepwise analytic methods, the psychometric
integrity of instrumentation, and presentation of statistically
nonsensical or impossible results. The errors are described using
references to works by other authors and small hypothetical data sets
to illustrate problems. Concrete examples of the errors as they occur
in dissertations are cited to make clear that the errors are not
hypothetical. Ten dissertations, completed since January of 1985, are
cited as examples, although pseudonyms are used. These common
methodological errors in dissertations may reflect poor training in
research. A 76-item list of references, 14 data tables, and three
figures are provided. A bibliography of the 89 dissertations from
which examples are taken are appended. (TJH)

COMMON METHODOLOGY MISTAKES IN DISSERTATIONS:

Improving Dissertation Quality

Bruce Thompson

University of New Orleans 70148

2

ABSTRACT

Dissertations are the cumulative, tangible "best evidence" of interests of doctoral faculty and students in serious and incisive scholarship. Thus, dissertations are thoroughly studied by the program review teams periodically hired by boards of higher education in most states. The present paper explores seven errors in quantitative analysis in both published literature and dissertations. The errors are explained in detail using references to works by other authors and small hypothetical data sets to illustrate problems. Concrete examples of the errors as they occur in dissertations are cited to make clear that the errors are not hypothetical. Ten dissertations, completed since January of 1985, are cited as examples, although pseudonyms are employed to avoid embarrasment of the students or their doctoral advisors. The discussion should also be useful to authors of published research who may wish to avoid the same errors.

Probably the most fundamental challenge confronting doctoral programs is maintaining dissertation quality while simultaneously respecting student prerogatives. Most accreditation regulations require that students be afforded substantial influence over the selection of advisors and dissertation committees. Although the the tension between expectations for quality and respect for student freedom creates a difficult dilemma for doctoral faculty, the dilemma is not one that faculty can afford to ignore. As Thompson (1987a, p. 1) notes,

> Even if a faculty member's own inherent interest
> in scholarship is not sufficient to warrant
> interest in the quality of dissertations being
> produced under the faculty member's own direction
> or under the direction of colleagues, interest in
> the survival of the program may itself warrant
> concern.

External review teams that make recommendations to state boards regarding program continuation or termination do tend to pay disproportionate attention to dissertation quality when making their recommendations.

Review teams quite reasonably feel that dissertations are the cumulative, tangible "best evidence" of faculty and student interest in serious and incisive scholarship. For example, a recent review of all education doctoral in Louisiana yielded conclusions that the programs at the University of New Orleans "produce the best dissertations" (Brown, Cooper, Griffiths, Howey & Lilly, 1985, p. 80) in the state. But the reviewers nevertheless offered the following observations regarding how

1

4

good the best dissertations in the state are:

> Dissertations [at UNC] are weak. With all of the improvements noted in the College, it is paradoxical that the dissertations remain so weak. (It is acknowledged that a few dissertations are excellent, but the majority are weak.) ...The problems with dissertations can be traced to poor training in research and inept supervision. (Brown et al., 1985, pp. 38-39)

The purpose of the present paper is not to propose mechanisms for improving dissertation quality; such proposals are available elsewhere (cf. Thompson, 1987a). Rather, the purpose of the paper is to identify common methodological errors in dissertations that may reflect "poor training in research." Seven errors are each explained. Examples of dissertations illustrating the errors are cited so as to moot any argument that the errors are purely hypothetical.

A Preliminary Caveat: The Role of Statistical Method in Inquiry

However, one preliminary caveat is in order--methodological integrity is not the ultimate sina qua non of research, published or otherwise. Certainly it is true that, "Although the quality of educational research is improving, evidence still indicates that much of the research published has important weaknesses" (Borg, 1983, p. 193). Empirical studies of methodological practice in published research confirm these general impressions (Persell, 1976; Wandt, 1965; Ward, Hall & Schramm, 1975). Some of the problems in the quality of the research literature can be

2

attributed to the journal review process, studied in an intriguing fashion by Peters and Ceci (1982). Nevertheless, as Glass (1979, p. 12) suggests, "Our research literature in education is not of the highest quality, but I suspect that it is good enough on most topics."

Even studies with methodological weaknesses can make noteworthy contributions to understanding of educational phenomena, i.e., to theory. Reasonable people can disagree about the role of theory in research. For example, Scriven (1980, p. 18) argues that, "In the practical sciences we are looking for solutions to problems, not just explanations of the failures that led to the problems... It does not take a theory."

But theory building is the ultimate objective of good science. As Kerlinger (1977, pp. 5-6) notes, "Science, then, really has no other purpose than theory, or understanding and explanation." As Gergen (1969, p. 13) explains, theoretically oriented research "not only satisfies our curiosity, but also has the advantage of maximum heuristic value. It leads to new investigations and suggests interesting links to other areas of concern." As Thompson (in press) argues,

> ...when Jenner discovered many years ago that
> milkmaids did not get smallpox if they had been
> exposed to cowpox, he had the basis for suggesting
> a possible cure for smallpox. But absent any
> understanding of the mechanics of the cure, if he
> had then attempted to identify a cure for polio,
> his original discovery would have been of no

3

6

assistance at all.

1. <u>Dissertations should reflect the limited contribution that statistical significance testing can be make to the interpretation of results</u>.

Few methodological offerings have sparked more controversy than Sir Ronald Fisher's promulgation of significance testing methods, methods that apparently were developed prior to Fisher's work (Carlson, 1976). The past 30 years have involved periodic efforts "to exorcise the null hypothesis" (Cronbach, 1975, p. 124). Morrison and Henkel (1970) and Carver (1978) provide historically important and incisive explanations of the limits of significance testing as an aid to interpretation. More recent informative treatments are available from Dar (1987), Huberty (1987), Kupfersmid (1988), and Thompson (1987b, 1988c).

Most researchers have been taught the statistical significance of results does not inform the researcher regarding the <u>importance</u> of outcomes. Shaver (1985, p. 58) makes this point in a concrete fashion in his contrived dialogue about significance testing:

> Chris: [Looking puzzled.] Well, as I said, it [my result] was statistically significant. You know, that means it wasn't likely to be just a chance occurrence... An unlikely occurrence like that <u>surely</u> must be important.

> Jean: Wait a minute, Chris. Remember the other day when you went into the office to call home? Just as you completed dialing the number,

4

7

your little boy picked up the phone to call someone. So you were connected and talking to one another without the phone ever ringing... Well, that must have been a truly important occurrence then?

Yet, in three ways actual behavior tends to belie a failure to really accept that significance testing does not inform decisions regarding the importance of results. First, journal editorial boards tend to perceive articles that report significant results more favorably than articles not reporting significant results (Atkinson, Furlong & Wampold, 1982). Second, readers of research findings tend to perceive more favorably those articles reporting statistically significant results (Cohen, 1979). Third, and most disturbing of all, authors tend not to submit manuscripts in which nonsignificant results must be reported, and even tend to abandon lines of inquiry on the basis of such results (Greenwald, 1975). These behaviors are too readily transmitted to doctoral students.

Too few researchers appreciate which study features contribute to statistical significance. Although significance is a function of at least seven interrelated features of a study (Schneider & Darcy, 1984), sample size is the primary influence on significance. Some example results may clarify the ways in which sample sizes affect significance tests.

Tables 1 and 2 present significance tests associated with varying sample sizes and either moderate (9.8%) or larger (33.6%) fixed effect sizes, respectively. The tables can be viewed as presenting results for either a multiple regression analysis

5

involving two predictor variables (in which case the "r sq" effect size would be called the squared multiple correlation coefficient, $\underline{R}$) or an analysis of variance involving an omnibus test of differences in three means in a one-way design (in which case the "r sq" effect size would be called the correlation ratio or eta squared).

INSERT TABLES 1 AND 2 HERE.

Each table presents results for fixed effect sizes but increasing sample sizes (4, 13, 23, 33, 43, 53, 63, or 123). For the fixed effect size of 9.8% involved in Table 1, the fixed effect size becomes statistically significant when there are somewhere between 53 and 63 subjects in the analysis. For the 33.6% effect size reported in Table 2, the result becomes statistically significant when there are somewhere between 13 and 23 subjects in the analysis.

For a fixed effect size, adding subjects to the analysis impacts statistical significance in two ways. First, as illustrated in Tables 1 and 2, the critical $\underline{F}$ at a fixed alpha gets smaller as degrees of freedom error increase. Second, as the degrees of freedom error increase, the mean square error gets smaller, and thus the calculated $\underline{F}$ gets larger.

The researcher who does not genuinely understand statistical significance would differentially interpret the effect size of 9.8% when there were 53 versus 63 subjects, and would differentially interpret the fixed effect size of 33.6% when there were 13 versus 23 subjects in the analysis. Yet the effect

sizes within each table are fixed. Empirical studies of research practice indicate that superficial understanding of significance testing has actually led to serious distortions. such as researchers interpreting significant results involving small effect sizes while ignoring nonsignificant results involving large effect sizes (Craig, Eison & Metze, 1976)!

Nor does significance testing typically inform the researcher regarding the likelihood that results will be replicated in future research (Carver, 1978). Researchers who wish to estimate the likely replicability of results should instead employ cross-validation logic (Campo, 1988), the "jackknife" logic developed by Tukey and his colleagues (Crask & Perreault, 1977), or the "bootstrap" logic developed by Efron and his associates (Diaconis & Efron, 1983).

Two aspects of significance testing interpretation in dissertations warrant attention. First, some students use language implying that they are interpreting significance tests as if they were effect sizes. But, as Kerlinger (1986, p. 214) emphasizes, "Tests of statistical significance like t and F unfortunately do not indicate the magnitude or strength of relations." Yet Kerlinger (1986) himself constantly refers to results being "highly significant" (cf. pp. 187, 248, 334), and other respected textbook authors do so as well (e.g., Cliff, 1987, p. 394). No wonder doctoral students such as Darlington#[1]

_____

[1] In order to minimize embarrassment to students and the members of their dissertation committees, pseudonyms designated with pound signs ("#") have been substituted for student names, and 2001 is cited as the date on which these dissertations were completed.

(2001, p. 80) find themselves reporting that "The results of this MANOVA was [sic verb agreement] highly significant."

A second problem in language, implying the interpretation of significance tests as effect sizes, involves the use of phrases such as "the results approached statistical significance." Robert Brown, former editor of the Journal of College Student Personnel, made the humorous but telling comment at a recent conference: "How do these authors know their results weren't trying to avoid statistical significance?" Yet dissertation students such as Spearman# (2001, p. 75) may find themselves reporting that, "The number of years of experience was not significant but did approach significance."

The most serious misinterpretations of significance testing tend to occur when sample size is small and effect sizes are large but are underinterpreted, or when sample sizes are commendably large and are statistically significant but effect sizes are modest and are are overinterpreted. Guilford# (2001) provides a thought provoking example of the latter case.

Guilford# (2001) administered a measure of self-esteem and a measure of achievement motivation to 1,401 subjects, and found that scores on the two measures had a statistically significant product-moment correlation of 0.449 (p. 83). Thus, the squared r effect size in the study was 0.202, or 20.2%. Guilford# (2001, p. 96) argued that,

> From the data collected in this study, it has been established that a relationship exists between self-esteem [Self-Esteem Inventory--SEI] and achievement motivation [Resultant Achievement

Motivation Scale--RAM] for vocational-technical

students. The existence of this relationship can

be both important and useful in vocational-

technical education.

Based on the two measures having only 20.2% of their variance in

common, but the result being statistically significant,

Guilford# (2001, pp. 97-98) suggests that "the educator can

choose either the RAM or the SEI and administer it [and get the

same information] since this study has established the existence

of a relationship between achievement motivation and self-esteem

in vocational-technical students"!

2. <u>Dissertations</u> <u>should</u> <u>reflect</u> <u>the</u> <u>fact</u> <u>that</u> <u>multivariate</u>
   <u>statistics</u> <u>are</u> <u>often</u> <u>vital</u> <u>in</u> <u>educational</u> <u>research</u>.

Multivariate statistics have been available to researchers

for many years, although even to ay "there are many articles in

the research literature in which multiple univariate statistics

are calculated rather than a single multivariate analysis; for

instance, one article may report 50 $\underline{t}$-tests rather than one

MANOVA" (Moore, 1983, p. 307). McMillan and Schumacher (1984)

isolated one reason why some researchers have hesitated to use

multivariate statistical methods:

The statistical procedures for analyzing many

variables at the same time have been available for

many years, but it has only been since the computer

age that researchers have been able to utilize these

procedures. There is thus ag in training of

researchers that has militated against the use of

9

these more sophisticated procedures. There are in
evidence more each year in journals, however... (p.
270)

Hinkle, Wiersma and Jurs (1979) concurred, noting that "it is
becoming increasingly important for behavioral scientists to
understand multivariate procedures even if they do not use them
in their own research." And recent empirical studies of research
practice do confirm that multivariate methods are employed with
some regularity in published behavioral research (Elmore &
Woehlke, 1988; Gaither & Glorfeld, 1985; Goodwin & Goodwin,
1985).

There are two reasons why multivariate methods are so
important in behavioral research, as noted by Thompson (1986b)
and by Fish (1988). First, multivariate methods control the
inf' tion of Type I "experimentwise" error rates. Most
researchers are familiar with "testwise" alpha. But while
"testwise" alpha refers to the probability of making a Type I
error for a given hypothesis test, "experimentwise" error rate
refers to the probability of having made a Type I error anywhere
within the study. When only one hypothesis is tested for a given
group of people in a study, "experimentwise" error rate will
exactly equal the "testwise" error rate.

But when more than one hypothesis is tested in a given
study, the two error rates will not be equal. Witte (1985, p.
236) explains the two error rates using an intuitively appealing
example involving a coin toss. If the toss of heads is equated
with a Type I error, and if a coin is tossed only once, then the
probability of a head on the one toss and of at least one head

10

13

within the set of one toss will both equal 50%. But if the coin is tossed three times, even though the "testwise" probability of a head on each given toss in 50%, the "experimentwise" probability that there will be at least one head in the whole set of three flips will be inflated to more than 50%. Researchers control "testwise" error rate by picking small values, usually 0.05, for the "testwise" alpha. "Experimentwise" error rate, on the other hand, can be controlled at the "testwise" level by employing multivariate statistics.

When researchers test several hypotheses in a given study, but do not use multivariate statistics, the "experimentwise" error rate will range somewhere between the "testwise" error rate and the ceiling calculated in the manner illustrated in Table 3. Where the experimentwise error rate will actually lie will depend upon the degree of correlation among the dependent variables in the study. Because the exact rate in a practical sense is readily estimated only when the dependent variables are perfectly correlated (and "experimentwise" error will equal the "testwise" error) or are perfectly uncorrelated (and "experimentwise" error will equal the ceiling calculated in the manner illustrated in Table 3), it is particularly disturbing that the researcher may not even be able to determine the exact "experimentwise" error rate in some studies!

INSERT TABLE 3 ABOUT HERE.

Paradoxically, although the use of several univariate tests in a single study can lead to too many hypotheses being spuriously rejected, as reflected in inflation of

"experimentwise" error rate, it is also possible that the failure
to employ multivariate methods can lead to a failure to identify
statistically significant results which actually exist. Fish
(1988) provides a data set illustrating this equally disturbing
possibility. The basis for this paradox is beyond the scope of
the present treatment, but involves the second major reason why
multivariate statistics are so important.

Multivariate methods are often vital in behavioral research
because <u>multivariate methods best honor the reality to which the</u>
<u>researcher is</u> / <u>purportedly trying to generalize</u>. Since
significance testing and error rates may not be the most
important aspect of research practice (Thompson, 1988c), this
second reason for employing multivariate statistics is actually
the more important of the two grounds for using these methods.
Thompson (1986b, p. 9) notes that the reality about which most
researchers wish to generalize is usually one "in which the
researcher cares about multiple outcomes, in which most outcomes
have multiple causes, and in which most causes have multiple
effects." As Hopkins (1980, p. 374) has emphasized:

> These multivariate methods allow understanding of
> relationships among several variables not possible
> with univariate analysis... Factor analysis,
> canonical correlation, and discriminant analysis--
> and modifications of each procedure--allow
> researchers to study complex data, particularly
> situations with many interrelated variables. Such is
> the case with questions based in the education of

human beings.

Similarly, McMillan and Schumacher (1984) argue that:

> Social scientists have realized for many years that
> human behavior can be understood only by examining
> many variables at the same time, not by dealing with
> one variable in one study, another variable in a
> second study, and so forth... These [univariate]
> procedures haved failed to reflect our current
> emphasis on the multiplicity of factors in human
> behavior... In the reality of complex social
> situations the researcher needs to examine many
> variables simultaneously. (pp. 269-270)

Unfortunately, dissertations do not always reflect a recognition that multivariate statistics are often vital in research. For example, Cronbach# (2001, pp. 78-97) reported 15 Pearson chi-square tests of contingency table data, each with degrees of freedom (95) that appear to be impossible for the data. Similarly, Spearman# (2001, pp. 54-66) reports 10 separate ANOVAs, each involving a factorial analysis, which maximally inflates experimentwise error rates.

But Spearman# (2001, p. 78) was primarily interested in interaction hypotheses, and was forced to report that "All null hypotheses failed to be rejected because no statistical differences [sic] were found in any of the groups tested for interactions." Paradoxically, different findings might have been isolated with the correct use of a multivariate method, as Fish (1988) illustrates, and perhaps statistically significant interactions would have resulted.

13

3. Dissertations should reflect the recognition that . discarding
   variance to conduct chi-square or OVA analyses can lead to
   serious distortions in interpretations, and that even when OVA
   methods are appropriate the methods should usually be
   implemented using regression approaches.

Cohen (1968, p. 441) has characterized the conversion of
intervally scaled variables down to the nominal level of scale as
the "squandering [of] much information." As Kerlinger (1986, p.
558) explains, this squandering can lead to distorted results:

> ...Partitioning a continuous variable into a
> dichotomy or trichotomy throws information away...
> To reduce a set of values with a relatively wide
> range to a dichotomy is to reduce its variance and
> thus its possible correlation with other
> variables.

Thompson (1988a, pp. 3-4) notes that

> Variance is the "stuff" of which all quantitative
> research studies are made... It is not usually
> sensible to invest serious effort in collecting
> reliable and valid continuous score data, and to
> then casually discard the information that we
> previously went to some trouble to collect.

Dissertation students frequently discard variance in order
to conduct either Pearson chi-square contingency table tests or
ANOVA, ANCOVA, MANOVA or MANCOVA (hereafter labelled OVA
methods). Certainly there are many problems with typical
applications of the chi-square contingency table test (Thompson,
1988b), but OVA methods are more frequently applied (Elmore &
Woehlke, 1988; Gaither & Glorfeld, 1985; Goodwin & Goodwin,

14

17

1985), and empirical research indicates that the use of OVA methods with variables that were originally intervally scaled does introduce distortions (Thompson, 1986a). Thus, Cliff (1987, p. 130) correctly criticizes the practice of discarding variance on intervally scaled predictor variables to perform OVA analyses:

> Such divisions are not infallible; think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less.

Furthermore, even when intervally scaled variables are naturally nominally scaled, regression approaches to OVA analyses still tend to be superior to classical OVA calculations (Thompson, 1985).

Most researchers employing OVA methods are aware that "A researcher cannot stop his analysis after getting a significant F" (Huck, Cormier & Bounds, 1974, p. 68). Gravetter and Wallnau (1985, p. 423) concur that "Reject Ho indicates that at least one difference exists among the treatments. With k [means] = 3 or more, the problem is to find where the differences are."

Many researchers employ unplanned (also called a posteriori or post hoc) multiple comparison tests (e.g., Sheffe, Tukey, or Duncan) to isolate which means are significantly different within OVA ways (also called factors) having more than two levels.

15

18

Textbook authors tend to discuss unplanned comparisons in somewhat prejorative terms. For example, several authors refer to the application of these comparisons as "data snooping" (Kirk, 1968, p. 73, 1984, p. 360; Pedhazur, 1982, p. 305). Keppel (1982, p. 150) makes reference to "milking" in his discussion of these tests. Similarly, Minium and Clarke (1982, p. 321) note that:

> Prior to running the experiment, the investigator
> in our example had no well-developed rationale for
> focusing on a particular comparison between means.
> His was a "fishing expedition"... Such comparisons
> are known as post hoc comparisons, because
> interest in them is developed "after the fact"--it
> is stimulated by the results obtained, not by any
> prior rationale.

Planned (also called a priori or focused) comparisons provide a valuable alternative to unplanned comparisons. Pedhazur (1982, chapter 9) and Loftus and Loftus (1982, chapter 15) provide readable explanations of these comparisons. Planned comparisons typically involve weighting data by sets of "contrasts" such as those presented by Thompson (1985) or those presented in Table 4. Other types of contrasts, those which test for trends in means, are provided by Fisher and Yates (1957, pp. 90-100) and by Hicks (1973).

INSERT TABLE 4 ABOUT HERE.

Contrasts are typically developed to sum to zero, as do all five contrasts presented for the data in Table 4. The data represent a hypothetical validity study conducted to determine

16

whether various clinical groups score differently on a psychological measure. Contrasts are uncorrelated or orthogonal (as are the hypotheses they represent or test) when the contrasts each sum to zero and when the sum of the cross-products of each pair of contrasts all sum to zero also. Thus, the contrasts presented in Table 4 are uncorrelated.

Some researchers do not believe that planned comparisons should necessarily be orthogonal. For example, Winer (1971, p. 175) argues that "whether these comparisons are orthogonal or not makes little or no difference." However, orthogonal planned comparisons do have special appeal, for statistical reasons delineated elsewhere (Kachigan, 1986, p. 309). But as Keppel (1982, p. 147) suggests:

> The value of orthogonal comparisons lies in the
> independence of _inferences,_ which, of course, is a
> desirable quality to achieve. That is, orthogonal
> comparisons are such that any decision concerning
> the null hypothesis representing one comparison is
> uninfluenced by the decision regarding any other
> orthogonal comparison. The potential difficulty with
> nonorthogonal comparisons, then, is interpreting the
> different outcomes. If we reject the null hypotheses
> for two nonorthogonal comparisons, which comparison
> represents the "true" reason for the observed
> differences?

There are two reasons why planned comparisons are usually superior to unplanned comparisons. First, as noted by numerous

17

20

researchers (Glasnapp & Poggio, 1985, p. 474; Hays, 1981, p. 438; Kirk, 1968, p. 95; Minium & Clarke, 1982, p. 322; Pedhazur, 1982, pp. 304-305; Sowell & Casey, 1982, p. 119), planned comparisons offer more power against Type II errors than do unplanned comparisons, for reasons explained elsewhere (Games, 1971a, 1971b). For example, for the data presented in Table 4, the omnibus test of differences among the six group means is not statistically significant ($F=1.5$, $df=5/6$, $p=.3155$). Furthermore, even if unplanned comparisons were conducted in violation of conventional practice (since the omnibus test was not statistically significant), statistically significant differences would not have been identified either. However, a planned comparison involving the mean of the two level-six subjects versus the mean of the remaining 10 subjects would have been statistically significant ($F=12.5$, $df=1/6$, $p=.0054$).

However, significance is not the end-all and be-all of research (Thompson, 1988c). The more important reason why planned comparisons are important is that planned comparisons tend to force the researcher to be more thoughtful in conducting research, since planned comparisons must be carefully formulated before data are collected and since typi lly only a limited number of planned comparisons can be stated in a given study. As Snodgrass, Levy-Berger and Haydon (1985, p. 386) suggest, "The experimenter who carries out post hoc comparisons often has a rather diffuse hypothesis about what the effects of the manipulation should be." As Keppel (1982, p. 165) notes,

> Planned comparisons are usually the motivating
> force behind an experiment. These comparisons are

targeted from the start of the investigation and

represent an interest in particular combinations

of conditions--not in the overall experiment.

Thus, as Kerlinger (1986, p. 219) suggests, "while post hoc tests

are important in actual research, especially for exploring one's

data and for getting leads for future research, the method of

planned comparisons is perhaps more important scientifically."

Wilks# (2001, p. 116) provides one of the more disturbing

examples of the use of OVA methods in a dissertation, even though

planned comparisons were applied in the study. In this study both

predictor variables, age and math anxiety, could have been

measured at the interval level of scale. Age was treated as a

trichotomy. Math anxiety data were actually collected at the

interval level of scale and were then converted into a dichotomy.

At least the cutoffs used in creating the dichomotomy (p. 84)

were not decided with the same arbitrariness employed in the

initial decision by Wilks# to discard variance on both interval

predictor variables.

4. <u>Dissertations</u> <u>should</u> <u>reflect</u> <u>a</u> <u>recognition</u> <u>that</u> <u>covariance</u>
   <u>statistical</u> <u>corrections</u> <u>are</u> <u>usually</u> <u>least</u> <u>helpful</u> <u>(and</u> <u>are</u>
   <u>most</u> <u>dangerous)</u> <u>when</u> <u>corrections</u> <u>are</u> <u>most</u> <u>needed</u>.

Many "statistical controls" can be invoked to adjust

posttest scores when the quantitative researcher believes that or

random assignment or design selection have failed to create

groups that were equivalent at the start of the experiment or

quasi-experiment. These statistical controls are available

throughout the entire gamut of quantitative methods. For example,

Gorsuch (1983, pp. 89-90) notes that the first factor extracted

19

in a factor analysis can be located to pass directly through a "covariate" variable in factor space. Since factors are uncorrelated, the effects of the first factor on all other factors will have been statistically controlled.

Though many of these statistical controls date back to the beginning of the century (Nunnally, 1975, p. 9), most of the controls have not enjoyed wide use. Analysis of covariance (ANCOVA), for example, has been used in about four percent of the recently published research (Goodwin & Goodwin, 1985, pp. 8-9; Willson, 1980, p. 7). As explained by McGuigan (1983, p. 230):

> Briefly this technique enables you to obtain a measure of what you think is a particularly relevant extraneous variable that you are not controlling. This usually involves some characteristics of your participants. For instance, if you are conducting a study of the effect of certain psychological variables on weight, you might use as your measure the weight of your participants before you administer your experimental treatments. Through analysis of covariance, you then can "statistically control" this variable--that is, you can remove the effect of initial weight from your dependent variable scores, thus decreasing your error variance.

One problem with statistical controls is that they assume very reliable measurement of the control variables. For example, Nunnally (1975, p. 10) notes that reliability will not usually

20

have an appreciable influence on the substantive interpretation of most statistical procedures as long as reliability of measurement is at least 0.70, but that "Measurement reliability becomes crucial... in employing statistical partialling operations, as in the analysis of covariance or in the use of partial correlational analysis." Cliff (1987, p. 129) concurs, noting that

> In general, partial correlation analysis is
> affected by any lack of reliability or validity in
> the variables. In many ways these effects resemble
> tuberculosis as it occurred a generation or two
> ago: They are widespread, the consequences are
> serious, the symptoms are easily overlooked, and
> most people are unaware of their etiology or
> treatment.

Unfortunately, too many researchers may not consider and certainly do not report the measurement error of their variables. As Willson (1980, p. 9) comments, "That reliability of instruments is unreported in almost half the published research is likewise inexcusable at this late date."

Statistical control has been particularly appealing to some quantitative researchers when random assignment was not performed. These researchers expect the statistical adjustments of ANCOVA to magically make groups equivalent.

However, the primary difficulty with statistical control performed to make groups equivalent involves the homogeneity or regression assumption of the methods. The methods assume that the relationship between the covariate and the dependent variable is

21

24

equivalent in all experimental groups. This assumption is necessary because the statistical control procedures are implemented by adjusting the dependent variable to the extent that the covariate and the dependent variable are correlated <u>when group membership information is completely ignored</u>.

Campbell and Erlebacher (1975) present a concrete illustration of how the use of statistical controls can seriously distort findings when the homogeneity of regression assumption is not met. ANCOVA has been very appealing in research investigating the effects of compensatory education programs. In these cases the treatment intervention is made available to all or most children who are eligible. The control group usually consists of children who were not eligible for the treatment and, therefore, the group is inherently different in its character than the treatment group. In these analyses both the dependent variable and the covariate are cognitive variables. The statistical control procedure assumes that the relationship between the two variables is the same in both groups, i.e., since correlation is a measure of the slope of the regression line for the two variables, that children who are eligible for and receive compensatory interventions <u>learn at the same rate</u> as children who are not eligible for the intervention.

The decision to blithely use the statistical control when the homogeneity of regression assumption is not met leads to "tragically 'misleading analyses" that actually "can mistakenly make compensatory education look harmful" (Campbell & Erlebacher, 1975, p. 597). Similarly, Cliff (1987, p. 273) argues that, "It

22

could be that the relationship bet⸱⸱en the dependent variable and the covariate is different under different treatments. Such occurrences tend to invalidate the interpretation of the simple partial correlations described above."

Persons who wish to use statistical controls of this type are usually trapped in a nasty dilemma. If the controls are not needed then they should not be used. But if statistical control is needed because the groups in a study are not equivalent, then often the homegeneity of regression assumption cannot be met and the use results in seriously distorted inferences.

It is interesting to note that many researchers do not recognize the paradox of testing both analytic assumptions and substantive hypotheses for statistical significance. Researchers frequently try to obtain as large a sample as possible, so that chances for "significance" of substantive tests are maximized. This practice also leads to greater likelihood that tests of homogeneity of variance or of regression will also be significant.

The fallacious use of statistical control in inappropriate ANCOVA applications needs to be recognized by more researchers, as some researchers have long warned of these various dangers (Elashoff, 1969; Lord, 1960). ANCOVA is a special case of regression analysis. As Cliff (1987, p. 275) notes, "We could say that we are fitting a single regression equation to the data for all the groups and then doing an anova of the deviation from the regression line."

Consider the hypothetical data presented in Table 5. The hypothetical study involves four children from a compensatory

23

program ("A") who have lower mean achievement (-.19) on the cognitive pretest ("ZX") than do their peers (mean=.19) from the noncompensatory group. Furthermore, as one might expect, and as illustrated in Figure 1 (which also presents the cognitive posttest ("ZY") scores of the eight children), the children in the two groups are learning at different rates.

INSERT TABLE 5 AND FIGURE 1 ABOUT HERE.

Nevertheless, the ANCOVA procedures employs the single beta weight (r = beta weight for two variable case = .81) derived by ignoring the group membership ("A" or "B") of the children, i.e., derived by ignoring the fact that the children are learning at different rates. This beta weight adjustment is presented in Figure 1 as the regression line for the variables, derived ignoring group membership. However, Figure 1 also indicates that the slopes of regression lines computed separately for the two groups are different, and that it is not reasonable to use the same adjustment for both groups.

Table 6 presents conventional ANOVA results for this data set when no covariance adjustments are implemented. Table 7 presents an ANCOVA utilizing pretest scores ("ZX") as a covariate. Table 8 presents an ANOVA performed on the residual raw scores ("YE" = "ZY" - "YHAT"); this analysis demonstrates that ANCOVA is an ANOVA on posttest scores once the posttest scores have been residualized with the covariate ("YE") in a regression analysis completely ignoring group membership information.

24

27

What many researchers do not understand is how ANCOVA can make the experimental intervention appear less effective. Figure 2 represents a case in which the covariate ("X") is associated with the dependent variable ("Y"), but not with the assignment to experimental conditions ("A"). In other words, the homogeneity of regression assumption is met.

Table 9 presents a one-way ANOVA corresponding to the Figure 2 Venn diagram. Table 10 presents the related ANCOVA. In this example all the adjustment involving the covariate involves variance in the dependent variable not associated with assignment to experimental conditions. Therefore, the sum of squares for the main effect remains unchanged, but the covariate does reduce the sum of squares for error. This results in a smaller mean square error, and thereby a larger calculated F for the main effect.

INSERT FIGURE 2 AND TABLES 9 AND 10 ABOUT HERE.

But Figure 3 presents a case where the homogeneity of regression assumption is not met. Tables 11 and 12 present the related ANOVA and ANCOVA results, respectively. Although the intervention does has some effect, the application of the covariate in this "worst case" example makes the intervention appear entirely ineffective. Clearly, covariance adjustments can have ^ffects that some researchers do not recognize.

INSERT FIGURE 3 AND TABLES 11 AND 12 ABOUT HERE.

The fact that ANCOVA is simply ANOVA on the residual raw

scores may also be disturbing from an interpretation point of
view. The researcher took a variable that presumably had some
meaning ("ZY"), made an adjustment on it, and was left with an
analysis of a residual raw score that, unlike the original
dependent variable, has little intrinsic meaning. The result
might be difficult to interpret even if the adjustment was
reasonable, i.e., if the homogeneity of regression assumption had
been met.

Too many researchers blindly apply ANCOVA absent an
understanding or either the method's logic or its pivotal
assumptions. As McGuigan (1983, p. 231) has observed, ANCOVA

> can be seriously misused, and one cannot be assured
> that it can "save" a shoddy experiment. Some
> researchers overuse this method as in the instance of
> a person I once overheard asking of a researcher,
> "Where is your an lysis of covariance?"--the
> understanding in his department was that it is <u>always</u>
> used in experimentation.

Of course, the preceeding discussion of the ANCOVA case
generalizes to the various types of statistical control that are
available to researchers.

ANCOVA is <u>not</u> robust to the violation of the homogeneity of
regression assumption, but dissertation students routinely
decline to evaluate this assumption. For example, Scheffe# (2001,
p. 109-110) did not test the assumption, but argued that "The
ANCOVA is a more powerful statistic that ANOVA since it is more
likely to detect true differences between groups (Huck, Cormier,

26
29

& Bounds, 1974)" (p. 107). Pearson# (2001, p. 93) suggested that, "An ANCOVA adds power to this analysis by controlling the within group variability related to teacher with administrator interactions," though it is not clear whether the assumption was tested in the ANCOVA application. Meehl# (2001, p. 52) similarly argued that ANCOVA is very useful but provided no test of homogeneity of regression.

Anastasi# (2001) provides a particularly noteworthy application of ANCOVA. Anastasi# employed one intact class of 30 fourth-grade students as the experimental group, and one intact class of 35 fourth-grade students as a control group. Pretest achievement data (Anastasi#, 2001, p. 79) indicated that the two groups differed by a standardized effect size of roughly 2.0 standard deviations--a huge difference! Fourteen of the 30 experimental group subjects (47%) had been retained in grade at least once (three students repeated grades twice); no control group subjects had been retained.

Anastasi# (2001, p. 78) explains the reason the groups were so systematically different: the students were homogeneously assigned to classes to guarantee (successfully) that the classes would be different. Anastasi# (2001, p. 78) also explains that, "test scores on all subjects were not available prior to the beginning of the study, so the extent of the differences between the experimental and control subjects was not known."

Anastasi# (2001, p. 101) decided that, "Since a statistically significant difference existed between the reading and language achievement levels of the experimental and control groups, an analysis of covariance was computed to equate the

27

groups." Actually, Anastasi# (2001, pp. 124-136) reports a series
of ANCOVAs, but no tests of the homogeneity of regression
assumption.

It is particularly intriguing that Anastasi# (2001, 124-136)
used four covariates, rather than one. As Cliff (1987, p. 278)
explains, "since this is really a form of regression, inferences
become slippier as the variables [covariates] increase" in
number. Furthermore, a "post hoc" analysis employing a t-test of
uncorrected dependent variable scores is somehow employed to
explore ANCOVA results associated with the corrected dependent
variable scores (Anastasi#, 2001, pp. 136-137).

5. <u>Dissertations should reflect the recognition that stepwise analytic methods can lead to seriously distorted interpretations</u>.

Stepwise analytic methods may be among the most popular
research practices employed in both substantive and validity
research. As commonly employed, these methods allow the entry of
predictor variables one step at a time, and at each step the
removal of previously entered variables is also considered. The
methods seem to be somewhat casually employed especially in
regression and discriminant analysis research, though variants
are also available when other techniques are used (cf. Thompson,
1984, pp. 47-51).

With respect to regression applications, Marascuilo and
Serlin (1988, p. 671) note that, "The most popular method in use
for selecting the fewest number of predictor variables necessary
to guarantee adequate prediction is based on a model referred to
as <u>stepwise regression</u>." Huberty (in press) concurs, suggesting

28

31

that "The conduct of analytical procedures in 'steps' is quite common... [These] procedures have enjoyed widespread use by social and behavioral researchers." Unfortunately, stepwise methods can lead to serious misinterpretations of results, and "social science research is replete with misinterpretations of this kind" (Pedhazur, 1982, p. 168).

Three problems with stepwise methods merit special emphasis. First, most researchers, thanks to "canned" computer programs, do not employ the correct degrees of freedom when evaluating changes in explained variance, i.e., usually changes in squared R or lambda. For example, in a stepwise regression analysis, the researcher at step two may add a second predictor variable into a prediction equation. The researcher might test the significance of the change in squared R by an F test using 1 and $n-q-1$ degrees of freedom, where q is the number of predictor variables in the last step. The numerator degrees of freedom reflects a premise that only one additional predictor variable was employed to yield the squared R change, but _ignores_ the fact that the added predictor was selected by consulting empirical sample results involving a larger set of candidates for entry into the prediction process. Thus, the process ignores that fact that, "in a sense, _all_ the variables are in the equation, even though some of them have [effectively] been given zero weights" (Cliff, 1987, p. 187). Consequently, Cliff (1987, p. 185) suggests that "most computer programs for [stepwise] multiple regression are positively satanic in their temptation toward Type I errors."

Second, some researchers incorrectly interpret stepwise

29

results in which q predictor variables have been selected as indicating that the predictor variables are the best variables to use if the predictor variable set is limited to size q. In fact, in a stepwise analysis in which three steps are conducted, and predictors $\underline{A}$, $\underline{B}$, and $\underline{C}$ are employed, it is entirely possible that three different predictors would represent the optimal predictor set of size three. Stepwise methods select the next-best predictor at each step, given the presence of previous predictors--this is not the same as selecting the optimal predictor variable set of size q. As Huberty (in press) notes, "It is generally understood by methodologists that the first q variables entered into either a regression analysis or a discriminant analysis do not necessarily constitute the 'best' subset of size q."

Third, some researchers incorrectly consult order of entry information to evaluate the importance of various predictor variables. As Huberty (in press) explains,

> The first variable entered with a stepwise
> regression analysis is determined by the
> correlation between each predictor variable and
> the criterion variable... The third, say, variable
> to be entered (and often considered to be the
> third most important) is dependent on the two
> variables already entered. If one or two the
> variables already entered would be changed, then
> the third variable entered may also be different.
> This dependence or conditionality truly makes
> var'able importance as determined by stepwise

30

analyses _very_ questionable.

The small data set for a population (_N_=12) presented in Table 13 can be employed to illustrate how sampling error can seriously distort the interpretation of stepwise results involved in predicting dependent variable _ZY_. Table 14 indicates that the three predictor variables share little variance with each other and that the order of predictor variable explanatory power is, respectively, _ZX1_, _ZX2_, _ZX3_, and _ZX4_.

INSERT TABLES 13 AND 14 ABOUT HERE.

Presume that the researcher draws a random sample of nine subjects from the population of 12 persons. Each of 55 random collections of nine subjects (omit subjects 1,2,3; omit 1,2,4; etc.) is equally probable. For these illustrative data, only eight samples (omit 1,2,5; 1,2,7; 2,3,7; 2,3,10; 3,4,5; 5,6,8; 7,8,9; and 7,8,12) enter the four predictor variables in the order that is known to be correct when the true population parameters are consulted.

Indeed, only 23 samples select predictor _ZX1_ as the first prediction entry. Sixteen samples select _ZX2_ as the first entry; 10 samples select _ZX3_ as the first variable entered; six samples select the worst predictor, _ZX4_, as the first or best predictor of _ZY_. For the sample omitting subjects 3, 4 and 9, the predictor variables are entered in the order: _ZX4_, _ZX2_, _ZX3_, and _ZX1_.

Clearly, sampling error can seriously distort stepwise results. As Kachigan (1986, p. 265) argues,

there is the danger that we might select variables

31

34

for inclusion in the regression equation based on chance relationships. Therefore, as stressed in our discussion of multiple correlation, we should apply our chosen regression equation to a fresh sample of objects to see how well it does in fact predict values on the criterion variable. This validation procedure is absolutely essential if we are to have any faith at all in the future applications of the regression equation.

Alternatively, the researcher might employ a cross-validation procedure such as the one recommended by Huck, Cormier and Bounds (1974, p. 159).

Given these considerations, Kerlinger (1986, p. 545) argues that "the research problem and the theory behind the problem [and not stepwise methods] should determine the order of entry of variables in multiple regression analysis." Researchers who choose to employ stepwise methods, particularly if they also fail to use replication or cross-validation methods, might best consider Cliff's (1987, pp. 120-121) argument that "a large proportion of the published results using this method probably present conclusions that are not supported by the data."

Dissertation students are not always aware of these subtleties. For example, Pearson# (2001, p. 92) reports a stepwise regression analysis. But Wilks# (2001, pp. 122-127) presents a more noteworthy application involving six steps of analysis. Wilks# somehow interprets the stepwise multiple regression results (p. 126) in comparison with a bivariate correlation matrix (p. 124) not involving the same subjects. But

32

the comparison is instructive in indicating how stepwise results can lead to interpretation errors. Wilks# (2001, p. 127) reports that, "The results of the stepwise regression indicate that computer experience [r=-0.288] and mathematics anxiety [r=-0.141] contribute significantly to the variance on computer anxiety" dependent variable. The importance of these two predictors was emphasized. Yet a third predictor variable, number of previous math classes, apparently had a larger r [-0.175] with the dependent variable than did math anxiety.

6. <u>Dissertations should reflect a recognition that instrumentation must have psychometric integrity if studies are to yield meaningful results</u>.

It is axiomatic that measurement integrity is vital in quantitative research. As Kerlinger (1986) explains with respect to reliability, for example,

> Since unreliable measurement is measurement overloaded with error, the determination of relations becomes a difficult and tenuous business. Is an obtained coefficient of determination between two variables low because one or both measures are unreliable? Is an analysis of variance F ratio not significant because the hypothesized relation does not exist or because the measure of the dependent variable is unreliable? ...High reliability is no guarantee of good scientific results, but there can be no good scientific results without reliability. (p. 415)

33

Too few authors of published research report measurement statistics regarding their instrumentation, as Willson (1980, p. 9) notes. So it not be too surprising that some dissertation students do not apparently consider these requirements.

For example, Cronbach# (2002, p. 66) provides the following description of the sole instrument employed in the dissertation:

> The survey instrument was designed by the author for this study. It consisted of two parts. The first part contained four demographic questions about the respondent and his/her institution. The second part contained 20 statements about the Consent Decree to which each participant would respond on a 5-point Likert scale. All responses were made on an answer sheet suitable for optical scanning to maximize accurate evaluation of the data.

No information regarding validity or reliability is presented.

Similarly, Cohen# (2001, pp. 129-133) developed an instrument that presumed faculty subjects would be aware of practices of other faculty. The full description of the investigation of this measure's psychometric integrity was rather terse: "Face and content validation of this instrument was obtained by a review of the literature in occupational therapy and jury review" (Cohen#, 2001, p. 52).

Cohen# (2001, p. 56) also developed a new genre of hypothesis substance and testing logic:

> Sub-hypothesis 1.2: There are no discernible

attitudes [sic] of occupational therapy faculty as they relate to the computer as a threat to society. Category two of the ATCQ-OT, Computer Threat to Society provided a basis for investigating this sub-hypothesis. The resulting mean response of 10.880 was within ten percent of the middle rating of 12 [the scale midpoint] on this factor, as measured by four items. Such an insignificant variance does not permit rejection of this null sub-hypothesis.

7. <u>Dissertations should reflect a recognition that it is not desirable to present statistically nonsensical or impossible results</u>.

Dissertations are more than a test of the student's ability to conduct original and independent research. Dissertations make critical contributions to the scholarly literature, for few contributions can reasonably be expected to involve as much thought and work or the pooled talent of as many scholars as are theoretically represented by the combination of the student and the members of the dissertation committee. Thus, it is not desirable to report nonsensical or impossible quantitative results that call into question the integrity of the remainder of a project as well.

But dissertations do occasionally report just such results. For example, Scheffe# (2001, p. 109) reported that,

> Since the smallest cell size in the initial ANCOVA
>
> was 85, 85 subjects were randomly selected for
>
> each cell. The results of the ANCOVA indicated a

35

significant difference among the adjusted mean scores of the four groups of subjects on desired knowledge about computers [$F(4,335) = 2.64$, $p<.05$].

The covariate effect size is not reported (p. 110). The researcher pooled the sum of squares for the main effect together with the sum of squares for the covariate, thus completely confounding interpretations of both effects. Furthermore, the confounded result appears to be interpreted as being solely due to the main effect.

But a more disturbing example is provided by Cronbach# (2001, pp. 102-104), who reported a factor analysis in which 15 factors were extracted, each ostensibly involving a different number of iterations. Regardless of whether iterations were employed in estimating communalities or in rotation, the number of iterations is one value for the entire solution. Thus, the result does not appear to be plausible.

Summary

Dissertations are the cumulative, tangible "best evidence" of interests of doctoral faculty and students in serious and incisive scholarship. Thus, dissertations are thoroughly studied by the program review teams periodically hired by boards of higher education in most states. The present paper explored seven errors in quantitative analysis in both published literature and dissertations. The errors are explained in detail using references to works by other authors and small hypothetical data sets to illustrate problems. Concrete examples of the errors as

36

they occur in dissertations are cited to make clear that the errors are not hypothetical. Ten dissertations, completed since January of 1985, are cited as examples, although pseudonyms are employed to avoid embarrasment of the students or their doctoral advisors. The discussion should also be useful to authors of published research who may wish to avoid the same errors.

References

Atkinson, D. R., Furlong, M. J., & Wampold, B. E. (1982). Statistical significance, reviewer evaluations, and the scientific process: Is there a (statistically) significant relationship? Journal of Counseling Psychology, 29, 189-194.

Borg, W.R. (1983). Educational research: An introduction (4th ed.). New York: Longman.

Brown, D., Cooper, J.M., Griffiths, D.E., Howey, K.R., & Lilly, S. (1985). Report to the Louisiana Board of Regents on the doctoral programs in education at the University of New Orleans, Northwestern State University, and Louisiana State University. Baton Rouge, LA: Louisiana Board of Regents.

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.

Campo, S.F. (1988, November). Alternative logics for estimating whether results will generalize. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY.

Carlson, R. (1976). Discussion: The logic of tests of significance. Philosophy of Science, 43, 116-128.

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

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

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, L. H. (1979). Clinical psychologists' judgments of the scientific merit and clinical relevance of psychotherapy outcome research. Journal of Consulting and Clinical Psychology, 47, 421-423.

Craig, J. R., Eison, C. L., & Metze, L. P. (1976). Significance tests and their interpretation: An example utilizing published research and omega-squared. Bulletin of the Psychonomic Society, 7, 280-282.

Crask, M.R., & Perreault, W.D., Jr. (1977). Validation of discriminant analysis in marketing research. Journal of Marketing Research, 14, 60-68.

Cronbach, L.J. (1975). Beyond the two disciplines of psychology. American Psychologist, 30, 116-127.

38

Dar, R. (1987). Another look at Meehl, Lakatos, and the scientific practices of psycholigists. American Psychologist, 42, 145-151.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-130.

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

Elmore, P.B., & Woehlke, P.L. (1988, April). Research methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

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

Fisher, R. A., & Yates, F. (1957). Statistical tables for biological, agricultural and medical research (5th ed.). New York: Hafner Publishing Co.

Gaither, N., & Glorfeld, L. (1985). An evaluation of the use of tests of significance in organizational behavior research. Academy of Management Review, 10, 787-793.

Games, P. A. (1971a). Errata for "Multiple comparisons on means," AERJ, 1971, 531-565. American Educational Research Journal, 8, 677-678.

Games, P. A. (1971b). Multiple comparisons on means. American Educational Research Journal, 8, 531-565.

Gergen, K.J. (1969). The psychology of behavior exchange. Reading: Addison-Wesley.

Glasnapp, D. R., & Poggio, J. P. (1985). Essentials of statistical analysis for the behavioral sciences. Columbus, OH: Merrill.

Glass, G.V. (1979). Policy for the unpredictable (uncertainty research and policy). Educational Researcher 8(9), 12-14.

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.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Gravetter, F. J., & Wallnau, L. B. (1985). Statistics for the

behavioral sciences. St. Paul, MN: West.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. Psychological Bulletin, 82, 1-20.

Guilford, J. P. (1965). Fundamental statistics in psychology and education (4th ed.). New York: McGraw-Hill.

Hays, W. L. (1981). Statistics (3rd ed). New York: Holt, Rinehart and Winston.

Hicks, C. R. (1973). Fundamental concepts in the design of experiments. New York: Holt, Rinehart and Winston.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1979). Applied statistics for the behavioral sciences. Chicago: Rand McNally.

Hopkins, C.D. (1980). Understanding educational research: An inquiry approach. New York: Holt, Rinehart and Winston.

Huberty, C.J. (1987). On statistical testing. Educational Researcher, 16(8), 4-9.

Huberty, C. J. (in press). Problems with stepwise methods--better alternatives. In B. Thompson (Ed.), Advances in social science methodology (vol. 1). Greenwich, CT: JAI Press.

Huck, S. W., Cormier, W. H., & Bounds, Jr., W. G. (1974). Reading statistics and research. New York: Harper and Row.

Kachigan, S. K. (1986). Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods (2nd ed.). New York: Radius Press.

Kepp ., G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Kerlinger, F.N. (1977). The influence of research on educational practice. Educational Researcher, 6(8), 5-12.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart & Winston.

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

Kirk, R. E. (1984). Elementary statistics. Monterey, CA: Brooks/Cole.

Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. American Psychologist, 43, 635-642.

Loftus, G. R., & Loftus, E. F. (1982). Essence of statistics. Monterey, CA: Brooks/Cole.

Lord, F. M. (1960). Large sample covariance analysis when the control variable is fallible. _Journal of the American Statistical Association, 55,_ 309-321.

Marascuilo, L. A., & Serlin, R. C. (1988). _Statistical methods for the social and behavioral sciences._ New York: W. H. Freeman.

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

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

Minium, W. W., & Clarke, R. B. (1982). _Elements of statistical reasoning._ New York: John Wiley and sons.

Moore, G.W. (1983). _Developing and evaluating educational research._ Boston: Little, Brown.

Morrison, D.E., & Henkel, R.E. (Eds.). (1970). _The significance test controversy._ Chicago: Aldine.

Nunnally, J. C. (1975). Psychometric theory--25 years ago and now. _Educational Researcher, 4_(10), 7-14, 19-20.

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

Persell, C.H. (1976). _Quality, careers and training in educational and social research._ Bayside, NY: General Hall.

Peters, D.P., & Ceci, S.J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. _The Behavioral and Brain Sciences, 5,_ 187-255.

Schneider, A. L., & Darcy, R. E. (1984). Policy implications of using significance tests in evaluation research. _Evaluation Review, 8,_ 573-582.

Scriven, M. (1980). Self-referent research. _Educational Researcher, 9_(6), 11-18, 30.

Shaver, J.P. (1985). Chance and nonsense. _Phi Delta Kappan, 67_(1), 57-60.

Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). _Human experimental psychology._ New York: Oxford University Press.

Sowell, E. J., & Casey, R. J. (1982). _Research methods in education._ Belmont, CA: Wadsworth.

Thompson, B. (1984). _Canonical correlation analysis: Uses and_

41

44

interpretation. Beverly Hills: SAGE.

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1986a). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

Thompson, B. (1986b, November). Two reasons why multivariate methods are usually vital. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis.

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, DC. (ERIC Document Reproduction Service No. ED 287 868)

Thompson, B. (1988a). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.

Thompson, B. (1988b). Misuse of chi-square contingency table test statistics. Educational and Psychological Research, 8, 39-49.

Thompson, B. (1988c). A note about significance testing. Measurement and Evaluation in Counseling and Development, 20, 146-147.

Thompson, B. (in press). The place of qualitative research in contemporary social science. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1). Greenwich, CT: JAI Press.

Wandt, E. (Ed.). (1965). A cross-section of educational research. New York: David McKay.

Ward, N.E.W., Hall, B.W., & Schramm, C.F. (1975). American Educational Research Journal, 12, 109-128.

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

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Witte, R.S. (1985). <u>Statistics</u> (2nd ed.). New York: Holt, Rinehart and Winston.

46

Table 1
Statistical Significance at Various Sample Sizes
for a Fixed Effect Size (Moderate Effect Size)

Source	SOS	r sq	df	MS	F calc	F crit	Decision
SOSexp	98.7	0.098473	2	49.35	0.054614	200.00	Not Rej
SOSunexp	903.6		1	903.6			
SOStot	1002.3		3	334.1			
SOSexp	98.7	0.098473	2	49.35	0.546148	4.10	Not Rej
SOSunexp	903.6		10	90.36			
SOStot	1002.3		12	83.525			
SOSexp	98.7	0.098473	2	49.35	1.092297	3.49	Not Rej
SOSunexp	903.6		20	45.18			
SOStot	1002.3		22	45.55909			
SOSexp	98.7	0.098473	2	49.35	1.638446	3.32	Not Rej
SOSunexp	903.6		30	30.12			
SOStot	1002.3		32	31.32187			
SOSexp	98.7	0.098473	2	49.35	2.184594	3.23	Not Rej
SOSunexp	903.6		40	22.59			
SOStot	1002.3		42	23.86428			
SOSexp	98.7	0.098473	2	49.35	2.730743	c3.19	Not Rej
SOSunexp	903.6		50	18.072			
SOStot	1002.3		52	19.275			
SOSexp	98.7	0.098473	2	49.35	3.276892	3.15	Rej
SOSunexp	903.6		60	15.06			
SOStot	1002.3		62	16.16612			
SOSexp	98.7	0.098473	2	49.35	6.553784	3.07	Rej
SOSunexp	903.6		120	7.53			
SOStot	1002.3		122	8.215573			

Table 2
Statistical Significance at Various Sample Sizes
for a Fixed Effect Size (Larger Effect Size)

Source	SOS	r sq	df	MS	F calc	F crit	Decision
SOSexp	337.2	0.336426	2	168.6	0.253495	200.00	Not Rej
SOSunexp	665.1		1	665.1			
SOStot	1002.3		3	334.1			
SOSexp	337.2	0.336426	2	168.6	2.534957	4.10	Not Rej
SOSunexp	665.1		10	66.51			
SOStot	1002.3		12	83.525			
SOSexp	337.2	0.336426	2	168.6	5.069914	3.49	Rej
SOSunexp	665.1		20	33.255			
SOStot	1002.3		22	45.55909			
SOSexp	337.2	0.336426	2	168.6	7.604871	3.32	Rej
SOSunexp	665.1		30	22.17			
SOStot	1002.3		32	31.32187			
SOSexp	337.2	0.336426	2	168.6	10.13982	3.23	Rej
SOSunexp	665.1		40	16.6275			
SOStot	1002.3		42	23.86428			
SOSexp	337.2	0.336426	2	168.6	12.67478	c3.19	Rej
SOSunexp	665.1		50	13.302			
SOStot	1002.3		52	19.275			
SOSexp	337.2	0.336426	2	168.6	15.20974	3.15	Rej
SOSunexp	665.1		60	11.085			
SOStot	1002.3		62	16.16612			
SOSexp	337.2	0.336426	2	168.6	30.41948	3.07	Rej
SOSunexp	665.1		120	5.5425			
SOStot	1002.3		122	8.215573			

Table 3
"Testwise" and "Experimentwise" Error Rates for Selected Studies

"Testwise" Rate	Minimum	"Experimentwise" Rate		n of Tests	Maximum
05.0%	05.0%	1 - (- 05.0%)	**	1	=
05.0%	05.0%	1 - (95.0%)	**	1	=
05.0%	05.0%	1 - 95.0%			= 05.00%
05.0%	05.0%	1 - (- 05.0%)	**	5	= 22.62%
05.0%	05.0%	1 - (- 05.0%)	**	10	= 40.13%
05.0%	05.0%	1 - (- 05.0%)	**	20	= 64.15%

Note. An alpha of 0.05 equals an alpha of 05.0%. "**" means "raised to the power of". The first several rows of the table illustrate the that "testwise" and "experimentwise" error rates are the same when only one test is conducted.

Table 4
Hypothetical Validity Study Data

Group	ID	DV	C1	C2	C3	C4	C5
1	1	10	-1	-1	-1	-1	-1
	2	20	-1	-1	-1	-1	-1
2	3	10	1	-1	-1	-1	-1
	4	20	1	-1	-1	-1	-1
3	5	10	0	2	-1	-1	-1
	6	20	0	2	-1	-1	-1
4	7	10	0	0	3	-1	-1
	8	20	0	0	3	-1	-1
5	9	10	0	0	0	4	-1
	10	20	0	0	0	4	-1
6	11	25	0	0	0	0	5
	12	35	0	0	0	0	5

Table 5
Hypothetical ANCOVA Data Set

Group	ZY	ZX	ZYZX	YHAT	YE
A	-.88	-1.68	1.48	-1.36	.48
A	-.44	-.68	.30	-.56	.11
A	.00	.31	.00	.25	-.25
A	.44	1.30	.57	1.06	-.62
B	-1.32	-.68	.90	-.56	-.77
B	-.44	-.19	.08	-.15	-.29
B	.88	.56	.49	.45	.43
B	1.76	1.06	1.86	.86	.91

Note. The beta weight for the covariance procedure (.813) equals the sum of the cross products (ZXZY) of ZX and ZY divided by $n-1$ (5.694/$n-1$). The predicted posttest score (YHAT) is each child's pretest (ZX) multiplied by the beta weight. The error in each prediction (YE) is equal to ZY minus YHAT.

Table 6
Conventional ANOVA Results

Source	Sum of Squares	df	Mean Squares	F	Effect Size
Treatment	.39	1	.39	.35	.056
"Error"	6.61	5	1.10		
Total	7.00	7	1.00		

Note. Effect size is a r squared analog.

Table 7
ANCOVA Results

Source	Sum of Squares	df	Mean Squares	F	Effect Size
Covariate	4.63	1	4.63	9.95	.661
Treatment	.04	1	.04	.08	.006
"Error"	2.33	5	.47		
Total	7.00	7	1.00		

<u>Note</u>. Effect size is a <u>r</u> squared analog.

Table 8
ANOVA Results Using YE as Dependent Variable

Source	Sum of Squares	df	Mean Squares	F	Effect Size
Treatment	.04	1	.04	.08	.006
"Error"	2.33	5	.47		
Total	2.37	6			

Table 9
ANOVA Associated with Figure 2

Source	Sum of Squares	df	Mean Squares	F Calc	F Crit
Treatment	35	1	35.00	4.85	5.12
"Error"	65	9	7.22		
Total	100	10			

Table 10
ANCOVA Associated with Figure 2

Source	Sum of Squares	df	Mean Squares	F Calc	F Crit
Covariate	20	1	20.0		
Treatment	35	1	35.00	6.22	5.32
"Error"	45	8	5.62		
Total	100	10			

Table 11
ANOVA Associated with Figure 3

Source	Sum of Squares	df	Mean Squares	F Calc	F Crit
Treatment	20	1	20.00	2.25	5.12
"Error"	80	9	8.89		
Total	100	10			

Table 12
ANCOVA Associated with Figure 3

Source	Sum of Squares	df	Mean Squares	F Calc	F Crit
Covariate	30	1	30.0		
Treatment	0	1	.00	.00	5.32
"Error"	70	8	8.75		
Total	100	10			

Table 13
Standardized Data for Five Variables

ID	ZY	ZX1	ZX2	ZX3	ZX4
1	.790	1.422	.350	.322	-.313
2	-1.589	.112	-1.239	-1.094	-.365
3	.127	-.265	.271	.201	-.060
4	-1.656	-2.167	-.498	-.970	.218
5	.176	-1.291	.153	2.393	.159
6	-.017	.636	-1.607	-.168	-1.746
7	-.397	-.173	.931	-.112	-1.704
8	-.594	.532	-.108	.092	.127
9	.846	.528	1.237	-.092	.035
10	.810	.642	-1.400	1.135	1.654
11	1.764	.373	1.290	-.543	1.005
12	-.260	.352	.620	-1.163	.989

Table 14
Bivariate Correlation Matrix

	ZY	ZX1	ZX2	ZX3	ZX4
ZY		.497	.444	.384	.319
ZX1	24.7%		.018	-.074	-.004
ZX2	19.7%	.0%		-.054	.099
ZX3	14.7%	.5%	.3%		.103
ZX4	10.2%	.0%	1.0%	1.1%	

Note. Bivariate r coefficients are presented above the diagonal. Common variance (squared r) percentages are presented below the diagonal.
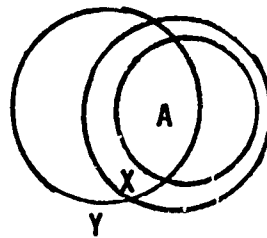
Figure 1
Scattergram of ANCOVA Data
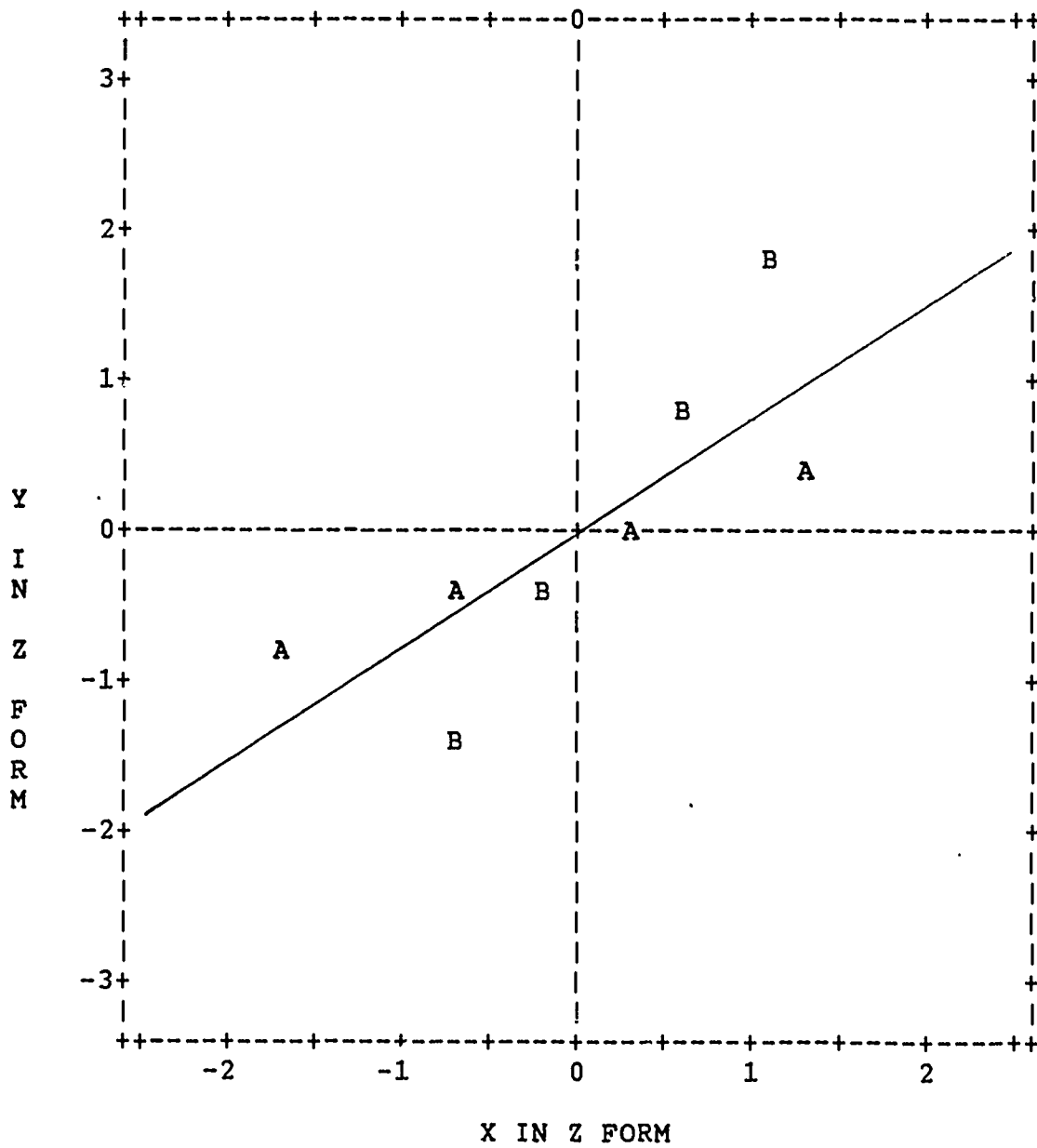
Figure 2

ANCOVA Best Case
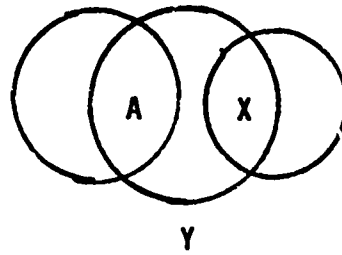
Figure 3

ANCOVA Worst Case

50

APPENDIX A:
The Population of Dissertations
from Which Examples were Selected

Alexis, R.N. (1986). A description and analysis of the perceived impact of the academic governance structure of higher education in the State of Louisiana after the adoption of the Constitution of 1974 (Doctoral dissertation, University of New Orleans, 1985). Dissertation Abstracts International, 46, 2200A. (University Microfilms No. 85-16,301)

Almerico, M.B.D. (1986). An investigation of educators' attitudes toward the Louisiana professional improvement program and its impact on education in Louisiana (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 1587A. (University Microfilms No. 86-17,837)

Andersson, B.V. (1982). Creativity as a mediating variable in inferential reading comprehension (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 2993A. (University Microfilms No. 81-29,909)

Aramburo, D.J. (1982). The efficacy of the structured interview in the selection of special education teachers (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 2604A. (University Microfilms No. 81-25,880)

Austin, P.J. (1987). The effects of process instruction and teacher feedback on revision of writing tasks by fourth grade students within the naturalistic context of the classroom (Doctoral dissertation, University of New Orleans).

Banbury, M.M. (1985). Parents of learning disabled students: Attitudes toward the academic and socio-emotional aspects of mainstreaming (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 45, 2062A. (University Microfilms No. 84-17,612)

Bartels, K.J. (1987). Effects of socio-dramatic play therapy on pragmatic language skills in children with down syndrome (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 2866B. (University Microfilms No. 86-25,147)

Barton, L.T. (1985). Leadership behavior and climate as they relate to the effectiveness of elementary schools (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 45, 2316A. (University Microfilms No. 84-25,266)

Beall, J.H. (1986). Brain periodization: An investigation of Epstein's distribution of Pragetian stages in five, six and

seven year old children (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 46, 2186A. (University Microfilms No. 85-07,561)

Becnel, S.A. (1983). Generalizability to school administrators of a professional concerns model (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 2168A. (University Microfilms No. 82-27,568)

Borrello, G.M. (1984). Personality and aptitude correlates of selected test-wiseness skills (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 45, 65A. (University Microfilms No. 84-08,136)

Brooks, C.R. (1984). Administrative decision-making and program evaluation use (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 44, 2301A. (University Microfilms No. 83-28,208)

Busenbark, M.M. (1985). Information use in educational decisions of administrators, teachers, and parents (Doctoral dissertation, University of New Orleans, 1985). Dissertation Abstracts International, 46, 1451A. (University Microfilms No. 85-16,302)

Carlton, M.B. (1982). The effects of an intraclass peer tutoring program on the sight-word recognition ability of mildly mentally retarded students (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 2606A. (University Microfilms No. 81-25,881)

Carter, M.M. (1985). The effect of implementing the Consent Decree on academic programs in Louisiana colleges and universities (A survey of the perceptions of administrators, faculty, and students) (Doctoral dissertation, University of New Orleans, 1985). Dissertation Abstracts International, 46, 1527A. (University Microfilms No. 85-16,303)

Casbergue, R.M. (1985). An investigation of the Cloze Reading Inventory as a qualitative and quantitative measure of the reading proficiency of selected suburban school children (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 46, 110A. (University Microfilms No. 85-07,562)

Chacko, H.E. (1987). Administrators' methods of upward influence, goals, motivational needs and perceptions of supervisors' leadership style (Doctoral dissertation, University of New Orleans).

Clausing, C.S. (1988). The effects of paper or CRT modality factors on reading rate and comprehension of eighth grade students (Doctoral dissertation, University of New Orleans).

52

Cunningham, S. (1988). An investigation of computer technology in occupational therapy education (Doctoral dissertation, University of New Orleans).

Currie, P.S. (1986). Reading and writing miscues and revisions of selected learning disabled students (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 1685A. (University Microfilms No. 86-17,838)

Dermody, M.M. (1987). Effects of metacognitive strategy training on fourth graders' reading comprehension (Doctoral dissertation, University of New Orleans, 1987). Dissertation Abstracts International, 48, 1159A. (University Microfilms No. 87-17,055)

Ducote, K.J. (1982). Correlates of selected test-wiseness skills (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 1888A. (University Microfilms No. 82-23,262)

Dumas, A.P. (1982). An investigation of the relationship between leadership style and philosophical orientation of elementary and secondary school principals (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 1872A. (University Microfilms No. 81-23,012)

Faust, F.L. (1981) An analysis of teacher and student acceptance of program innovations through evaluation of attitudes toward curriculum change and schedule change (Doctoral dissertation, University of New Orleans, 1980). Dissertation Abstracts International, 41, 4956A. (University Microfilms No. 81-12,898)

Feola, E.K. (1983). An investigation of the relationship between teacher perceived leader behavior and principal sex role characteristics (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 2842A. (University Microfilms No. 82-23,263)

Frere, R.M. (1987). Teachers' perceptions of principals' logic of confidence and leadership styles and structural coping in schools (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 4250A. (University Microfilms No. 87-07,940)

Garcia-Otero, M.D.L. (1987). Knowledge of learning styles and the effect on the clinical performance of nurse anesthesiology students (Doctoral dissertation, University of New Orleans).

Garlington, W.K. (1984). An examination of the relationship among a student's self-concept, level of anxiety, and attitude toward school (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 45,

5.?

1342A. (University Microfilms No. 84-17,613)

Gaudin, F.A. (1982). The effect of alternative biology-chemistry sequences on ACT performance of college-bound students (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 1914A. (University Microfilms No. 82-25,264)

Green, R.B. (1985). The relationship of teachers' perceptions of their principal's managerial style to teachers' self-reported pupil control ideology and students' perceptions of teachers' pupil control behavior (Doctoral dissertation, University of New Orleans, 1985). Dissertation Abstracts International, 46, 1458A. (University Microfilms No. 85-16,304)

Greene, P.J.F. (1982). The morpheme conceptualization barrier: Adult non-inflected language speakers' acquisition of English morpheme structures (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 4816A. (University Microfilms No. 82-08,060)

Guice, L.J. (1986). A study of the relationship between self-esteem and achievement motivation in vocational-technical school students (Doctoral dissertation, University of New Orleans, 1985). Dissertation Abstracts International, 46, 2904A. (University Microfilms No. 85-27,069)

Harris, P.W. (1984). An investigation of the attitudes of business administration graduates of Southern University in New Orleans from 1978-1981 toward their undergraduate preparation (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 45, 4594A. (University Microfilms No. 84-08,137)

Harwood, W. (1987). The relationship between job satisfaction and life satisfaction of educators in an urban public school system (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 2378A. (University Microfilms No. 86-25,149)

Heausler, N.L. (1988). The effects of dance/movement as a learning medium on the acquisition of selected word analysis concepts and the development of creativity of kindergarten and second grade children (Doctoral dissertation, University of New Orleans, 1987). Dissertation Abstracts International, 48, 2233A. (University Microfilms No. 87-26,761)

Jackson, S.C. (1984). Analysis of social-perceptual ability in learning disabled and nonhandicapped adolescents (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 45, 147A. (University Microfilms No. 84-09,292)

Jeansonne, C.J. (1982). The relationship between student perceptions of teachers and teacher concerns and anxieties

54

(Doctoral dissertation, University of New Orleans, 1981). *Dissertation Abstracts International*, 42, 2502A. (University Microfilms No. 81-25,882)

Kappelman, S.K. (1982). Teachers' perceptions of principals' bases of power in relation to principals' styles of leadership (Doctoral dissertation, University of New Orleans, 1981). *Dissertation Abstracts International*, 42, 2405A. (University Microfilms No. 81-25,884)

Kaul, C.J.E., Jr. (1987). Access to privacy of school records: An investigation of the implementation of the family educational rights and privacy act (PL93-380) in Louisiana (Doctoral dissertation, University of New Orleans, 1986). *Dissertation Abstracts International*, 47, 2398A. (University Microfilms No. 86-25,150)

Kelley, R.H. (1984). Differences and patterns of speech and language development in selected high-risk infants (Doctoral dissertation, University of New Orleans, 1983). *Dissertation Abstracts International*, 44, 2717B. (University Microfilms No. 83-28,209)

Kirby, P.C. (1987). Reflective practice as a predictor of teacher effectiveness (Doctoral dissertation, University of New Orleans).

Kreig, P.E. (1986). Effects of systematic training for effective parenting (STEP) on perception and attitude of parents of learning disabled children (Doctoral dissertation, University of New Orleans, 1985). *Dissertation Abstracts International*, 46, 1922A. (University Microfilms No. 85-16,305)

Lane, E.R. (1981). The relationship between selected demographic variables and teacher strike participation and between teachers' needs levels and their inclination to strike (Doctoral dissertation, University of New Orleans, 1980). *Dissertation Abstracts International*, 41, 4913A. (University Microfilms No. 81-12,899)

Levitov, J.E. (1981). A computer-based semantic content analysis of examplars of five counseling theories (Doctoral dissertation, University of New Orleans, 1980). *Dissertation Abstracts International*, 42, 143A. (University Microfilms No. 81-13,580)

Lumpkin, W.C. (1981). The relationships of marital communication and marital adjustment with self-esteem and self-actualization (Doctoral dissertation, University of New Orleans, 1981). *Dissertation Abstracts International*, 42, 2883A. (University Microfilms No. 81-26,222)

McCulloch, J.M. (1982). An examination of the concerns theory and its application in physical therapy education (Doctoral dissertation, University of New Orleans, 1981). *Dissertation*

55

Abstracts International, 42, 2528A. (University Microfilms No. 81-25,885)

McDonough, D.D. (1987). A comparison of spoken language abilities in children with aphasia, autism, learning disabilities, and normal language abilities (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 2542A. (University Microfilms No. 86-25,151)

McLin, J. (1987). Coherence and cohesion in the writing of eighth grade students (Doctoral dissertation, University of New Orleans).

McMurray, M.A. (1988). Biology analytical problem solving interventions and field independence as predictors of critical thinking changes (Doctoral dissertation, University of New Orleans).

Melancon, J.G. (1981). Teachers' and principals' perceptions of teacher brinkmanship behaviors (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 1882A. (University Microfilms No. 81-23,013)

Melville, G.A. (1985). Shaping on-task-behavior of artistic students: A combined multiple baseline/changing criterion design (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 46, 123A. (University Microfilms No. 85-07,563)

Meza, J., Jr. (1982). The relationship of stressful teaching events, teacher anxiety and teacher behavior as perceived by students (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 1882A. (University Microfilms No. 81-23,017)

Miederhoff, P.A. (1985). The effect of personality and work values on career choice and job satisfaction in pharmacy. Unpublished doctoral dissertation, University of New Orleans.

Miller, J.H. (1935). Demographic personality and experiential correlates of death anxiety of educators (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 46, 128A. (University Microfilms No. 85-07,564)

Miller, W.J. (1983). A study of principal contingency leadership in elementary school organizations (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 2189A. (University Microfilms No. 82-27,570)

Morris, P.D. (1987). Comparison of teacher-assisted and computer-assisted instruction with learning disabled high school students (Doctoral dissertation, University of New Orleans,

56

1986). <u>Dissertation Abstracts International</u>, <u>47</u>, 2543A. (University Microfilms No. 86-25,15?)

Mowen, L.O. (1987). The effects of instruction and practice in test-taking techniques upon the test anxiety assessment levels and standardized reading achievement scores of selected fifth grade students (Doctoral dissertation, University of New Orleans, 1986). <u>Dissertation Abstracts International</u>, <u>47</u>, 4261A. (University Microfilms No. 87-05,866)

O'Connor, N.M. (1988). Effects of four instructional conditions on generalized shopping skills of students with moderate/severe handicaps (Doctoral dissertation, University of New Orleans. 1987). <u>Dissertation Abstracts International</u>, <u>48</u>, 2312A. (University Microfilms No. 87-26,762)

Phillips, K.D. (1982). An investigation of the Louisiana school board presidents' and Louisiana superintendents' perceptions of the superintendent's ideal and actual leadership role (Doctoral dissertation, University of New Orleans, 1981). <u>Dissertation Abstracts International</u>, <u>42</u>, 4676A. (University Microfilms No. 82-08,061)

Pitts, M.M. (1982). Effects of congitive styles on inferential reading comprehension (Doctoral dissertation, University of New Orleans, 1982). <u>Dissertation Abstracts International</u>, <u>43</u>, 1910A. (University Microfilms No. 82-23,266)

Poole, M.G. (1987). Implementing change: The effects of teacher efficacy and interactions among educators (Doctoral dissertation, University of New Orleans).

Quain, W.J. (1982). <u>An examination of an integrated confrontation model for use by food service supervisors in the performance appraisal of employees</u>. Unpublished doctoral dissertation, University of New Orleans.

Rhoda, D.M. (1987). A meta-analysis of the effects of demographic, academic and non-academic factors on student retention 1970-1984 (Doctoral dissertation, University of New Orleans, 1986). <u>Dissertation Abstracts International</u>, <u>48</u>, 275A. (University Microfilms No. 87-08,059)

Richards, J.C. (1986). Theoretical orientation and first and third grade teachers' reading instruction (Doctoral dissertation, University of New Orleans, 1985). <u>Dissertation Abstracts International</u>, <u>46</u>, 2983A. (University Microfilms No. 85-27,070)

Richardson, R.C. (1981). Ego development and power base reliance of school principals (Doctoral dissertation, University of New Orleans, 1981). <u>Dissertation Abstracts International</u>, <u>42</u>, 2421A. (University Microfilms No. 81-25,886)

Rosenberg, R.H. (1987). Text mapping as procedural facilitation

57

for teaching argument to basic writers in a community college (Doctoral dissertation, University of New Orleans).

Sabrio, L.D. (1982). The effectiveness of parent training on the language skills of preschool handicapped children (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 1832A. (University Microfilms No. 82-23,267)

Scafidel, R.J. (1983). An analysis of the relationship between teachers perceived leadership style of the principal and the implementation of continuous progress (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 2850A. (University Microfilms No. 82-23,268)

Schulz, I.L. (1986). A study of the relationship of teachers' job satisfaction and their perceptions of principals' use of power and organizational effectiveness. Unpublished doctoral dissertation, University of New Orleans.

Scott, C. (1987). Student, teacher and principal expectations and attributed responsibility as predictors of achievement: A casual modeling approach (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 4261A. (University Microfilms No. 87-07,942)

Shannon, M.T. (1987). Attitudes toward computers and desired level of knowledge about computers: A comparative study of educators and students in master, baccalaureate, and associate nursing programs (Doctoral dissertation, University of New Orleans).

Sisterhen, D.H. (1986). Auditory, visual, and multi-sensory nonverbal social perception in learning disabled and nonhandicapped adolescents (Doctoral dissertation, University of New Orleans, 1985). Dissertation Abstracts International, 46, 1597A. (University Microfilms No. 85-16,307)

Slate, N.M. (1982). The adaptive behavior inventory for children, the behavior rating profile, and the Vineland Social Maturity Scale in nonbiased assessment of mentally retarded children (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 2628A. (University Microfilms No. 81-25,887)

St. Peter, S.M. (1987). Investigating the effectiveness of classroom instruction for adolescents/young adults with autism and/or severe language disorders: A comparison of computer-assisted versus human instruction (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 2544A. (University Microfilms No. 86-25,155)

Stack, E.G. (1987). The effects of comprehension of nonverbal

communication, self-concept, and race in learning disabled children (Doctoral dissertation, University of New Orleans).

Steedley, J.C. (1981). Facilities operation and maintenance: A study in the use of formula funding for the operation and maintenance of physical plant organizations in public institutions of higher education in Louisiana (Doctoral dissertation, University of New Orleans, 1980). Dissertation Abstracts International, 42, 46A. (University Microfilms No. 81-13,581)

Stuart, J.L. (1986). A comparison of the stages of moral development of white female gifted students, white female regular students, and teachers' perceptions of students' stages of moral development (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 1683A. (University Microfilms No. 86-16,910)

Sullivan, V.G. (1984). Standardization of the Cloze Reading Inventory as a measure of reading comprehension in the odd grades 3 through 11 (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 44, 2425A. (University Microfilms No. 83-28,214)

Taylor, J.A. (1982). A study of the relationship of the self-concepts of high school seniors to college-bound and work-bound plans (Doctoral dissertation, University of New Orleans, 1981). Dissertation Abstracts International, 42, 2061A. (University Microfilms No. 81-23,015)

Texidor, M.S. (1984). Knowledge, attitudes and behaviors: The effects of a seminar in wellness, holistic assessment, and collaborative approach on professional school social workers (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 44, 2369A. (University Microfilms No. 83-28,215)

Theriot, J.G. (1983). Factors affecting the vocational home economics teachers integration of Future Homemakers of America into the home economics curriculum (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 3828A. (University Microfilms No. 83-10,891)

Thieneman, L.M. (1987). Managerial behaviors of secondary social studies and English teachers and their relationship to student on-task behavior (Doctoral dissertation, University of New Orleans, 1986). Dissertation Abstracts International, 47, 2442A. (University Microfilms No. 86-25,156)

Thomas, P.A. (1984). An investigation of the knowledge of wellness, holistic assessment, and collaborative approach among public school counselors (Doctoral dissertation, University of New Orleans, 1983). Dissertation Abstracts International, 44, 2682A. (University Microfilms No. 83-

59

28,216)

Torrey, C.C. (1987). Environmental effects on the social play behavior of handicapped and nonhandicapped preschoolers (Doctoral dissertation, University of New Orleans).

Townsend, J., II (1988). Media selection in medical education: The development of a frame of reference (Doctoral dissertation, University of New Orleans, 1987). Dissertation Abstracts International, 48, 2237A. (University Microfilms No. 87-26,757)

Virgilio, I.R. (1987). An examination of the relationships among school effectiveness, time-on-task, and teacher effectiveness in elementary and junior high schools (Doctoral dissertation, University of New Orleans).

Williams, A. (1985). An analysis of perceived effectiveness of written communication from central office administrators to elementary level administrators in a metropolitan school district (Doctoral dissertation, University of New Orleans, 1984). Dissertation Abstracts International, 46, 44A. (University Microfilms No. 85-07,566)

Wilson, B.A. (1987). Prevalence of computer anxiety and factors which affect computer anxiety in associate and baccalaureate nursing students (Doctoral dissertation, University of New Orleans).

Zeigler, V.F. (1983). A study of the relationship of principals' self-report humor scores and their leadership styles as perceived by teachers (Doctoral dissertation, University of New Orleans, 1982). Dissertation Abstracts International, 43, 2197A. (University Microfilms No. 82-27,571)