

INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.
- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.
- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

8706188

Meltzer, Richard J.

THE EFFICACY OF WITHIN-SUBJECT MEASUREMENT

Wayne State University

PH.D. 1986

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1986

by

Meltzer, Richard J.

All Rights Reserved

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark .

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other _____

University
Microfilms
International

© COPYRIGHT BY
RICHARD J. MELTZER
1986
All Rights Reserved

TABLE OF CONTENTS

	<u>Page</u>
LIST OF TABLES	iv
LIST OF FIGURES.....	v
 CHAPTER	
I. STATEMENT OF THE PROBLEM.....	1
Introduction.....	1
Purpose.....	9
II. REVIEW OF THE LITERATURE.....	15
Summary.....	35
III. RATIONALE OF THE DEPENDENT MEASURE.....	37
IV. METHODS AND PROCEDURES.....	42
Introduction.....	42
General Procedure.....	42
Excitation Threshold Methodology.....	47
Experimental Design.....	48
Factor Analysis.....	50
F-wave Conduction Latency Methodology.....	55
Statistical Analysis.....	56
V. RESULTS AND DISCUSSION.....	69
Principal Components Analysis.....	74
Discriminant Analysis and	
One-Way ANOVA.....	81
Repeated-Measures Analysis,	
Using All Trials.....	86
Repeated-Measures Analysis,	
Dropping the Baseline.....	98
Repeated-Measures Analysis,	
Using Baseline as Covariate.....	101
Summary and Conclusions.....	105
Recommendations.....	108
 APPENDICES	
A. Control Commands for	
Repeated-Measures Analyses.....	111
B. Rules for Determining the	
Appropriate Design.....	112
C. Rules for Degrees of Freedom.....	114

TABLE OF CONTENTS (continued)

D. Rules for Determining the Expected Mean Squares (EMS).....	115
E. Rules for Determining Error Terms.....	116
F. Repeated-Measures Analysis by Method of First Differences.....	117
BIBLIOGRAPHY.....	118
ABSTRACT.....	131
AUTOBIOGRAPHICAL STATEMENT.....	133

LIST OF TABLES

<u>Table Number</u>		<u>Page</u>
4.1	CONCEPTUALIZATION OF FACTORS AND DEGREES OF FREEDOM OF DESIGN.....	51
5.1	MEANS AND STANDARD DEVIATIONS OF EXCITATION THRESHOLD (IN VOLTS).....	71
5.2	MEANS AND STANDARD DEVIATIONS OF F-WAVE RESPONSE (IN VOLTS).....	73
5.3	COMMUNALITIES OF PHYSIOLOGICAL MEASURES AND EIGENVALUES AND PERCENT OF ACCOUNTED-FOR VARIANCE OF FACTORS.....	76
5.4	VARIMAX ROTATED FACTOR MATRIX.....	77
5.5	COMMUNALITIES OF THRESHOLD RESPONSE MEASURES AND EIGENVALUES AND PERCENT OF ACCOUNTED-FOR VARIANCE OF FACTORS.....	80
5.6	COMMUNALITIES OF F-WAVE RESPONSE MEASURES AND EIGENVALUES AND PERCENT OF ACCOUNTED-FOR VARIANCE OF FACTORS.....	82
5.7	CANONICAL DISCRIMINANT FUNCTION FOR TWO-GROUP DISCRIMINANT ANALYSIS USING INITIAL THRESHOLD RESPONSE.....	84
5.8	1-WAY ANALYSIS OF VARIANCE OF BASELINE MEASURE.....	85
5.9	REPEATED-MEASURES ANALYSIS OF VARIANCE, ALL TRIALS.....	87
5.10	REPEATED-MEASURES ANALYSIS OF VARIANCE, DROPPING BASELINE.....	100
5.11	REPEATED-MEASURES ANALYSIS OF VARIANCE, USING COVARIATE.....	103

LIST OF FIGURES

<u>Figure Number</u>		<u>Page</u>
4.1	Variation of Magnetic Flux Density of Drum-Shaped Samarium-Cobalt Magnetic Element.....	45
4.2	Magnetic Field Lines of a Drum- Shaped Magnetic Necklace Element.....	46
4.3	Design Conceptualization.....	49
4.4	Conceptualization of Factors and Error Terms.....	52
4.5	Format of Data File.....	59
5.1	Electromyograms of Supraspinatus Muscle in Response to Subthreshold, Threshold, and Suprathreshold Stimulation.....	70
5.2	F-waves Recorded From Hypothenar Muscle Following Supramaximal Stimulation of the Ulnar Nerve.....	72
5.3	Conceptualization of Pain Factor as a Main Effect.....	88
5.4	Conceptualization of Magnet Factor as a Main Effect.....	90
5.5	Conceptualization of Pain by Magnet Interaction.....	92
5.6	Conceptualization of Trials Factor as a Main Effect.....	94
5.7	Conceptualization of Pain by Trials Interaction.....	95
5.8	Conceptualization of Magnet by Trials Interaction.....	97

CHAPTER 1

STATEMENT OF THE PROBLEM

Introduction

Educational, social science, and medical investigators are sometimes confronted with data that have been collected from the same subject measured repeatedly over successive trials. Often such data are derived from experimental paradigms intrinsically concerned with the repeated-measures of individual performance across some time domain. This is of special interest when the time function itself is under examination. Since "repeated-measures" or "within-subject" designs have extensive application within classroom settings, they should be of value to the educational researcher. If the correlated nature of such data goes unrecognized, however, the researcher is prone to perform an inappropriate between-groups analysis of variance. This dissertation will examine the efficacy of such repeated-measures, within-subject designs. It will also serve to demonstrate the proper conceptualization and methods of analysis for data sets comprised of correlated measurements.

The classroom teacher/researcher probably has been exposed to analysis of variance and regression methodologies in their college education. Those with graduate training and a higher level of sophistication might recall the importance of blocking and matching and apply these

techniques within their research efforts. Such methods are routinely used for between-groups analysis. Note, however, that analyses become considerably more complex when the correlated nature of the repeated-measures is properly taken into account. Unfortunately, educational researchers are often only marginally aware of the value and applicability of repeated-measures designs. A lack of familiarity with various nested designs presents an obstacle for engaging in more complex forms of research. Greater exposure to within-subject measurement strategies would improve the caliber of future educational studies. This, in turn, would lead to more meaningful insights into educational phenomena.

Succinctly, "repeated-measures" describes a set of circumstances whereby each subject or experimental unit is measured successively on the same criterion on two or more occasions (e.g. different times, trials, ages, or conditions) or on different criteria that utilize a uniform scale. The more general case of which repeated-measures is a member is known as "multiple measures" which is simply characterized by measuring each subject (or experimental unit) on two or more different variables. By contrast, traditional analysis of variance methodology (ANOVA) assigns one group of subjects to one treatment and another group to a different treatment, etc. This is typically called a "between-groups" or "randomized-groups" design. In the fully crossed, repeated-measures design, sometimes called a "within-subjects" design, subjects are exposed to multiple

treatment conditions.

The elegance and value of the repeated-measures design can best be appreciated when the following three points, as discussed by Dayton (1970), are considered. First, since "matching" of subjects in any design tends to increase the homogeneity of groups of subjects, the ideal match for any given experimental subject would be that same subject. Thus, repeated-measures allows for comparisons involving highly homogeneous material which increases the precision of the experiment. This can result in a reduction in the sum of squares attributable to the error terms which are the denominators of the "F" test for effects. Since there is a high probability that the measurement errors attributable to a given subject will be correlated, the total error sum of squares in a repeated-measures design should be less than for a factorial design in which the subject factor is not crossed with any other factor.

In the randomized groups design, many more subjects are necessary to generate the same number of observations since each subject is measured on the dependent variable only once. Consequently, since scores are independent, experimental error is uncorrelated and additive. It should be realized that the reduction in repeated-measures error sums of squares is gained at the expense of losing degrees of freedom. There are more degrees of freedom associated with the single error term of a full factorial design than with any of the multiple, partitioned error terms of a

repeated-measures design. Whether this trade-off between sums of squares and degrees of freedom benefits the researcher is context dependent.

This leads to the second noteworthy feature of repeated-measures designs. Traditionally, educational researchers have been plagued by the dilemma of sample size. A large sample is desirable in experiments to increase the power of statistical methods testing the null hypothesis. There is also a desire to use numerous intact classrooms so as to retain the ability to generalize the research findings into as broad a parent population as possible. Moreover, classrooms as a unit of measure help contain the external variance that inevitably encroaches into any experiment due to different instructors, locations, demographics of the students, test conditions, and an assortment of other contaminants. This, however, produces the contradictory circumstance of reducing large samples of individuals into small samples of classroom means. Furthermore, large scale attempts to measure classrooms face difficulties that can be insurmountable, i.e. gaining administrative approval, coordination of efforts, minimization of disruptive effects on daily curriculum instruction, etc. Since, in a repeated-measures design, each subject is measured successively, there is a substantial savings in terms of cost, time, and administrative effort. Similarly, it becomes considerably easier to increase the relative sample size since any increase in absolute sample size (number of subjects)

results in a multiplicative increase in the total number of observations. In the context of classroom research, the above points are extremely relevant since we now recognize that the classroom is a far larger sample than is typically perceived. A within-subject view of the classroom should facilitate research as a result of operational ease and increased sample size. Though generalizability would be compromised by abandoning the concept of classroom means, it is argued that this is more than off-set by the potential increase in research activities.

Thirdly, there are certain generic experiments that are intrinsically concerned with the passage of time. Such experiments attempt to assess effects which are time-linked, e.g. maturation and child development studies, learning studies, and medical studies concerned with effects of new treatments including drugs (clinical trials). Repeated-measures is inherently dictated within these paradigms. The ability of within-subject designs to isolate effects due to repetitions (repeated trials) is a valuable attribute. Often there is a latent linear assumption on the part of the researcher. That is, there is a tendency to assume that if there is an effect, it will "grow" in a linear fashion over time. However, for some investigations this may be an overly simplistic and unwarranted assumption. In fact, there may be a complex relationship between the phenomenon under investigation and its manifestation over time. The maximal effect may not necessarily appear at the last trial.

Consequently, a repeated-measure analysis can provide valuable insight into such circumstances. This last feature is important for the purpose of this dissertation since an analysis of treatment over time will be investigated.

Gentile, et al. (1982) has argued that the between-groups design (traditional ANOVA) lacks psychological power as opposed to statistical power which can be compensated for by increasing sample size. According to Bevan (1968) there is a meaningful psychological difference between experiencing only one experimental treatment versus all experimental treatments. Whether presented successively or simultaneously, each treatment provides a context for understanding, interpreting, and therefore responding to other treatments. If subjects have not been exposed to all treatments, they cannot contrast them and no amount of randomizing subjects to groups can compensate for this shortcoming since such experience resides in the individual. It is interesting to note at this point that analysis of variance was originally derived in an agricultural context and was concerned primarily with increasing yields per acre and producing sturdier crops. Plants, of course, are unable to appreciate the context of an experiment and are not subject to biases and other foibles of the human condition. Gibson (1967), in discussing such context effects, said that distinguishing between stimuli was dependent on such stimuli acquiring distinctive features through being contrasted among themselves.

Pavlov (1927) may have been the first to demonstrate a classically conditioned contrast effect which he called positive induction. Bower (1961) has been cited for showing contrast effects in the field of instrumental conditioning. Furthermore, contrast effects have been investigated for human and animal subjects in probability learning (Lipkin, 1965), and in incentive magnitude and stimulus intensity (Grice and Hunter, 1964; Hulse, Egeth, and Deese, 1980; and Pubols, 1960).

These studies indicate that different designs produce different results. Given the same independent and dependent variables, within-subject designs typically yield larger, and sometimes different, effects than between-subject designs. As Gentile, et al. (1982) state, "if there is a real difference between two stimuli or...between two treatments, then that difference is more likely to manifest itself when subjects have been exposed to both" (p. 55). Gentile adds that though this phenomenon is well known in psychology, it is a "well-kept secret" among educational investigators. With regard to educational research, he says, "It is no wonder that the major conclusion drawn about studies comparing different educational treatments - or indeed about the interactions of treatments with aptitudes - is that there is no difference" (p. 55).

That simple random assignment of subjects to treatments avoids solving real-life problems is discussed by Sidman (1960) in the context of extinction paradigms.

Uncontaminated extinction data obtained from separate groups will yield a functional relation that has no counterpart in the behavior of the individual. The function obtained from the individual is the result of an interactive process that extends from one segment of the subject's behavior to another. (p. 53)

The kind of research questions asked in education often are concerned with how individuals are affected by some curriculum, administration, or teacher behavior. Such treatments happen over time and often, like medical treatment, are progressive and cumulative (not reversible). Meaningful real-life answers to such questions can only be obtained by using the methodology of repeated-measures on individuals under various treatment conditions. Sidman continues,

If it proves impossible to obtain an uncontaminated relation between number of reinforcements and resistance to extinction in a single subject, because of the fact that successive extinctions interact with each other, then the "pure" relation simply does not exist. The solution of our problem is to cease trying to discover such a pure relation, and to direct our research toward the study of behavior as it actually exists. If reversibility does not exist in nature, it does not exist in the laboratory...The student should not be deceived into [thinking] that the group type of experiment in any way provides a more adequate controlled or more generalizable substitute for individual data. (p. 53)

Gentile, et al. (1982), in agreement with Sidman, suggest specific shortcomings result from the use of between-subject designs (as opposed to repeated-measures/within-subject designs) in the investigation of

nonreversible educational phenomenon (p. 56). First, since subjects cannot contrast treatments, differences between groups will be negligible. Secondly, if large samples are used to increase the power of the experiment, statistically significant effects will be educationally trivial. Thirdly, educationally significant results will not be generalizable to individuals within schools since the results relate to group, not individual, performance and because results do not encompass the day-to-day reality whereby individuals are exposed to sequences of treatments.

Purpose

Upon consideration, it becomes apparent that the educational experience is inextricably bound-up with the notion of repeated-measures. Both the areas of development and learning reflect progressive growth. In the continual unfolding of childhood, the educator, much as a physician, looks for milestone accomplishments. It is commonly understood, largely due to the work of Jean Piaget (1954, 1963, 1967), that children have a unique view of their world and events at different stages of development. The ideas of norms, standardized tests, and percentile rankings, are all part of a child's early educational experience. School administrative records are kept in order to document academic and psychological progress. The true picture of a child's progression can only be fully appreciated from an overview entailing a comprehensive evaluation of the

"repeated-measures."

Taking the previous discussion into account, the justification for classroom application of repeated-measures should be clear. Much of what is measured in the classroom is, in fact, within-subjects. The gains in sample size and precision have been previously noted. Additionally, this under-utilized methodology allows for examination of questions and issues that often are either overlooked or avoided due to a perceived difficulty in handling the experimental design considerations. This, in fact, may be the context that allows for the most meaningful "payoff" for future educational research. As useful as traditional analysis of variance has been, it has the limitation of establishing a time-constrained view of phenomena. This has the effect of influencing the kind of research questions asked. For example, reading method "A" is posited as superior to reading method "B" on some theoretical basis. Typically, a researcher conducts an experiment and is prepared to make some statement regarding the differential effects, if any, of the two (or more) different reading methodologies. Historically, this line of inquiry has been well utilized and very fruitful. However, the previous analysis, if not supplemented, has precluded the asking of other fundamentally important questions. It may be rooted in the learning process that either or both reading methodologies only reveal their true contribution with a long term (over time) analysis.

Interestingly, there emerges a commonality between educational research and the medical model of research often termed "clinical trials." It is generally understood among medical researchers that evaluation of a new drug, for instance, is inherently rooted in a progressive time frame. Clinical improvement is not expected to take place miraculously. Rather, beneficial effects (as well as deleterious effects) reveal themselves only after a suitable time has transpired. Repeated-measures therefore allow asking research questions that are directed at an analysis of effect over time. This should be seen to include differential effects over time.

The medical model described above has further import within educational research. At many universities, various medical curricula are either housed within a College of Education or are shared between this college and a College of Medicine. Physical therapy and programs for the developmentally and occupationally disabled are two examples. Similarly, most medical schools have research units concerned with curriculum evaluation and the continuing development of superior training materials for physicians. Heuristic and synergistic effects can be expected by greater communication of research findings between the disciplines of education and medicine.

This dissertation will take the form of a demonstration project. The nature of the data will be subjugated to the role of methodology. The author hopes, by way of example,

to show the applicability of a repeated-measures analysis to a data set characterized by multiple measurements collected from a group of subjects (experimental units). The first goal is to have those unfamiliar with such techniques recognize the distinction between measures taken from independent subjects, and measures that are no longer independent since they emanate from the same subject. This distinction is critical since it is not uncommon for such a data set to be incorrectly analyzed by using the traditional analysis of variance model.

The second goal of this research will be to demonstrate and evaluate the effects of using the initial baseline measure as a covariate. Often investigators are uncertain as to whether the baseline measure should be viewed as one of the repeated-measures of the dependent variable. This results in confusion as to whether the baseline measure should be included in the analysis at all. Note that, by definition, the baseline measure has not been subjected to the experimental treatment. To include it with the repeated-measures dilutes the treatment effect, if such an effect exists. Alternatively, simply to ignore the baseline and exclude it from the analysis gives up valuable information. Consequently, this dissertation will compare and contrast all three types of analyses: (1) repeated-measures including the baseline as the first repeated measure; (2) repeated-measures ignoring the baseline completely; and (3) repeated-measures using the baseline as

a covariate.

The high speed digital computer has had profound impact upon the research process in most academic disciplines. Education is no exception. The ability to ask complex research questions and to be able to perform iterative analyses (ask another question) based on the results of the first, allows for a dramatic increase in the quantity, and hopefully in the quality, of both applied and theoretical research. With this in mind, the third goal of this dissertation will be a demonstration of the proper view of the experimental design and the manner in which data must be coded in order to facilitate a computer analysis. Standard procedures for identifying factors, interactions, and their respective error terms will be discussed. These techniques are especially useful for full appreciation of the sources of all experimental variation. According to this view there is no residual variation since all sources are identified. This provides conceptual clarity for the researcher.

Much of what has inhibited researchers in the past from doing more complex analyses - based on more complex experimental designs - was the tedium of the mathematics required to complete the analysis. Supplemental analyses are rarely done if large expenditures of time are required to perform voluminous calculations. Due to computers and application software commonly available today, educational researchers no longer need be shackled by such difficulties. In place of these considerations, the contemporary

educational researcher must acquire knowledge of data processing and data manipulation. Similarly, facility with application software, including interfaces between packages (products), become necessary requirements for the productive researcher. Such proficiency is an important consideration with regard to this dissertation. Therefore, the fourth goal of this dissertation will be to demonstrate how to use appropriate computer application software to solve repeated measure problem situations.

Lastly, the complexities of the computer output will be discussed. Often, the volume of reported statistics calculated and displayed as a result of computer analysis can be intimidating to the unsophisticated. Interpretation of the myriad factors will be delineated so as to encourage this type of analysis among those who, in the past, have lacked a suitable model for their own research endeavors.

CHAPTER 2

REVIEW OF THE LITERATURE

According to Hotelling (1943), the earliest work on the development of the analysis of variance, which is the more general case of repeated-measures, can be traced back to 1826 when Gauss first described the concepts of degrees of freedom and least squares. Wilks (1932) cites Helmert, an astronomer and mathematician, as having derived the exact distribution of the sum of squared deviations of a normal variate about the population mean in 1876. The Rothamsted Experimental Station in England was established by Lawes in 1841. From 1843 to the present day a plethora of longitudinal data originates from this research facility. It was here that Fisher developed analysis of variance in the 1920s. Prior to this he developed the sampling distribution for the "t" statistic (1915) and later showed its application to experimental data (1925). Fisher based this work on the earlier development by Gossett (1908), who using the pseudonym Student, suggested the form of the distribution of the sum of squared deviations about a sample (as opposed to a population) mean. Snedecor (1934) is credited with production of the first table of mean square ratios which he called "F" in tribute to Fisher. Fisher (1935) demonstrated his cognizance of the relationship between research design and analysis when he stated that

"statistical procedure and experimental design are only two different aspects of the same whole, and that whole is the logical requirement of the complete process of adding to natural knowledge by experimentation" (p. 3).

An examination of historical materials specifically chronicling the use of repeated-measures reveals that virtually all early work was done in the field of experimental psychology. The early researchers had to conceptually struggle with the role and importance ascribed to subjects within their experiments. Subjects as a discrete factor was only fully appreciated years later as the result of continuing evolution of research design considerations. In the early years, subject variation was grouped with residual variation. There was little interest in fully partitioning the sums of squares into the smallest additive components. That progress was slow in changing this view is attributable to the fact that subject differences were seen as no more than a source of extraneous variance. Lovie (1981), who contributes greatly to this overview, states,

The subject effect [needed] to emerge from the error term before a proper understanding of repeated measure designs as mixed effects models became possible...The history of such designs charts the painful way in which psychology began to bring subjects qua subjects into the framework of experimental design. No less painful was the realization that what were then viewed as accepted factors differed conceptually from the use of subjects as a factor. (p. 13)

Published papers by Robinson and Bills (1926) and

Telford (1931) made use of repeated-measures designs even before analysis of variance (ANOVA) began appearing in journals (late 1930s). Robinson and Bills investigated the differences among various levels of a factor over trials, the second factor, in several multi-factor experiments. They were concerned with fatigue effects in repetitive work situations such as number of fingers used in a typing experiment or the number of letters in a writing task. Their work should be seen as conceptually advanced, since factor interaction was a focal point of their inquiry.

Telford, in investigating what is known in the perceptual literature as the refractory effect, showed each of his subjects sequences of two lines. Subjects in the experiment had to distinguish between line length with a second factor being time variation between exposure to the pair of lines. Thus he utilized a two-factor design with line length and time between exposures as factors. Typical of his era, Telford's analysis was rather informal concentrating on one factor at a time at the expense of failing to reveal any possible interaction.

Into the next decade, the work of Garrett (1940), Hackman (1940), and Baxter (1942) all show concern for partitioning separate subject terms in their analyses. Baxter contrasted the two designs he discussed by noting that the separate subject term was only appropriate in the one design since the other involved a degree of confounding between subjects and factors. Garrett noted that for the

single factor repeated-measure design the interaction between subjects and the factor was the proper error term for testing both the subjects and factor terms instead of the usual within-levels estimate. Garrett and Zubin (1943) took account of dividing subject variance into between- and within-subject effects in their discussion of repeated-measure designs.

Further evidence of this new view of subject variances is reflected in Baxter's use of the term "individual differences." Once identified, these differences can be evaluated in the analysis. Similarly, Hackman discussed how such individual differences paralleled differences between the means of subjects which "have been shown to enter into and obscure results" (p. 555). Garrett, too, had computed between-subject effects based on individual differences. Lovie (1981) has summarized this view.

It seems, therefore, that subjects were viewed less as a factor in their own right and more as a nuisance variance whose effects needed to be extracted. Alternatively they were considered to be equivalent to the plots or blocks of classical design in agriculture and biology which formed the medium through which the effects of the various factors were revealed. (p. 2)

We see then that the attitudes towards subjects and their role in the overall structure and analysis of the design reflects a historical progression. Many of these early studies were concerned with learning or fatigue effects and had a naturally occurring repeated-measure that was recognized as a separate orthogonal factor. Crude,

though some of these designs were, they show concern for order effects and a willingness to refine experimental design in an attempt to evaluate such effects.

In a milestone paper viewed as significantly advancing repeated-measure methodology, L.G. Humphreys (1943) described a three factor conditioning and extinction study utilizing repetitions on one factor. He stated that

the usual procedures of the analysis of variance, as outlined by Snedecor (Statistical Methods, 1937) and Lindquist (Statistical Analysis in Education Research, 1940) are not applicable in the present instance [since] certain of the variables and interactions involve correlated means. (p. 104)

Consequently Humphreys subdivided the analysis into two parts each using a different error term. First, the between-subject factors were partitioned into the sums of squares of all the independent factors and interactions and a residual which Humphreys called "net between-subjects." Supposedly, this last term reflects uncontrolled inter-subject variability. Winer (1971) has called this "subjects within groups variability." Edwards (1950a) used the phrase "variability between subjects in the same group." Secondly, Humphreys analyzed the within-subject factor and its interactions with the non-repeated factor by calculating what he called "the sums of squares within-subjects" from which the sums of squares of the factor and its interactions are subtracted. This left a residual called the "sums of squares for net within-subject variability" and which Humphreys says represents uncontrolled intra-subject

variation. More currently, Lovie (1981) takes issue with this view citing that in modern repeated-measure analysis this term would be said to reflect not differences within a subject, but rather the interaction between subjects within a group and the repeated factor. This being the only interaction with subjects that could be extracted due to the pattern of nesting.

Edwards (1950b) in his Experimental Design in Psychological Research differs from this early line of Humphreys. Whereas Humphreys viewed the experiment as a split-plot design with subjects as plots, Edwards thought of the design more as a multifactor one with subjects being considered a factor. This evolving view allowed for extraction of as many orthogonal subject-by-factor interactions as the pattern of nesting permitted as well as a between-subject effect. Winer (1971) comments that it is possible to develop split-plot designs that copy the structure of many repeated-measure designs, but the structural models and expected mean squares ($E[MS]$) differ. The two points of significance regarding Humphreys' split-plot analysis are: (1) his use of separate treatments for the independent and correlated factors with different error terms; and (2) he had not yet seen subjects as a separate factor. Interestingly, as previously noted, Humphreys does refer to Lindquist's (1940) discussion of the difficulties associated with translating agricultural terms into their educational equivalents. This is one of the earliest

references to an educational application of such methodology.

The next wave of papers on repeated-measure designs was attentive to more than choice of error terms. There was also interest in extending the analysis by making use of the extra information about the performance of subjects over the repeated factor. Alexander (1946, 1947), Lindquist (1947), and Brozek and Alexander (1947) all dealt with repeated-measure factorial designs, either single or two factor with repetitions on all factors. Lindquist (1947), Kogen (1948), and Edwards (1950a) all utilized two factors with repetitions on one factor. There is a marked emphasis on learning and performance studies within this early experimental psychology literature. This seems due to the fact that many of the repeated factors were trials, or some other time-linked variable. Though more recently Poulton (1975) has detected order effects in such designs, these early studies were relatively immune to such concerns since subjects were usually faced with repetitions of a single or a homogeneous series of experimental situations. Two important points emerge from this series of papers: (1) the awareness that subjects could now be thought of as a factor; and (2) the assumption of either a variance-covariance matrix which exhibited compound symmetry or pair-wise independence between the repeated levels. Regarding limitations of design that had not yet been resolved, Lovie (1981) states that "one of the problems of repeated measure

designs, that is, correlated levels, was usually avoided by assumptional fiat in order to solve the other problems, for example, choice of error terms and the presence of nesting" (p. 4).

As might be expected, much of the early work of this era (late 1940s) is conceptually incomplete with regard to the proper partitioning and testing of effects. Most researchers of the day would have been familiar with Fisher's Design of Experiments (1935) which contains a section on interactions as error terms. This can be seen as the basis for Lindquist's mixed effects analysis providing: (1) that subjects are seen as a factor; (2) that they can be viewed as plots; and (3) that they are chosen at random. Additionally, both Snedecor's (1937) and Goulden's (1939) discussions of the choice of error terms in ANOVA were commonly known.

Lindquist's paper (1947) on trend analysis in both independent and repeated-measure designs deserves further attention since it is commonly regarded as a "tour de force." Unfortunately, the source of Lindquist's treatment of the designs remains unclear as no references are given in any of the cases quoted. Therefore, it is difficult to offer more than speculation about the impact previous researchers had on his work. He, unlike others previously mentioned, viewed the single-factor repeated-measure design as a two-factor structure with subjects as the second factor. His "F" test of the factor utilized the mean square

interaction between subjects and the factor in the denominator. Though advanced in his thinking, no justification is given for using the interaction term as the appropriate comparison for the factor or even why the subjects have become a factor.

Lindquist showed considerable insight into the structure and analysis of a particular case study involving two factors with repetitions on both factors. Here the interaction between the two factors is compared with the factor-one by factor-two by subjects interaction, while both main effects are properly compared with the appropriate factor by subject interaction. Lindquist states, "In terms of the methods of analysis of variance...[this study] is a factorial design involving the three factors: treatments, intervals and subjects" (p. 78). Note his view of subjects as the third factor. According to Eisenhart (1947) the analysis of this design had the further significance of making considerable headway in the attempt to satisfy the need for "more general methods...for interpreting mixed analysis of variance tables, particularly in regard to tests of significance for individual factors" which were then lacking (p. 21).

Lovie (1979) has argued that psychologists differentiate between salient and background factors. Their lack of any interest in, or even ability to provide, explanations of behavior based on individual differences alone made Lindquist and his contemporaries treat subjects

as a background variable providing only a basis for inferences about the effects of the other factors. This criticism is substantiated by further appraisal of Lindquist's work. Though he treated subjects as a factor, he did not utilize tests for between-subjects or for any of the interactions between subjects and the other factors in the design. In analyzing repeated-measure designs, the researcher must make the difficult choice between univariate and multivariate techniques, with one of the main determinants being the possible presence, and the pattern, of intercorrelation between repeated levels. Lindquist apparently did not consider the possibility that intercorrelations between the levels of the repeated factor might effect the analysis. Lovie notes that this convenient oversight stems from his statistical assumptions that subjects' regression lines had zero slope, or were parallel with equal variances. Humphreys, as noted earlier, attended to this difficulty by separating treatment results into within- and between-plot (subject) analyses. Lindquist, by failing to modify his univariate analysis, simply discounted the possibility of intercorrelation within his data.

In illuminating the fallacy of such blind assumptions, Lovie (1981) cites even more unwanted implications than Lindquist might have imagined. If the slopes of the subject regression lines were all zero then this might mean there was no difference between the levels of the trials factor, while the weaker assumption of parallelism could imply a

lack of subjects by trials interaction. This would change the comparison variance from that of interaction to that of residual error.

Evidence for further evolution of these design considerations is given in a paper by Alexander (1947). Here, he borrows a table of $E(MS)$ from Jellinek (1936) who had studied oxygen consumption rates in schizophrenic patients. More significantly, Alexander derived reliability formulae for single-factor repeated-measure designs. There seems to be some confusion in the $E(MS)$ table, but since Alexander's model assumed no significant interaction, the tests on the repeated and nonrepeated factors were unaffected. If the interaction was shown or assumed to be significant however, then the model and the $E(MS)$ s would have been different and only the "F" test on the repeated factor would have been possible. This highlights the historical gap since this latter model would be the present day model of choice.

Alexander seems to have then taken a step backwards, however, since in his next paper (Brozek and Alexander, 1947) his viewpoint apparently changed when he failed to appreciate his single-factor repeated-measure design as a two-factor structure allowing for an interaction between the subject factor and the repeated factor. It should be noted that one need not necessarily view the design as encompassing interaction especially since it is easier to compute components of variance by eliminating interaction

from consideration. If, however, Brozek and Alexander had been more conscious of residuals and interaction (even if confounded), they might have more fully appreciated structure. As Lovie (1981) has stated, "ambivalence about the nature of the residual variance reveals a lack of insight into the possible structures in [an] experiment" (p. 7). This confusion might have stemmed from the fact that Brozek and Alexander had employed what they called "random variance" as the comparison variance used to evaluate the between-subjects and between-trials variances. They conceptualized this as the variation left when differences between subjects and trials had been extracted. In his previous paper, Alexander had termed this "interaction variance" which was more in keeping with the emerging modern view.

Kogen (1948) provided a link between earlier and later work. Building upon the work of Lindquist and Alexander, he explored the rationale for different tests of significance taking particular note of Alexander's assumption of independence between trials, or levels of the repeated factor. Kogen, using a two-factor design with repetitions on one factor (trials), did three analyses in an attempt to come to grips with the problems of intercorrelation between the level scores on the "F" tests. In clearly differentiating between inter- and intra-subject effects as well as main and interaction effects, he demonstrated greater insight into the structure of the design than either

Lindquist or Alexander. Lovie (1981) has detailed Kogen's work. In Kogen's first analysis, he correctly argues against the use of the methods by trials interaction as an error term for the two main effects, since the interaction term, because it reflects average intercorrelation, is not appropriate for the independent factor. He adds that its use as an error term for the trials factor is inappropriate since one loses the opportunity to use the within-subject replications as an estimate of error. Unfortunately, it seemed to have gone unnoticed that this also precluded testing the methods-by-trials interaction. This may have been due to the view held by Kogen, based on Lindquist, that the effect of trials was considered to be a background factor that held little inherent interest. As a final point regarding Kogen's first analysis, he rejected what he called the residual term due to its nested form and opted to partition this source of variation into a within-subject term, which he called "inter-subject effect" and trials-by-within-subject interaction which he confusingly referred to as an "intra-subject effect."

Kogen's second analysis utilized proper tests for both methods and methods-by-trials though he failed to properly evaluate the methods mean square apparently due to some misconception over the structure of the design and the proper allocation of degrees of freedom. However, the greater point here is that building on his first analysis, Kogen's second analysis utilized separate error terms for

the various effects. Lovie points out that it was of historical significance that Kogen had abandoned the futile search for a single error term for all the effects in the design (p. 7). Unfortunately, Kogen, like Humphreys (1943) and Lindquist (1947) before him, was inconsistent in his use of the terms inter- and intra-subject effects in his first two analyses. Additionally, he was unable to detect that his intra-subject term was really an interaction of trials by within-subject effects. This was perhaps his greatest oversight.

Kogen's third analysis fails to make a convincing case regarding the error term for the methods factor where the possibility of correlation between the levels of the repeated factor must be considered. Kogen felt that if there is a high ratio of inter- and intra-subject variances then the average intercorrelation is high and the methods factor should be tested against the inter-subject term. Conversely, if the ratio is low the test is against a pooled inter- and intra-subject variance. However, Fertig (1936) years before had a cogent discussion of the close relationship between inter- and intra-subject effects and correlation for the case of a two-level repeated-measure design in which he cautioned against the approach that was used by Kogen as being too dependent on the homogeneity of the intercorrelations. Furthermore, Lovie (1981) states that "the unfortunate consequences of Kogen's second rule might be to obscure a real methods effect since the decision

whether or not to pool is based not on the absolute value of the variances concerned but upon their relative weights" (p. 8). Lovie adds in summary that Kogen's work is seen as significant due to his insight into the structure of design and his alternative decompositions of the total sum of squares into components of the design. Lastly, he is credited with tackling the correlated nature of the data through conceptual justification of the use of interactions as error terms because they estimate average intercorrelations of the repeated factors.

Edwards (1950a) can be seen as the next historical successor to the early work of Fisher (1935), Goulden (1939), and Snedecor (1937) in as much as he attempted to devise proper tests for the fixed factors in a three-factor mixed effects model. Edwards states that "when the categories or classifications of one of the variables may be regarded as a random selection from the population being sampled" there is justification for using interactions, rather than residuals, as error terms (p. 215). He further suggests criteria to aid this decision making process. If the highest-order interaction is insignificant then residual error should be used to test lower-order interactions which, if also insignificant, should be pooled for tests on lower-order interactions and main effects. Lovie, however, cautions against any cavalier attitude towards pooling because of the complexity of the $E(MS)$ for the main effects and interactions (p. 10). Though Edwards' work described

above was based on the three-factor mixed effects model, he utilized similar thinking in his description of a two-factor design with repetitions on one factor. Edwards' real contribution can be seen in the importance he placed on the differentiation between fixed and random factors.

Any discussion of repeated-measures would not be complete without presenting the contrary view held by some investigators who have shunned this methodology, almost to the point of disdain. The main issue of contention is whether order effects, also known as context or range effects, confound the experiment to such a degree as to render the results meaningless. Babington-Smith (1950) was among the first to note the implications of subjects remembering stimuli and responses over time. He argued the order in which stimuli are presented is important since each unique order serves as a different stimuli producing differential effects. That is, given multiple treatments, different orders of presentation in a repeated-measure design can produce different effects. Poulton (1973) attempted to document range (context) effects within the literature of experimental psychology. Though previously known, the full import of this effect was perhaps not fully appreciated since different sub-disciplines used different terms to describe it. In motor performance, range effects are considered within the domain of "transfer of training" while in magnitude estimation they are called "adaptation level" effects. For reaction time studies, they come under

the heading of "uncertainty" or "probability." With regard to such effects and their relationship to repeated-measure designs, Poulton (1973) argues that "the day should come then when no reputable psychologist will use a within-subject design, except for a special purpose, without combining it with a separate-groups design" (p. 119). That such effects exist, is not to be disputed. However, to suggest that repeated-measure designs be summarily dismissed as a consequence, as Poulton argues, is simply foolish and reflective of a myopic viewpoint.

Greenwald (1976), in response to Poulton, suggests that context effects generated by a within-subjects design could be subdivided under three component headings: (1) practice effects; (2) sensitization effects; and (3) carry-over effects. Greenwald proceeds to offer advice for each of these situations. For practice effects he states that a within-subject design be avoided if the researcher is interested in the effects of the treatments in the absence of practice where practice is likely to affect performance either as a main effect of successive tests or as an interaction of successive tests with treatments. Importantly, he notes that undesired practice effects may sometimes be controlled by either counterbalancing treatment order or by providing extensive practice sessions prior to treatment. He further states that a within-subjects design is appropriate when the practice effect itself is under investigation.

The issue of sensitization is discussed in the context of an illumination study where the lighting level is altered at periodic intervals in counterbalanced order across subjects. Here, the concern is that a subject might discriminate the illumination differences and therefore be more sensitive and responsive to illumination than if there were exposure to only one of several illumination level treatments as would be the case for a between-subjects design. Greenwald offers strategies that can be utilized to obviate this difficulty. First, changes in illumination can take place so gradually as to go unnoticed by the subject. Secondly, several extraneous variables may be systematically altered so as to draw attention away from the critical treatment variable. Greenwald admits that a within-subjects design should be avoided when juxtaposition of treatments facilitates perception of treatment variations if such perceptions interfere with the processes being studied. However, the fact that perception of stimuli differences may be enhanced by their juxtaposition in a within-subjects design may actually aid research when the concern is the subject's capacity to discriminate such differences.

According to Greenwald "a carry-over effect occurs when the effect of one treatment persists in some fashion at the time of measurement of the effect of another" (p. 318). He adds that the preferred method of reducing such effects is to separate treatments in time since counterbalancing provides only a partial solution. In fact, a within-

subjects design may be inappropriate when treatments have persistent effects since, in this context, the effect of one treatment may "carry-over" and confound the measurement of another treatment. Within-subject designs are appropriate when studying the sequence in which treatments are administered since their temporal proximity are themselves often of psychological interest. As an unintended benefit Greenwald adds,

the fact that intertreatment carry-overs are likely to be a major source of serendipitous findings should not be overlooked as one of the virtues of employing within-subjects designs in which treatments that would otherwise not be examined in near temporal proximity are juxtaposed. (p. 318)

Lastly, Greenwald raises the issue of external validity with regard to within-subject designs. He notes that often there is a trade-off since practice, sensitization, and carry-over effects can be at odds with external (ecological) validity. The specific examples cited by him and further summarized by Poulton generally occur in the area of experimental psychology. Though, as previously discussed, there are remedies for handling such problems, the relevant issue is the one previously raised by Gentile. He argued that "context" provided the basis for understanding, judging, and responding to other treatments. That this view is more reflective of human behavior in the "real-world" is justification for asserting repeated-measures (within-subject) designs to have superior external validity than the between-subjects counterpart. Criticism of repeated-measure

methodology, as presented by Babington-Smith and Poulton, should be placed in proper perspective and viewed simply as a caution in the same manner that virtually all statistical techniques have cautions regarding violations of assumptions.

Knapp (1982) is persistent in his arguments against repeated-measure, within-subject designs. However, a close examination of the specifics he has singled out shows that his assertions can be used as easily to defend this methodology as to attack it. For instance, he believes that the feature of having each subject serve as their own control, as is the case with repeated-measures, is much overrated. Instead, he prefers the advantages gained by a completely randomized design with independence of measurement virtually assured. Likewise, he asks "why do repeated-measures advocates want to 'milk subjects dry'?...Does it really matter whether 8 subjects come and stay for all 4 treatments or 32 subjects come in for their treatment and then take off?" (p. 63). It is as if Knapp has chosen to evaluate the features of repeated-measure designs in a vacuum.

It has previously been discussed that such within-subject designs are not necessarily appropriate in all experimental situations and that the use of this methodology should be applied with some degree of caution regarding statistical violations. The real failure of Knapp and others, such as Babington-Smith and Poulton, is to

appreciate those experimental contexts where repeated-measure designs can, in fact, make a real contribution. In many cases this methodology allows for the practice of research where, prior, none was possible, at least in the practical sense. One such context is, of course, the classroom where sample size is typically limited and where often the substantive issue under examination has to do with investigations of change over time. To this end, repeated-measures designs have utility because they use subjects as their own control and because they attempt to "milk the subjects dry."

Summary

A review of the literature serves to inform the reader of two fundamental points with regard to within-subject, repeated-measure experimental designs. First, this methodology did not have a discreet, precise origin; it evolved rather than having been invented. Secondly, not all researchers are in accord when discussing the usefulness and appropriateness of this design.

The 1920s-1930s saw R. A. Fisher's development of the analysis of variance and the propagation of this technique into the psychological and educational research communities. Much of the early work utilizing repeated-measure ANOVA was done by experimental psychologists, often because of the intrinsic nature of the phenomenon under investigation, i.e. fatigue effects due to a subject's exposure to a repetitive

stimulus.

Through the next two decades, the continuing evolution of repeated-measure, within-subject designs was marked by a gradual appreciation of subjects as a factor in their own right rather than simply being perceived as background noise. Furthermore, a sophisticated view of designs was being fostered which allowed for a more meaningful partitioning of effects and degrees of freedom. This, in turn, permitted a clearer delineation of interactions and error terms appropriate for testing effects. Hence, by the mid-1950s this methodology was refined and commonly utilized.

Though some investigators, most notably Poulton and Babington-Smith, have leveled criticism at within-subject designs primarily due to context effects, others such as Greenwald have quickly come to their defense. With specific regard to educational classroom research, the repeated-measure, within-subject design should still be viewed as a useful research tool.

CHAPTER 3

RATIONALE OF THE DEPENDENT MEASURE

Nerve conduction measurements have long been recognized as a valuable diagnostic tool for diseases or injuries which might otherwise be difficult to diagnose (Buchthal and Rosenfalck, 1966; Eaton and Lambert, 1957; Gilliatt and Sears, 1958). They are now routinely used as part of the electrodiagnostic examination for many abnormalities associated with muscles and nerves. In the carpal tunnel syndrome, a condition where the median nerve's passage through the wrist's carpal tunnel is obstructed, a prolonged distal conduction latency is obtained upon stimulation of the median motor nerve at the wrist (Goodgold and Eberstein, 1972). When the brachial plexus is involved in compression of the median or ulnar nerve, conduction studies usually reveal low velocity and long latency (Urschel, et al. 1971; Jebesen, 1967). Furthermore, one of the earliest indications of peripheral nerve lesion is an increase in the threshold of excitability of the nerve to electrical stimulation (Wynn-Parry, 1969).

Physical examination of patients with chronic neck and shoulder pain often reveal spasms in the supraspinatus muscles, usually with bilateral involvement. Studies involving spastic paralysis lesions commonly show prolonged conduction latency and slowed conduction velocity in nerve

fibers to proximal muscles (Redford, 1964; Kimura and Butzer 1975; Wu and Stratigos, 1976). The amplitudes of the response are also lower and have greater variability. It is therefore of interest to examine changes in conduction time and threshold of stimulation of the proximal segment of the motor nerve (brachial plexus) that innervates the supraspinatus muscle.

The effects of static magnetic fields on the structure and function of biological tissue have long been a subject of scientific investigation (Barnothy, 1964, 1969; Kholodov, 1966; Aceto, et al., 1970; Dubrov, 1974; Silver and Tobias, 1974; Tenforde, 1978). In surveying the broad range of biological effects that have been reported, it becomes apparent that several molecular systems and lower organisms exhibit fairly well-defined and reproducible responses to static magnetic fields. For example, fields of the same order of magnitude as the geomagnetic field (0.3-0.6 gauss) have been shown to produce orientational effects on the movement of certain bacteria (Blakemore, 1975; Kalmijn and Blakemore, 1977), insects (Lindauer and Martin, 1968; Martin and Lindauer, 1977), fishes (Kalmijn, 1977, 1978) and birds (Keeton, 1969, 1971; Moore, 1977; Walcott, et al., 1979). However, the paucity of well defined magnetic field effects becomes increasingly evident as one reviews the literature regarding organisms of higher complexity. With the exception of magnetic phosphene (Barlow, et al., 1947; Oster, 1970) most observations relating magnetic field

effects on higher animals have been clouded by contradictory reports. Although static magnetic fields have been reported by many investigators to affect the structure and function of excitable tissues, the reported responses have often been controversial.

It is generally felt that such discrepancies result from a lack of precision in measurement technique. Therefore, for the purpose of this study, EMG applications were carefully examined by a staff of experienced EMG physicians and technicians. Electrode alignment was adjusted weekly for each subject. Furthermore, the reliability and validity of the previously collected test measures will be evaluated statistically. Several of the contradictory findings are presented below in order to highlight the difficult and tenuous nature of this experimental methodology.

Exposure to a static magnetic field (10,000 gauss) has been reported to significantly increase the conduction velocity and excitation threshold of isolated nerves (Erdman, 1955; Reno, 1969). Others, however, have indicated that an equivalent magnetic field had no detectable influence on the conduction velocity and excitation threshold of similar preparations (Lieberman, et al. 1959; Schwartz, 1978). A study by Vovk and Tkach (1971) showed that when isolated skeletal muscle was placed in a 2000 gauss constant magnetic field, the fluctuations in the threshold of stimulation increased although the mean

threshold stayed unchanged.

Effects on the electroencephalogram of rabbits have been produced by magnetic fields as low as 800 gauss (Kholodov, 1964, 1966, 1969). A highly significant increase in the number of slow high-amplitude waves was interpreted to indicate an inhibitory effect of the static magnetic field on the central nervous system. A study by Beischer and Knepton (1966) however, noted an increase in the frequency and amplitude of electroencephalographic patterns from squirrel monkeys exposed to magnetic fields up to 90,000 gauss. Others have found that exposure of amphibians to a magnetic field near 2,500 gauss reduced consciousness and altered the electroencephalographic pattern to one resembling moderate to deep anesthesia, i.e., from that of a moderate-amplitude alpha wave to a slow high-amplitude delta wave (Becker, 1961, 1963). Experiments involving humans have shown that field strengths in the same order of magnitude as geomagnetic fields could alter electroencephalographic patterns and reduce heart rate (Dubrov, 1974), while others have shown that fields up to 100 gauss had no consistent alterations in either electroencephalographic pattern or simple reaction-time performance (Friedman, et al., 1967).

A few papers also have reported beneficial therapeutic effects from a static magnetic field on cancer patients, on phantom pains following amputation of the extremities, and on a case of thrombophlebitis (Kholodov, 1971). Moreover, a

hygienic study on industrial workers exposed to a magnetic field (30-5,000 gauss) has revealed a noticeable increase in the pain threshold (Vyalov, 1971).

Magnetic therapeutic devices have been widely used in Japan, and more recently in the United States, for the treatment of a variety of musculo-skeletal disorders. Several clinical reports have indicated that these devices are highly effective in alleviating subjective symptoms such as neck, shoulder, and other muscular pains (Nakagawa, 1974, 1975, 1976; Nambu, et al., 1973). A controlled clinical laboratory investigation however, has yet to appear in the literature. Therefore, this study attempted to document, through electromyographic (EMG) techniques, the efficacy of a static magnetic field as produced by a magnetic device (necklace) on neck and shoulder pains of volunteer subjects.

CHAPTER 4

METHODS AND PROCEDURES

Introduction

This dissertation is concerned primarily with the methodology and application of a repeated-measures design. In order to demonstrate the efficacy of such an experimental design, the author will utilize an existing data set comprised of repeated physiological measurements taken from subjects over a period of four clinical trials. As was previously discussed, such a medical model is a natural candidate for the repeated-measures design. The substantive issue being examined during this investigation is the effectiveness of a magnetic device (necklace) in the treatment of neck and shoulder pain. The dependent measure is the excitation threshold of the suprascapular nerve that innervates the supraspinatus muscle. This was determined by electromyographic (EMG) examination conducted in an out-patient clinic by medical personnel.

General Procedure

Subjects (N=100) were recruited by notices that were posted throughout Wayne State University and the Detroit Medical Center. These notices requested both subjects who experience neck and shoulder pain or stiffness and those with little or no such experience. A nominal monetary

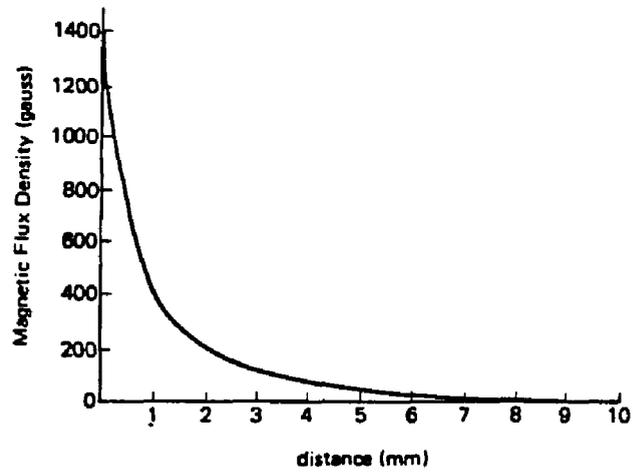
incentive was offered for participation through completion of the project. Group assignment (pain versus no-pain) was based on an initial interview conducted by a physician who took the subject's medical history. The physician determined group assignment on the basis of both the frequency and intensity of reported pain. Individuals who exhibited either an erratic pain history or simply had low-level pain were rejected as subjects due to difficulty in making a clear assignment. Consequently, all subjects were either completely pain free with regard to upper neck and shoulder musculature or had a history of reasonably severe to pronounced constant discomfort in these regions.

Subjects were seen weekly for five consecutive weeks. The first session comprised the interview where a medical history was taken and consent forms were reviewed and signed by subjects willing to participate. During the next four consecutive weeks EMG recordings were taken from each subject in an electromyographic laboratory located in the Medical Center. During the first EMG session (second week), the pretreatment excitation thresholds were obtained so as to serve as a baseline comparison. Due to the demands placed upon the EMG lab, including the time of the cooperating physicians, the 100 subjects were broken up into 4 waves of 25 subjects, each wave lasting the five week interval. Therefore, the total elapsed run time of the experiment was 20 weeks.

Prior to beginning treatment, magnetic and non-magnetic

devices (necklaces) were randomly assigned for both pain and no-pain groups using a table of random numbers. Thus a "double-blind" procedure was established whereby (1) neither the experimenter nor subject was aware of the type of necklace (magnetic versus non-magnetic) used; and (2) the experimenter was not aware of the medical history of the subject. This created a two by two analysis of variance with trials (EMG excitation threshold) as the repeated-measure. The first factor had two levels, pain versus no-pain. Likewise, the second factor had two levels, magnetic versus non-magnetic conditions. Trials, the repeated-measure, had four replications, the first trial serving as a baseline measurement.

Subjects were instructed to keep the necklace in continuous contact with the skin, the only exception being removal of the necklace during the weekly EMG examination. The length of each necklace was adjusted to fit the individual subject. The mean number of magnets or non-magnetized elements was 9 with a range of 7 to 11. Each magnetic necklace consisted of samarium cobalt magnets (TDK REC-20) with brass chains and was plated with either gold or rhodium. The drum-shaped magnetic element was 3 mm in length by 2.2 mm in diameter and had a surface flux density of approximately 1300 gauss that decreased rapidly away from the surface. This is depicted in Figure 4.1 (TDK, 1978). The direction of magnetic field lines is illustrated in Figure 4.2 (TDK, 1978). The non-magnetic necklaces used in



**Figure 4.1 Variation of Magnetic Flux Density
of Drum-Shaped Samarium-Cobalt
Magnetic Element**

Direction of magnetic field lines

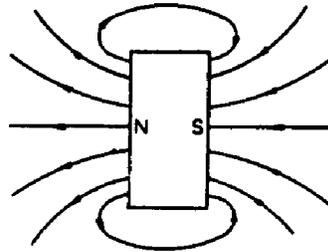


Figure 4.2 Magnetic Field Lines of a Drum-Shaped Magnetic Necklace Element

the study had the same characteristics as the magnetic ones except the samarium cobalt elements were not magnetized.

Excitation Threshold Methodology

The suprascapular nerve is derived from the upper trunk of the brachial plexus and passes downward, laterally and posteriorly across the lower part of the posterior triangle of the neck to supply both the supraspinatus and infraspinatus muscles. The excitation threshold was determined by stimulating the suprascapular nerve with a surface bar electrode placed on the lower portion of the posterior triangle of the neck. The negative electrode was located near the margin of the trapezius muscle and was about 1.2 cm above the clavicle. The positive electrode was placed on a line between the negative electrode and the ear lobe and was about 2 cm above the negative electrode. The suprascapular nerve was stimulated with a 0.1 msec wide square pulse (TECA-4 EMG machine). The neck was in a neutral position during stimulation.

Each session consisted of three successive measurements of the stimulus voltage required for a threshold response. The stimulation voltage was gradually increased from zero until a perceptible response was recorded from the supraspinatus muscle. The stimulation voltage was measured with a Tektronix oscilloscope having a 2-volt resolution. The stimulus was then set about 10 volts above threshold level and slowly lowered until the action potential

disappeared. The stimulation voltage was again measured. Finally, the stimulation voltage was returned to zero and gradually raised to threshold values. The three measurements were averaged to give a mean threshold stimulation voltage. At the end of each session, a small water-fast stain was put on the neck of each subject to mark the point of attachment for the negative stimulation electrode so as to insure that recordings were made from the same point in the weeks that followed. As was noted earlier, the necklace was removed during the examination.

Experimental Design

The experimental design utilized by this study was a 2 x 2 analysis of variance with a third repeated factor, trials. This is depicted in Figure 4.3. There were two levels for each of the two grouping factors (no-pain versus pain and non-magnetic versus magnetic conditions). The third factor, trials, had four repetitions. The replicates (experimental units or subjects) were nested within specific levels of the pain and magnetic factors. Each of the four combinations of the two grouping factors had twenty-five subjects, each subject being exposed to the four repeated trials. The four combinations were: (1) no-pain, non-magnetic; (2) no-pain, magnetic; (3) pain, non-magnetic; and (4) pain, magnetic. Thus each combination of the two grouping factors had one hundred observations (25 subjects x 4 trials). Given four combinations of the two grouping

	P1 No-Pain				P2 Pain			
	T1	T2	T3	T4	T1	T2	T3	T4
M1 Non- Magnetic	R 1- 25	R 1- 25	R 1- 25	R 1- 25	R 26- 50	R 26- 50	R 26- 50	R 26- 50
M2 Magnetic	R 51- 75	R 51- 75	R 51- 75	R 51- 75	R 76- 100	R 76- 100	R 76- 100	R 76- 100

Figure 4.3 Design Conceptualization

factors, there were a total of 400 observations (100 subjects observed across four trials).

Table 4.1 includes a list of all factors, nested factors, and interactions. It is vitally important to be able to specify all such terms in order to correctly allocate the proper degrees of freedom and to confirm the conceptualization of the design. Knowing that there are a total of 400 observations allows for 399 degrees of freedom. If the design is properly laid-out, factors correctly identified, and factor levels specified, simple arithmetic should yield this result. The reader will see this displayed in the table. Also included in the same table is a cross-check on the arithmetic under the heading of "count." By specifying each factor including replicates, a simple multiplication of each of their levels produces a check on the total number of observations. Note that interactions are omitted.

Figure 4.4 depicts the proper error term for testing each of the factors and interactions in the design. A description of the procedure used for this determination, as described by Millman and Glass (1967), appears in the appendices. The error terms for tests have been determined based on the factors pain, magnet, and trials being fixed while subjects are considered random.

Factor Analysis

Factor analysis is a powerful analytical tool that can

TABLE 4.1
CONCEPTUALIZATION OF FACTORS AND DEGREES
OF FREEDOM OF DESIGN

Factor	Calculated Degrees of Freedom (df)	Degrees of Freedom (df)
P	(2-1)	1
M	(2-1)	1
T	(4-1)	3
PM	(2-1)(2-1)	1
MT	(2-1)(4-1)	3
PT	(2-1)(4-1)	3
PMT	(2-1)(2-1)(4-1)	3
R/PM	(25-1)(2)(2)	96
RT/PM	(25-1)(4-1)(2)(2)	<u>288</u>
		399 Total

COUNT

p=2
m=2
t=4
r=25

Total Observations = $2 \times 2 \times 4 \times 25 = 400$

	RT/PM	R/PM	PMT	PT	MT	PM	T	M	P	Error
P	P	P	P	P		P			P	R/PM
M	M	M	M		M	M		M		R/PM
T	T		T	T	T		T			RT/PM
PM	PM	PM	PM			PM				R/PM
MT	MT		MT		MT					RT/PM
PT	PT		PT	PT						RT/PM
PMT	PMT		PMT							RT/PM
R/PM	R/PM	R/PM								No Test
RT/PM	RT/PM									No Test

P, M, and T are fixed
R is random

Figure 4.4 Conceptualization of Factors and Error Terms

be useful to the educational researcher for a varied number of applications. Principal components analysis (PCA) is usually subsumed under factor analysis though there are some conceptual differences between the two techniques. Principal components analysis will be employed within this research investigation for the following two purposes. First, to help demonstrate the construct validity of the dependent variable, EMG threshold response, and secondly as a means by which one can calculate the reliability of the dependent variable across repeated trials.

As Cronback and Meehl (1955) state: "Construct validity must be investigated whenever no criterion or universe of context is accepted as entirely adequate to define the quality to be measured" (p. 282). In discussing the basic rationale they add,

Construct validation takes place when an investigator believes his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses which are a means of confirming or disconfirming the claim. (p. 290)

More specifically, construct validity is concerned with the assessment of whether a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured. The EMG technique utilized in this study is a relatively new approach to the quantification of nerve and muscle pathology with regard to subjective neck and shoulder distress. This being the case, one basis for suggesting the

validity of the dependent measure is the experience of several physicians who specialize in EMG recording and in the clinical treatment of such related pathologies. Such reliance upon professional, judgmental evaluation reflects another form of validity, the content validity of the dependent measure. It was generally agreed prior to the start of this study that the suprascapular nerve and the supraspinatus and infraspinatus muscles were pathways that should reveal abnormality in the threshold response given true neck and shoulder discomfort in the subject.

To further test and demonstrate the construct validity of the dependent measure, another EMG measure called the F-wave response (discussed below) was collected from the subjects on a weekly basis (four repeated trials). The F-wave response is believed to measure different properties than the EMG threshold response, the dependent measure. Consequently, a principal components analysis will be useful to determine if, in fact, two distinctly different measures were collected. If so, this would serve as further evidence as to the specificity of the threshold response. The attempt to further validate the dependent measure using the F-wave response was implemented after the study had already begun. As a result, only 75 of the 100 participating subjects had F-wave measures collected.

The second purpose for which factor analysis (PCA) will be employed within this study is to demonstrate the reliability of the dependent measure (threshold response)

across the four trials. The physiological recording techniques utilized were very complex and the ability of the cooperating physicians to both stimulate the correct nerve and capture the proper response can not be taken for granted. Therefore, the reliability of the measurements will be determined using principal components analysis to calculate coefficient theta. This technique has been discussed by Zeller and Carmines (1980). Coefficient theta may be considered a maximized alpha coefficient which is the average of all possible correlations of the four trials (pp. 56, 62). This method will be extended to the F-wave response as well since this measure is itself being utilized for validation purposes.

F-Wave Conduction Latency Methodology

F-wave conduction latency was determined using standard techniques (Kimura, 1975; Wu and Stratigos, 1976). The ulnar nerve was stimulated supramaximally at the wrist and at the axilla with surface electrodes with a 0.1 msec wide square pulse (TECA-M EMG machine). Muscle action potentials were then obtained through a surface disc electrode placed over the hypothenar muscle of the right upper extremity about 7 cm from the stimulating electrodes.

A burst of eight stimuli was first delivered at the wrist, and muscle responses successively recorded with a storage oscilloscope. The F-wave with the shortest time to the initiation of the responses was measured to compute the

conduction latency. This procedure was repeated five times during each session. Polaroid films were used to document the responses. A stimulus was then applied at a point on the arm (axilla) about 25 cm from the sternal notch. The response (M-response) at the wrist was recorded on the oscilloscope and the M-wave latency measured. The axillary F-loop latency (AFL) was then calculated using the following equation (Wu and Stratigos, 1976):

$$\text{AFL} = (\text{FW} + \text{MW}) - 2(\text{MA})$$

where:

FW = F-Wave latency for stimulating at the wrist;

MW = M-Wave latency for stimulating at the wrist;

MA = M-Wave latency for stimulating at the axilla.

Statistical Analysis

The EMG threshold response data from this experiment will be comparatively analyzed using three variations of a repeated-measures analysis of variance (ANOVA). All analyses implicitly take into account the non-independent, correlated nature of the data. The primary difference in the three analyses concerns the manner in which the first trial measurement is handled. This first trial is a baseline measurement that had not been subjected to experimental treatment. To ignore this distinction potentially dilutes the effect of treatment. None the less, since many researchers are prone to include the baseline with the other trials, the first of the three repeated-

measures analyses will do likewise.

Regardless of the number of trials considered, the design consisted of three factors with repeated-measures on only one of the factors (trials). Each of the groups was observed under all levels of the trials factor, but each group was assigned to only one combination of the pain and magnet factors. R_{ij} describes the group of subjects assigned to treatment combination ab_{ij} , where "a" represents the pain factor and "b" represents the magnet factor. A subject within group R_{ij} is identified by the subscript $m(ij)$. This notation indicates that the subject effect is nested within both factors pain and magnet. Winer (1971) gives the structural model on which the statistical analyses are based as:

$$X_{ijkm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_{m(ij)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \gamma_{km(ij)} + \epsilon_{o(ijkm)}.$$

Since the subject factor is nested within treatment combinations, there can be no interaction between the pain and magnet factors and the subject factor. Winer also notes that this model implies homogeneity of the variance-covariance matrices associated with the repeated-measures.

The Statistical Package for the Social Sciences (SPSS, version 10) will be used for all repeated-measures analyses. The MANOVA procedure allows for the identification of two grouping factors (pain and magnet conditions) and one trials factor (the repetitions of the dependent measure). MANOVA is a generalized multivariate analysis of variance and

covariance program that performs univariate and multivariate linear estimation and tests of hypotheses for any crossed and/or nested design with or without covariates. The user has complete control over model specification. One advantage of MANOVA is that the procedure will read repeated-measures data in the straight-forward manner depicted in Figure 4.5. Given the repeated measurements, SPSS refers to this format as "the multivariate data setup." That is, all scores for the same subject should reside on the same case. This is contrasted with the "classical" setup where every measure is represented as a separate line in a data file. SPSS will correctly identify multivariate data as repeated-measures only if the control phrase "ANALYSIS(REPEATED)/" is used within the MANOVA procedure. (See Appendix A for the complete control card set-up for the three repeated-measures analyses.)

The user must include the "WSFACTORS" subcommand in combination with the multivariate data setup. WSFACTORS provides the names and number of levels for within-subject factors, which in this case is "TRIALS" the repeated-measure. The user must also specify a within-subjects model so that SPSS can create a within-subjects transformation matrix from the basis matrices corresponding to the contrast matrices specified within any previous "CONTRAST" statement. If contrasts are not requested, then deviation contrasts are assumed by default. The within-subjects model is specified using the "WSDESIGN" subcommand. The last subcommand for a

<u>COLUMNS</u>	<u>FIELD</u>
1-3	Subject ID
4	Pain Condition 1=no-pain 2=pain
5	Magnetic Condition 1=non-magnetic 2=magnetic
6-10	First threshold response measure
11-15	Second threshold response measure
16-20	Third threshold response measure
21-25	Fourth threshold response measure
26-29	First F-wave measure
30-33	Second F-wave measure
34-37	Third F-wave measure
38-41	Fourth F-wave measure

Figure 4.5 Format of Data File (one record per case)

given analysis must be "DESIGN." This subcommand specifies the between-subjects model for the analysis. For the current study, "DESIGN/" requests a full factorial analysis. Normally the procedure output produces the familiar ANOVA F test table only for grouping factor main effects (including treatment) and their interactions. In order to obtain an ANOVA table for the repeated-measure, its interactions with the main effects, and all higher-order interactions the user must code "SIGNIF(AVERF)" within the "PRINT" subcommand.

The second repeated-measures analysis is predicated on the notion that since the baseline measure dilutes treatment effect, it is best to eliminate it from the analysis completely. This supposition will be examined recognizing that dropping the first threshold measure gives up valuable information. The second analysis easily follows from the first. The only changes that need be made in the SPSS control statements are: (1) to indicate that there are now three repetitions of the trials factor rather than the previous four; and (2) to delete the first threshold measure from the list of dependent variables.

The third analysis attempts to use all the available information generated by the experiment. Here, the baseline will be used as a covariate so as to remove initial differences present at the first trial from the subsequent three repeated trials. In effect, this establishes a uniform "starting point" for all subjects given the individual physiological differences they bring to the

experiment. Such a standardization allows for a cleaner, more precise evaluation of the experimental effect. Consequently, this last of the three repeated-measures analyses offers the investigator the greatest utility for evaluating such an experimental paradigm.

Analysis of covariance (ANCOVA) is a method of statistical rather than experimental control. As such it is a general technique for increasing the precision of any experiment. This is accomplished by adjusting the dependent variable in terms of one or more covariates. Dayton (1970) states that "the adjustment of criterion scores will effect a reduction in the size of experimental error and produce within-cell measures of greater homogeneity" (p. 304). He adds that the success of the adjustment depends on the degree of linear correlation between the covariate(s) and the dependent measure.

The analysis of covariance, like ANOVA, is attributable to R. A. Fisher. Cox and McCullagh (1982) cite Eden and Fisher as first giving the decomposition of a sum of products and using the corresponding correlation coefficients descriptively. They state that the procedure for using analysis of covariance for precision improvement described by Fisher in Statistical Methods for Research Workers (1932) was first used by Sanders (1930) at Fisher's suggestion. He applied standard ANOVA to the term $Y - bX$ where Y is the dependent variable, X is the covariate, and b is the regression coefficient estimated from the regression

of Y on X . This provided an approximate F test of treatments. Papers by Wishart (1934) and Wilsdon (1934) presented to the Royal Statistical Society both had contributions by Fisher who by that point had worked out an exact F test. Cox and McCullagh cite E. S. Pearson's Appendix to Wilsdon's paper as setting out detailed calculations for the description and decomposition of regression relations. They further note that the idea of components of regression was developed by Tukey (1951).

There is a fundamental relationship between linear regression and the analysis of covariance (ANCOVA). In ANCOVA the influence of the uncontrolled variable, also called the covariate or concomitant variable, is removed from the dependent variable by linear regression. The residual sum of squares represents the portion of variance in the dependent variable that cannot be predicted from the covariate. The residual sum of squares is used to provide variance estimates which in turn are used to make tests of significance.

The nature of regression and ANCOVA will be further explored in order to demonstrate the role of the residual sum of squares in the analysis. Much of the following exposition comes from Ferguson (1971). For the remainder of this discussion the author will adopt Ferguson's notation of representing the dependent variable as X and the covariate as Y .

Given paired data comprised of X a dependent variable,

and Y a covariate, if the total, within-groups, and between-groups sums of cross-products are divided by the respective sums of squares, three regression coefficients are obtained. From this it follows that three regression lines can be identified. The first line represents the total over-all regression for predicting X from a knowledge of Y using all observations. The slope of this line is defined as:

$$b_t = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X})(Y_{ij} - \bar{Y})}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2}$$

A second regression coefficient (line) represents the overall within-groups regression. In predicting X from a knowledge of Y each of the k groups may be considered separately. Each group has its own within-group regression line with slope b_j . Ferguson states that these separate regression lines may be pooled thus producing an over-all groups regression line with slope:

$$b_w = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2}$$

The numerator is the within-groups sum of cross-products and the denominator is the within-groups sum of squares for the covariate. The pooling process assumes a homogeneity of slope among the groups.

A third regression coefficient is obtained by dividing the between-groups sum of cross-products by the between-groups sum of squares for the covariate. The slope is given

as:

$$b_y = \frac{\sum_{j=1}^k n_j (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^k n_j (Y_j - \bar{Y})^2}$$

Ferguson notes that the within-groups regression equation is of particular interest in ANCOVA. That equation has the form:

$$X'_{ij} = b_w(Y_{ij} - \bar{Y}_j) + \bar{X}_j$$

ANCOVA makes use of the sum of squares of residuals of the dependent variable about this regression line.

The total sum of squares for both the dependent variable and the covariate can be partitioned into within-group and between-group sums of squares. Similarly, the total sum of cross-products can be partitioned into within-groups and between-groups components. The general form of the equation for the linear prediction of the dependent variable X , given the covariate Y is:

$$X'_i = b_{xy}(Y_i - \bar{Y}) + \bar{X}$$

where b_{xy} is the slope of the line. The sum of squares of residuals about this line is:

$$\sum_{i=1}^N (X_i - X'_i)^2.$$

By substituting $b_{xy}(Y_i - \bar{Y}) + \bar{X}$ for X'_i it follows that the total sums of squares of residuals is:

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - X'_{ij})^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} [(X_{ij} - \bar{X}_j) - b_y(Y_{ij} - \bar{Y}_j)]^2 \\ &= \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 - \frac{\left[\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j) \right]^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2} \end{aligned}$$

This quantity is also called a reduced or adjusted total sum of squares. Note that b_t the regression coefficient for all observations is used above.

An adjusted within-groups sum of squares has a similar form but utilizes b_w as the regression coefficient and within-groups means. It is given by:

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - X'_{ij})^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} [(X_{ij} - \bar{X}_j) - b_w(Y_{ij} - \bar{Y}_j)]^2 \\ &= \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 - \frac{\left[\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j) \right]^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2} \end{aligned}$$

The adjusted sum of squares for between-groups can be obtained by subtraction. It is the difference between the adjusted total sum of squares and the adjusted within-groups sum of squares on the dependent variable.

The number of degrees of freedom associated with the adjusted total sum of squares on the dependent variable is $N-2$ since one degree of freedom is lost due to the covariate in the adjustment process. The adjusted within-groups sum of squares has $N-k-1$ degrees of freedom, where k is the number of groups. The adjusted between-groups sum of squares has $k-1$ degrees of freedom. Variance estimates for between-groups and within-groups are obtained by dividing the adjusted sums of squares by the respective degrees of freedom. Ferguson states that these variance estimates are interpreted as in ANOVA, except that the null hypothesis relates to adjusted treatment means that are free of the linear effect of the covariate.

It should be noted that though the general precepts of ANCOVA are as outlined above, there is some controversy in the literature regarding this technique. Bingham and Fienberg (1982) have discussed whether partial sums of squares should be computed hierarchically or whether they should be computed after all other model terms, including higher-order interactions, have been fitted for both balanced and unbalanced models. Though they prefer the hierarchical approach to ANCOVA, their primary concern is the omission of this discussion from standard texts.

Searle and Hudson (1982) have examined the manner in which different statistical packages compute ANCOVA. They have found different results especially for unbalanced models. They state "computer output for analysis of covariance is not all that it is made out to be by its labeling. Values with labels that appear to be the same can be quite different because they do in fact represent different calculations" (p. 744). Researchers must exercise caution in the interpretation of results. The running of known sample problems can be useful for confirming the interpretation of statistical software output.

For the third repeated-measures analysis, the first threshold measure must be designated as the covariate. After listing the dependent variables (threshld2 to threshld4) and the grouping factors (pain and magnet and their levels) the covariate is indicated by use of the "WITH" qualifier. The MANOVA procedure, however, assigns

the covariate on a per-trial basis. Therefore, two additional covariates, each equivalent to the first threshold measure, must be incorporated into the last analysis (see Appendix A).

Once completed, all three repeated-measures analyses will be compared to see whether they vary in their statistical evaluation of this experiment. Again, the issue being investigated is whether there has been significant improvement in the EMG threshold response for those who suffer neck and shoulder pain and were exposed to the experimental treatment. Intrinsically, improvement is defined in comparative terms. Since the design utilizes a 2 x 2 ANOVA, there are four combinations of the pain and magnetic conditions: (1) no-pain, non-magnetic; (2) no-pain, magnetic; (3) pain, non-magnetic; and (4) pain, magnetic. The evaluation is concerned with whether the pain x magnetic group differs after treatment relative to the other groups. Thus it is the interaction of these two factors that is of greatest interest rather than either main effect. Also of interest will be the trials factor. The substantive question here is whether there are significant differences in treatment effect between the trials (repeated-measures).

It is believed that the analysis utilizing the covariate will be the most cogent of the three since initial individual differences will have been removed. However, in order to more fully evaluate the results of this study, the author intends to perform both a one-way ANOVA and a two-

group discriminant analysis for the purpose of determining if there was a statistically significant difference between the pain and the no-pain subjects prior to any experimental treatment. The dependent measure for both these supplemental analyses will be the initial EMG threshold response baseline measure. If the two repeated-measures analyses that do not utilize the covariate find differences in the pain main effect, this potentially might simply reflect a "no-change" situation since there may have been initial differences between pain and no-pain subjects. The one-way ANOVA and the two-group discriminant analyses could provide useful supplemental evidence as to the ability of the repeated-measures covariate technique to remove such differences. This approach serves to caution other investigators against performing a repeated-measures analysis without regard to such initial differences. The first two analyses, neither of which removes the effect due to the covariate, conceivably might produce spurious results.

CHAPTER 5

RESULTS AND DISCUSSION

The dependent measure of this study is the EMG threshold response. A typical threshold response waveform, as recorded from an experimental subject, is displayed in Figure 5.1. This figure graphically depicts the supraspinatus muscle responses to subthreshold, threshold, and suprathreshold stimulation of its peripheral nerves. The mean excitation threshold, accompanied by the respective standard deviations, for each of the four groups of 25 subjects is presented in Table 5.1.

F-wave response data were collected during the course of this study in the attempt to document the construct validity of the EMG threshold response. An example of a typical F-wave, as recorded from a subject's hypothenar muscle in the hand upon supramaximal stimulation of the peripheral (ulnar) nerve, is shown in Figure 5.2. The conduction latency means and standard deviations for each of the four groups appears in Table 5.2.

The research data will be subjected to three variations of repeated-measures analysis. The three methods to be utilized are: (1) repeated-measures including the baseline as the first repeated-measure; (2) repeated-measures ignoring the baseline completely; and (3) repeated-measures using the baseline as a covariate. These three approaches

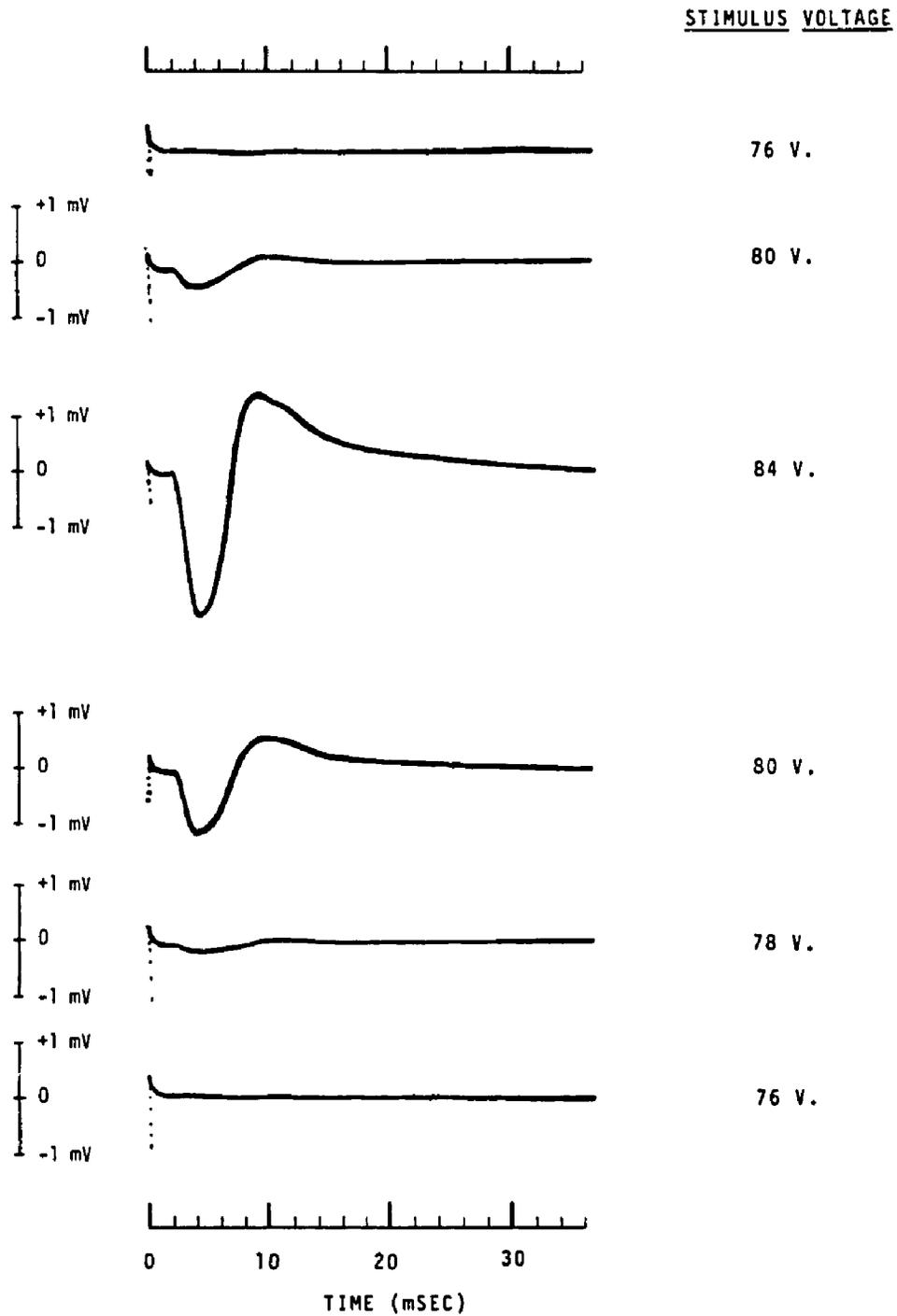


Figure 5.1 Electromyograms of Supraspinatus Muscle Response to Subthreshold, Threshold, and Superathreshold Stimulation

TABLE 5.1
MEANS AND STANDARD DEVIATIONS OF
EXCITATION THRESHOLD (IN VOLTS)

	MAGNETIC		NON-MAGNETIC	
	Mean	S.D.	Mean	S.D.
<u>PAIN</u>				
Trial 1 (Baseline)	68.50	25.20	64.73	19.89
Trial 2	68.07	24.52	65.28	23.62
Trial 3	65.86	23.72	62.34	23.61
Trial 4	63.41	26.72	65.45	23.34
<u>NO-PAIN</u>				
Trial 1 (Baseline)	53.56	24.22	53.71	23.78
Trial 2	51.87	18.07	53.87	23.04
Trial 3	53.20	23.38	58.40	26.06
Trial 4	52.47	19.37	57.05	27.26

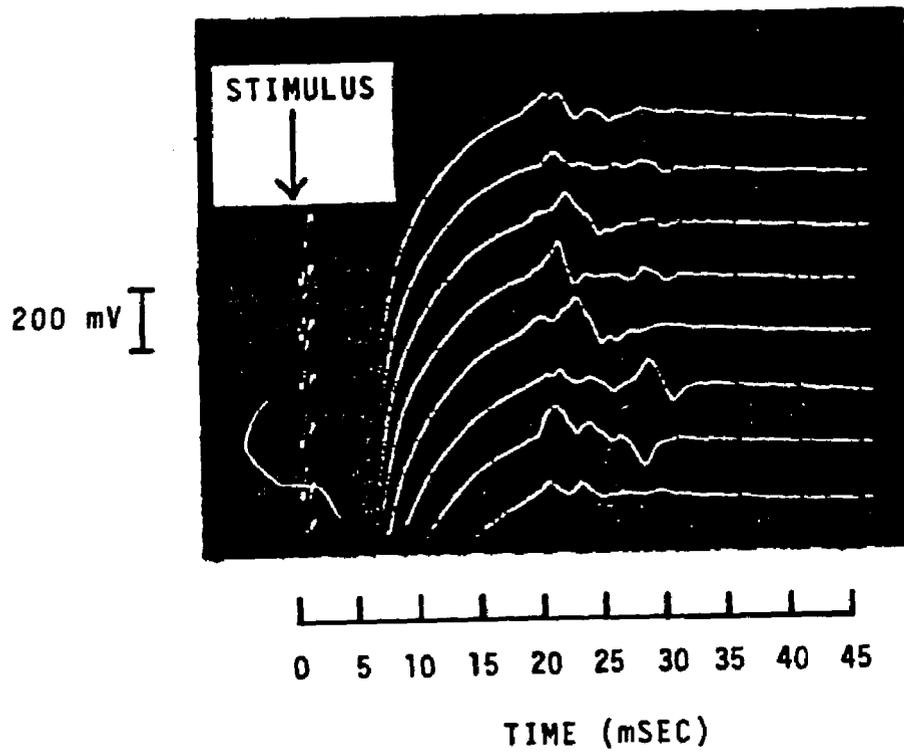


Figure 5.2 F-waves Recorded From Hypothenar Muscle Following Supramaximal Stimulation of the Ulnar Nerve

TABLE 5.2
MEANS AND STANDARD DEVIATIONS OF
F-WAVE RESPONSE (IN VOLTS)

	MAGNETIC		NON-MAGNETIC	
	Mean	S.D.	Mean	S.D.
<u>PAIN</u>				
Trial 1	9.41	0.80	9.18	1.25
Trial 2	9.46	0.92	9.13	1.19
Trial 3	9.18	1.11	9.05	1.07
Trial 4	9.12	1.10	8.99	1.12
<u>NO-PAIN</u>				
Trial 1	9.04	0.88	9.34	0.63
Trial 2	9.30	0.94	9.20	0.81
Trial 3	8.81	1.02	9.46	1.06
Trial 4	8.74	0.88	9.30	0.91

will be compared and contrasted so as to examine the differential outcomes resulting from the role of the baseline.

Inspection of summary Table 5.1, which gives the means and standard deviations of the dependent measure for each group across all four trials, reveals a noticeable difference on the pain dimension. All pain subjects, regardless of magnetic condition or trial level, appear to have a higher threshold response level than no-pain subjects. The mean EMG response for all pain subjects across all trials was 65.45. In contrast, the mean EMG response for all no-pain subjects across all trials was 54.27. It is further noted that the pain-magnetic group showed some decline in EMG response across trials. However, it is also noted that the no-pain - non-magnetic group showed an increase in EMG response across trials. These empirical observations will be statistically tested and discussed later in this section.

Principal Components Analysis

As was discussed in the previous chapter, principal components analysis is employed in this study for the purpose of evaluating the construct validity and reliability of the dependent measure, the EMG threshold response. By introducing another response measure, the F-wave, whose properties are known to differ from the dependent measure, one expects the factors to differentiate between these

supposedly different concepts. Though by no means offering conclusive proof as to the validity of the EMG threshold response, the results of the factor analysis clearly demonstrate that two unrelated concepts (factors) can be identified. Using the SPSS PA2 Factor Program (version H, release 9.1) simple structure was obtained using a two factor solution. The PA2 procedure starts with communality estimates on the main diagonal of the correlation matrix. These estimates represent the squared multiple correlation between a particular variable and the remaining variables in the correlation matrix. The program then successively iterates. It extracts the specified factors and calculates the accounted-for variance. It then replaces these values on the diagonal of the matrix. When there is no further improvement in the communality estimates, the procedure terminates.

Table 5.3 shows final communalities associated with each variable. In addition, the eigenvalue and percent of variance accounted for by each factor is also displayed. In this two factor solution each factor accounts for a large percentage of the variance within the data. Factor 1, representing the F-wave, accounts for 53.1% of the total variance while factor 2, the EMG threshold response, accounts for 46.9%. Table 5.4 shows factor loadings for each variable after a varimax rotation has been completed. Using an orthogonal rotation such as varimax facilitates attaining "simple structure." All the EMG threshold

TABLE 5.3

COMMUNALITIES OF PHYSIOLOGICAL MEASURES AND
EIGENVALUES AND PERCENT OF ACCOUNTED-FOR
VARIANCE OF FACTORS

VARIABLE	ESTIMATED COMMUNALITY	FINAL COMMUNALITY
THRSLD1	0.39988	0.43506
THRSLD2	0.54882	0.68108
THRSLD3	0.54246	0.66478
THRSLD4	0.46004	0.51969
F1	0.50085	0.55080
F2	0.5115	0.59156
F3	0.56816	0.57440
F4	0.65206	0.75612

FACTOR	FINAL EIGENVALE	PERCENT OF VARIANCE	CUMULATIVE PERCENTAGE
1 (F-wave)	2.53502	53.1	53.1
2 (Thrslid)	2.23840	46.9	100.0

TABLE 5.4
VARIMAX ROTATED FACTOR MATRIX

VARIABLE	FACTOR 1	FACTOR 2
THRSLD1	-0.23800	0.61515
THRSLD2	0.11234	0.81759
THRSLD3	0.09137	0.81021
THRSLD4	0.01970	0.72062
F1	0.73215	-0.12150
F2	0.76884	-0.02099
F3	0.75334	0.08295
F4	0.86593	0.07922

measures load heavily on factor 2 and negligibly on factor 1. Conversely, all the F-wave variables load heavily on factor 1 and negligibly on factor 2. Consequently one can infer factor 1 to be a discrete F-wave factor. Similarly, factor 2 can be inferred to be a discrete threshold factor.

The idea of simple structure goes back to Thurstone (1947) who thought that most data sets actually contained only several obscure factors each of which would be associated with several of many variables. It was anticipated that each variable in the data set would load heavily on only one or perhaps a few of the numerous factors. In order to aid interpretation, the goal was to find clusters of variables, each of which defined only one factor. Furthermore, interpretation is facilitated if simple structure is obtained using an orthogonal rotation in as much as the factors are then uncorrelated. Such is the case with the F-wave and EMG threshold factors. This provides conceptual clarity and evidence for the specificity of the dependent measure, the EMG threshold response. The content validity, as determined by the professional judgement of the participating physicians, in addition to the attainment of "simple structure" described above, give reasonable support for the validity of the dependent measure.

Another benefit of principal components analysis is the ability to determine the reliability of the four repeated

EMG trials by calculating coefficient theta. As previously discussed, this is done in the manner described by Zeller and Carmines (1980). Table 5.5 lists the communality associated with each of the four threshold trials variables and the eigenvalues associated with each of the specified four factors. These values are derived from a separate analysis containing only the four threshold measurements (trials). No rotation was performed since this application was not concerned with data reduction.

Coefficient theta is akin to a maximized alpha coefficient which is the average of all possible correlations of the four trials. Theta is calculated according to the formula:

$$\text{Theta} = (N/(N-1)) * (1-(1/E))$$

where N = number of variables

E = largest eigenvalue

Solving the above equation produces a conservative EMG theta value of .74 using eigenvalues based on iterated, final communalities. If, in the more conventional manner, eigenvalues based on initial estimates of communalities are used, the value of theta increases to .83. In either case the reliability value can be considered good and is ample evidence as to the repeatability of the physiological measurement over trials. This provides confidence that whatever was being measured in the name of the EMG threshold response, was being measured consistently.

Since the F-wave measure is being used to help validate

TABLE 5.5
COMMUNALITIES, EIGENVALUES, AND PERCENT OF
ACCOUNTED-FOR VARIANCE OF THRESHOLD
RESPONSE FACTORS

VARIABLE	ESTIMATED COMMUNALITY	FINAL COMMUNALITY	
THRSLD1	0.30148	0.40305	
THRSLD2	0.52720	0.72237	
THRSLD3	0.53823	0.71676	
THRSLD4	0.44941	0.58695	

FACTOR	FINAL EIGENVALUE	PERCENT OF VARIANCE	CUMULATIVE PERCENTAGE
1	2.26327	93.2	93.2
2	0.14863	6.1	99.4
3	0.01647	0.7	100.0
4	-0.00077	0.0	100.0

the EMG threshold response, its reliability warrants examination. Coefficient theta is therefore calculated for the F-wave since a measure that is valid but not reliable is of little worth. This supplemental analysis incorporated only the four F-wave measurements (trials) for the purpose of determining each trial's eigenvalue. The calculated conservative theta for the F-wave is .81 using final, iterated communalities (Table 5.6). Initial communalities produce a value of .86. One can safely conclude the F-wave to be a very reliable measure.

Discriminant Analysis and One-Way ANOVA

A two-group discriminant analysis was utilized for the purpose of detecting initial differences on the dependent measure between pain and no-pain subjects prior to any exposure to treatment. There is much research evidence suggesting that the experience of pain is influenced by subjective factors. It is recalled that during the initial screening of potential subjects considerable efforts were taken to make a careful determination as to the medical history of those individuals who claimed to suffer with neck and shoulder discomfort. Individuals for whom a clear determination could not be made were eliminated from the study. Of course, these efforts alone do not guarantee all participants to be bonafide sufferers of the pathology under examination. Pain remains a subjective experience.

The two-group discriminant analysis attempts to obviate

TABLE 5.6
COMMUNALITIES, EIGENVALUES, AND PERCENT OF
ACCOUNTED-FOR VARIANCE OF F-WAVE
RESPONSE FACTORS

VARIABLE	ESTIMATED COMMUNALITY	FINAL COMMUNALITY	
F1	0.47251	0.64816	
F2	0.49632	0.67843	
F3	0.56578	0.72140	
F4	0.64497	0.84515	

FACTOR	FINAL EIGENVALUE	PERCENT OF VARIANCE	CUMULATIVE PERCENTAGE
1	2.54818	88.1	88.1
2	0.27318	9.4	97.6
3	0.07075	2.4	100.0
4	-0.00102	0.0	100.0

a subjective evaluation by examining group differences in objective terms; namely, measurable differences in physiology. To find a statistically significant difference between the pain and no-pain groups at the onset would give further supporting evidence to the validity of the dependent measure. The EMG threshold response had been selected as the dependent measure in this study precisely because it is believed to reflect the underlying pathology of interest. Furthermore, since two of the three repeated-measures analyses to be reported fail to compensate for initial individual (and therefore group) differences on the dependent measure, the outcome of the two-group discriminant analysis and one-way ANOVA should provide valuable information with regard to this issue.

Table 5.7 presents the results of the first analysis. In fact, the canonical discriminant function is significant ($p=0.006$). The inference is drawn that pain subjects were statistically different from no-pain subjects when measured on the EMG threshold response at the first trial prior to being exposed to any treatment.

For further confirmation of the disparity between pain and no-pain subjects on the dependent measure at the start of the study, a one-way ANOVA was performed to supplement the above analysis. Like the two-group discriminant function, the one-way ANOVA results indicate a significant initial difference on the pain dimension between pain and no-pain subjects (Table 5.8, $p=0.006$). The supposition that

TABLE 5.7

CANONICAL DISCRIMINANT FUNCTION FOR TWO-GROUP
DISCRIMINANT ANALYSIS USING INITIAL
THRESHOLD RESPONSE

FUNCTION	EIGENVALUE	RELATIVE VARIANCE	CUMULATIVE PERCENT	CANONICAL CORRELATION
1	0.08012	100.00	100.00	0.2723607

AFTER FUNCTION	WILK'S LAMBDA	CHI-SQ.	D.F.	SIG.
0	0.9258196	7.5149	1	0.0061

TABLE 5.8
1-WAY ANALYSIS OF VARIANCE OF
BASELINE MEASURE

SOURCE	SS	DF	MS	F	PROB.
PAIN	4212.14	1	4212.14	7.85	0.006
RESIDUAL	52570.27	98	536.43		

the EMG threshold response has validity as a dependent measure is supported since the results of the discriminant analysis and the one-way ANOVA demonstrate the effectiveness of the EMG measure to differentiate pain from no-pain subjects.

Repeated-Measures Analysis Using All Trials

Having presented reasonable evidence for both the validity and reliability of the EMG threshold response, the focus of attention now turns to the first of the three comparative repeated-measures analyses. Table 5.9 displays the repeated-measures ANOVA results (all trials). Inspection of the F values and associated probabilities reveals a significant main effect for the pain dimension. This is interpreted to mean that regardless of magnetic condition, there is a statistically significant difference ($p=0.004$) between pain subjects and no-pain subjects across all four trials, conceptualizing all four trials as being pooled. This can be visualized by referring to Figure 5.3, which is derived from Figure 4.3, Design Conceptualization. Note that both magnet and trials factors are absent. The pain main effect is only concerned with mean differences between the left-half and right-half sides of the table. The difference between the mean (marginal) EMG threshold value for the pain side of the table (65.45) and the mean (marginal) EMG threshold value for the no-pain side of the table (54.27) only has a probability of 0.004 of occurring

TABLE 5.9
REPEATED-MEASURES ANALYSIS OF VARIANCE,
ALL TRIALS

SOURCE	SS	DF	MS	F	PROB.
PAIN	12517.13	1	12517.13	8.81	0.004
MAGNET	23.52	1	23.52	0.02	0.898
PAIN X MAGNET	622.91	1	622.90	0.44	0.509
R/PM	136351.47	96	1420.33	ERROR 1	
TRIALS	15.75	3	5.25	0.02	0.996
PAIN X TRIALS	518.93	3	172.98	0.64	0.590
MAGNET X TRIALS	352.56	3	117.52	0.44	0.728
PN X MAG X TRLS	132.05	3	44.02	0.16	0.921
RT/PM	77777.87	288	270.06	ERROR 2	

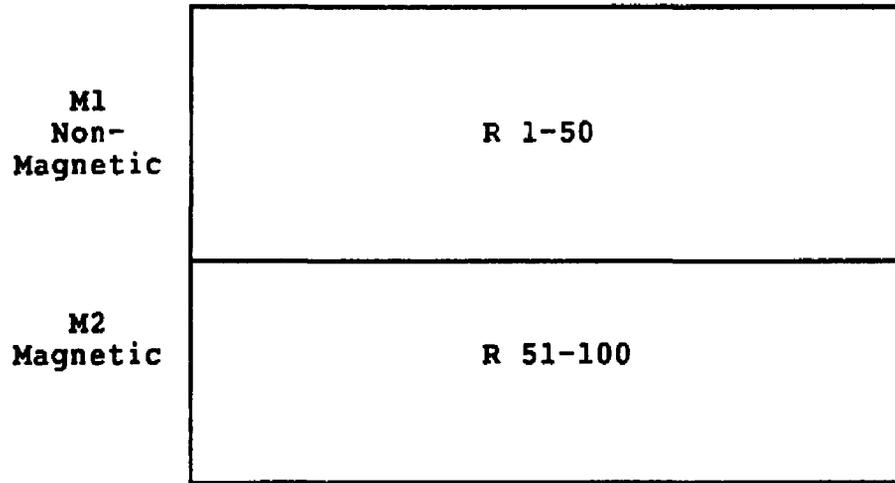
P1 No-Pain	P2 Pain
R 1-25 & R 51-75	R 26-50 & R 76-100

**Figure 5.3 Conceptualization of Pain Factor
as a Main Effect**

by chance alone.

The previously reported two-group discriminant analysis and one-way ANOVA demonstrated that pain and no-pain subjects were significantly different at the onset of the study prior to any treatment. The significant pain main effect indicates this circumstance has not changed after completion of treatment. Therefore we have a no-change situation whereby pain subjects were statistically different from no-pain subjects both prior to and upon completion of treatment.

Further inspection of results in Table 5.9 shows a nonsignificant result for the magnet factor ($p=0.898$). Referring to Figure 5.4, the graphic interpretation is that the marginal mean value for the top-half of the table (60.10) is not statistically different than the bottom-half of the table (59.62). Across all treatment trials, there is no statistically significant difference between the magnetic and non-magnetic conditions regardless of pain condition or trial. No comparison of magnetic and non-magnetic initial conditions was planned since it was assumed that random assignment of subjects to magnetic condition would preclude any systematic effect. It is argued that subjects do not "bring" any magnet condition, predisposition, or magnetic bias to the experiment. However, one can extract from Table 5.1 enough information (means) to make an informal evaluation of this assumption. Calculating relevant means on only the first threshold measure reveals a value of 59.22



**Figure 5.4 Conceptualization of Magnet Factor
as a Main Effect**

for the non-magnetic condition marginal and a value of 61.03 for the magnetic condition marginal. This suggests, as expected given random assignment, that there is no difference in threshold response attributable to magnetic condition assignment. Considering all trial data, as before there is a no-change outcome.

The next topic to be discussed is the pain x magnet interaction. It should be recalled that this pain x magnet interaction is the outcome of primary interest regarding the efficacy of the experimental magnetic therapy. If there is a treatment effect, it has been anticipated to manifest itself differentially across the pain dimension. Graphically, it is the lower right portion of Figure 5.5 (Conceptualization of Pain x Magnet Interaction) that should reflect any effect within its group mean (pain-magnetic group combination). Though Table 5.1 shows some change in EMG response for this group, Table 5.9 indicates that the pain x magnet interaction is not statistically significant ($p=0.509$). Within the context of this all-trials repeated-measures analysis, one must accept the null hypothesis of no differential effect and conclude that after four trials (baseline and three exposures to treatment) the experimental treatment had no discernable effect. This all-trials analysis fails to offer any supporting evidence for the contention that magnetic treatment is effective in the relief of neck and shoulder discomfort.

Though the main experimental question has been answered

	P1 No-Pain	P2 Pain
M1 Non- Magnetic	R 1-25	R 26-50
M2 Magnetic	R 51-75	R 76-100

**Figure 5.5 Conceptualization of Pain
 by Magnet Interaction**

(at least for the all-trials analysis), there is another factor and several interactions that deserve attention due to their importance within the context of repeated-measure designs. Testing the trials factor investigates whether, regardless of group, there are any differential effects among the trials. This is really an examination of the relationship between treatment and time. This can be visualized in Figure 5.6. Conceptually, this table represents a complete collapsing of the original design such that all 100 subjects now appear within each trial. The design can now be viewed as four vertical "stacks" analogous to a one-way ANOVA with four groups. Results indicate that there are no statistically significant differences between the trials ($p=0.996$). In fact, the differences in the trial means are so small that statistical testing becomes superfluous (60.13, 59.77, 59.95, 59.59).

The trials x pain interaction examines whether there are differential effects among the trials across the pain dimension. The visualization of this interaction is displayed in Figure 5.7. Note that the middle horizontal dividing line present in the original design conception (Figure 4.3) is now absent. This allows pooling of subjects in a vertical direction eliminating any distinction attributable to the magnet dimension. Therefore, the top-half of the table is collapsed or overlaid on the bottom-half of the table. Testing this interaction reveals no statistically significant differential effects among the

T1	T2	T3	T4
R 1-100	R 1-100	R 1-100	R 1-100

**Figure 5.6 Conceptualization of Trials Factor
as a Main Effect**

**P1
No-Pain**

**P2
Pain**

P1 No-Pain				P2 Pain			
T1	T2	T3	T4	T1	T2	T3	T4
R 1- 25	R 1- 25	R 1- 25	R 1- 25	R 26- 50	R 26- 50	R 26- 50	R 26- 50
&	&	&	&	&	&	&	&
R 51- 75	R 51- 75	R 51- 75	R 51- 75	R 76- 100	R 76- 100	R 76- 100	R 76- 100

**Figure 5.7 Conceptualization of Pain
by Trials Interaction**

trials when compared between pain and no-pain conditions ($p=0.590$).

Similarly, the trials x magnet interaction is also not statistically significant ($p=0.728$). The conceptualization of this interaction removes the vertical dividing line that partitions pain subjects from no-pain subjects. The left-half side of Figure 4.3 can then be collapsed over the right-half side of the table eliminating the distinction between pain and no-pain subjects while preserving the distinction between magnetic and non-magnetic conditions. This interaction is concerned with differential effects for trials dependent on magnetic condition. The conceptualization of this interaction is shown in Figure 5.8. For illustrative purposes the magnet dimension has been reoriented to the top of the table to most effectively display its relationship to the trials factor.

Lastly, the trials x pain x magnet 3-way interaction requires no conceptual alteration of the original design scheme (Figure 4.3). That figure, in fact, is a visual representation of the three factors pain, magnet, and trials. Here the notion of interaction investigates whether there are any differential effects among the trials dependent on which major quadrant of the table is examined. Each major quadrant represents a different combination of the pain and magnet conditions. Results show this interaction to be not statistically significant ($p=0.921$).

It should be noted that the pain and magnet main

M1 Non-Magnetic				M2 Magnetic			
T1	T2	T3	T4	T1	T2	T3	T4
R 1- 50	R 1- 50	R 1- 50	R 1- 50	R 51- 100	R 51- 100	R 51- 100	R 51- 100

Figure 5.8 Conceptualization of Magnet
by Trials Interaction

effects, as well as the pain x magnet interaction, were tested using R/PM as the error term. That is, variance attributable to individuals nested within a major quadrant of the table served as the denominator for the above mentioned F tests. The remaining factor (trials) and all other interactions were tested using RT/PM as the error term. This second error term calculates variances within each trial condition nested within each major quadrant. Figure 4.4 lists the proper test for each factor. The general procedures for determining the test error term are covered in the appendices.

Repeated-Measures Analysis Dropping the Baseline

The second repeated-measures analysis excludes the initial baseline measurement completely. It is recalled that the justification for this omission is that since the baseline was not subjected to treatment, it might only serve to dilute any possible treatment effect. This second repeated-measures analysis examines this contention. The design conceptualization is altered at this point so that where previously there were four trials, there are now three. This change produces fewer observations and therefore fewer degrees of freedom within the analysis. This reduction is reflected in the ANOVA results tables. Previous factor and interaction conceptualization illustrations will not be repeated to reflect this change since there is no difference in the way factors and

interactions are interpreted. Readers are simply advised that the current analysis, as well as the one to follow, utilize three, rather than four, measurement trials.

The ANOVA results (Table 5.10) show a reduction in sums of squares for the pain and trials factors and all interactions when contrasted with the first all-trials analysis. Interestingly, the magnet factor shows an increase in sums of squares after dropping the baseline suggesting that the first trial had a pronounced effect on magnetic condition means.

Further inspection of Table 5.10 again reveals a significant main effect for the pain factor ($p=0.011$). Dropping the baseline measure from the analysis has not altered the condition whereby pain subjects were statistically different from no-pain subjects both prior to and upon completion of treatment. It is interesting to note that there has been a noticeable decrease in probability associated with the pain main effect as a result of excluding the baseline measure. In the prior analysis the probability of the pain main effect was 0.004 versus 0.011 in the current analysis. This suggests some decrease in the degree of association between the dependent variable and the pain dimension. Though there has been a diminution, the "p" value remains highly significant indicating that the difference in means between the pain and no-pain groups had a low likelihood of occurring by chance.

The pain x magnet interaction is not significant

TABLE 5.10
 REPEATED-MEASURES ANALYSIS OF VARIANCE,
 DROPPING BASELINE

SOURCE	SS	DF	MS	F	PROB.
PAIN	8412.06	1	8412.06	6.72	0.011
MAGNET	117.00	1	117.00	0.09	0.761
PAIN X MAGNET	536.56	1	536.56	0.43	0.514
R/PM	120197.90	96	1252.06	ERROR 1	
TRIALS	6.30	2	3.15	0.01	0.986
PAIN X TRIALS	411.86	2	205.93	0.95	0.388
MAGNET X TRIALS	177.45	2	88.73	0.41	0.664
PN X MAG X TRLS	122.45	2	61.22	0.28	0.754
RT/PM	41538.75	192	216.35	ERROR 2	

($p=0.514$). It is recalled that this interaction is of major concern with regard to treatment effect. The probability associated with the pain x magnet interaction differed little from the first analysis ($p=0.509$). Furthermore, magnet and trials factors and all other interactions are also not significant in this second analysis. Their "p" values (see Table 5.10) are again little different than in the first analysis.

In this particular study, dropping the baseline measurement has little effect upon the results of the analysis. It should be kept in mind though, that within each of the four groups, the baseline measure was only marginally different from the remaining three trail measures (refer to Table 5.1). For circumstances where more variance exists between trials, dropping the baseline may have a more dramatic effect.

Repeated-Measures Analysis Using Baseline as Covariate

By using the baseline as a covariate in this third and final analysis, total, between-group, and within-group sums of squares are adjusted for the linear effect of the covariate. The purpose of this methodology is to extract from all trials subjected to treatment, that portion of each score that can be explained by the initial baseline measure. In this way, any initial differences between pain and no-pain subjects are removed. This analysis, like the second, has fewer degrees of freedom than the first analysis as a

result of having only three repeated trials. Additionally, another degree of freedom is lost due to the covariate.

Inspection of Table 5.11 reveals a dramatic reduction in the pain factor sum of squares when compared with the last analysis. Whereas previously this factor had been significant ($p=0.004$ and $p=0.011$), utilizing the covariate has rendered the pain factor main effect statistically non-significant ($p=0.174$). This attests to the covariate's power to remove initial differences between pain and no-pain subjects. It had earlier been reported that a two-group discriminant analysis was successful in differentiating between pain and no-pain subjects at the onset of the study. This was further confirmed by a one-way ANOVA. That initial difference has been effectively negated allowing for a statistically equitable comparison of the pain and no-pain groups as a result of utilizing the covariate.

As was true in the previous two analyses, the magnet main effect is not significant ($p=0.578$). This is consistent with the belief that there was no basis for any expectation of a systematic effect based on magnetic condition alone. Rather, it is the pain x magnet interaction that is of greatest interest. Here, in the third analysis, this interaction remains not significant ($p=0.609$). Since this interaction is not significant after employing the covariate, one must conclude that this study offers no supporting evidence to suggest that magnetic treatment is effective in the relief of chronic neck and

TABLE 5.11
 REPEATED-MEASURES ANALYSIS OF VARIANCE,
 USING COVARIATE

SOURCE	SS	DF	MS	F	PROB.
PAIN	1821.78	1	1821.78	1.87	0.174
MAGNET	302.59	1	302.59	0.31	0.578
PAIN X MAGNET	256.09	1	256.09	0.26	0.609
R/PM	92307.68	95	971.66	ERROR 1	
TRIALS	6.30	2	3.15	0.01	0.986
PAIN X TRIALS	411.86	2	205.93	0.95	0.388
MAGNET X TRIALS	177.45	2	88.73	0.41	0.664
PN X MAG X TRLS	122.45	2	61.22	0.28	0.754
RT/PM	41538.75	192	216.35	ERROR 2	

shoulder discomfort.

The trials main effect examines whether there were any differences among the trials ignoring any distinctions among the pain and magnet conditions. This forces a collapsing of the design table such that all 100 subjects occur at each trial (see Figure 5.6). The conceptualization reduces down to a one-way ANOVA with three trial conditions. Results show there are only trivial differences for the trial means when all subjects are considered. The means for the three trials conditions are 59.77, 59.95, 59.59 with a grand mean of 59.77. Such miniscule variation among trials produces an extremely small "F" value. All other interactions are also not statistically significant.

This third analysis does not differ from the previous two analyses in failing to support the hypothesis that magnetic treatment is effective in the treatment of neck and shoulder discomfort. This last analysis, however, does serve to demonstrate the power and utility of using a covariate. The application of this technique clearly eliminates the initial differences between pain and no-pain subjects. This beginning difference was statistically documented through the use of a two-group discriminant analysis and a one-way ANOVA. By employing a covariate in the last analysis, the two groups were statistically equated.

Summary and Conclusions

Three types of repeated-measures analyses were used in the attempt to document the effect of magnetic treatment on neck and shoulder discomfort. Since measurements were taken successively from the same subjects, these measures were not independent as is required in the traditional analysis of variance model. This data dependency was the justification for use of the repeated-measures model of ANOVA.

Though results showed some change in the EMG response of the pain-magnet group across trials, none of the three analyses offered statistical evidence to support the experimental hypothesis of pain reduction as a consequence of magnetic treatment. The second analysis dropped the baseline measure on the contention that this first trial had not been subjected to treatment and therefore might only "dilute" any treatment effect. The second analysis did show some decline in the level of significance of the pain factor, but otherwise differed little from the first, all-trials, analysis. This, in part, may be attributable to the lack of variance among the trial means. Dropping the first trial had minimal effect when all trial measurements were virtually identical.

The third repeated-measures analysis used a covariate in the attempt to remove initial differences between pain and no-pain subjects. The last analysis did serve to demonstrate the power and utility of this technique. The application of this methodology clearly eliminated the

statistical difference between pain and no-pain subjects that was present at the start of this study. This beginning distinction was documented through the use of a two-group discriminant analysis and a one-way ANOVA.

Results indicated that there was very little variance among the trial means. This can be seen as evidence for the reliability of the EMG measure. There was further evidence for the dependent measure's reliability through the use of coefficient theta calculated from eigenvalues produced by a principal components analysis. Principal components analysis was also used to establish the validity of the measurement technique.

Of the three designs examined in this evaluation, the repeated-measures analysis utilizing a covariate has the greatest value for the behavioral researcher. This is attributed to the covariate's ability to remove extraneous variance and thereby statistically equate experimental groups. Both the second analysis (dropping the baseline) and the third analysis (covariate) showed a reduction in total sums of squares and degrees of freedom as a consequence of fewer observations compared with the first analysis (all-trials). Moreover, employing the baseline measure as a covariate in the third design (analysis) resulted in a further reduction in the sums of squares for the pain main effect, the pain x magnet interaction, and error 1 (R/PM) when compared with the second design.

For both the second and third designs virtually all the

lost degrees of freedom were dropped from error 2 (RT/PM) while error 1 (R/PM) was unaltered (except for one degree of freedom for the covariate). In both cases the mean square for error 1 was reduced due to fewer sums of squares while retaining the same degrees of freedom. However, the third design showed a further decline in the error 1 mean square proportionate to the degree that the covariate was successful at removing extraneous variance from the design. In this study the error 1 mean square was 1420.33 for the all-trials design, 1252.06 for the design without the first trial, and 971.66 for the covariate design. F tests employing the error 1 mean square in the denominator were effected by this reduction. Note that besides a diminution in sums of squares, some factors, interactions, and error terms were undergoing concurrent reductions in degrees of freedom due to fewer observations, and in the case of the third design, the use of the covariate.

The second analysis, which dropped the baseline from consideration, had the same number of degrees of freedom associated with all factors, interactions, and error terms as the third analysis (except for the one degree of freedom lost due to the covariate). The second and third analyses had exactly the same sums of squares for Error 2, the trials factor, and all its interactions. However, though the second analysis did show a reduction in sums of squares when compared with the first analysis (all-trials) as a result of fewer measurements, the reduction was not as pronounced as

in the covariate analysis.

Use of the second analysis is discouraged since it does not maximize the utility of the first trial. Though identical to the third analysis for some tests (including the trials factor), other tests do not match the gains in precision offered by the last analysis. It is always bothersome to discard any useable information collected during a study. Therefore, if the first trial is not to be used as a covariate it should be incorporated into the design so as to provide more information about the trials factor. For this reason the first, all-trials, analysis is preferred over the second analysis which excluded the baseline measure from consideration.

Recommendations

Though the small degree of variance among the trial means suggested reliable measurement, it is also possible, however, that the current methodology is not valid and therefore was insensitive to changes that may have taken place. Principal components analysis indicated the technique was valid but essentially only by being differentiated with a theoretically dissimilar measure, the F-wave. Further validation of the EMG technique is an issue that should be more fully addressed by bio-medical investigators.

This physiological approach was a first attempt at this type of investigation. As such, the techniques utilized

might require refinement. For future consideration, other factors such as length of time that subjects received treatment may need to become variables in future studies. Perhaps the four week duration was inadequate. The role of the measurement technique itself may need further evaluation. The EMG and F-wave evoked potentials were in response to applied electrical stimulation. What effect this stimulation had on subjects and treatment warrants examination.

By failing to support the experimental hypothesis, this study casts doubt on previous claims asserting the effectiveness of magnetic therapy for pain reduction. Until such time that this and related studies are replicated, no definitive conclusion regarding the merits, if any, of this therapy can be drawn. There is a need to amass a greater body of literature regarding this issue.

In a repeated-measures design a decrease in error sums of squares is accompanied by a reduction in degrees of freedom. There is little in the literature to indicate under what conditions tests of effects either benefit or suffer as a result of this trade-off when compared with a full-factorial design. For future consideration, a Monte Carlo simulation study would be useful for comparing the impact of sample size and sample variance on tests of significance within these two designs.

For educational researchers, the technique of incorporating a covariate into repeated-measures analysis

has shown itself to be both practical and useful. A covariate is commonly used in education studies when dealing with pre-test/post-test scores or in a simple ANOVA model. The technique discussed in this dissertation allows researchers to take advantage of a similar strategy in more complex nested designs. It is advocated that repeated-measures methodology be employed by educators for the analysis of longitudinal, within-subject data derived from repeated classroom measurement or archived historical records.

APPENDIX A

Control Commands for Repeated-Measures Analyses

```
TITLE THREE REPEATED-MEASURES ANALYSES USING SPSS
FILE HANDLE SERCOM NAME='*SINK*'
FILE HANDLE DATA1 NAME='DISS.DATA'
DATA LIST FILE=DATA1/
  PAIN 4 MAGNET 5 THRSLD1 TO THRSLD4 6-25 (2)
VARIABLE LABELS
  PAIN      'PAIN CONDITION'
  MAGNET    'MAGNET CONDITION'
  THRSLD1  '1ST EMG TRIAL'
  THRSLD2  '2ND EMG TRIAL'
  THRSLD3  '3RD EMG TRIAL'
  THRSLD4  '4TH EMG TRIAL'
VALUE LABELS
  PAIN      1 'NO PAIN'          2 'PAIN' /
  MAGNET    1 'NON-MAGNETIC'    2 'MAGNETIC'
COMMENT    THE ALL-TRIALS ANALYSIS FOLLOWS:
MANOVA     THRSLD1 TO THRSLD4 BY PAIN MAGNET (1,2)/
           WSFACTORS=TRIAL(4)/
           WSDSIGN=TRIAL/
           PRINT=OMEANS(VARIABLE(THRSLD1 TO THRSLD4),
           TABLES(PAIN,MAGNET,PAIN BY MAGNET)),
           SIGNIF(AVERF)/
           ANALYSIS(REPEATED)/
           DESIGN/
COMMENT    THE ANALYSIS DROPPING THE BASELINE FOLLOWS:
MANOVA     THRSLD2 TO THRSLD4 BY PAIN MAGNET (1,2)/
           WSFACTORS=TRIAL(3)/
           WSDSIGN=TRIAL/
           PRINT=SIGNIF(AVERF)/
           ANALYSIS(REPEATED)/
           DESIGN/
COMMENT    THE ANALYSIS UTILIZING THE COVARIATE FOLLOWS:
COMPUTE X2=THRSLD1
COMPUTE X3=THRSLD1
MANOVA     THRSLD2 TO THRSLD4 BY PAIN MAGNET (1,2)
           WITH THRSLD1, X2, X3/
           WSFACTORS=TRIAL(3)/
           WSDSIGN=TRIAL/
           PRINT=SIGNIF(AVERF)/
           ANALYSIS(REPEATED)/
           DESIGN/
FINISH
```

APPENDIX B

Rules for Determining the Appropriate Design

1. Use a capital letter to designate each factor (including a factor for replications).
2. Write the letter for the factor that has the smallest number of levels.
3. Write the letter for the factor that has the second smallest number of levels.
4. Compare the second factor with the first to see whether they are crossed or nested. Two factors are crossed if every level of one factor appears with every level of the other factor. A factor is nested within a second factor if each level of the first factor appears at exactly one level of the second. Perhaps more simply, if two factors are not crossed, they are nested. The factor which has more levels is said to be nested within the other. For example: If A represents a factor with two levels, and if A and B are not crossed, then B is said to be nested within A and is written B/A. (A slash is always read, "nested within.") Note that there are many different notations used for nesting, depending upon the text or computer program used. For example if C is nested in S one of the following notations is used: S, C nested within S; S, CwS; S, C:S, S, C(S); or S, C/S. In this section the slash notation is used.
5. Write the letter for the factor that has the third smallest number of levels.
6. Compare this factor with all previous factors that have been written. If D was a factor that was nested within both M and R, then write D/MR or D/RM. It should be noted that nesting is not a transitive relationship. That is, if C is nested in B and B is nested in A, then it does not follow that C is nested in A. The ordering of the letters on one side of a slash is irrelevant.
7. Continue until all factors have been written including a factor for replications.
8. After all main-effect factors have been written, the interaction effects are written. These are obtained by multiplying terms (consisting of both factors and interaction terms that have just been determined) together, pairwise, in an algebraic manner. Letters appearing on the left of the slash in a factor appear on the left of the slash in the interaction. If there is no slash the letters

are assumed to be on the left. Letters appearing on the right of the slash in a factor appear on the right of the slash in the interaction. If a letter appears twice in a term, the term is deleted. For example: the main-effect factors might be A, B, R/B. The following could then be written: $A \times B = AB$; $A \times R/B = AR/B$; $B \times R/B = BR/B$ (this is deleted); $B \times AB = BAB$ (deleted); $B \times AR/B = BAR/B$ (deleted). Continuing these tries, terms like $AB \times AR/B = ABAR/B$ (deleted) could be written but the following would finally be accepted as being the complete design: A, B, R/B, AB, AR/B. With a little practice this process becomes intuitive and bad choices are avoided. In large designs it is easy to overlook an interaction term. The error is discovered later, when the checks in Appendix C...are performed.

(Reprinted with author's permission from W. L. Landrum, 1972)

APPENDIX C

Rules for Degrees of Freedom

1. Each factor has been designated with a capital letter. The count of the factor is now defined and is designated with the corresponding lower case letter. For example, the factor A has the count a.

2. If a factor is not nested within any other factor, its count equals its number of levels.

3. If a factor is nested within one or more factors, its count equals the number of levels that it has at one level of each of the factors. For example, for a so-called two factor (2x3) completely crossed design with five replications per cell and 30 subjects, the design would be A, B, AB, R/AB (where R refers to replications) with $a=2$, $b=3$, and $r=5$. The count for R is five since five subjects appear both in level one of A and level one of B.

4. To designate the degrees of freedom for a term, the capital letter designating the factor is replaced with its count if the letter is on the right of the slash, or with the count minus one if the letter is on the left of the slash. For example, the degrees of freedom for the term ABP/CD could be equal to $(a-1)(b-1)(p-1)(c)(d)$.

5. For balanced designs the total number of observations is always equal to the product of the counts and the sum of the degrees of freedom is always equal to the product of the counts minus one. For example, any balanced design that used just the factors A, B, C, and R would have the total number of observations equal to $abcr$ and the sum of the degrees of freedom would be $abcr-1$. These are good checks to detect careless errors in formulating the design.

(Reprinted with author's permission from W. L. Landrum, 1972)

APPENDIX D

Rules for Determining the Expected Mean Squares (EMS)

1. A variance component is listed for each source of variance designating that source. For example: For the source of variance R/AB one would list $\text{VAR}(R/AB)$.

2. If all the letters in one source of variance are included in the letters for a second source of variance, the variance component of the second is added to that of the first. For example: Given the design: A, B, AB, R/AB, we would list $\text{VAR}(A)$ as a variance component for A. To that one would add $\text{VAR}(AB)$ and $\text{VAR}(R/AB)$ so that we now have $\text{EMS}(A) = \text{VAR}(A) + \text{VAR}(AB) + \text{VAR}(R/AB)$.

3. If a letter representing a factor does not appear as a subscript in a variance component term then its corresponding count appears as a coefficient of the term. In the previous example we would have: $\text{MS}(A) = br\text{VAR}(A) + r\text{VAR}(AB) + \text{VAR}(R/AB)$.

4. The preceding rules will generate the expected mean squares for a fully random model. If some factors are fixed, the following rules should be used to delete some of the variance component terms.

5. All letters appearing to the left of the slash in the subscript of a variance component are called "essential." For example in the term $akc\text{VAR}(MQP/RD)$; M, Q, and P are essential.

6. Examine the subscripts of a variance component term. After considering only the essential letters and disregarding any letters necessary to denote the source of variance being examined, if any of the remaining letters represent fixed factors, delete the term just examined. For example: Suppose the $\text{EMS}(A)$ has a term $cw\text{VAR}(ABR/K)$. We start with ABR/K, keep only the essential letters ABR, disregard the A as it is the source of variance being examined, and finally examine B and R. Suppose B is fixed. We then delete the term $cw\text{VAR}(ABR/K)$ from the expression for $\text{EMS}(A)$.

See Appendix E for an example.

(Reprinted with author's permission from W. L. Landrum, 1972)

APPENDIX E

Rules for Determining Error Terms

1. To calculate the error term for a source of variance, select a second source of variance whose expected mean square is identical to the expected mean square of the first source of variance when the term designating the first source of variance is ignored. For example: If the expected mean square for $A = brVAR(A) + rVAR(AB) + VAR(R/AB)$, then we would choose a source of variance whose expected mean square was $rVAR(AB) + VAR(R/AB)$. Probably the expected mean square of $AB = rVAR(AB) + VAR(R/AB)$.

2. It is never possible to test all the sources of variation. Usually the terms involving subjects (or replications) are not testable and are referred to as error by most statistical texts. If an important source of variance is not testable, it is frequently possible to perform an approximate test using techniques discussed in many of the standard statistical texts.

3. A full example of what a design might look like is given below. See if you can figure out the terms on your own and then compare.

<u>Source</u>	<u>EMS[fully random]</u>	<u>Error Term</u>
A	$VAR(R/AB) + rVAR(AB) + brVAR(A)$	EMS(AB)
B	$VAR(R/AB) + rVAR(AB) + arVAR(B)$	EMS(AB)
AB	$VAR(R/AB) + rVAR(AB)$	EMS(R/AB)
R/AB	$VAR(R/AB)$	NO TEST

<u>Source</u>	<u>EMS[assuming A fixed]</u>	<u>Error Term</u>
A	$VAR(R/AB) + rVAR(AB) + brVAR(A)$	EMS(AB)
B	$VAR(R/AB) + arVAR(B)$	EMS(R/AB)
AB	$VAR(R/AB) + rVAR(AB)$	EMS(R/AB)
R/AB	$VAR(R/AB)$	NO TEST

(Reprinted with author's permission from W. L. Landrum, 1972)

APPENDIX F

Repeated-Measures Analysis by
Method of First Differences

SOURCE	SS	DF	MS	F	PROB.
PAIN	91.49	1	91.49	0.37	0.543
MAGNET	217.65	1	217.65	0.89	0.349
PAIN X MAGNET	3.97	1	3.97	0.02	0.899
R/PM	23577.28	96	245.60	ERROR 1	
TRIALS	18.92	2	9.46	0.02	0.983
TRIALS X PAIN	732.36	2	366.18	0.66	0.519
TRIALS X MAGNET	22.00	2	11.00	0.02	0.980
TRLS X PN X MAG	336.47	2	168.24	0.30	0.739
RT/PM	106720.32	192	555.84	ERROR 2	

BIBLIOGRAPHY

- Aceto, H., Tobias, C. A. and Silver, I. L. Some studies on the biological effects of magnetic fields. IEEE Transactions on Magnetics, 1970, 6, 368-373.
- Alexander, H. W. A general test for trend. Psychological Bulletin, 1946, 43, 533-557.
- Alexander, H. W. The estimation of reliability when several trials are available. Psychometrika, 1947, 12, 79-99.
- Babington-Smith, B. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. British Journal of Psychology, 1950, 40, 250-268.
- Baldwin, L., Medley, D. M., & MacDougall, M. A. A comparison of analysis of covariance to within-class regression in the analysis of non-equivalent groups. Journal of Experimental Education, 1984, 52(2), 68-76.
- Barlow, H., Kohr, H. I. and Walsh, C. A. Visual sensation aroused by magnetic fields. American Journal of Physiology, 1947, 148, 372-375.
- Barnothy, M. F. (Ed). Biological effects of magnetic fields, Vol. 1 and 2, New York: 1964, 1969.
- Baxter, B. The application of factorial design to a psychological problem. Psychological Review, 1940, 47, 494-500.
- Baxter, B. Problems in the planning of psychological experiments. American Journal of Psychology, 1941, 54, 270-280.
- Baxter, B. A study of reaction time using factorial design. Journal of Experimental Psychology, 1942, 31, 430-437.
- Becker, R. O. Search for evidence of axial current flow in peripheral nerves of salamanders. Science, 1961, 134, 101-102.
- Becker, R. O. Geomagnetic environment and its relationship to human biology. New York State Journal of Medicine, 1963, 2215-2219.
- Beischer, D. E. and Knepton, J. C. The electroencephalogram of the squirrel monkey in a very high magnetic field, NASA Report NAMI-972. Aero Space Medicine, 1966, 37,

287.

Bevan, W. The contextual basis of behavior. American Psychologist, 1968, 23, 701-714.

Bingham, C. & Fienberg, S. E. Textbook analysis of covariance - Is it correct? Biometrics, 1982, 38, 747-753.

Blakemore, R. P. Magnetotactic bacteria. Science, 1975, 190, 377-379.

Bloomers, P. and Lindquist, E. F. Experimental and statistical studies: Application of newer statistical techniques. Review of Educational Research, 1942, 12, 501-520.

Bower, G. H. A contrast effect in differential conditioning. Journal of Experimental Psychology, 1961, 62, 196-199.

Brozek, J. and Alexander, H. A. A note on estimation of the components of variation in a two-way table. American Journal of Psychology, 1947, 60, 629-637.

Buchtal, F. and Rosenflack, A. Evoked action potentials and nerve conduction velocity in human sensory nerves. Brain Research, 1966, 3, 1-22.

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

Cochran, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics, 1947, 3, 22-28.

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

Cochran, W. G. and Cox, G. M. Experimental designs. New York: Wiley, 1950.

Coombs, C. H. The role of correlation in analysis of variance. Psychometrika, 1948, 13, 233-243.

Cox, D. R. & McCullagh, P. Some aspects of analysis of covariance. Biometrics, 1982, 38, 541-561.

Cronbach, L. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Crutchfield, R. S. Efficient factorial design and analysis of variance illustrated in psychological experimentation.

Journal of Psychology, 1938, 5, 339-346.

Crutchfield, R. S., and Tolman, E. C. Multiple-variable design for experiments involving interaction of behavior. Psychological Review, 1940, 47, 38-42.

Davenport, C. B. and Ekas, M. P. Statistical methods in biology, medicine, and psychology. New York: Wiley, 1936.

Davidson, M. L. Univariate versus multivariate tests in repeated-measures experiments. Psychological Bulletin, 1972, 77, 446-452.

Dayton, C. M. Design of educational experiments. New York: McGraw-Hill, 1970.

Dubrov, A. P. Geomagnetic field and life, 1974, Gidrometeoizdat, Leningrad (American Edition, Plenum, New York, 1978).

Dunlap, J. W. Applications of analysis of variance to educational problems. Journal of Educational Research, 1940, 33, 434-442.

Dunlap, J. W. Recent advances in statistical theory and applications. American Journal of Psychology, 1941, 54, 583-601.

Eaton, C. M. and Lambert, E. H. Electromyography electrical stimulation of nerves in diseases of motor unit. Journal of the American Medical Association, 1957, 163, 1117-1124.

Edwards, A. L. Statistical analysis for students in Psychology and Education. New York: Rinehart, 1946.

Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1950.

Edwards, A. L. On the use of interactions as "error terms" in the analysis of variance. Educational and Psychological Measurement, 1950, 10, 214-223.

Edwards, A. L., and Horst, P. The calculation of sums of squares for interactions in the analysis of variance. Psychometrika, 1950, 14, 17-24.

Eisenhart, C. The assumptions underlying the analysis of variance. Biometrics, 1947, 3, 1-21.

Eisenhart, C., Hastay, M. W., and Wallis, W. A. Selected techniques of statistical analysis. New York: McGraw-Hill, 1947.

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

Engelhart, M. D. The analysis of variance and covariance techniques in relation to the conventional formulas for the standard error of difference. Psychometrika, 1941, 6, 221-234.

Erdman, G. M. The problem of the effect of a constant magnetic field on nerve. Akademiya Nauk S.S.S.R. Institut Biologicheske Fizika, 1955, 1, 35-39.

Erlebacher, A. Design and analysis of experiments contrasting the within-and between-subjects manipulation of the independent variable. Psychological Bulletin, 1977, 84, 212-219.

Federer, W. T., and Balaam, L. N. Bibliography on experimental and treatment design--pre 1968. New York: Hafner, 1973.

Feldt, L. S. A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika, 1958, 23, 335-353.

Ferguson, G. A. Statistical analysis in Psychology and Education. New York: McGraw-Hill, 1971.

Fertig, J. W. The use of interaction in the removal of correlated variation. Biometric Bulletin, 1936, 1, 1-14.

Fisher, R. A. On the distribution of the standard deviation of small samples: Appendix 1. to papers by "Student" and R. A. Fisher. Biometrika, 1915, 10, 522-529.

Fisher, R. A. Applications of "Student's" distribution. Metron, 1925, 5, 90-104.

Fisher, R. A. Statistical methods for research workers. London: Oliver and Boyd, 1925.

Fisher, R. A. The design of experiments. London: Oliver and Boyd, 1935.

Friedman, H., Becker, R. O., and Bachman, C. H. Effect of magnetic fields on reaction-time performance. Nature, 1967, 213, 949-956.

Garrett, H. E. Variability in learning under massed and spaced practice. Journal of Experimental Psychology, 1940, 26, 547-567.

Garrett, H. E. ANOVA in psychological research. Journal of Educational Research, 1943, 36, 631-632.

Garrett, H. E. Statistics in Psychology and Education. New York: Longmans, Green, 1947.

Garrett, H. E. and Zubin, J. The analysis of variance in psychological research. Psychological Bulletin, 1943, 40, 233-267.

Gentile, J. R. Significance of single-subject studies (and repeated-measures designs). Educational Psychologist, 1982, 17, No. 1, 54-60.

Gentile, J. R., Roden, A. H., & Klein, R. D. An analysis-of- variance model for the intrasubject replication design. Journal of Applied Behavior Analysis. 1972, 5, 193-198.

Gibson, E. J. Principles of perceptual learning and development. New York: Appleton-Century-Crofts, 1967.

Gilliatt, R. W. and Sears, T. A. Sensory nerve action potentials in patients with peripheral nerve lesions. Journal of Neurology and Neuro-surgical Psychiatry, 1958, 21, 109-118.

Goodgold, J. and Eberstein, A. Electrodiagnosis of neuromuscular diseases, Baltimore: William and Wilkins, 1972.

Gossett, W. S. (pseudonym, Student). The probable error of a mean. Biometrika, 1908, 6, 1-25.

Goulden, C. H. Methods of statistical analysis. Minneapolis, MN: Burgess, 1939.

Greenwald, A. G. Within-subject designs: To use or not to use. Psychological Bulletin. 1976, 83, 314-320.

Grice, G. R. Dependence of empirical laws upon the source of experimental variation. Psychological Bulletin, 1966, 66, 488-498.

Grice, R., & Hunter, J. J. Stimulus intensity effects depend upon the type of experimental design. Psychological Review, 1964, 71, 247-256.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill, 1950.

Hackman, R. B. An experimental study of variability in ocular latency. Journal of Experimental Psychology, 1940, 27, 546-558.

Hemmerle, W. J. Analysis of covariance algorithms. Biometrics, 1982, 38, 725-736.

Hotelling, H. The generalization of "Student's," ratio. Annals of Mathematical Statistics, 1931, 2, 360-378.

Hotelling, H. Review of Snedecor, G. W.: Calculation and interpretation of the analysis of variance and covariance. Journal of the American Statistical Association, 1935, 30, 118.

Hotelling, H. Dr. Peters' criticism of Fisher's statistics. Journal of Educational Research, 1943, 36, 707-711.

Hotelling, H. The impact of R. A. Fisher on statistics. Journal of the American Statistical Association, 1951, 46, 35-46.

Hulse, S. H., Egeth, H., and Deese, J. The Psychology of Learning. New York: McGraw-Hill, 1980.

Humphreys, L. G. The strength of a Thorndikian response as a function of the number of practice trials. Journal of Comparative Psychology, 1943, 35, 101-110.

Isaac, S. and Michael, W. B. Handbook in research and evaluation. New York: Robert A. Knapp, 1971.

Jackson, R. W. B. Application of the analysis of variance and covariance methods to educational problems. Department of Educational Research, University of Toronto Bulletin, 1940, 11, 67-74.

Jebsen, R. H. Motor conduction velocities in median and ulnar nerves. Archives of Physical Medicine and Rehabilitation, 1967, 48, 185-194.

Jellinek, E. M. Measurements of the consistency of fasting oxygen consumption rates in schizophrenic patients and normal controls. Biometric Bulletin, 1936, 1, 15-43.

Johnson, P. O., & Tsao, F. Factorial design and covariance in the study of individual educational development. Psychometrika, 1945, 10, 133-162.

Kalmijn, A. J. The electric and magnetic sense of sharks, skates and rays. Oceanus, 1977, 20, 45-52.

Kalmijn, A. J. Experimental evidence of geomagnetic orientation in dasmobranch fishes. Proceedings in Life Sciences, Schmidt-Konig, Ed., New York: Springer-Verlag, 1978, 384-354.

Kalmijn, A. J. and Blakemore, R. P. Geomagnetic orientation in marine mud bacteria. Proceedings of the International Union of Physiological Sciences, 1977, 13, 364.

Karabinus, R. A. The use of ancova, multiple regression, repeated anova, and effect size. Evaluation Review, 1983, 7, 841-850.

Keeton, W. T. Orientation by pigeons: is the sun necessary? Science, 1969, 165, 922-928.

Keeton, W. T. Magnetic interferences with pigeon homing. Proceedings of the National Academy of Sciences, USA, 1971, 68, 102-106.

Kerlinger, F. N. Foundations of behavioral research, (2nd ed.), New York: Holt, Rinehart, and Winston, 1973.

Kerlinger, F. N. & Pedhazur, E.J. Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston, 1973.

Kholodov, Yu. A. Effects on the central nervous system, in Biological Effects of Magnetic Fields, Barnothy, Ed., New York: Plenum, 1964, 196-200.

Kholodov, Yu. A. The effect of electromagnetic and magnetic fields on the central nervous system, Izadetelstvo, Moscow, 1966. (NASA Technical Translation, TTF-465, 1967. Available from NTIA, Springfield, Virginia, 22151).

Kholodov, Yu. A. Influence of magnetic fields on biological subjects, Moscow, 1971, (Available from National Technological Information Service, Springfield, VA, JPRS 63038, 1974).

Kholodov, Yu. A., Alexandrovskay, M. M., Lukjanova, S. N. and Ndarova, N. S. Investigation of the reactions of the mammalian brain to static magnetic fields, in Biological effects of magnetic fields, Barothy, M. F., Ed., Plenum, New York: 1969, 215-225.

Kimura, J. and Butzer, J. F. F-wave conduction in Guillain-Barre syndrome. Archves of Neurology, 1975, 32, 524-529.

Knapp, T. R. A case against the single-sample repeated-measures experiment. Educational Psychologist, 1982, 17, 61-65.

Koch, G. G., Amara, I. A., Davis, G. W., & Gillings, D. B. A review of some statistical methods for covariance analysis

of categorical data. Biometrics, 1982, 38, 563-595.

Kogan, L. S. Analysis of variance--Repeated measurements. Psychological Bulletin, 1948, 45, 131-143.

Kogan, L. S. Variance designs in psychological research. Psychological Bulletin, 1953, 50, 1-40.

Landrum, W. L. The efficacy of teaching analysis of variance by rules of thumb (doctoral dissertation, University of Georgia), Ann Arbor, MI: University Microfilms, 1972. Number 34-105.

Lane, P. W. & Nelder, J. A. Analysis of covariance and standardization as instances of prediction. Biometrics, 1982, 38, 613-621.

Liberman, E. A., Vaintsvaig, M. N. and Tsofina, L. M. The effect of constant magnetic field on the excitation threshold of isolated frog nerve. Biofizika, 1959, 4, 505-506.

Lindauer, M. and Martin, H. The earth magnetic field affects the orientation of honeybees in the gravity field. Zeitschrift fuer Vergleichende Physiologie, 1968, 60, 219-243.

Lindquist, E. F. Statistical analysis in educational research. Boston: Houghton-Mifflin, 1940.

Lindquist, E. F. Goodness of fit and trend curves and significance of trend differences. Psychometrika, 1947, 12, 65-78.

Lipkin, S. G., Schnorr, J. A., Suydam, M. M., & Meyers, J.L. Effects of risk as a with-in subject variable in probability learning. Psychonomic Science, 1965, 2, 145-146.

Lovie, A. D. The analysis of variance in experimental psychology: 1934-1945. British Journal of Mathematical and Statistical Psychology, 1979, 32, 151-178.

Lovie, A. D. On the early history of ANOVA in the analysis of repeated measure designs in psychology. British Journal of Mathematical and Statistical Psychology, 1981, 34, 1-15.

Martin, H. and Lindauer, M. The effect of earth magnetic field on gravity orientation in the honeybee. Journal of Comparative Physiology, 1977, 122, 145-187.

Melton, A. W. The methodology of experimental studies of human learning and retention: I. The functions of a

methodology and the available criteria for evaluating different experimental methods. Psychological Bulletin, 1936, 33, 305-394.

Michael, J. Statistical inference for individual organism research: Mixed blessing or curse? Journal of Applied Behavior Analysis, 1974, 7, 647-653.

Millman, J. and Glass, G. V. Rules of thumb for writing the ANOVA table. Journal of Educational Measurement, 1967, 4, 41-51.

Moore, F. R. Geomagnetic disturbance and the orientation of nocturnally migrating birds. Science, 1977, 196, 682-684.

Morrison, D. F. Multivariate statistical methods, 2nd ed. New York: McGraw-Hill, 1976.

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

Nakagawa, K. Clinical application of magnetic field, Japanese, J. Society of Non-traditional Technology, 1974, 66, 6-17.

Nakagawa, K. A study on clinical effects of the magnetic necklace, Isuzu Hospital, Tokyo, Japan, 1975.

Nakagawa, K. Magnetic field deficient syndrome and magnetic treatment. Japan Medical Journal, 1976, 2745, 24-25.

Nambu, S., Tomari, Y. Kawahara, Y. and Narazake, M. Effect of magnetic band on "so-called stiff shoulders", in Magnetism and Living Bodies (in Japanese), Nakagawa, K., Ed., Nikkan Kogyo Shumbun Sha, Japan, 1973, 125-131.

Oster, G. Phosphenes. Scientific American, 1970, 222, 82-87.

Pavlov, I. P. Conditioned reflexes (G.V. Anrep. Trans.). London: Oxford, 1927.

Peters, C. C. Misuses of the Fisher statistics. Journal of Educational Research, 1943, 36, 546-549.

Peters, C. C. Interaction in analysis of variance interpreted as intercorrelation. Psychological Bulletin, 1944, 41, 287-299.

Piaget, J. The construction of reality in the child. New York: Basic Books, 1954.

Piaget, J. The origins of intelligence in children. New York: Norton, 1963.

Piaget, J. Six psychological studies. New York: Random House, 1967.

Poor, D. D. S. Analysis of variance for repeated measures designs: Two approaches. Psychological Bulletin, 1973, 80, 204-209.

Poulton, E. C. Unwanted range effects from using within-subjects experimental designs. Psychological Bulletin, 1973, 80, 113-121.

Poulton, E. C. Range effects are characteristics of a person serving in a within-subjects experimental design--A reply to Rothstein. Psychological Bulletin, 1974, 81, 201-203.

Poulton, E. C. Range effects in experiments on people. American Journal of Psychology, 1975, 88, 3-32.

Pubols, B. H., Jr. Incentive magnitude, learning, and performance in animals, Psychological Bulletin, 1960, 57, 89-115.

Redford, J. B. Nerve conduction in motor fibers to the anterior tibial muscle in peroneal palsy. Archives of Physical Medicine and Rehabilitation, 1964, 45, 500-504.

Reno, V. R. Conduction velocity in nerve exposed to a high magnetic field, NASA Report NAMI-1089, Pensacola, Florida, 1969.

Robinson, E. S. and Bills, A. G. Two factors in the work decrement. Journal of Experimental Psychology, 1926, 9, 415-443.

Rothstein, L. D. Reply to Poulton. Psychological Bulletin, 1974, 81, 199-201.

Rucci, A. J. & Tweney, R. D. Analysis of variance and the and the 'second discipline' of scientific psychology: A historical account. Psychological Bulletin, 1980, 87, 166-184.

Schwartz, J. L. Influence of a constant magnetic field on nervous tissues: 1. Nerve conduction velocity studies. IEEE Transactions on Biomedical Engineering, 1978, 25, 467-471.

Searle, S. R. & Hudson, G. F. S. Some distinctive features of output from statistical computing packages for analysis

of covariance. Biometrics, 1982, 38, 737-745.

Shen, E. Experimental design and statistical treatment in educational research. Journal of Experimental Education, 1940, 8, 346-353.

Sidman, M. Tactics of scientific research. New York: Basic Books, 1960.

Silver, I. L. and Tobias, C. A. Magnetic fields and their biological effects, in Space Biology and Related Topics, Tobias, C.A. and Todd, P., Eds., New York: Academic Press, 1974, 257-292.

Smith, H. F. Interpretation of adjusted treatment means and regressions in analysis of covariance. Biometrics, 1957, 13, 282-308.

Snedecor, G. W. Calculation and interpretation of the analysis of variance and covariance. Ames, Iowa: Collegiate Press, 1934.

Snedecor, G. W. Statistical methods. Ames, Iowa: Collegiate Press, 1937.

SPSS user's guide. Version 10 (2nd. ed.) New York: McGraw-Hill, 1986.

Stanley, J. C. The influence of Fisher's "The design of experiments" on educational research thirty years later. American Educational Research Journal, 1966, 3, 223-229.

Stanley, J. C., Editor Improving experimental design and statistical analysis.4, Seventh Annual Phi Delta Kappa Symposium on Educational Research. Chicago: Rand McNally and Company, 1967.

TDK Magnetics Corporation technical communication no. 223, Magnetic fields and applications, 1978, Beverly Hills, CA.

Telford, C. W. The refractory phase of voluntary and associative responses. Journal of Experimental Psychology, 1931, 14, 1-36.

Tenforde, T. S., Ed., Proceedings of the biomagnetic effects workshop, Lawrence Berkeley Laboratory, University of California, Berkley, 1978.

Thompson, G. H. The use of the Latin square in designing educational experiments. British Journal of Educational Psychology, 1941, 11, 135-137.

Thurstone, L. L. Multiple factor analysis, Chicago:

University of Chicago Press, 1947.

Tukey, J. W. Components in regression. Biometrics, 1951, 7, 33-69.

Urschel, H. C., Razzuk, M. A., Wood, R. E., Parekh, M. and Paulsen, D. L. Objective diagnosis (ulnar nerve conduction velocity) and current therapy of the thoracic outlet syndrome. Annals of Thoracic Surgery, 1971, 12, 608-620.

Vovk, M. I. and Tkach, Y. K. Influence of a permanent magnetic field on the fluctuations in the threshold of stimulation of isolated skeletal muscle. Biofizika, 1971, 16, 833-836.

Vyalov, A. M. Clinico-hygienic and experimental data on the effects of magnetic fields under industrial conditions, in Influence of Magnetic Fields in Biological Objects, Kholodov, Yu. A. Ed., Moscow, 1971, 163-174 (JPRS 63038, 1974).

Walcott, C., Gould, J. L. and Krischvink, J. L. Pigeons have magnets. Science, 1979, 205, 1027-1029.

Walker, H. M. Degrees of freedom. Journal of Educational Psychology, 1940, 31, 253-269.

Walker, H. M. Certain unsolved statistical problems of importance to psychological research. Harvard Educational Review, 1947, 17, 297-304.

Wilks, S. S. Certain generalizations in ANOVA. Biometrika, 1932, 24, 471-474.

Wilsdon, B. H. Discrimination by specification statistically considered and illustrated by the standard specification for Portland cement (with discussion). Journal of the Royal Statistical Society, Supplement, 1934, 1, 152-206.

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

Wishart, J. Statistics in agricultural research (with discussion). Journal of the Royal Statistical Society, Supplement, 1934, 1, 26-61.

Wu, Y. and Stratigos, J. S. Axillary F-loop latency in proximal neuropathy and myelopathy. Archives of Physical Medicine and Rehabilitation, 1976, 57, 569.

Wynn-Parry, C. B. Technique of neuromuscular stimulation and their clinical application, in Disorders of Voluntary

Muscles, Walton, J. N., Ed., London: J & A Churchill, 1969,
763-784.

Zeller, R. A. and Carmines, E. G. Measurement in the
social sciences, Cambridge: Cambridge University Press,
1980.

ABSTRACT

THE EFFICACY OF WITHIN-SUBJECT MEASUREMENT

by

RICHARD J. MELTZER

December, 1986

Adviser: Dr. Donald Marcotte

Major: Evaluation and Research

Degree: Doctor of Philosophy

50 subjects suffering from chronic neck and shoulder pain were evenly divided into two groups. One group was exposed to a magnetic treatment while the other group received a placebo treatment. 50 control subjects (no-pain) were similarly divided and exposed to treatment. Duration of treatment was three weeks.

At the onset of the study a baseline physiological response (EMG threshold response) was recorded from all subjects. This was followed by repeated measurements of this response at weekly intervals over the course of treatment. This objective physiological response, which served as the dependent variable in this study, was believed to be related to the subjective experience of pain.

A repeated-measures design is appropriate when using

correlated "within-subject" data. Often, however, the initial baseline measure is included in the analysis in spite of being collected prior to subjects' exposure to treatment. The relationship between repeated-measures analysis and the baseline measure was of substantive interest to this investigation. Consequently, three different repeated-measures analyses were compared: (1) repeated-measures including the baseline as the first repeated measure; (2) repeated-measures ignoring the baseline completely; and (3) repeated-measures using the baseline measure as a covariate.

Results of the first two analyses showed little difference. Each failed to support the hypothesis of pain reduction as a consequence of magnetic treatment. In both cases, however, the pain factor was significant thereby reflecting initial differences between pain and no-pain subjects at the start of the study. The final analysis utilized the baseline measure as a covariate in order to statistically equate all subjects. Though the third analysis also failed to support the experimental hypothesis, it did eliminate the initial statistical difference between pain and no-pain subjects.

AUTOBIOGRAPHICAL STATEMENT

Richard J. Meltzer was born and raised in Detroit, Michigan. After graduating from Mumford High School he attended Wayne State University where he earned a Bachelor of Arts degree majoring in Psychology. In both his Junior and Senior years the author was awarded a Board of Governor's scholarship. Also in his Senior year he was a recipient of an Academic Achievement award from the Psychology Department.

The author earned a Master of Science degree from Eastern Michigan University majoring in Psychology. At Wayne State University he has worked as a bio-medical researcher for the Colleges of Medicine and Engineering and as the staff research and computer consultant for the School of Business Administration.

Presently the author is pursuing the degree of Doctor of Philosophy in Evaluation and Research at Wayne State University. The author was the recipient of a Graduate Professional Scholarship while taking coursework for this degree. He is currently employed in industry as a statistical and research programmer/analyst. He is a member of the Association for Computing Machinery, the American Statistical Association, and the Metro Detroit SAS Users Group.