8701336

Tritschler, Kathleen Ann

USE OF STATISTICS IN RECENTLY-PUBLISHED PHYSICAL EDUCATION RESEARCH

The University of North Carolina at Greensboro Ed.D. 1985

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

USE OF STATISTICS IN RECENTLY-PUBLISHED PHYSICAL EDUCATION RESEARCH

by

Kathleen Ann Tritschler

A Dissertation Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Greensboro
1985

Approved by

Pearl Berlin

Dissertation Adviser

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of the Graduate School at The University of North Carolina at Greensboro.

Dissertation Adviser Pearl Berlin

Committee Members Richard M. Jaeger

Rosemary McGee

Kate R. Barrett

Richard H. Weller

November 13, 1985

Date of Acceptance by Committee

August 23, 1985

Date of Final Oral Examination

TRITSCHLER, KATHLEEN A., Ed.D. Use of Statistics in Recently-Published Physical Education Research. (1985) Directed by Dr. Pearl Berlin. 185 pp.

The types and frequencies of statistical techniques reported in recently-published physical education research were studied. Also investigated were: (a) complexity of the data that were analyzed, (b) frequency and levels of significance testing, assumption testing, and data transformation, and (c) characteristics of the reporting of statistical analyses. Stratified random sampling with proportional allocation from seven physical education research journals was used to identify a sample of 233 quantitative research reports.

Content analyses revealed that a wide variety of statistical techniques were employed in the sample of reports. Descriptive statistics were reported most frequently; a majority of the reports did, however, employ at least one inferential analysis. A *p* value of .05 was the most commonly reported alpha level for significance testing, although most studies failed to state a criterion alpha level. Among the inferential studies, 98% reported statistical "significance" of their findings.

The research investigations were found to be complex in terms of the number of variables studied, but less complex when one considers the number of variables simultaneously analyzed and the sample sizes employed. Multivariate analyses

were employed in 25.8% of the research reports; it was suggested, however, that multivariate and repeated measures analyses should have been used more frequently than they were. Researcher writers generally did not provide readers with "help" in understanding statistics. The data analysis revealed limited observations of justifications for selection of a particular statistical technique or citations of statistical references. Very seldom did writers identify the data analysis program that was utilized.

The types and frequencies of statistical techniques employed were analyzed according to the subspecialty focus of the research. Multivariate techniques were reported most frequently in Measurement & Evaluation research and least frequently in investigations of the Functional Effects of Physical Activity. Management Theory & Practice researchers employed nonparametric techniques more frequently than did other subspecialty investigators; no nonparametric analyses were observed in research classified as Functional Effects of Physical Activity and/or Mechanical & Muscular Analysis of Motor Skills.

Findings were discussed in relation to pertinent sources cited in the review of literature. Additionally, suggestions were made for improving the quality of published physical education research and for academic preparation in statistics.

ACKNOWLEDGMENTS

The author extends a very big "THANK YOU" to all of the persons who financially and emotionally supported her during the production of this dissertation.

Dr. Pearl Berlin is credited with far more than mere guidance of the dissertation. She has taught the author the techniques and values of scholarly inquiry, and has been a model of professionalism in physical education. She has taught me things I have not yet figured out how to express in words. Thank you!

Special thanks also goes to Dr. Richard Jaeger, without whom I would not have had the skills and knowledge to conduct a survey-style investigation. Jacqueline Gillis provided me with expert advice on use of the Statistical Analysis System for computer analysis of my data. Thanks also to Dr. Kate Barrett, Dr. Rosemary McGee, and Dr. Richard Weller for their critical commentaries at several stages of the dissertation process.

Final thanks are extended to the alumni, administrators, faculty, staff, and student body of The University of North Carolina at Greensboro for creating such a supportive environment within which to pursue doctoral study.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

CHAPTER I

NATURE AND SCOPE OF THE PROBLEM

Over two decades ago, in what is now considered to be a landmark treatise, Franklin M. Henry challenged physical educators to look critically at their field to determine if it was a true academic discipline (Henry, 1964). In doing so, Henry paved the way for serious introspection into the knowledge base of physical education. Vanderzwaag's (1973) introductory remarks to a Quest philosophic position paper reminded readers of the seemingly obvious, "Before something can be studied there must be something to study"! He explained that research, therefore, commands a key role in any academic enterprise. "Without research, the body of knowledge will remain a speculative construct and not become a reality" (p. 78). If, however, knowledge deriving from research is to be valuable to academicians within a discipline and to professionals who seek to apply such knowledge, the research must be read and understood. Kroll (1982) suggested that it is not at all clear how much basic knowledge and how much research competence is required to be an adequate consumer of research, but that it is very clear that "one must know something of research techniques. . ." (p. 19).

Physical education research belongs to the larger family of behavioral research. Kerlinger (1979) claimed that it is virtually impossible to conceive of modern behavioral research without statistical understanding (p. 308). Indeed, both sophisticated and unsophisticated readers of physical education research are aware that a vast majority of published research is quantitative and statistical in nature (Berlin, 1973; Clarke & Clarke, 1984; Teraslinna, 1967). Thus, it seems reasonable to propose that statistical knowledge is an important competence for the consumers of physical education research.

The Delegate Assembly of NASPE, the National Association for Sport and Physical Education, apparently concurs with the above proposition. In the "NASPE Accreditation and Interpretation of Standards for the Master's Degree Program in Physical Education" (1984), coursework in "analytical method" was deemed essential to a quality master's degree program. Left unspecified by NASPE, however, was the exact nature of the analytical knowledge.

In 1968, Brady determined that the most popular statistical techniques reported in physical education doctoral dissertations were the mean, standard deviation, Pearson's product moment correlation, and analysis of variance. Burkhardt (1969) found among researchers publishing in the Research Quarterly that 42% used techniques of hypothesis testing about means, variances, or proportions, 24% invoked correlation and/or regression analyses, 15% used reliability and/or validity techniques, 13% treated data according to descriptive statistical methods, 5% used nonparametric statistics, and a mere 1% factor analyzed their data. A decade later, Kerlinger (1979) posited that knowledge of advanced multivariate statistics was mandatory to appreciate research in psychology, sociology, and education. Kerlinger's comment suggests the following questions. What statistical knowledge is needed to read the body of physical education research produced today? What statistical techniques are employed by physical education researchers?

If Van Doren and Heit's (1973) assertion is true that "academic journals mirror the direction of research and serve as a medium for a discipline's communication" (p. 67), the answers to the above questions can be found by examining the scientific journals that publish original physical education research. It is readily acknowledged that the Research Quarterly for Exercise and Sport is the primary publication for original reports of physical education research (Gensemer, 1985; Kroll, 1982; Montoye & Washburn, 1980). Crase (1978) noted, however, that in any consideration of scholarship in physical education, one must recognize the emergence of the "subdisciplines" which have added "to the breadth of the mother discipline. . ." (p. 23). He also noted that the subdisciplines have "promising journals" which serve as an indication of the growing knowledge base in the subspecialties of physical education. An investigation of published physical education research would be incomplete without consideration of subspecialty journals.

An obvious concern for the consumer of research who encounters an unfamiliar statistic while reading a research report is the availability of information to help him/her to understand the statistic. In a guide to research writing, Leedy (1985) stressed the importance of providing a justification for one's selection of statistical procedures. As he explained in the text, "Where the data were subjected to statistical analysis, a rationale for employing the particular statistical approach should be presented. It is important to know, for example, not that one employs a particular correlational technique, but why one has done it" (p. 231). The same concern is echoed by both Clarke and Clarke (1984) and Isaac and Michael (1982) in their research texts. Isaac and Michael additionally stressed the importance of reporting and justifying the selection of data processing procedures. One is led to wonder regarding the extent to which physical education researchers provide justification for the statistical treatments of their data.

Slater-Hammel (1965a) hinted at one other piece of information that can aid the research reader who encounters an unfamiliar statistical technique, i.e., a good statistical reference. He criticized the carelessness with which physical education researchers cited statistical references, claiming that

> physical education researchers are frequently dependent upon antiquated introductory texts on statistics. . . . It is not unusual, for example, to find doctoral dissertations and research reports completed since 1960 citing a text of 1930-1940 vintage as the authoritative source. . . . (p. 212)

Are today's researchers still guilty of such neglect in statistical referencing?

Physical education research has also been criticized for inappropriate use of statistical techniques and for neglecting basic rules of statistical inference (Baumgartner, 1969; Cox & Serfass, 1981; Gould, 1982; Karpman, 1981; Kenyon, 1965; Korell & Safrit, 1977; Levine, 1977; Montoye, 1955; Morrow & Frankiewicz, 1979; Pierson, 1960; Schutz, 1972; Schutz, Smoll, & Gessaroli, 1983; Singer, 1966; Teraslinna, 1967). Use of inadequate sample sizes (Baumgartner, 1974; Dotson, 1980; Schutz, 1973) and an overconcern for statistical significance (Nelson & Hurst, 1963; Schutz, 1973) have also been areas of expressed concern in published physical education research. These criticisms suggest a question as to the current practices of researchers concerning appropriateness of statistical techniques, sample sizes employed, and extent of significance testing. Both Baumgartner (1974) and Slater-Hammel (1965a) pointed out that poor research practices are especially undesirable when published research serves as a model for the novice researcher. They feared that the beginner would repeat the errors he/she has seen in print.

Statement of the Problem

The purposes of this study were (a) to analyze selected aspects of statistical use in a sample of recently-published original reports of physical education research in selected physical education journals, and (b) to make inferences to the population of such physical education research. More specifically, the following research questions were addressed:

Type and Frequency of Statistical Techniques

1. What statistical techniques were used in a sample of physical education research reports?
    - 1a. What was the estimated population proportion of physical education research reports that used each of the identified statistical techniques? What are 95% confidence intervals for the estimated population proportions?
        - 1a-1. For each physical education subspecialty, what proportion of the sampled research reports used each statistical technique?

1b. What was the rank order of statistical techniques used in the sample of research reports?

1b-1. For each subspecialty, what was the rank order of statistical techniques used?

1b-1(a). What was the extent of concordance in the rankings for all pairs of the subspecialties?

1c. When statistical analyses were classified by similar purpose, what was the estimated population proportion of the total number of analyses that were classified into each category of statistical techniques? What was the rank order of use by category of statistical analyses?

1c-1. For each subspecialty, what proportion of the total number of analyses were classified into each category of statistical techniques? What was the rank order of use by category of statistical analyses?

Analytic Complexity

2. What was the estimated population proportion of physical education research reports that used as their most complex analysis a 1-variable statistical technique? ...a 2-variable technique? ...a multiple-variable technique? ...a multivariate technique?

2a. What multivariate statistical techniques were used in the sample of physical education research reports?

2a-1. What is the estimated population proportion of reports that used each of the identified multivariate techniques?

2a-1(a). For each subspecialty, what proportion of the sampled reports employed a multivariate statistical technique?

<u>Type of Generalization and Significance Levels</u>

3. What was the estimated population proportion of physical education research reports that used inferential statistical techniques? ...only descriptive techniques?

3a. For each subspecialty, what proportion of the sample research reports used inferential statistical techniques? ...only descriptive techniques?

4. What was the estimated population proportion of physical education research reports that employed a *p* value of .05 for tests of significance? ...reported exact *p* values?

4a. For each subspecialty, what proportion of sample research reports employed a *p* value of .05? ...reported exact *p* values?

Number of Variables

5. What was the estimated population median number of variables studied in physical education research? ...of dependent variables studied?

    5a. For each subspecialty, what was the median number of variables studied? ...of dependent variables studied?

6. What was the estimated population median for the largest number of variables simultaneously analyzed in a single statistical analysis? ...of dependent variables simultaneously analyzed?

    6a. For each subspecialty, what was the median for the largest number of variables simultaneously analyzed? ...of dependent variables simultaneously analyzed?

Number of Subjects

7. What was the estimated population median number of subjects employed in physical education research?

    7a. For each subspecialty, what was the median number of subjects employed?

8. What was the estimated population proportion of physical education research reports that used a within-subjects or mixed design?

    8a. For each subspecialty, what proportion of the sampled research reports used a within-subjects or mixed design?

Statistical Assumptions, Transformations, Nonparametrics

9. What was the estimated population proportion of physical education research that tested for a statistical assumption?

   9a. For each subspecialty, what proportion of the sampled research reports tested for assumptions?

10. What was the estimated population proportion of physical education research reports that transformed data prior to statistical analysis?

   10a. For each subspecialty, what proportion of the sampled research reports transformed data?

11. What was the estimated population proportion of physical education research reports that used a nonparametric statistical technique?

   11a. For each subspecialty, what was the proportion of the sampled research reports that used a nonparametric statistical technique?

Reporting of Statistical Analyses

12. What was the estimated population proportion of physical education research reports that provided justification for use of a particular statistical technique? ...that reported data analysis methods? ...that cited a statistical reference?

12a. For each subspecialty, what proportion of the sampled research reports provided justification? ...reported data analysis methods? ...cited a statistical reference?

Definitions of Terms

For purposes of interpretation, the following meanings were designated for use in this study:

Analytic Complexity. Classification of statistical techniques based upon the numbers of variables simultaneously entering the analysis. Categories are: (a) 1-variable statistic (e.g., mean, standard deviation), (b) 2-variable statistic (e.g., Pearson's r, t-test), (c) multiple-variable statistic (e.g., factorial ANOVA, multiple regression), and (d) multivariate statistic (e.g., canonical correlation, MANOVA). Both a multiple-variable statistic and a multivariate statistic simultaneously analyze three or more variables, but a multiple-variable statistic has a single dependent variable and a multivariate statistic has two or more dependent variables. Additionally, multivariate statistical techniques measure, explain, or predict relationships among variates, i.e., weighted combinations of variables.

Descriptive Statistical Technique. A statistical procedure that yields an index that summarizes and describes distributions and/or relationships of variables, without implying generalizations beyond the present sample.

Highest-Level Statistical Technique. Classification according to the highest of the four ordered categories of analytic complexity. A research study that employed only mean and standard deviation is said to have used a 1-variable analysis as its "highest-level statistic." A report that employed both Pearson's r and multiple linear correlation is classified as multiple-variable research.

Inferential Statistical Technique. A statistical procedure that yields a numerical index that summarizes sample data and generalizes to a population parameter with a stated level of probability.

Nonparametric Statistical Technique. An inferential statistical procedure used to test a hypothesis or define a confidence interval that does not depend on the form of the underlying distribution (Kendall & Buckland, 1971).

Original Research Report. A document written by the principal investigator(s) that provides complete and detailed information describing all aspects of the research endeavor.

Physical Education Research. Research inquiries seeking to answer questions germane to the study of purposeful human movement, broadly including anatomical-physiological and motoric studies, social and behavioral studies, and historical-philosophical studies (Haag, 1979). "Physical education" shall be understood to include nomenclature such as sport studies, sport science, exercise science, and pedagogy.

Quantitative Research. Research that employs measurement in data collection. "Measurement" includes frequency counts of nominal and/or ordinal data, rank ordering, and precise scoring to assess relative and/or absolute quantities of subject characteristics.

Recently-Published. Having a publication date between July 1, 1977 and June 30, 1984.

Research. An investigation that generates knowledge in response to an identified problem. It has the characteristics of being (a) systematic, (b) logical, (c) empirical, (d) reductive, and (e) replicable (Tuckman, 1978).

Similar Purpose. A classification of statistical techniques into the categories of: Central Tendency, Dispersion, t-Test, ANOVA, ANCOVA, MANOVA, MANCOVA, Correlation, Regression, Association, Factor Analysis, Multiple Comparison, Goodness-of-Fit, and Reliability. This classification focuses on the intended purpose of the statistical technique rather than the mathematical procedures of calculating the statistic.

Statistic. Numerical index that is generated in the summarizing, analyzing, or interpreting of an aggregate of units of measurement (Kendall & Buckland, 1971).

Statistical Analysis or Technique. A specific procedure for examining an aggregate of units of observations that results in generation of a statistic.

Subspecialty within Physical Education. Classification of areas of scholarly study and research into: (a) Background, Meaning, and Significance, (b) Functional Effects of Physical Activity, (c) Sociocultural and Behavioral Aspects, (d) Motor Learning and Development, (e) Mechanical and Muscular Analysis of Motor Skills, (f) Management Theory and Practice, (g) Program Development, and (h) Measurement and Evaluation (Zeigler, 1983).

Delimitations

The boundaries for the present study were largely established by the criteria invoked in sampling published physical education research. Materials for sampling were delimited to:

1. Research reports published between June 30, 1977 and July 1, 1984.

2. Quantitative research reports; excluded were reports that were non-numerical and/or qualitative, such as many historical, philosophical, and anthropological research papers.

3. Original research reports; excluded were summaries, abstracts, and research notes. Also excluded were traditional reviews of literature, articles presenting theory or model formulation, and articles presenting methodological suggestions or criticism.

4. Research reports published in journals; excluded were theses, dissertations, government reports, and oral reports.

5. Reports published in journals having the following characteristics:

a. Journals specializing in physical education with the Library of Congress call letters of "GV"; excluded were journals that publish physical education research but do not have the "GV" clasification such as Perceptual and Motor Skills.

b. Journals that specialize in reporting original physical education research; excluded were topical journals such as the Journal of Physical Education, Recreation, and Dance, current-event journals such as the NAIA News, and theoretical issue journals such as Quest.

c. Journals that primarily publish quantitative research reports; excluded were journals specializing in historical and philosophical research such as the Journal of Sport History and the Journal of Philosophy of Sport.

d. Journals published in the United States; excluded were foreign publications such as the Canadian Journal of Applied Sport Sciences.

e. Journals with national readership; excluded were state physical education research journals such as the North Carolina Association for Health, Physical Education, Recreation, and Dance Journal.

f. Journals that were periodic in publication; excluded were monographs and supplements.

The journals included in the study on the basis of the above criteria were: Dance Research Journal, Journal of Sport and Social Issues, Journal of Sport Behavior, Journal of Sport Psychology, Journal of Teaching in Physical Education, Research Quarterly for Exercise and Sport, and Review of Sport and Leisure.

Research Assumptions

The present investigation was based on two research assumptions. That is to say, the following two propositions were accepted as given and not examined as part of the research per se:

1. Journal selection procedures yielded a representative sample that permitted generalizations to the population of recently-published quantitative physical education research.

2. The published reports of research accurately reflected the actual processes and results of the investigations.

Limitations

The limitations of this study derive primarily from the nature of analyzing written archival data. It must be remembered that reports of research constituted the data of the present study. The thoroughness and accuracy of the written reports limited the accuracy of the present investigation.

There were also limitations inherent in the use of content analysis as a research method. The main concern was the definition of categories for the coding of data. Zeigler's (1983) classification scheme was used for the identification of subspecialties within physical education. Generalizations must accordingly be restricted to the definitions of subspecialties as defined by the scheme.

Statistical techniques were initially identified as they were presented by the author of the research report. Subsequently, the techniques were classified into various categories of interest by the present investigator. The accuracy of the resultant classification was, therefore, limited by her knowledge of statistics.

Significance of the Study

The results of this study provided a picture of the current use of statistics in published research in physical education. This knowledge is valuable in its own right in that it may help physical educators better understand the nature of their use of statistics as research tools. Additionally, knowledge of the types and frequencies of statistical procedures one encounters most often in physical education research may influence decisions relating to the content of statistics and research methods courses for physical educators.

The study also suggested that general statistical knowledge may not be appropriate for all physical educators. Subdiscipline specialists may need more-specific statistical competencies. For example, the statistics used most frequently in a specialized area such as Program Development differed markedly from statistics used in studying the Functional Effects of Physical Activity. Different coursework for persons studying in these two areas might be recommended.

Although the quality of physical education research is not assessed directly in this investigation, portions of the research findings addressed issues that are associated with quality. For example, in 1973 it was determined that 89% of the inferential research published in the Research Quarterly reported statistical significance (Schutz, 1973). This was interpreted as an unenlightened prejudice against the null hypothesis in physical education research. The incidence of reporting of statistical significance was one of the variables addressed in the present study. Also addressed were size of *p* values employed, numbers of subjects and variables studied, and prevalence of testing for statistical assumptions. It is believed that the present investigation of these variables provided insight into the quality of published physical education research.

CHAPTER II

REVIEW OF RELATED LITERATURE

There is a voluminous and diverse body of literature relating to research. The present review is limited to writings highlighting (a) the role of research in physical education, (b) the use of statistics in physical education research, and (c) selected methodological issues that relate to the use of content analysis and survey sampling. These topics were studied in order to provide valuable background information and justification for the present research, to clarify the research questions, and to allow comparisons with previous research.

Research in Physical Education

Physical educators are not newcomers to research. They have been interested in "research and the scientific side of physical education" since the early days of the profession (Clarke, 1938, p. 25). One wonders, then, why physical educators value research. What do they think will be gained from conducting research in physical education?

Importance to Physical Educators

McCloy (1930), whose influence on the profession is acknowledged by physical education historians, assured readers of the very first volume of the Research Quarterly that research was necessary for continued progress in physical education. He explained that,

> Professional progress in any science leads through at least four stages. The first is that of trial and error. . . . In the second stage "leaders" of the past are quoted. . . . The third stage is that of speculation and argumentation. . . . The fourth stage is that of hypothesis and experimentation. (p. 63)

Many years later, Vanderzwaag (1973) reiterated this theme to Quest readers. He asserted that research is a key factor in any "truly academic enterprise" (p. 78). Crase (1978) regarded research as an "index of scholarly production" (p. 23).

Physical educators have expressed the belief that research not only contributes to the discipline, but also contributes to the profession of physical education.

> The findings of research and the attitude of research are essential to any profession if it is to retain its vitality and have as its purpose an effective service to mankind and not the indoctrination of single ideologies, methods, or programs. (Lloyd, 1938, p. 33)

Application of research knowledge to solve humankind's problems is central to the role of the "physical educator as researcher" (Massey, 1966).

More recently, Silva and Parkhouse (1982) stated that accountability becomes increasingly vital as sport and exercise are viewed in the greater economic, social, and political contexts of society. They further suggested that research can provide accountability for the decisions made by physical educators in their various roles.

Steinhaus (1949) used a woodchopping analogy to express his idea of how research contributes to physical education. As woodchopping provides both wood and a better woodchopper, so does research provide to physical education "the building materials of accurate facts and principles with which to construct sound practice and wise philosophy, while concomitantly supplying "ideas to kindle enthusiasm in our professional ranks and, in the public mind, a warm reception for our programs"(p. 18).

Research Publications

Essential to the discovery of new knowledge by researchers is the dissemination of such knowledge. The primary vehicle for communication of research knowledge is the research journal. But obviously, "interest in a research journal does not operate in a vacuum" (Park, 1980, p. 2). There must be individuals and activities sufficient to warrant a special publication. The Research Quarterly began publication in 1930 and has since served as the single most important outlet for reporting research in physical education in the United States (Gensemer, 1985; Kroll, 1982; Montoye & Washburn, 1980).

The field of physical education, perhaps never was, and clearly now is not a discipline/profession with a unitary purpose of teaching sport, games, and exercise. Rather, it is a field comprised of many "areas of scholarly study and research" with associated "sub-disciplinary aspects" and "sub-professional aspects" (Zeigler, 1983). Crase (1978) asserted that the emergence of subspecialties added to the breadth of the mother discipline of physical education. With the increased interest in specializations within physical education came the development and expansion of organizations, many of which began to publish their own journals.

Today, there are numerous research journals serving the broad field of physical education both nationally and internationally (AAHPERD, 1982; Crase, 1979; Haag, 1979; Park, 1980; Sachs, 1978). The majority of these journals publish research reports focused on a subspecialty area such as sport psychology, exercise physiology, or teaching in physical education. The Research Quarterly is one of the few research journals that publishes research across the full spectrum of physical education interests. In 1979, the title of this journal changed to the Research Quarterly for Exercise and Sport, but there was purportedly no change in the focus of the journal. Additionally, a section editor arrangement with 14 section areas was initiated. Park (1980) believed this to be "graphic acknowledgement of the diverse nature of physical education and the maturation of research in the many areas which comprise the field" (p. 21).

Van Doren and Heit (1973) reasoned that since journals "mirror the direction of research" and "serve as a medium for a discipline's communication," they should be monitored "from time to time in order to recognize trends and to critically appraise their contributions to highly structured disciplines" (p. 67). This idea is consistent with Cureton's (1944) earlier comment that much could be learned from the current research literature of a field that could not be learned in any other way. Textbooks traditionally lag behind the current research lierature; each time they are rewritten they "catch up some of the outstanding work in the theses, articles, and research bulletins" (p. 150).

The Reading of Research Journals

Clarke and Clarke (1984) contended that it is very important that research be read by the professionals in a field of study.

> Inasmuch as all academic fields are becoming more technical and detailed, the need for informed practitioners is readily apparent. The ability to read and to evaluate critically the scientific literature of the field is a primary requisite in physical education. (p. 18)

Sharp (1976) studied the professional periodical reading habits of college and university physical educators. Included in his sample of 12 professional journals were three research journals, the Journal of Applied Physiology, Medicine and Science in Sports, and Research Quarterly. It was found that among the physical educators studied, none of the research journals was read with any degree of regularity. The Research Quarterly, primary research publication of the field, was reportedly read by only 26% of the college or university physical educators.

It has been suggested that poor reading habits have contributed to the creation of a "gap" between research and practice in physical education (Locke, 1972). But another consideration is the possibility that many physical educators understand little of the content of a research journal. One of the main stumbling blocks to understanding may be the systematic, symbolic language in which original research reports are written (Gabert, 1976; Locke, 1972; Puhl, 1982).

Statistics in Physical Education Research

Much of the published research in physical education has been analyzed and interpreted in a statistical frame of reference (Berlin, 1973; Clarke & Clarke, 1984, Teraslinna, 1967). Thus, the consumer of physical education research must know statistics to understand, interpret, and evaluate the research literature. This is a prerequisite for the acceptance or rejection of the conclusions of quantitative research investigations (Gephart, 1969; Good, 1933; Kroll, 1982).

One might wonder why it is that a reader of research cannot simply trust the conclusions as stated in a research report. It must be remembered that research is a product, and "like any other product the quality may range from excellent to shoddy" (Gensemer, 1985, p. 64). There are several texts written to warn the naive reader of research of the dangers of blindly trusting the interpretations of statistics proffered by the writers of research. These texts have intriguing titles such as How to Lie with Statistics (Huff, 1954), How to Use (and Misuse) Statistics (Kimble, 1978), How to Tell the Liars from the Statisticians (Hooke, 1983), and Statistics: A Spectator Sport (Jaeger, 1983).

The editors of Research Quarterly have been reprimanded for their failure to eliminate flawed research reports from the journal (Slater-Hammel, 1965b). And although "there is general agreement that the Research Quarterly has recently [since the late 1970's] enjoyed marked scholarly improvements" (Park, 1980, p. 21), there is still concern for the errors in the older lierature and for errors continuing to be made today. Erroneous conclusions may not only hinder new research efforts, but it is also possible that beginning researchers may repeat errors they have seen in published research (Slater-Hammel, 1965b). Criticisms of published research in physical education have attacked such problems as poorly conceived research questions (Locke, 1969; Pelton, 1976; Van Dalen, 1962), design errors (Callahan & Ziegler, 1980; Fellingham, Bryce, & Carter, 1978), and methodological errors (Baumgartner, 1969a; Landers, 1973; Martens, 1973; Singer, 1973; Slater-Hammel, 1959; Williams, 1973). More relevant to the present investigation are the numerous criticisms that have been leveled against inappropriate use of statistics (Baumgartner, 1969; Cox & Serfass, 1981; Gould, 1982; Karpman, 1981; Kenyon, 1965; Korell & Safrit, 1977; Levine, 1977; Montoye, 1955; Morrow & Frankiewicz, 1979; Pierson, 1960; Schutz, 1972; Schutz, Smoll, & Gessaroli, 1983; Singer, 1966; Teraslinna, 1967), inappropriate sample sizes employed (Baumgartner, 1974; Dotson, 1980; Schutz, 1973), and overconcern for statistical significance (Nelson & Hurst, 1963; Schutz, 1973).

Research on the Use of Statistics

Before today's widespread accessibility of high-speed computers, statistics in educational research were largely restricted to t-tests, one-way and two-way analyses of variance (ANOVAs), correlations, and chi square analyses. Today, there is evidence of increased use of multiple-variable statistics appearing in a variety of scholarly journals (Kerlinger, 1979; McMillan & Brown, 1984). One would suspect that this is the case for physical education journals. But, there has been little reported research accounting for the type of statistics presently in use in physical education research.

In 1968, Brady determined that the mean, standard deviation, Pearson's product moment correlation, and analysis of variance were the most commonly used statistical techniques in doctoral dissertations. His panel of 29 research experts believed, however, that factor analytic, nonparametric, and other correlational statistical techniques should be given greater emphasis in physical education research preparation. In the following year, Burkhart (1969) identified and categorized the statistical techniques reported in the Research Quarterly from 1962 to 1966. He found that of the 382 studies that employed statistical analyses, 42% used techniques of hypothesis testing about means, variances, or proportions, 24% used correlation and/or regression techniques, 15% used reliability and/or validity techniques, 13% relied on descriptive statistics, 5% used nonparametric statistics, and 1% of the studies used factor analyses.

The above figures can be contrasted with those reported by Van Doren and Heit (1973) who studied the statistical analyses reported in the first three years of The Journal of Leisure Research, the primary research publication for the allied profession of leisure. Regression and correlation techniques were the most commonly reported category of statistics. They were used in 41% of the research articles. Other statistical categories and the percentage of their use were: (a) chi square analysis, 24%; (b) analysis of variance, 15%; (c) factor analysis, 13%; (d) Kolmogorov-Smirnov tests, 4%; and (e) time series and differential equations, 1.5% each.

Use of multivariate statistics. Morrow and Frankiewicz (1979) reviewed Research Quarterly articles published in 1976 and 1977 and found that many of the researchers considered more than one dependent measure. Morrow and Frankiewicz supported the practice noting that it was "certainly a justified procedure in that experimental or treatment effects are seldom confined to one dependent variable or constrained to one point in time" (p. 297). However, the same investigators discovered that few of the studies they reviewed employed appropriate multivariate or repeated measures analyses. They explained that there is often a lag between the theoretical development of new statistical procedures and the time when the techniques become available to practitioners. Morrow and Frankiewicz predicted, however,

> The current state of development of multivariate statistical techniques, the advent of computers, and the availability of reasonable computing algorithms in packaged programs suggest that these more appropriate tests are no longer beyond the grasp of researchers. (p. 302)

However, in a 1983 investigation, Schutz, Smoll, and Gessaroli found that multivariate statistical techniques were still relatively uncommon in published physical education research. They surveyed 188 quantitative research reports published in Journal of Motor Behavior, Journal of Sport Psychology, Medicine and Science in Sports, and Research Quarterly for Exercise and Sport. Content analysis revealed that approximately 70% of the studies employed more than one dependent variable, and over 50% used five or more dependent variables. However, consistent with the earlier finding by Morrow and Frankiewicz (1979), it was found that only 40% of the studies that reported multiple dependent variables used multivariate analyses.

In recent years, there have been several treatises written to encourage the use of more complex statistical analyses in physical education research. Korell and Safrit (1977) and Levine (1977) wrote articles to explain and stimulate use of multidimensional scaling. The 1981 Symposium Consortium Papers addressed technical aspects of using selected multivariate and repeated measure analyses (Cox & Serfass, 1981). Karpman (1981) discussed an issue related to improving interpretability in use of canonical correlation analysis. Schutz, Smoll, and Gessaroli (1983) provided readers of Research Quarterly for Exercise and Sport with a "self-test" and "guide" to the utilization of six "useful" multivariate procedures: Hotelling's T-square, MANOVA, discriminant analysis, canonical correlation, automatic interaction detection, and multiple regression. Each of the statistical techniques was presented via a sample problem, then illustrated by a worked example, contextual explanation, and several statistical references.

Criticisms of the Use of Statistics

Statistical significance. A statistically significant result is one that is unlikely to occur by chance. Significance is tested at a certain level of probability, represented by the *p* value. Conventionally, this *p* value is set at a conservative level such as .05 to avoid making a Type I error, i.e., falsely rejecting a true null hypothesis. Schutz (1973) suggested that physical education researchers and their journal editors have been guilty of what Greenwald (1975) referred to as "prejudice against the null hypothesis." In an informal survey of two issues of the Research Quarterly, Schutz discovered that 16 of the 18 studies involving significance testing reported statistically significant findings. He posited that it is very unlikely that 89% of all research conducted yields significant findings. Nor did Schutz believe that reports indicating significance were necessarily superior in their research methods and/or theoretical rationale. Schutz concluded that many studies were judged valuable primarily on the basis of the level of significance attained. In a 1980 review of physical education research, Dotson reached the same conclusions.

Nelson and Hurst (1963) were also concerned with the issue of significance. They believed that physical education researchers often made the naive error of judging a piece of research by inspecting the magnitude of the p value at which significance was claimed. The practice has been termed "worship of 'p'" (Schulman, Kupst, & Suran, 1977). "A p value of .05 puts you in the courtroom, at .01 you are an honored star of the bar, and at .001 you are a veritable Clarence Darrow" (p. 40). However, Isaac and Michael (1982) cautioned that the practice of demanding conservative p values of .05 or less may be counterproductive in certain situations such as an educational setting in which learning outcomes may result from many complex factors acting independently and jointly. In much educational and psychological research, differences between group means and magnitudes of correlation coefficients are most likely to be low. The researcher who sets a restrictive value in an attempt to avoid making a Type I error, may be increasing the risk of making a Type II error, i.e., retaining a false null hypothesis. Statisticians suggest that the p value for a test of significance should be selected with care in light of the nature of the investigation. The conventional p value of .05 may be used too often by physical education researchers.

Sample size. "Ensuring that acceptance of the null hypothesis does not represent a Type II error calls for attention to the selection of samples that are adequate to permit the true state of affairs to reveal itself" (Dotson, 1980, p. 29). Sadly, however, it has been concluded that most physical education research is conducted with inadequate sample sizes (Baumgartner, 1974; Dotson, 1980; Schutz, 1973). The consequence of using inadequate samples is an inadvertent reduction in statistical power, thus making it harder to attain the "worshipped" statistical significance.

Jones and Brewer (1972) reviewed the power of t and ANOVA statistical tests reported in the Research Quarterly between October 1969 and May 1971. They found that sample sizes for the 106 articles analyzed ranged from 3 to 1200; but the median sample size was between 51 and 75 subjects. Powers were calculated for different effect sizes, i.e., the difference between group means expressed in standard deviation units (Glass & Hopkins, 1984). The resulting powers were "disturbingly low." Jones and Brewer recommended that physical education researchers (a) increase sample sizes for fixed alpha and effect sizes, and (b) not be so "greedy" in striving for very conservative *p* values when a more liberal *p* is adequate.

In 1977, Christensen and Christensen conducted a similar analysis of the power of statistical tests in the Research Quarterly. They calculated powers for different effect sizes for t-tests, ANOVAs, ANCOVAs, Pearson correlations, and chi square analyses. It was found that, "on the average, the chances of detecting anything less than a large effect size in the population was less than one-to-one" (p. 207). Thus, they concurred with Jones and Brewer in recommending that more attention be given to sample sizes and the concept of statistical power.

King (1978) focused criticism specifically on the inappropriate sample sizes used in survey research in physical education. He believed that most survey researchers selected sample sizes in an arbitrary manner without regard for the desired precision of parameter estimates.

Roundy (1968) explained that sample size cannot be determined in an arbitary way; it is best determined in relation to practical significance. Ideally, a researcher needs to (a) calculate the smallest difference, or correlation, or the like, that is of practical importance, and (b) obtain an estimate of the population variance. Then the researcher can use appropriate formulas and statistical tables to select the needed sample size.

Tolson (1980) extended the concept of practical significance by recommending that physical education researchers calculate and report the omega squared value. He reiterated the notion that a statistically significant result says nothing about the practical significance of the association. The omega squared statistic estimates the degree of association, or percent variance accounted for, between the independent and dependent variables. The larger the value of omega squared, the more homogeneous are the observations within clases relative to between classes (p. 580). Tolson illustrated his idea by calculating omega squared for three articles previously published in the Research Quarterly. He clearly demonstrated that statistical and practical significance are not synonymous concepts. One of the most recent research textbooks written for health, physical education, recreation, and dance (Thomas & Nelson, 1985) included a strong plea for increased use of the omega squared statistic.

Appropriateness of the statistic. One problem with using an inappropriate statistic lies in its effect on probabilistic conclusions. Conclusions lose their intended precision because the actual alpha levels may be greater than or less than nominal alpha levels (Anderson et al., 1975; Cox & Serfass, 1981; Pruzek, 1973). If the particular statistic is robust, the approximation is generally reasonable. However, in some cases, an erroneous conclusion is reached.

Slater-Hammel (1969) labelled the use of a wrong statistical model in evaluation of data a "vulgar error." Baumgartner (1969b) agreed, "If assumptions underlying a model are not approximately met then there is no justification for using the statistical test and the results mean little" (p. 863). Dotson (1980) also addressed the relationship between development of a research strategy and the emphasis on underlying statistical requirements of the model. He mourned, "It is inexcusable that such assumptions are not attended to by the majority of contributors to the Research Quarterly" (p. 27).

There are several varieties of assumption violations noted in the physical education research literature. The assumption of interval level of measurement has been questioned for the analysis of attitude scale scores (Petrie, 1969). The assumption of randomization of subjects to treatments for use of *t* or *F* tests has been challenged because of the frequent use of intact groups in physical education research (Baumgartner, 1969b; Rosemier, 1968, 1969; Slater-Hammel, 1968). And, the assumption of normality in distribution of scores has also been challenged (Baumgartner, 1974; Berenson & Wolf, 1977; Slater-Hammel, 1968). Berenson and Wolf (1977) suggested that more physical education researchers should consider use of data transformations or use of appropriate nonparametric statistical procedures when data do not meet the assumptions for analysis of variance procedures.

Criticisms have also been leveled at researchers who failed to distinguish among designs in their statistical analyses. Especially problematic seem to be randomized block designs (Henry, 1977; Slater-Hammel, 1969) and repeated measures designs (Cox & Serfass, 1981; Stamm & Safrit, 1975). And, as discussed earlier, many researchers fail to treat multiple-dependent-variable designs with appropriate multivariate statistical procedures.

Another concern is the inappropriate use of post hoc procedures following a significant statistical test. Kenyon (1965) and Singer (1966) both addressed the problem of using multiple t tests of all possible cell pairs as a post hoc test for significant analysis of variance findings. More recently, Mihevic and Spray (1979) explained the advantages of using a simultaneous confidence interval procedure for post hoc analysis of significant multivariate analyses of variance.

Reporting of statistical analyses. Physical education research has been criticized for the manner in which statistical analyses have been reported. Nelson and Hurst (1963) complained that there is often too much journal space devoted to the reporting of intermediate statistical calculations. They also suggested that exact probability values be reported rather than merely indicating significance with asterisks. Slater-Hammel (1965a) criticized physical education researchers for citing outdated statistical references. Teraslinna (1967) complained that, "When sophisticated statistical procedures are used, they are worthless without careful discussion and interpretation of the results" (p.156). Cox and Snell (1981) agreed with Teraslinna, "Effort spent in trying to present in a simple way the conclusions of complex analyses is almost always worthwhile" (p. 6).

Research textbooks and style guides for research writing are logical references for the selection of what to include in a research report. Isaac and Michael (1982) recommended including text to justify the appropriateness of both statistical treatment and data processing procedures. Leedy (1985) also stressed the importance of presenting a rationale for use of a particular statistic. The current publication manual of the American Psychological Association (1983), in the guidelines for statistical copy, specified that authors (a) are responsible for selection of statistical method and all supporting data, (b) should give references for less common statistics, especially those that are not yet incorporated into textbooks, (c) should give formulas for uncommon statistics, and (d) should give the value of a statistic, the degrees of freedom, the exact probability level, and "any other descriptive statistics to clarify the nature of an effect" (p. 80).

Content Analysis and Survey Sampling

Newspapers, periodicals, and textbooks provide written evidence of the activities and interests of persons involved in physical education and sport. The research methodology which is primarily associated with systematic investigations of such written media is content analysis.

Content Analysis as a Research Method

Most persons are aware of the "tricks" that their attentions and memories can play when they read impressionistically (Carney, 1972, p. xv). Obviously, impressionistic reading is unacceptable for any conscientious research effort. Content analysis, however, is a tool by which written and oral communications can be studied in a way that is consistent with the rigorous demands of research.

Definitions of content analysis reflect how the methodology evolved to what it is today. Berelson (1952) defined it as "a research technique for the objective, systematic and quantitative description of the manifest content of communication" (p. 18). The terms "objective" and "systematic" in Berelson's definition implied replicability of the process. "Quantitative" denoted simplification of complex data via frequency counts, ranking, or rating. And, "manifest content" related content analysis to the overt aspects of communications. Newer conceptions of content analysis have extended its domain to qualitative analyses and to latent aspects of communications. Holsti's (1969) definition reflected a broader idea. "Content analysis is any technique for making inferences by objectively and systematically identifying specified characteristics of messages" (p. 14). Krippendorff (1980) also acknowledged the function of inference by defining content analysis as "a research technique for making replicable and valid inferences from data to their context" (p. 21). The latter definition also considered the context within which a communication occurs; a researcher may interpret the meaning of a message in relation to the sender's intent, to the effects on the receiver, or to the cultural institution within which it is exchanged.

<u>Uses of content analysis</u>. Journalists and sociologists have, perhaps, used content analysis more than any other specialists. However, its popularity is widespread, as evidenced by formal research in anthropology, education, history, literature, philosophy, psychology, and religion (Stone, 1966). Anderson et al. (1975) termed content

analysis a "general" technique because it is modifiable to so many different settings and purposes.

One of the more popular uses of content analysis is for investigation of data reported in newspapers and periodicals. Researchers allege that the technique is especially good as a means to monitor social change because the available space in newspapers and periodicals represents a "closed system" (Naisbitt, 1984). There are choices which must be made about what is permitted to enter the analysis due to space constraints. Thus, analyses of what has been published tells something about values.

Advantages and limitations. Every research technique has identifiable advantages and limitations. Probably the greatest advantage of content analysis is its wide range of uses. Other advantages stem from characteristics of recorded messages. Such messages are (a) "static," and thus, can be copied and shared with other researchers, (b) can be re-analyzed several times to ensure accuracy and to collect data on several dependent variables from one record, and (c) may be re-used for other research purposes (Stone, 1966). Records of communications for content analysis are also readily available; written and oral communications are an enduring part of human culture. Content analysis is also credited with being (a) an unobtrusive technique, (b) context sensitive, (c) able to cope with large volumes of data, as well as (d) accepting unstructured data (Krippendorff, 1980).

Content analyses do, however, have limitations. Category construction is widely regarded as the most crucial aspect of content analysis (Berelson, 1952; Stone et al., 1966). Knowledge and familiarity with the field under investigation is considered the best protection against problems of category construction. Content analysis often involves sampling from a larger population, thus sampling and statistical inference problems may add to difficulty in using the technique.

Reliability and validity are obviously essential to any research effort. If the population of data is identified with care and sampling is performed meticulously, validity generally is not a major problem in content analysis. However, reliability may be a greater problem. In order for a content analysis to be reliable, data should be "reproducible, by independent researchers, at different locations, and at different times, using the same instructions for coding the same set of data" (Krippendorff, 1980, p. 132). Krippendorff asserted that errors are often made in the determination of reliability coefficients in content analyses. Reliability should be expressed as a function of the agreement, above and beyond chance, achieved among coders for the assignment of data units to categories.

Content analyses in physical education. As noted earlier, content analysis has been used by researchers in many disciplines. Physical educators used the technique for varied research purposes and with different types of materials. For example, physical education researchers have performed content analyses of children's diaries (Cowell, 1937), autobiographies of baseball players (Haerle, cited in Loy & Segrave, 1973), transcripts of interviews with physical education teachers (Earls, 1981), elementary physical education textbooks (Hildreth, 1979), doctoral dissertations (Cureton, 1949), non-research physical education publications (Hirsch, 1980; Lock, 1975), lay sports periodicals (Condor & Anderson, 1984; Hart, 1972; Holtzworth, 1977; Reid & Soley, 1979), and the sports pages of newspapers (Lau & Russell, 1980; Novak, 1942; Pearman, 1978).

Of particular relevance to the present investigation are the relatively few studies that content analyzed physical education research publications for the purpose of studying the use of statistics. Such analyses were conducted by Burkhardt (1969), Morrow and Frankiewicz (1979), and Schutz, Smoll, and Gessaroli (1983). Van Doren and Heit (1973) performed a content analysis for the leisure profession that also examined statistical usage.

The present search of the literature also revealed several content analyses of physical education research publications that were conducted to study topics besides that of statistics use. Loucks (1952) analyzed 20 years of Research Quarterly, considering (a) topic, (b) field, (c) "area of thought," (d) sex of author, and (e) geographical region from which the author came. He found that (a) 72% of the articles were from the field of physical education, (b) tests and measurement was the most popular topic, (c) education was the most popular area of thought, (d) male authors outnumbered female authors by a ratio of more than two to one, and (e) that more authors came from the Eastern region than any other geographical area. He also reported that a sizable number of different authors from nearly 300 secondary schools and institutions of higher learning contributed to the Research Quarterly during the time period studied.

Russell (1962) content analyzed only health research reports published in Research Quarterly during a 10-year period. He found (a) the most popular research topic was health education curriculum, (b) 66% of the research reports were by a single author, and (c) male authors outnumbered females by nearly three to one. Russell also performed a frequency analysis of the research methodologies that had been used. He reported that one-third of all health research had employed some form of normative survey.

The first seven years of the International Journal of Sport Psychology were content analyzed for the purpose of identifying major trends in the social-psychological realm of sport (Groves, Heekin, & Banks, 1978). In addition to the stated purpose, the researchers also studied reference citations. It was found that journal articles and reference materials such as books and theses were cited with approximately equal frequency. Additionally, the authors reported that the single most frequently cited journal was the Research Quarterly.

Two other content analyses using physical education research journals were found. They were performed to learn more about research in youth sports (Gould, 1982; Weiss & Bredemeier, 1983). Both studies reviewed research articles from a variety of periodicals. Weiss and Bredemeier searched the youth sport empirical research and review papers in Research Quarterly, Journal of Sport Psychology, Review of Sport and Leisure, Journal of Sport Behavior, Perceptual and Motor Skills, Journal of Motor Behavior, The Physical Educator, and Psychology of Motor Behavior and Sport. They found (a) that approximately 80% of the articles were empirical studies and 20% were reviews of some kind, and (b) that only about 10% of the articles were written from a "developmental perspective." Gould used content analysis to identify critical research questions in youth sports research. He did not, however, specify the periodicals which he reviewed.

Survey Sampling in Research

According to Kerlinger (1973), "Survey research studies large and small populations . . . by selecting and studying samples chosen from the populations to discover the relative incidence, distribution, and interrelations of sociological and psychological variables" (p. 410). There are two general components of a sampling design: (a) a selection process that describes the rules and operations which determine the members of the population to be included in the sample, and (b) an estimation process for computation of sample statistics and their associated error variances which estimate the population parameters (Schutz, 1973).

Simple random sampling assumes that all elements in the population have an equal chance of being selected and that this probability of selection is constant at any draw. Such a selection procedure provides unbiased estimates of population means and totals (Jaeger, 1984). Many times, however, the variable of interest may be related to another variable for which information is readily available. The population can be partitioned into strata on the basis of the concomitant variable and simple random sampling can then be applied within each stratum. Alternatively, the strata can be defined on the basis of naturally occurring boundaries. In such cases, the size of the sample taken from each stratum might be directly proportional to the total number of elements within each stratum. Stratified random sampling with proportional allocation guarantees estimation precision that is at least as good as that derived from simple random sampling (Jaeger, 1984). In fact, there can be a marked increase in statistical efficiency for estimation of population means and totals. There may be only small gains in efficiency when population proportions are estimated (Jaeger, 1984).

Sample size in survey research. Determination of the size of the desired sample is related to the concept of statistical significance. No longer is it believed that samples should be as large as economically and practically feasible. Rather, sample size should be determined in light of desired precision and level of confidence for the parameters being estimated. An estimate of the population variance is necessary in order to calculate the required sample size. When estimating population proportions, as is commonly done in survey research, it has been recommended that .50 can be used conservatively as a hypothesized value of the population proportion (Issac & Michael, 1981; Jaeger, 1984).

Survey sampling in physical education. In a comprehensive review of research from the subspecialty of sport sociology, Loy and Segrave (1973) indicated that sample surveys constituted the main means of data collection. However, they also determined that "relatively few investigators have given adequate attention to the problems of sampling" (p. 305). Schutz (1973) and King (1978) echoed this concern for the field of physical education and offered help to the researcher interested in using survey sampling techniques. Schutz provided a broad overview of theoretical and practical knowledge about sampling procedures. He also identified several problem areas which arise in the study of sport and physical activity, and proposed methods for reducing the severity of the problems. King demonstrated sampling errors in three recently-published Research Quarterly articles, then presented a nomogram to assist in determination of sample size for populations containing fewer than 2000 elements.

Summary

The review of literature presented has illustrated the extent of research interest among physical educators. There are now numerous journals for the purpose of publishing original research reports in physical education or one of the subspecialty areas. A substantive amount of physical education research deals with quantitative information. Whether or not the reports have meaning to the readers depends, in part, on the understanding they have of statistics.

Physical education research has been criticized for numerous flaws. Among the more serious flaws are errors in the use of statistics. The nature and implications of these errors were discussed briefly.

Content analyses and surveys have been used widely in physical education as research methodologies. However, careful attention has not always been given to the limitations of these methods. The above review considered techniques for effective execution of both methods.

CHAPTER III

PROCEDURES

The purpose of this study was to investigate the current use of statistics in physical education research. Particular types of statistical analyses used and issues relating to the use and reporting of statistical analyses in published research reports were systematically examined. This chapter contains information concerning the methods used in carrying out the research.

Sampling Procedures

Stratified random sampling of articles with proportional allocation was the sampling strategy used to guarantee representation of selected physical education research journals. Jaeger's (1984) detailed steps were followed in order to accurately draw the sample.

Journals Selected for Analysis

The present investigation was delimited to major American physical education research journals that (a) report full-length, original quantitative physical education research, (b) have Library of Congress call letters of "GV," (c) are published in the United States, and (d) have national readership. Accordingly, the following seven journals were identified for inclusion in the study: Dance Research Journal, Journal of Sport and Social Issues, Journal of Sport Behavior, Journal of Sport Psychology, Journal of Teaching in Physical Education, Research Quarterly for Exercise and Sport, and Review of Sport and Leisure.

Determination of Sample Size

For the present study, it was decided that proportions estimated to within ±.05 of the true population proportion with a 95% level of confidence were acceptable. Total sample size was calculated by substituting appropriate values into the formula given by Jaeger (1984, p. 59):

$$\text{Sample Size} = \frac{(t/E)^2 \, P \, (1 - P)}{1 + (1/N) \, [(t/E)^2 \, P \, (1 - P) - 1]}$$

where t was the standard normal deviate for the desired level of confidence, E was the allowable estimation error, P was a hypothesized value of the population proportion for estimation purposes, and N was the total population size. This formula assumes simple random sampling, and this investigator assumed no reduction in error variance would be gained by stratifying by journal.

The population size for the present study was determined by identifying and counting all articles published in the selected journals between July 1, 1977 and June 30, 1984 that met the criteria stated above. A total of 582 articles constituted the population. All articles were listed separately by journal to create the sampling frames for the study. A value of .50 was used as the hypothesized P value to give a conservative estimate of the necessary sample size. Thus, substituting these values into the above formula, i.e., $t = 1.96$, $E = .05$, $N = 582$, and $P = .50$, the desired sample size for the present study was calculated to be 232 research articles.

The sample size for each of the seven journals was determined so that the number of articles selected from each of them was directly proportional to the size of the population of quantitative research articles published therein. A constant sampling fraction of .4 was used for the seven journals. The resultant sample sizes for each journal are given in Table 1. The *Research Quarterly for Exercise and Sport* sample required 122, the largest number of articles. Only 2 articles were required for the *Dance Research Journal* sample. Due to rounding of fractions, the final sample sizes totaled 233 articles.

Articles Selected for Analysis

Using lists of random numbers (Rand Corporation, 1955) and procedures specified by Jaeger (1984), simple random samples of articles were chosen from each journal list. Appendix A provides a complete listing of the research articles from each journal that were selected by using the above procedures.

Table 1

Sample Sizes for Physical Education Research Journals

Stratum (Journal)	Population Size	Sampling Fraction	Sample Size
Dance Research Journal	5	.4	2
Journal of Sport and Social Issues	12	.4	5
Journal of Sport Behavior	75	.4	30
Journal of Sport Psychology	116	.4	46
Journal of Teaching in Physical Education	32	.4	13
Research Quarterly for Exercise and Sport	305	.4	122
Review of Sport and Leisure	37	.4	15

Specification of Research Questions

Most of the research questions for the present investigation were formulated to reflect problems suggested by the review of literature. Several additional research questions were designed in response to a scheme for evaluating and understanding statistical analyses presented by Cox (1979) and Cox and Snell (1981). The scheme consisted of several categories by which statistics used in research studies can be compared and contrasted. One category considered variations in the characteristics of the statistical technique itself, such as (a) the "type of answer" it provides, i.e., descriptive vs. probabilistic, (b) its "conceptual complexity," (c) its mathematical or "numerical analytic complexity," and (d) the "sensitivity" of the technique to detect differences and associations. Another of the framework categories considered variations in the "complexity and quantity" of the data analyzed, i.e., the number of variables and the sample sizes employed. Another category described and compared statistical analyses in terms of the "computational load" defined by data transformations and special programming effort required. Thus, the refined problem statement dictated the procedures executed in carrying out each step in the investigation.

Content Analysis of the Selected Articles

Each of the 233 selected articles was content analyzed to determine the statistical techniques reported in the respective data analyses. Statistical procedures identified within the Methods, Results, and/or Discussion sections of each article were recorded using a sign system, i.e., a statistic was tallied only once per article no matter how many times the technique was applied or reported. Reports of simple frequency counts were not considered. Nor were distinctions made for ANOVAs, ANCOVAs, MANOVAs, and MANCOVAs that were more complex than three-way analyses.

Other information necessary to answer the research questions was also considered: (a) classification of subspecialty focus, (b) number of subjects and basic research design, (c) number of variables studied and simultaneously analyzed by using a single statistical technique, (d) selected details of significance tests, and (e) reporting of tests of statistical assumptions and data transformations. Also noted for each article were the presence or absence of explanations for the statistical analyses employed, citations of statistical references, and reports of computer data analysis procedures. Coding was conducted over a period of several weeks; the articles were analyzed in the same random order in which they were selected.

Data Collection

A code sheet was developed to collect the necessary data for each research report. The code sheet is presented in Appendix B. Because category definitions are essential to the validity of content analyses, operational rules and explanations for each category on the code sheet were also developed. The definitions, rules, and explanations employed in the present study follow.

Question A. For each article, the main physical education subspecialty focus was identified. The subspecialty categorization followed Zeigler's (1983) classification scheme. Zeigler's scheme is unique in identifying both sub-disciplinary and sub-professional aspects associated with each of the named areas of scholarly study and research. Figure 1 presents the scheme in summary form.

The best clue for determining the subspecialty focus of an article for analysis was its stated research purpose. Also helpful was the content of the supporting literature reviewed in the report. In every case, each article was coded according to its primary focus. For example, research on perceived exertion was classified as "Functional Effects of Physical Activity" if its primary focus was on physiological variables. But it was classified as "Sociocultural & Behavioral Aspects" if its main focus was on psychological variables. When an article was encountered that was particularly difficult to classify, reference was made to

Areas of Scholarly Study & Research	Sub-Disciplinary Aspects	Sub-Professional Aspects
Background, Meaning, & Intercultural Significance	History Philosophy International & comparative study	International relations Professional ethics
Functional Effects of Physical Activity	Exercise physiology Anthropometry & body composition	Fitness & health appraisal Exercise therapy
Socio-cultural & Behavioral Aspects	Sociology Psychology (individual & social) Anthropology Political science Geography Economics	Application of theory to practice
Motor Learning & Development	Psycho-motor learning Physical growth & motor development	Application of theory to practice
Mechanical & Muscular Analysis of Motor Skills	Biomechanics Neuro-skeletal musculature	Application of theory to practice
Management Theory & Practice	Theory about the management function	Application of theory to practice
Program Development	Theory about program development (General education; professional preparation; intramural sports & recreation; intercollegiate athletics; programs for the handicapped--including curriclum & instructional methodology)	Application of theory to practice
Evaluation & Measurement	Theory about the measurement function	Application of theory to practice

Figure 1. Subspecialty foci within physical education (Zeigler, 1983, p. 58).

Zeigler's (1982) text, Physical Education and Sport: An Introduction in which separate chapters were devoted to each subdiscipline.

Question B. The number of subjects were determined for each article. "Subjects" were usually persons, but were occasionally units of analysis such as classes or teams. Only the subjects from whom data were actually used in subsequent analyses were included. The present analysis did not acknowledge those subjects for whom data were not analyzed because they were incomplete or because the obtained data failed to meet a specified criterion level established by the researcher.

Question Bl required classification of each article according to its research design. The main consideration for the decision in the present analysis was whether or not data were collected on the same subjects under more than one condition or at more than one point in time. For example, a study was classified as "Completely Randomized" if different subjects were measured in all treatment conditions and if the subjects were measured only once per variable. A study was classified as "Pure Within-S" if all subjects were measured in all conditions and/or measured more than once per variable. A "Mixed" classification was designated for those studies that measured some, but not all, subjects in more than one condition and/or more than once per variable. Reference was made to Keppel (1982) if further clarification was needed.

Question C. The total number of variables that were statistically analyzed were recorded for each article. Included were independent, dependent, clasification, and control variables. Variables that were mentioned in an article but did not receive any statistical analysis were not considered. Levels of categorical variables were not counted separately.

The number of dependent variables was recorded for Question C1. In each case, the total number of variables that were treated in analyses as dependent variables were recorded. For example, if a separate ANOVA was executed for each of three different dependent variables, the coding was "3." For statistical techniques that do not specify a dependent variable, special coding conventions were followed. Simple correlations were coded as having "1" dependent variable, factor analyses were coded as if all variables were dependent, and discriminant function analyses were coded as if all predictor variables were dependent variables used to predict level of a single categorical variable. As used in the sampled studies, discriminant function analysis selected subjects on the basis of their membership to a criterion variable such as athlete versus non-athlete. Thus, the criterion variable acts as an independent variable (Thomas & Nelson, 1985).

Question C2 called for recording the largest number of variables that were simultaneously analyzed. Number of levels for categorical variables were ignored in this analysis. For example, a 2 X 2 X 8 analysis of variance was coded as "4", counting the three categorical variables and the one dependent variable.

For Question C2a the largest number of dependent variables that were simultaneously analyzed was tabulated. Again, the coding conventions for simple correlations, factor analyses, and discriminant function analyses were followed.

Question D. All statistical techniques employed in analysis of the data for each article were listed using a sign system. Only statistical procedures were listed, not all of the statistics derived from the procedures. For example, "multiple regression" was recorded while "standardized beta coefficient" was not. Statistical techniques used to describe the sample and the data collection instruments were listed only if the analysis was performed as part of the reported study.

Statistical procedures were most often located in the text of the Methods, Results, and Discussion sections of the research reports. However, tables, illustrations, and footnotes were also studied for statistical symbols that indicated use of a particular technique.

The listing of statistical procedures was as specific as possible. For example, a complete listing would be

"2 X 3 X 2 analysis of variance with repeated measures on the last factor." Whenever possible, names of statistical procedures were directly quoted from the article. In cases where the statistical technique was not clearly specified, the procedure was listed in parentheses if it could be determined from supporting text.

Question E. The coding of data for Question E was "Yes" if a single inferential statistical procedure was reported within an article. For Question E1 "Yes" was coded if there was a single reporting of statistical significance for a research question.

The criterion alpha, or p value, used for tests of significance throughout the article, was tabulated for Question E2. Categories included ".10", ".05", ".01" and "Other." "No" was coded if the p value was not recorded or clearly implied by listing in statistical tables. "No" was also coded if the p value changed from analysis to analysis. "Yes" was coded for Question E3 if exact p values were reported for any of the analyses conducted within an article.

Question F. "Yes" was coded when a justification or rationale for use of a particular statistical technique was proffered within an article. The explanation could be brief, but had to be included in the article itself. A reference citation not accompanied by any explanation was coded as "No." The nature of the justification and the page of the article on which it was given was specified in the space provided on the code sheet.

Question G. "Yes" or "No" was coded for each article to indicate whether or not a statistical reference was cited within an article. In order to be coded "Yes", a reference had to imply intent to explain a particular statistical technique. Citation of another research study that merely employed the same statistical technique was not considered a statistical reference. In each case, the reference itself and the page number of the article in which it was cited were specified on the code sheet.

Question H. Question H was coded to indicate whether or not a statistical assumption was reportedly tested within an article. "Tested" was a key word in the coding of this variable. For example, the mere reporting of standard deviation values was not considered to constitute testing for homogeneity of variance. The assumption tested, the test employed, and the page where reported were specified on the coding sheet.

Question I. "Yes" or "No" was recorded to indicate initial data transformation required for a particular statistical technique. Both linear and nonlinear transformations were considered for this question. The type of transformation and the page where it was reported were specified.

Question J. Question J was coded to indicate whether or not data analysis methods were reported in an article. "Report" was a key word in the coding of this variable. An article was coded "Yes" only if a specific computer program or other analysis procedure was named. The data analysis method and the page where it was reported were specified on the code sheet.

Determination of Reliability

The criterion level for reliability of coding between the principal investigator and an independent coder for each variable was established at a level of 80% or better than chance agreement (Krippendorff, 1980). The reliability check was performed on a proportional random sample of 10% of all articles analyzed. The articles checked for reliability are indicated in Appendix A with an asterisk.

Agreement reliability for categorical variables derived from Questions A, B1, and D through J, were computed using procedures for content analysis outlined by Krippendorff (1980). Reliability for ratio level data originating with Questions B, C, C1, C2, and C2a, were computed by procedures prescribed by Winer (1971). Appendix C gives a computational example of each of these procedures.

Obtained reliability coefficients for all variables are given in Table 2. The interjudge reliability coefficients ranged from .81 for Question F for justification of

Table 2

Reliability Coefficients for Coding Variables*

Variable	Reliability Coefficient
A. Subspecialty Focus of Article	.90 (K)
B. Number of Subjects	.87 (W)
B1. Design	.92 (K)
C. Number of Variables	.87 (W)
C1. Number of Dependent Variables	.84 (W)
C2. Number of Variables Entered into a Single Statistical Analysis	1.00 (W)
C2a. Number of Dependent Variables Entered into a Single Analysis	1.00 (W)
D. Statistical Analyses Used	.91 (K)
E. Use of Significance Testing	1.00 (K)
E1. Reporting of Significance	1.00 (K)
E2. Reporting of Alpha	.94 (K)
E3. Reporting of Exact *p* Values	1.00 (K)
F. Justification for Statistical Analysis	.81 (K)
G. Citation of Statistical Reference	.84 (K)
H. Testing of Statistical Assumption	1.00 (K)
I. Reporting of Data Transformation	.84 (K)
J. Reporting of Data Analysis Method	1.00 (K)

* Calculated according to Krippendorff procedure = K
Calculated according to Winer procedure = W

statistical analyses to 1.00 for Question C2, C2a, E, E1, E3, H, and J. All coefficients exceeded the criterion level set for reliability.

Statistical Analyses

Data were entered from the code sheets into a computer data file. These data were subsequently submitted to a series of analyses using the "PROC FREQ", "PROC UNIVARIATE", and "PROC CHART" programs of the Statistical Analysis System (SAS, 1982). These programs yielded frequency counts, percentages, means, and standard deviations necessary to answer the research questions based upon sample data. Computing was carried out at the Academic Computer Center of the University of North Carolina at Greensboro.

Population Proportions and Confidence Intervals

Population proportions were estimated for research questions 1a, 1c, 2, 3a, 4, 5, 9, 10, 11, 12, and 13. Unbiased estimators of population proportions were calculated by:

$$p_{st} = \frac{\sum_{k=1}^{k} \sum_{i=1}^{n_k} y_{ik}}{n}$$

where $y_{ik} = 1$ or 0 and denoted the value of the sampling variable for the *i*th element sampled from the *k*th stratum and n is the overall sample size (Jaeger, 1984, p. 82).

A 95% confidence interval for each estimated population proportion was calculated by:

$$\text{C.I.}_{.95\,p_{st}} = p_{st} \pm 1.96 \sqrt{v\,(p_{st})}$$

where $v\,(p_{st})$ was an unbiased estimator of the variance of the proportion. This value was computed by:

$$v\,(p_{st}) = \frac{(1 - f)}{n\,N} \cdot \sum_{k=1}^{k} \frac{N_k^2\, p_k\,(1 - p_k)}{(N_k - 1)}$$

where N denoted the overall population size, n denoted the overall sample size, N_k represented the population size for stratum k, f was the sampling fraction, and p_k was the sample proportion for stratum k (Jaeger, 1984, p. 83).

Rank Order

Research questions 1b, 1b-1, 1c, and 1c-1 required determination of rank in the use of particular statistical analyses. Rankings were based upon estimated population proportions from high to low, e.g., a statistical analysis which had the highest *p* value received a rank of 1. Tied ranks shared the value of the ranks proportionally.

Concordance in Rankings

Research question 1b-1(a) asked for the extent of concordance between pairs of rankings. Kendall's tau was calculated by procedures described in Daniels (1978) for all possible pairs of rankings for the subspecialty categories. Significance of associations in rankings were tested with alpha set at the .05 level using a one-tailed positive test. A .05 level was selected because of the nature of the data, i.e., written reports that remained constant from reading to reading.

Population Means

Research Questions 5, 6, and 7 required estimation of a population mean. With proportional allocation and simple random sampling from each journal, the arithmetic average of the variable under investigation for all sampled articles served as an unbiased estimator of the population mean (Jaeger, 1984).

Summary

Methods described in this chapter included the procedures for sampling, data collection, and data analysis. Specific formulas were cited for calculations used in sampling and analysis. Details were given to clarify the coding decisions made by the researcher in analyzing the research reports.

CHAPTER IV

FINDINGS

This chapter presents the findings derived from content analyzing 233 research articles sampled from seven different physical education research journals. Each report was analyzed in order to answer research questions concerning: (a) the use of particular statistical techniques, (b) the nature of the quantitative data that were analyzed, and (c) selected characteristics of the written descriptions reporting the statistical analyses employed. Appendix D indicates the definitions of abbreviations used in tables presented in this chapter.

Statistical Techniques Employed in Physical Education

Type and Frequency of Statistical Techniques

A total of 87 different statistical techniques were used in the 233 research reports. Many of the techniques, such as Pearson's product moment correlation coefficient and the analysis of variance, were well-known. Others, such as the Behrens-Fisher statistic and the Pearson-Filon test, were comparatively unknown. Reported frequencies of observed use in the sample of papers studied ranged from 183 uses of the arithmetic mean to a single reported use of 26 different statistical techniques. Table 3 lists the 87 observed

Table 3

Frequencies of Statistical Techniques Used in 233 Selected Physical Education Research Reports

Statistical Technique	Frequency of Use in Sample	Percent of Sample Reports Using Technique
Mean	183	78.5%
Standard Deviation	112	48.1%
Pearson's Product Moment Correlation	80	34.3%
Proportion	70	30.0%
Range	42	18.0%
Two-Way ANOVA	39	16.7%
One-Way ANOVA	38	16.3%
Independent t-Test	26	11.2%
Two-Way Repeated Measures ANOVA	25	10.7%
Three-Way Repeated Measures ANOVA	24	10.3%
Discriminant Function Analysis	24	10.3%
Newman-Keuls Multiple Comparison	20	8.6%
Multiple Regression	19	8.2%
Dependent t-Test	16	6.9%
Scheffé Multiple Comparison	16	6.9%
One-Way Chi Square Analysis	15	6.4%
Two-Way Chi Square Analysis	15	6.4%
Principal Components Factor Analysis	15	6.4%
Two-Way MANOVA	14	6.0%
Median	14	6.0%
Standard Error of the Mean	14	6.0%
Tukey HSD Test	12	5.2%
Three-Way ANOVA	11	4.7%
Variance	9	3.9%
Two-Way ANCOVA	8	3.4%
Linear Regression	8	3.4%
One-Way Repeated Measures ANOVA	7	3.0%
Duncan New Multiple Range Test	7	3.0%
One-Way MANOVA	7	3.0%
Three-Way MANOVA	6	2.6%
One-Way ANCOVA	5	2.1%
Partial Correlation	5	2.1%
Canonical Correlation	4	1.7%
Cronbach's Coefficient Alpha	4	1.7%
Hotelling's T-Square	4	1.7%

Table 3 (continued)

Statistical Technique	Frequency of Use in Sample	Percent of Sample Reports Using Technique
Path Analysis	4	1.7%
Spearman Rank Order Correlation	4	1.7%
Trend Analysis	4	1.7%
z-Test	4	1.7%
Three-Way ANCOVA	3	1.3%
Biserial Correlation	3	1.3%
Goodman-Kruskal's Gamma	3	1.3%
Intraclass Correlation Coefficient	3	1.3%
Kendall's Tau Beta	3	1.3%
Kruskal-Wallis ANOVA by Ranks	3	1.3%
Mode	3	1.3%
Alpha Factor Analysis	2	0.9%
Bonferroni t Multiple Comparison	2	0.9%
Canonical Factor Analysis	2	0.9%
Cluster Analysis	2	0.9%
Difference in Proportions Test	2	0.9%
Eta-Squared Correlation	2	0.9%
Fisher's LSD Multiple Comparison	2	0.9%
Image Factor Analysis	2	0.9%
Kendall's Coefficient of Concordance	2	0.9%
Multiple Classification Analysis	2	0.9%
Three-Way MANCOVA	2	0.9%
Maximum Likelihood Factor Analysis	2	0.9%
Omega Squared	2	0.9%
Scored-Interval Agreement Method	2	0.9%
Tukey's Omega Multiple Comparison	2	0.9%
One-Way Repeated Measures ANCOVA	1	0.4%
Two-Way Repeated Measures ANCOVA	1	0.4%
Three-Way Repeated Measures ANCOVA	1	0.4%
Behrens-Fisher Statistic	1	0.4%
Confusion Matrix	1	0.4%
Coefficient of Variation	1	0.4%
Duncan-Bonner Multiple Comparison	1	0.4%
Generalizability Coefficient	1	0.4%
Kendall's tau C	1	0.4%
Kuder-Richardson 20	1	0.4%
Likelihood Goodness-of-Fit Test	1	0.4%
Log Linear Analysis	1	0.4%
One-Way MANCOVA	1	0.4%
Two-Way MANCOVA	1	0.4%
One-Way Repeated Measures MANOVA	1	0.4%

Table 3 (continued)

Statistical Technique	Frequency of Use in Sample	Percent of Sample Reports Using Technique
Three-Way Repeated Measures MANOVA	1	0.4%
Mann-Whitney U Test	1	0.4%
Pearson-Filon Test	1	0.4%
Profile Analysis	1	0.4%
Point Biserial Correlation	1	0.4%
R-Type Factor Analysis	1	0.4%
Tetrachoric Correlation	1	0.4%
Tukey's Alpha Procedure	1	0.4%
Tukey's Beta Procedure	1	0.4%
Tukey's WSD Multiple Comparison	1	0.4%
Wilcoxon t-Test	1	0.4%

statistical techniques ordered by frequency of use in the sample of papers.

Most common within the research reports were the descriptive techniques of mean, standard deviation, Pearson's product moment correlation coefficient, proportions, and range. The arithmetic mean was used in a vast majority of the published research reports, i.e., 78.5% of the sampled articles; standard deviation was used in nearly half of the reports, i.e., 48.1% of the articles.

Only 11 statistical techniques were used in at least 10% of the sampled articles. Eleven more analytic techniques were identified when one considered 5% of all articles, i.e., at least 12 of the 233 articles. The remaining 65 statistical techniques were employed in fewer than 5% of the sampled research reports. The lesser-used statistics included several relatively new and more complex procedures such as log linear analysis and three-way multiple analysis of covariance. Surprisingly, the list of lesser-used statistics included several relatively familiar and simple statistical procedures such as computation of the mode and linear regression. A number of nonparametric techniques, e.g., Goodman-Kruskal's gamma; multivariate techniques, e.g., Hotelling's T-square; correlational techniques, e.g., tetrachoric correlation; and multiple comparison tests, e.g., Bonferroni's $\underline{t}$, were also used infrequently within the sampled research articles.

Because the research reports were sampled using proportional allocation with simple random sampling from each journal, estimators of population proportions for use of statistics were computationally equivalent to the observed proportions for the total sample. Ninety-five percent confidence intervals were constructed around estimated population proportions. These confidence intervals vary as a function of the observed proportion of use for each technique within each journal. Table 4 presents the journal proportions and the resultant confidence intervals for the 26 statistical techniques for which the confidence interval included or exceeded the 5% usage level.

Statistic use by subspecialties. Table 5 presents the results of a crossbreak analysis of the 26 most-frequently-used statistics by seven physical education subspecialties. Within parentheses in Table 5 are the ordered ranks of statistics, used by each subspecialty. Examination of Table 5 reveals a number of zero frequencies. This occurred for subspecialties for which a large number of articles were sampled as well as those with fewer articles sampled.

Comparison across subspecialties demonstrates few similarities in the rank ordering of the 26 statistics. Exceptions are the mean and standard deviation which share top rankings across all subspecialties, and variance and two-way analysis of covariance which are consistently the

Table 4

Sample Proportions of Statistics Used by Journal, and

95% Confidence Intervals on Overall Population Proportions

STAT	RQES	JOSP	JOSB	ROSL	JTPE	JSSI	DNRJ	95% Confidence Interval
MEAN	.92	.76	.63	.40	.62	.40	.50	.747 to .823
SDEV	.61	.41	.40	.07	.31	.20	.50	.434 to .528
PPMC	.37	.41	.23	.20	.38	.20	.50	.296 to .390
PROP	.16	.33	.43	.67	.54	1.00	.50	.258 to .342
RANG	.18	.15	.17	.20	.31	.00	.00	.142 to .218
ANV2	.17	.22	.13	.07	.15	.00	.50	.130 to .204
ANV1	.14	.22	.27	.00	.23	.00	.00	.127 to .199
INTT	.09	.13	.13	.00	.23	.20	.50	.081 to .143
AVR2	.19	.02	.00	.00	.00	.00	.00	.078 to .136
AVR3	.15	.11	.03	.00	.00	.00	.00	.073 to .133
DFNA	.07	.17	.13	.00	.23	.00	.00	.074 to .132
MREG	.07	.15	.07	.13	.00	.00	.00	.055 to .109
NKPH	.11	.13	.03	.00	.00	.00	.00	.058 to .114
DETT	.09	.04	.07	.00	.08	.00	.00	.044 to .094
SCHE	.11	.00	.10	.00	.00	.00	.00	.044 to .094
PCFA	.05	.11	.07	.07	.08	.00	.00	.039 to .089
CSQ1	.04	.09	.07	.13	.08	.20	.00	.040 to .088
CSQ2	.04	.09	.17	.07	.00	.00	.00	.041 to .087
MEDI	.03	.11	.07	.20	.00	.00	.00	.037 to .083
SERM	.11	.00	.00	.00	.00	.20	.00	.037 to .083
MAN2	.04	.13	.07	.00	.00	.00	.00	.038 to .082
THSD	.05	.07	.03	.07	.00	.00	.00	.031 to .073
ANV3	.06	.11	.00	.00	.00	.00	.00	.025 to .069
VARI	.03	.07	.03	.00	.08	.00	.00	.020 to .058
LREG	.07	.00	.00	.00	.00	.00	.00	.016 to .052
ACV2	.02	.09	.03	.00	.00	.00	.00	.017 to .051

Table 5

Frequencies and (Ranks) of Statistical Techniques Used,

by Physical Education Subspecialties

STAT	Func Effects	Socio & Beh	M Lng & Dev	M & M Analy	Mgmt T & P	Prgm Dev	Meas & Eval
MEAN	35(1)	65(1)	29(1)	16(1)	8(1)	23(1)	7(2)
SDEV	27(2)	37(3)	14(2)	11(2)	6(2.5)	12(2)	5(3)
PPMC	18(3)	27(4)	10(3.5)	6(4)	2(8)	9(4)	8(1)
PROP	3(13.5)	39(2)	6(6.5)	2(7.5)	6(2.5)	11(3)	3(5)
RANG	4(11)	15(8)	6(6.5)	8(3)	1(11.5)	6(7)	2(10)
ANV2	4(11)	16(6.5)	9(5)	1(11.5)	3(4.5)	6(7)	0(23)
ANV1	5(9)	17(5)	4(11.5)	2(7.5)	0(19.5)	8(5)	2(10)
INTT	7(6.5)	9(12)	2(16.5)	0(20.5)	0(19.5)	6(7)	2(10)
AVR2	11(5)	1(24.5)	5(9)	4(5)	0(19.5)	2(15.5)	2(10)
AVR3	2(16.5)	8(14.5)	10(3.5)	0(20.5)	0(19.5)	2(15.5)	2(10)
DFNA	0(23)	16(6.5)	1(20)	0(20.5)	1(11.5)	4(9.5)	2(10)
NKPH	4(11)	8(14.5)	5(9)	1(11.5)	0(19.5)	1(19.5)	1(16.5)
MREG	3(13.5)	12(9)	1(20)	0(20.5)	2(8)	0(23.5)	1(16.5)
DETT	6(8)	4(20)	2(16.5)	1(11.5)	0(19.5)	2(15.5)	1(16.5)
SCHE	0(23)	5(18.5)	4(11.5)	3(6)	0(19.5)	3(11.5)	1(16.5)
CSQ1	0(23)	10(10)	0(24.5)	0(20.5)	3(4.5)	2(15.5)	0(23)
CSQ2	0(23)	9(12)	1(20)	0(20.5)	2(8)	1(19.5)	2(10)
PCFA	0(23)	7(16.5)	0(24.5)	1(11.5)	0(19.5)	4(9.5)	3(5)
MAN2	1(19)	9(12)	1(20)	0(20.5)	0(19.5)	3(11.5)	0(23)
MEDI	0(23)	7(16.5)	3(14)	1(11.5)	2(8)	0(23.5)	1(16.5)
SERM	12(4)	1(24.5)	0(24.5)	1(11.5)	0(19.5)	0(23.5)	0(23)
THSD	2(16.5)	3(21.5)	3(14)	0(20.5)	2(8)	2(15.5)	0(23)
ANV3	0(23)	3(21.5)	5(9)	0(20.5)	0(19.5)	0(23.5)	3(5)
VARI	2(16.5)	1(24.5)	3(14)	0(20.5)	0(19.5)	2(15.5)	1(16.5)
ACV2	2(16.5)	5(18.5)	1(20)	0(20.5)	0(19.5)	0(23.5)	0(23)
LREG	7(6.5)	1(24.5)	0(24.5)	0(20.5)	0(19.5)	0(23.5)	0(23)

lowest ranked. None of the other statistics is consistent in rank across all seven subspecialties. For example, proportions were a popularly used statistical technique among the studies classified as Sociocultural & Behavioral Aspects and Program Development, i.e., 2nd and 3rd in rank, respectively. But neither statistical technique was used as often by researchers from the other subspecialties. Another inconsistently used statistic was the standard error of the mean. It ranked fourth in use among the Functional Effects of Physical Activity group of studies, but it was observed only once or not at all within the other subspecialties.

Kendall's tau values were calculated to quantify the extent of concordance in the rankings of statistic use for each pair of the seven subspecialty rankings. Table 6 presents the complete matrix of Kendall's tau values. The lowest observed tau value was +.07 between Functional Effects of Physical Activity and Management Theory & Practice. The highest value was +.62 for the association between use of statistics in papers classified from the subspecialties Sociocultural & Behavioral Aspects and Program Development. Generally, the Kendall's tau values were in the +.30 to +.50 range.

Kendall's tau can be interpreted as an inferential statistic to test for significance of association between

Table 6

Matrix of Kendall's tau Associations, by Physical Education Subspecialties

	Func Effets	Socio & Beh	M Lng & Dev	M & M Analy	Mgmt T & P	Prgm Dev	Meas & Eval
Func Effects		.15	.35*	.44*	.07	.30*	.15
Socio & Beh	.15		.33*	.30*	.57*	.62*	.39*
M Lng & Dev	.35*	.33*		.52*	.28*	.44*	.47*
M & M Analy	.44*	.30*	.52*		.25*	.49*	.43*
Mgmt T & P	.07	.57*	.28*	.25*		.37*	.21
Prgm Dev	.30*	.62*	.44*	.49*	.37*		.47*
Meas & Eval	.15	.39*	.47*	.43*	.21	.47*	

*Significant at .05 level for one-tailed test (critical value = +.237) two sets of rankings. Setting alpha at the .05 level and performing a one-tailed positive test of significance, four associations failed to meet the criterion level for significance. The ranking of statistics use for Functional Effects of Physical Activity was not significantly associated with the rankings for Sociocultural & Behavioral Aspects, Management Theory & Practice, or Measurement & Evaluation. Neither was there significant agreement in the rankings for Management Theory & Practice and Measurement & Evaluation. The significant agreements between rankings for the various subdisciplines are indicated by asterisks in Table 6.

Classification of statistics by similar purpose. Consistent with the findings for use of individual statistical techniques, the analysis of statistical techniques classified by similar purpose in Table 7 also showed that descriptive techniques were most commonly used in the published research. Statistics that describe dispersion and central tendency ranked first and second in frequency of use as a percentage of the total number of analyses performed. Inferential statistical techniques then followed. In decreasing order of use were: ANOVA, Correlation, Multiple Comparison, Regression, t-Test, MANOVA, Association, Factor Analysis, Goodness-of-Fit, ANCOVA, Reliability, and MANCOVA.

Table 8 presents the results of a crossbreak analysis of the same statistical technique categories by seven physical education subspecialties. In contrast to the

Table 7

Frequencies and (Ranks) of Statistical Techniques Used, Classified by Similar Purpose

Category Statistical Technique	Frequency and (Rank) of Use in Sample	Percent of Total Statistics Used*
Dispersion	248 (1)	24.5%
Central Tendency	200 (2)	19.8%
ANOVA	151 (3)	14.9%
Correlation	96 (4)	9.5%
Multiple Comparison	66 (5)	6.5%
Regression	62 (6)	6.1%
t-Test	47 (7)	4.6%
MANOVA	34 (8)	3.4%
Association	26 (9.5)	2.6%
Factor Analysis	26 (9.5)	2.6%
Goodness-of-Fit	21 (11)	2.1%
ANCOVA	19 (12)	1.9%
Reliability	11 (13)	1.1%
MANCOVA	4 (14)	0.4%

*Refers to percentage of the total 1011 statistical analyses employed in the sample of research reports.

Table 8

Frequencies and (Ranks) of Statistical Technique Categories, by Physical Education Subspecialties

Category Statistical Technique	Func Effects	Socio & Beh	M Lng & Dev	M & M Analy	Mgmt T & P	Prgm Dev	Meas & Eval
Dispersion	49(1)	93(1)	29(3)	22(1)	13(1)	31(1)	11(1)
Central Tendency	35(2)	75(2)	32(2)	17(2)	10(2)	23(2)	8(5)
ANOVA	27(3)	48(3)	35(1)	9(3)	5(3)	18(3)	9(3.5)
Correlation	21(4)	34(4)	11(5)	7(4)	3(5.5)	10(4.5)	10(2)
Multiple Comparison	10(6.5)	21(6)	18(4)	5(5)	2(7.5)	8(6)	2(11)
Regression	10(6.5)	33(5)	4(7.5)	0(11.5)	4(4)	6(8)	5(6)
t-Test	13(5)	14(8.5)	6(6)	1(7.5)	0(11.5)	10(4.5)	3(9)
MANOVA	4(8.5)	19(7)	3(9)	0(11.5)	0(11.5)	5(9.5)	3(9)
Association	0(13)	13(10)	1(11)	0(11.5)	2(7.5)	7(7)	3(9)
Factor Analysis	0(13)	10(11.5)	0(13.5)	2(6)	0(11.5)	5(9.5)	9(3.5)
Goodness-of-Fit	0(13)	14(8.5)	1(11)	0(11.5)	3(5.5)	3(11)	0(13)
ANCOVA	4(8.5)	10(11.5)	4(7.5)	0(11.5)	0(11.5)	1(13.5)	0(13)
Reliability	1(10.5)	2(13.5)	1(11)	1(7.5)	0(11.5)	2(12)	4(7)
MANCOVA	1(10.5)	2(13.5)	0(13.5)	0(11.5)	0(11.5)	1(13.5)	0(13)
Total Stats Used	175	388	145	64	42	130	67
No. Reports Sampled	35	95	32	18	12	28	13
Mean Stats per Report	5.0	4.1	4.5	3.6	3.5	4.6	5.2

subspecialty rankings for individual statistical techniques, the category rankings show greater agreement. Nonetheless, there are clear differences across some of the categories. For example, Factor Analysis ranks high for studies in Measurement & Evaluation and Mechanical & Muscular Analysis of Motor Skills, yet the technique ranks near the bottom for the other subspecialties.

Analytic Complexity

The analysis of the complexity of statistical techniques called for classification of each report as 1-variable, 2-variable, multiple-variable, or multivariate according to the most complex analysis employed within the report. It was found that the most complex analysis used in 21 reports, representing 9.0% of the total sample, was a 1-variable statistical technique such as the mean, standard deviation, or proportions. A 2-variable statistical technique such as Pearson's product moment correlation coefficient or a t-test, was employed as the most complex analysis in 60, or 25.8%, of the reports. Ninety-three of the reports, i.e., 39.9%, used a multiple-variable procedure such as ANOVA or multiple regression as the most complex analysis. Multivariate techniques such as MANOVA or canonical correlation were reported in 59 of the 233 research reports. This represented 25.3% of the total sample.

Use of multivariate statistics. Table 9 presents the overall and subspecialty frequencies of use of multivariate statistical techniques. Twenty-one different multivariate procedures were used, contributing 9.4% to the total number of statistical techniques employed within the sample. The most commonly used multivariate statistical techniques were discriminant function analysis, principal components factor analysis, and two-way multiple analysis of variance. These techniques were used, respectively, in 24, 15, and 14 research reports.

Analysis of use of multivariate statistics within the various subspecialties revealed the greatest use among studies classified as Measurement & Evaluation. Multivariate techniques contributed 22.4% to the total number of statistical techniques observed within the sample for Measurement & Evaluaton. They contributed 13.1% and 12.6%, respectively, to the statistics used within the subspecialties of Program Development and Sociocultural & Behavioral Aspects.

Type of Generalization and Significance Levels

Relatively few of the 233 research reports were purely descriptive in nature. The vast majority, i.e., 87.6%, of the published reports employed one or more inferential statistics to test for the significance of sample findings. And in all but four of the inferential studies, significance was reported. Analysis according to subspecialties revealed

Table 9

Frequencies and (Ranks) of Multivariate Statistical Techniques, by Physical Education Subspecialties

STAT	Func Effects	Socio & Beh	M Lng & Dev	M & M Analy	Mgmt T & P	Prgm Dev	Meas & Eval	Total
MAN1	0	5	0	0	0	2	0	7 (4)
MAN2	1	9	1	0	0	3	0	14 (3)
MAN3	0	3	1	0	0	0	2	6 (5)
MVR1	0	1	0	0	0	0	0	1 (18)
MVR3	1	0	0	0	0	0	0	1 (18)
MCV1	0	0	0	0	0	1	0	1 (18)
MCV2	1	0	0	0	0	0	0	1 (18)
MCV3	0	2	0	0	0	0	0	2 (11)
HTSQ	1	1	1	0	0	0	1	4 (6.5)
CANC	0	1	1	0	0	2	0	4 (6.5)
DFNA	0	16	1	0	1	4	2	24 (1)
MCLA	0	1	1	0	0	0	0	2 (11)
PRAN	1	0	0	0	0	0	0	1 (18)
CONM	0	0	0	0	0	0	1	1 (18)
PCFA	0	7	0.	1	0	4	3	15 (2)
ALFA	0	0	0	0	0	0	2	2 (11)
CAFA	0	0	0	0	0	0	2	2 (11)
IMFA	0	1	0	0	0	0	1	2 (11)
CLFA	0	1	0	0	0	1	0	2 (11)
MLFA	0	0	0	1	0	0	1	2 (11)
RTFA	0	1	0	0	0	0	0	1 (18)
Mvariate Stats Used	5	49	6	2	1	17	15	95
Total Stats Used	175	388	145	64	42	130	67	1011
Percent Mvariate Used	2.9%	12.6%	4.1%	3.1%	2.4%	13.1%	22.4%	9.4%
Rank Mvariate	(7)	(3)	(4)	(5)	(6)	(2)	(1)	
Number Reports Sampled	35	95	32	18	12	28	13	

that the greatest percentage of purely descriptive studies compared to inferential studies, were in the categories of Program Development and Mechanical & Muscular Analysis of Motor Skills. Table 10 summarizes the results of the analysis.

Level of significance. Testing for significance of a statistical result requires determination of an alpha level. In over half of the research reports, or 59.3%, no alpha level was either reported or clearly implied by consistent use in tables. In the articles reporting an alpha level, the majority employed a nominal alpha at the .05 level. Only 11 articles employed a *p* value other than .05. Of interest, however, was the relatively prevalent reporting of exact *p* values. Exact *p*'s were indicated in 15.7% of the 204 inferential studies. Table 11 presents the frequencies of the analyses for the subspecialties. Exact *p* values were reported in Program Development articles most often.

Nature of the Data and Related Statistical Issues

Several analyses were necessary in order to investigate issues that permitted understanding of the nature of the data upon which statistical analyses were performed. These included determining (a) the number of variables studied and analyzed, (b) the number of subjects involved in the data collection, and (c) the extent of assumption testing, use of data transformations, and use of nonparametric statistical techniques.

Table 10

Frequencies of Significance Testing and Significance Reporting, by Physical Education Subspecialties

Subspecialty	Significance Tested	Not Tested	Significance Reported	Not Reported
Program Development	20	8	20	0
Mechanical & Muscular Analysis of Motor Skills	13	5	13	0
Sociocultural & Behavioral Aspects	82	13	80	2
Management Theory & Practice	11	1	12	0
Measurement & Evaluation	12	1	12	0
Functional Effects of Physical Activity	34	1	33	1
Motor Learning & Development	32	0	31	1
Total Sample	204	29	200	4
Percent of Total Sample	87.6%	12.4%	98.0%	2.0%

Table 11

Frequencies of Alpha Levels and p Values Reported, by Physical Education Subspecialties

Subspecialty	.05 Alpha	Other Alpha	Alpha Not Reported	Exact p Values	p Values Not Reported
Management Theory & Practice	2	0	9	2	9
Measurement & Evaluation	2	1	9	1	11
Sociocultural & Behavioral Aspects	24	0	55	13	69
Motor Learning & Development	12	0	20	4	28
Program Development	4	4	12	7	13
Mechanical & Muscular Analysis of Motor Skills	5	2	6	1	12
Functional Effects of Physical Activity	24	0	10	4	30
Total Sample	72	11	121	32	172
Percent of Total Sample	35.3%	5.4%	59.3%	15.7%	84.3%

Number of Variables

The physical education research studies investigated a median number of 10 variables per study. However, there was great variability across studies, as evidenced by the semi-interquartile deviation, 6.25, which is more than half as large as the median number of variables. The tally of the number of variables studied ranged from 2 to 99. Within subspecialties, Functional Effects of Physical Activity research averaged a median number of 13 variables per study. For Motor Learning & Development research a median of 7 variables was identified. Table 12 presents the median number of variables studied for each of the seven subspecialties. Figure 2 presents box-and-whisker plots (Chambers, Cleveland, Kleiner, & Tukey, 1983) for number of variables studied by each of the physical education subspecialties.

Table 12 also indicates the median values for number of dependent variables studied, number of variables simultaneously analyzed in the most complex statistical analysis employed, and the number of dependent variables simultaneously analyzed in the most complex analysis. These values are given for the total sample and for subspecialties. Figures 3, 4, and 5 present box-and-whisker plots for these variables.

For the total sample, it was observed that although a median of 4 dependent variables were studied per published investigation, very few studies simultaneously analyzed

Table 12

Median Number of Variables Studied and Simultaneously Analyzed, by Physical Education Subspecialties

Subspecialty		Total Variables Studied	Dependent Variables Studied	Variables Simultaneously Analyzed	Dep. Variables Simultaneously Analyzed
Prgm Dev		12	9	3	1
Meas & Eval		10	9	7	5
Func Effects		13	4	3	1
Mgmt T & P		8.5	3.5	2.5	1
Socio & Beh		9	3	4	1
M & M Analy		9.5	5.5	2	1
M Lng & Dev		7	3	4	1
Total	Md =	10	4	3	1
Sample	Q =	6.25	3.5	1.5	0

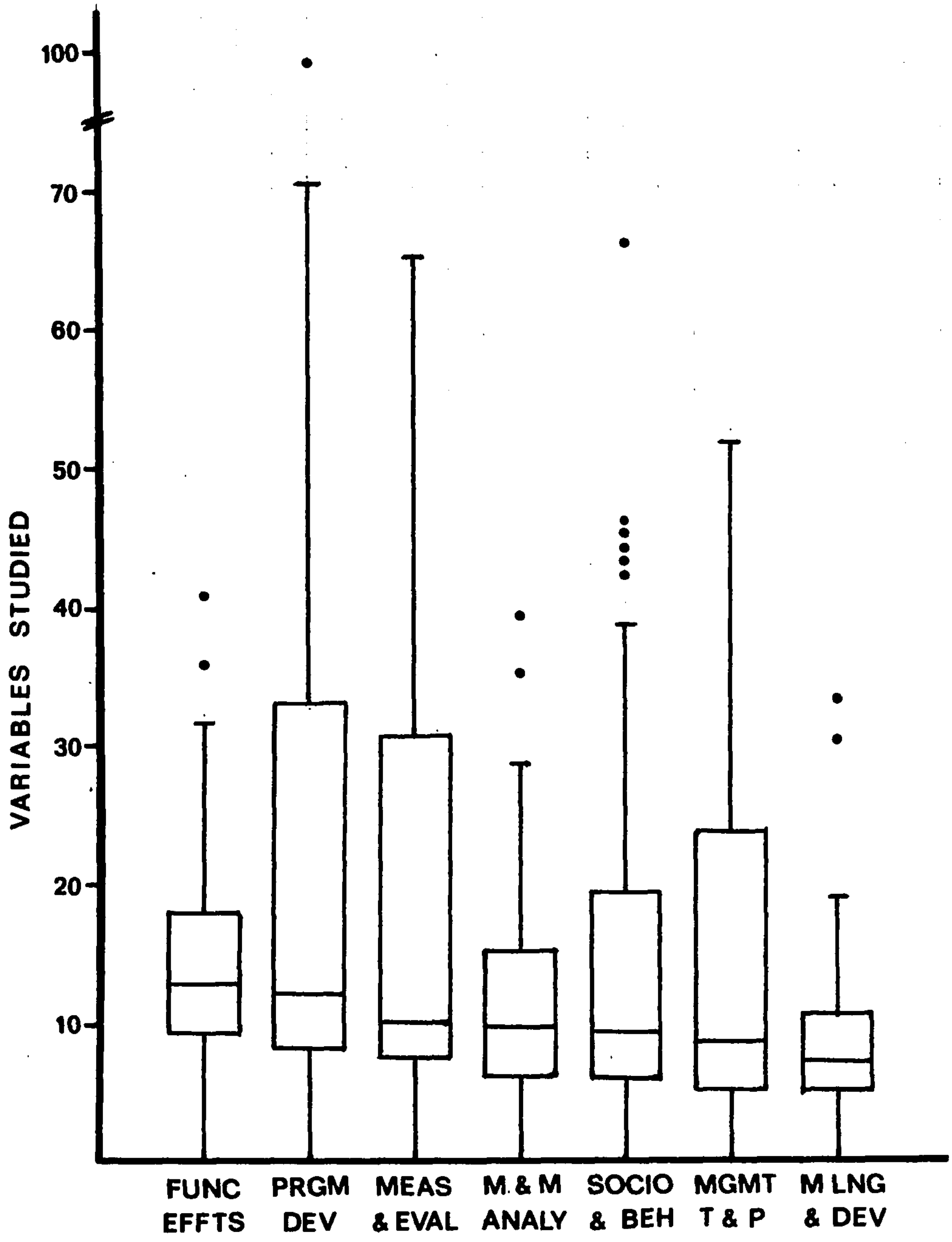

Figure 2. Box-and-whisker plots for total variables studied, by physical education subspecialties.

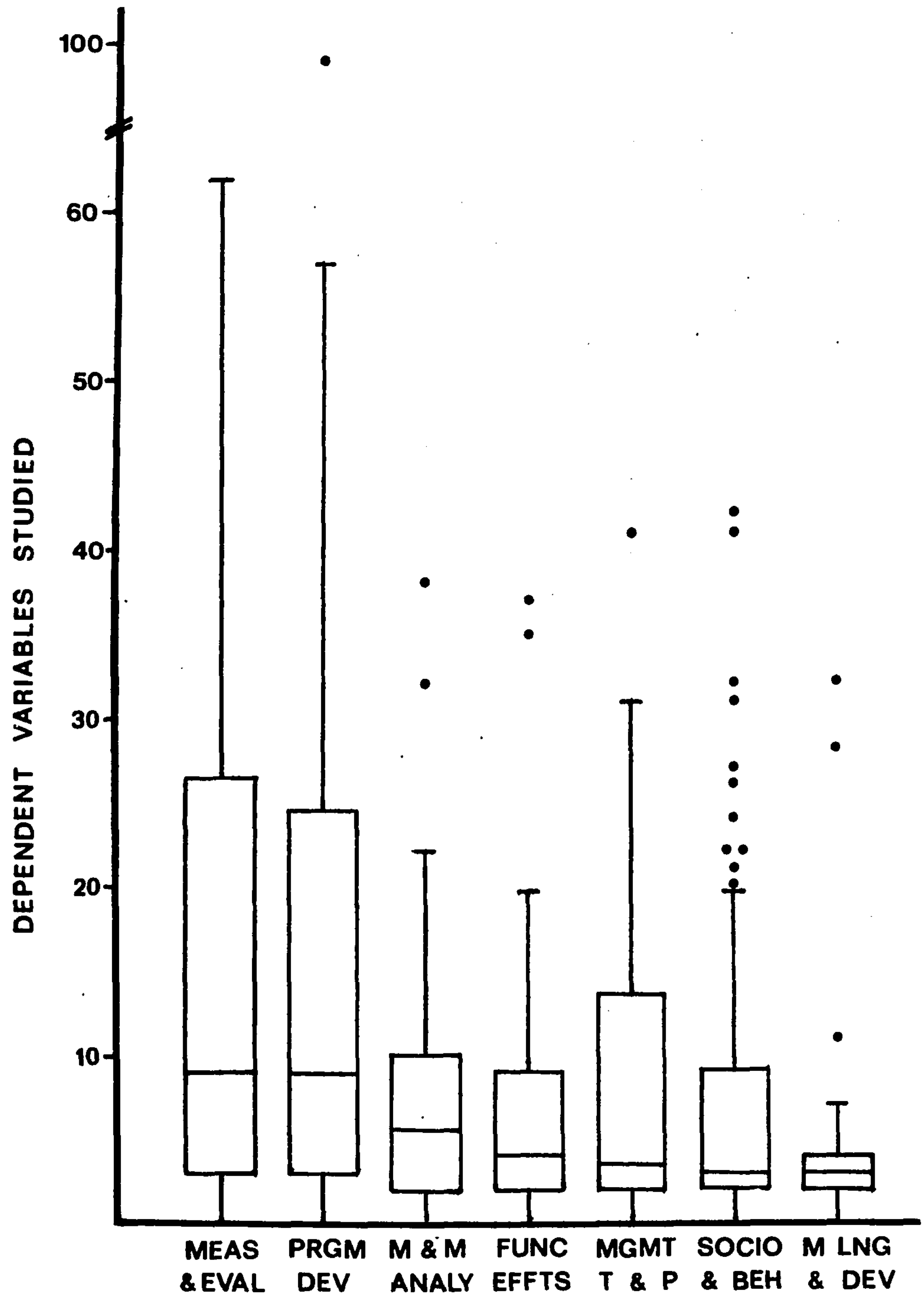

Figure 3. Box-and-whisker plots for dependent variables studied, by physical education subspecialties.

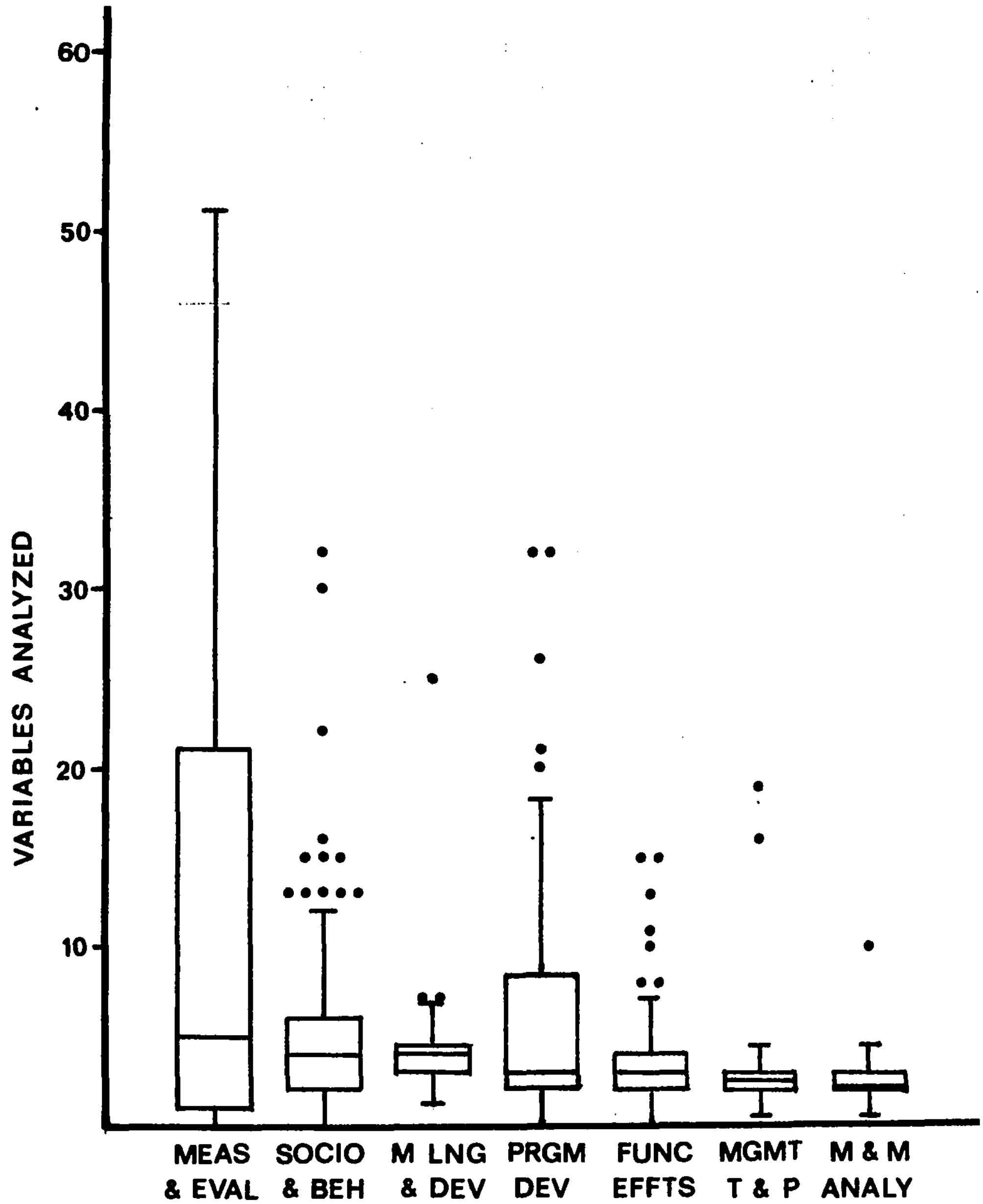

Figure 4. Box-and-whisker plots for variables simultaneously analyzed, by physical education subspecialties.

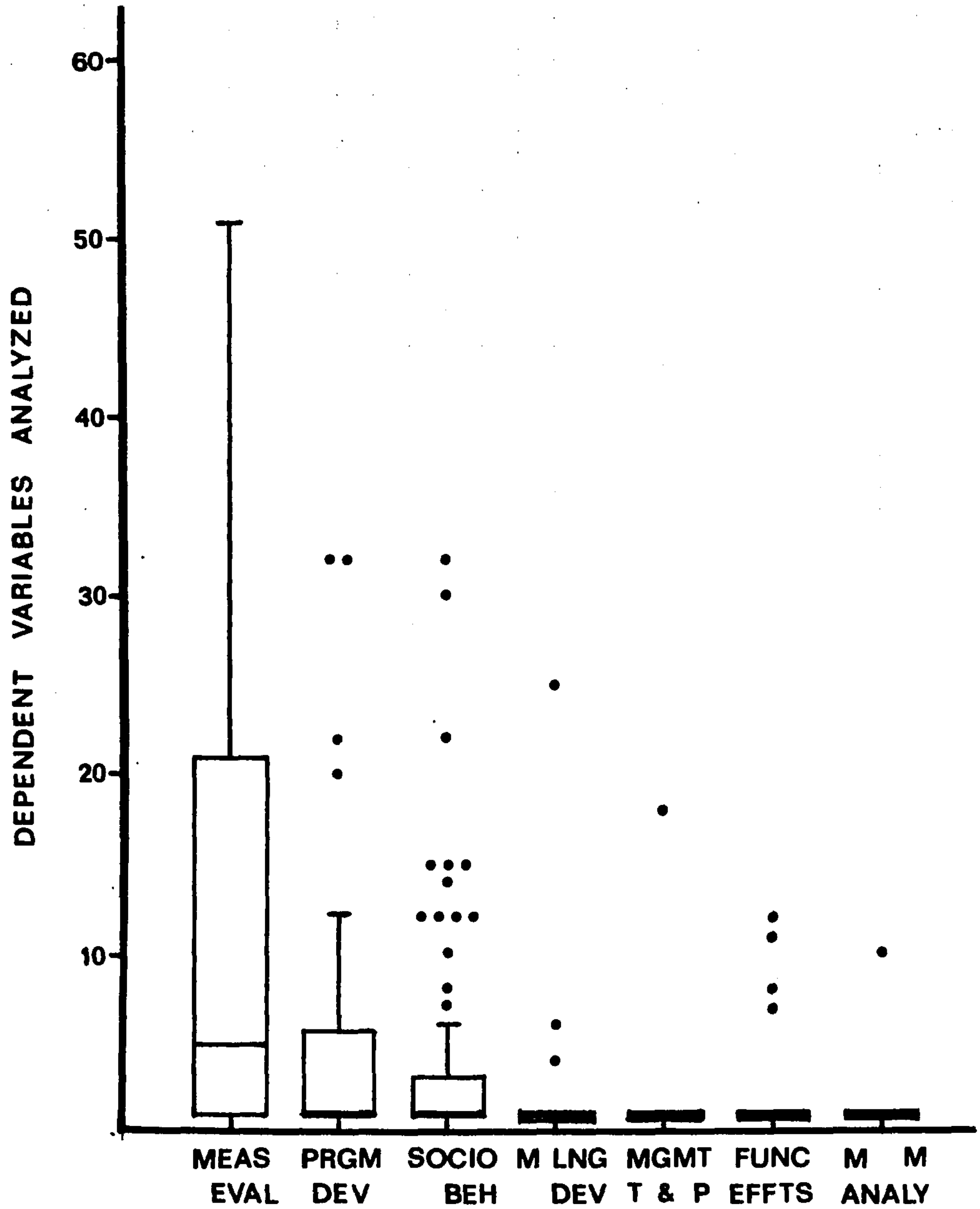

Figure 5. Box-and-whisker plots for dependent variables simultaneously analyzed, by physical education subspecialties.

more than one dependent variable. The semi-interquartile deviation for number of dependent variables simultaneously analyzed was 0. Within the subspecialties, the most striking contrast between the median values for dependent variables studied and simultaneously analyzed was for Program Development research. Within this area, a median 9 dependent variables were studied, but only 1 dependent variable was analyzed. Measurement & Evaluation research simultaneously analyzed a median 5 dependent variables; however, this value was inflated by the large number of factor analyses performed in this subspecialty.

Number of Subjects

Descriptive statistics were calculated for the number of subjects used in the 233 sampled research reports. The median number of subjects was 64, with a semi-interquartile deviation of 65.5. Numbers of subjects ranged from a low of 1 to a high of 9433. The distribution was positively skewed with a value of 8.6. Because of the skewness, the median value of 64 gives a more representative central tendency value for sample size employed in the sampled physical education research than does the mean value.

Table 13 presents the median sample sizes employed by the various physical education subspecialties. The median values varied greatly. The largest median of 272 subjects was for research conducted in Management Theory & Practice.

Table 13

Medians and Semi-Interquartile Deviations for Number of Subjects, by Physical Education Subspecialties

Subspecialty		Subjects Median		Semi-Interquartile Deviation
Management Theory & Practice		272		120.75
Sociocultural & Behavioral Aspects		108		142.5
Program Development		83		37.75
Measurement & Evaluation		76		55.5
Motor Learning & Development		52.5		27.25
Mechanical & Muscular Analysis of Motor Skills		26.5		21.5
Functional Effects of Physical Activity		20		15
Total Sample	Md =	64	Q =	65.5

The smallest medians recorded were in the areas of Mechanical & Muscular Analysis of Motor Skills, i.e., 26.5 subjects, and Functional Effects of Physical Activity, i.e., 20 subjects. Figure 6 presents box-and-whisker plots of sample sizes employed by the physical education subspecialties.

Assessing the appropriateness of a particular sample size depends on several factors, one of which is the basic design of the study. Of particular concern is whether measures were repeated on the same subjects or whether different subjects served in all of the various conditions of the study. Almost exactly half of all the research reports employed either a pure repeated measures or mixed design. The other half of the studies were completely randomized. Two of the subspecialties employed many more repeated measures and/or mixed designs than randomized designs. These were Functional Effects of Physical Activity and Motor Learning & Development. Table 14 presents the frequencies resulting from the analysis.

Statistical Assumptions, Transformations, Nonparametrics

Not all statistical techniques require testing of assumptions. However, it was interesting to find that statistical assumptions were tested in only 16 of the 233 reports analyzed. The number represented 6.9% of the total sample of research articles. Most commonly tested were

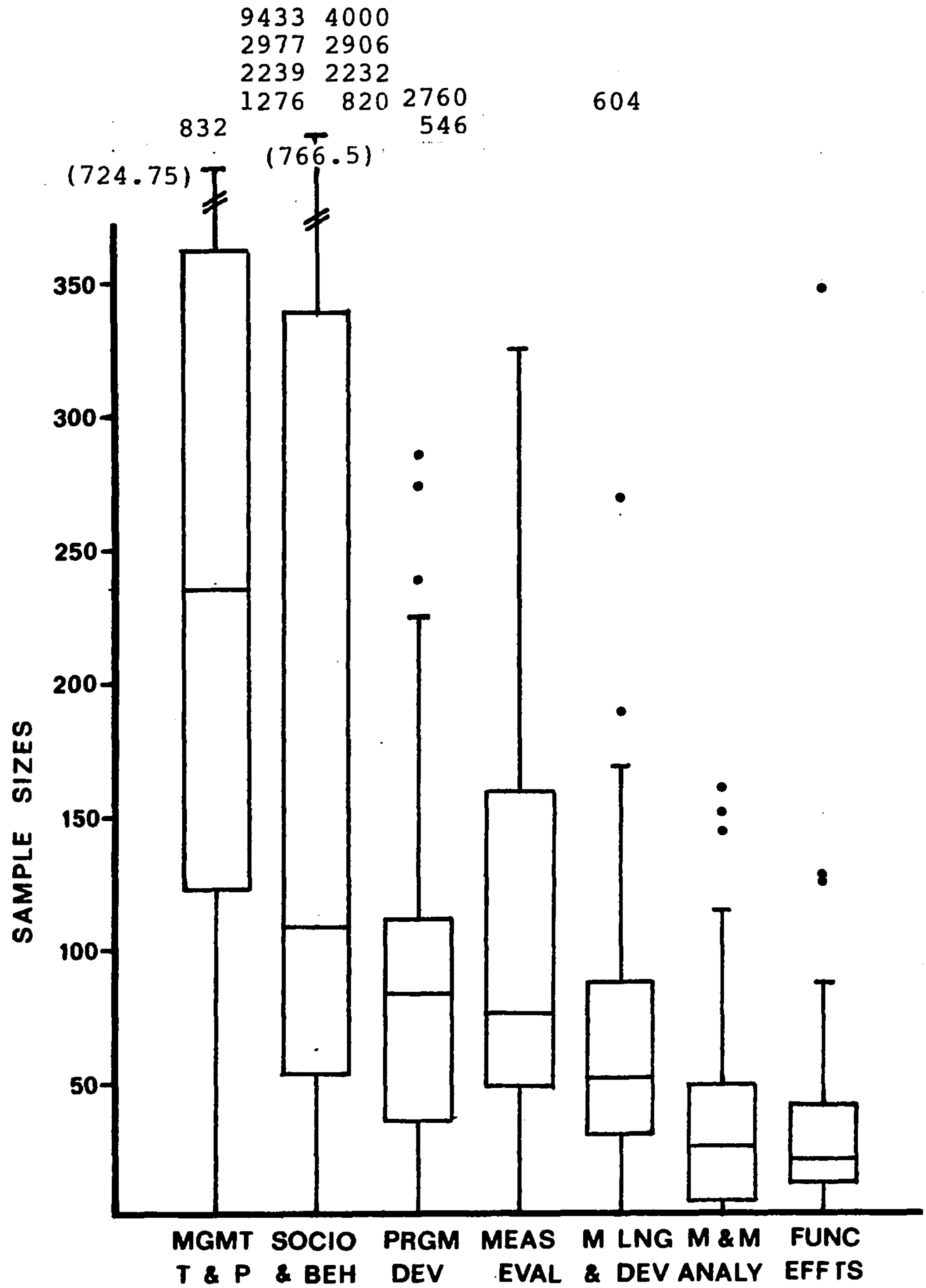

Figure 6. Box-and-whisker plots of sample sizes employed, by physical education subspecialties.

Table 14

Frequencies of Designs Employed, by Physical Education Subspecialties

Subspecialty	Completely Randomized	Within-Subjects	Mixed
Sociocultural & Behavioral Aspects	59	6	30
Management Theory & Practice	11	0	1
Program Development	15	3	10
Measurement & Evaluation	8	1	4
Motor Learning & Development	8	4	20
Functional Effects of Physical Activity	8	15	12
Mechanical & Muscular Analysis of Motor Skills	8	9	1
Total Sample	117	38	78
Percent of Total Sample	50.2%	16.3%	33.5%

assumptions regarding homogeneity of variance and intercorrelations among variables. Also tested were normality of distributions, symmetry of covariance matrices, equality of slopes for analysis of covariance, and expected cell frequencies for chi square analyses. Table 15 shows the distribution of assumption testing across the physical education subspecialties. Assumptions were tested most frequently, in 12.5% of the articles, by Motor Learning & Development researchers.

Transformations. It was observed that 12% of the physical education research studies employed some type of data transformation prior to statistical analysis. One transformation was recorded for each of the subdisciplines of Motor Learning & Development and Program Development, 2 each for Management Theory & Practice and Measurement & Evaluation, 6 each for Functional Effects of Physical Activity and Mechanical & Muscular Analysis of Motor Skills, and 10 for Sociocultural & Behavioral Aspects. It should be noted, however, that the vast majority of these transformations were linear, e.g., transformations to z-scores or ranks, or creation of derived variables from raw scores. Only 3 instances were observed of nonlinear transformations to normalize extremely skewed data, one each for papers classified as Functional Effets of Physical Activity, Mechanical & Muscular Analysis of Motor Skills, and Sociocultural & Behavioral Aspects.

Table 15

Frequencies of Testing for Statistical Assumptions, by Physical Education Subspecialties

Subspecialty	Assumption(s) Tested	Not Tested
Management Theory & Practice	1	34
Motor Learning & Development	4	28
Program Development	3	25
Measurement & Evaluation	1	12
Sociocultural & Behavioral Aspects	5	90
Functional Effects of Physical Activity	1	34
Mechanical & Muscular Analysis of Motor Skills	0	18
Total Sample	16	217
Percent of Total Sample	6.9%	93.1%

Nonparametric statistics. Another analysis of interest concerned the type and frequency of use of nonparametric statistical techniques. Table 16 provides both overall and subspecialty frequency counts for the 10 nonparametric statistical techniques observed in the sample of research reports. Overall, nonparametric statistics contributed a mere 4.6% to the total number of statistical techniques employed and were distributed across 37 of the 233 sampled articles. That is to say, 15.9% of the research studies employed at least one nonparametric statistical analysis. The most common of the nonparametric techniques employed were one-way and two-way chi square analyses.

When examining use among the various subspecialty groupings, it was observed that researchers in Management Theory and Practice employed nonparametric statistics more often than did researchers associated with the other subspecialties. Nonparametric statistics contributed 11.9% to the total number of statistical techniques observed within the sample of Management Theory & Practice research reports. At the other extreme, it was observed that nonparametric statistics were not used at all in the sample of reports from Functional Effects of Physical Activity or Mechanical & Muscular Analysis of Motor Skills.

Table 16

Frequencies and (Ranks) of Nonparametric Statistical Techniques, by Physical Education Subspecialties

Stat	Func Effects	Socio & Beh	M Lng & Dev	M & M Analy	Mgmt T & P	Prgm Dev	Meas & Eval	Total
CSQ1	0	10	0	0	3	2	0	15 (1)
CSQ2	0	9	1	0	2	1	0	13 (2)
SROC	0	2	0	0	0	1	1	4 (3)
GKGM	0	2	0	0	0	1	0	3 (5)
KTAB	0	0	0	0	0	3	0	3 (5)
KWAR	0	2	1	0	0	0	0	3 (5)
KCCW	0	0	0	0	0	2	0	2 (7)
LKGF	0	1	0	0	0	0	0	1 (9)
MWUT	0	0	1	0	0	0	0	1 (9)
WILT	0	0	0	0	0	1	0	1 (9)
Nonpara Stats Used	0	27	3	0	5	11	1	47
Total Stats Used	175	388	145	64	42	130	67	1011
Percent Nonpara Used	0%	7.0%	2.1%	0%	11.9%	8.5%	1.5%	4.6%
Rank Nonpara	(6.5)	(3)	(4)	(6.5)	(1)	(2)	(5)	
Number Reports	35	95	32	18	12	28	13	

Reporting of Statistical Analyses

The final analysis of the published research reports was concerned with the extent to which the researchers "helped" the reader by (a) providing a justification for use of a particular statistical technique, (b) identifying specific data analysis programs, and (c) citing a statistical reference. Findings suggest that these were not very common practices among the writers of physical education research reports.

Table 17 summarizes the data for reporting of statistical analyses for the total sample and the subspecialties. There was one noted exception to the general failure to report explanations, data analysis programs, and statistical references. The articles sampled in Measurement & Evaluation generally did include this information in their reports.

Justifications were often single sentence explanations such as, "Data were evaluated by the chi square test of significance, considered the most appropriate measure, because . . . categories were mutually exclusive and collectively exhaustive and each observation was discrete" (Evans, 1979, p. 3). Occasionally the explanations were much more comprehensive, such as the 16 line rationale offered for use of discriminant function analysis as a post hoc procedure in an article published in the Journal of Sport Psychology (Rejeski, Darracott, & Hutslar, 1979). A good example of

Table 17

Frequencies of Selected Reporting Practices,

by Physical Education Subspecialties

Subspecialty	Explanation Provided	Not Provided	Data Analysis Identified	Not Identified	Reference Cited	Not Cited
Meas & Eval	9	4	4	9	7	6
Mgmt T & P	5	7	1	11	4	8
Prgm Dev	8	20	2	26	9	19
Socio & Beh	27	68	2	93	28	67
M Lng & Dev	9	23	1	31	3	29
Func Effects	3	32	0	35	4	31
M & M Analy	0	18	4	14	1	17
Total Sample	61	172	14	219	56	177
Percent Total Sample	26.2%	73.8%	6.0%	94.0%	24.0%	76.0%

reporting the data analysis method is from McAuley and Gill (1983), ". . . a confirmatory factor analysis was implemented using the Lisrel V program (Joreskog & Sorbom, 1981). This program employs a measurement model to test the hypothetical factor structure" (p. 413). Belka and Williams (1980) provided readers with four current references for canonical correlation analysis. A poor example of statistical referencing was the citing of two 1960 texts to explain the need for a post hoc multiple comparison test after a significant analysis of variance (Crompton, Lamb, & Vedlitz, 1979).

Summary

Analysis of the data revealed that many different statistical techniques were employed in the sample of recently-published physical education research reports. It also revealed that the frequency of use varied according to the subspecialty focus of the reports.

Physical education research was discovered to be complex in terms of the number of variables studied, but relatively less complex in terms of the number of variables simultaneously analyzed. Sample sizes appeared to be no larger than they were two decades ago. The vast majority of published research reports were inferential; and nearly all inferential studies claimed statistical significance using a p value of .05. Few physical education research studies tested for statistical assumptions, performed data transformations, or employed nonparametric analyses.

Research writers generally did not provide readers with "help" in understanding statistical usage. The data analysis revealed limited observations of justifications for selection of a particular statistical technique or citations of statistical references. Very seldom did writers comment on their specific data analysis methods.

CHAPTER V

DISCUSSION

This chapter is organized into two major sections. The first section discusses and interprets major findings of the present investigation and relates them to the outcomes of research and opinions of others. The second section addresses the implications of the findings for the field of physical education.

Findings of the Present Investigation

Statistical Techniques Used in Physical Education Research

Within the sample of research reports, a total of 87 different statistical techniques were used. The analysis presented in Table 3 suggested that (a) many different statistical techniques are being employed by physical education researchers, and (b) that many of the statistical techniques are being used very infrequently. This implies that the nature of physical education investigations must be diverse, thus requiring many different approaches to data analysis. It also implies that *if* a consumer of research needs to know the reported statistical techniques in a study in order to fully understand and appreciate the research, physical educators must have a broad statistics background to read and accurately interpret recently-published physical education research.

One way to consider the need for a working knowledge of statistics is to identify the statistical techniques which a "typical" reader of research might know. Assume that a particular research consumer does know (a) the descriptive statistics of mean, median, mode, standard deviation, variance, range, standard error of the mean, and proportions; (b) the nonparametric inferential statistics of Spearman's rank order correlation, one-sample chi square analysis, and two-sample chi square analysis, (c) the parametric inferential statistics of dependent t-test, independent t-test, one-way ANOVA, two-way ANOVA, and linear regression; and (d) the multiple comparison techniques of the Newman-Keuls test, the Scheffé test, and the Tukey HSD test. Knowledge of the above 20 statistical techniques would permit the reader to understand 72.8% of all of the different analyses employed in the present sample of research reports. While the percentage tends to give the impression that the consumer could successfully read a majority of the published physical education research literature, this may not be the case.

The dispersion of statistical techniques within an article and across several reports also warrants consideration. In the present sample of research, the average number of different statistical techniques used per study was between

4 and 5, i.e., mean = 4.3, range = 1 to 9. Typically, a study reported one or two descriptive statistics, one or two inferential statistics, and a post hoc test of some type. Although the research reader might understand the majority of the statistics used in a study, there is the possibility that he or she would encounter at least one unfamiliar statistic when reading a research report. A consumer who knows only the 20 above-identified statistical techniques could read and interpret only 29.6% of the present sample of research reports, i.e., 69 of the 233 reports, without encountering an unfamiliar statistic. The implications of the "readability" of almost one-third of the published research of the field are startling. How many physical education research consumers have statistical knowledge that permits them to read and interpret only a portion of the research literature? Moreover, how can the purported role of research be fulfilled for the field of study if the research cannot be read and interpreted by members of the discipline and profession?

A somewhat different perspective of this finding is gained when one considers categories of statistical techniques reported rather than individual statistical techniques. Burkhardt (1969) classified statistical techniques employed in a sample of research reports from Research Quarterly into six categories: (a) descriptive statistics, (b) techniques of hypothesis testing of means, variances, and proportions,

(c) correlation and regression, (d) factor analysis, (e) nonparametric tests, and (f) reliability techniques. Although the present study identified 14 classifications, by combining categories it was possible to compare the frequencies observed in Burkhardt's 1962-1966 sample with the findings of the present study. When combined, it was revealed that in every category of statistics there was greater use in the present sample. This suggests that more statistical techniques were used per research report within the past 7 years than were used 20 years ago. Perhaps present-day authors rely on the help of high speed computers to generate a series of statistics that might have taken hours to calculate by hand a couple of decades ago. Another possible explanation is that there is now greater sophistication and competence in statistical knowledge among physical education researchers than in prior years.

Nonparametric statistical analyses. Uses of both nonparametric and multivariate statistical techniques were investigated in the present study. Burkhardt (1969) reported use of chi square analyses in 3.9% of his 382 article sample; "other" nonparametrics were used in 3.6% of his sample. The present sample revealed use of chi square analyses in 10.7% of the 233 articles; "other" nonparametrics were used in 5.2% of the sampled articles. This suggests that there may be greater familiarity and acceptance of nonparametric statistics. However, nonparametric techniques other than chi square analyses still seem to be used in only a limited number of studies. Is the infrequent use primarily a function of unfamiliarity, because nonparametric statistics traditionally receive little attention in graduate research and statistics courses? Is it because most physical education data can appropriately be analyzed by parametric procedures? Or is it possible that there is a prejudice against nonparametric statistics among physical education researchers because of the lesser power of nonparametric techniques?

Multivariate statistical analyses. The only multivariate statistical technique mentioned in Burkhardt's classification system was factor analysis. It is possible, however, that his "other" categories under hypothesis testing or correlation and regression could have included some multivariate statistics. Factor analysis was used in 4.3% of Burkhardt's sample and was used in about 8.2% of the current sample. However, "other" multivariate statistics were employed in 18.9% of the current sample of research reports.

Morrow and Frankiewicz (1979) discovered that many of the articles published in Research Quarterly in 1976 and 1977 considered more than one dependent measure. However, they noted that few of the studies employed appropriate multivariate or repeated measures analyses. A similar conclusion was reached by Schutz, Smoll, and Gessaroli (1983) as a result of a content analysis of four sport and physical activity research journals. Both sets of authors strongly encouraged physical education researchers to increase their use of multivariate and/or repeated measures analyses. Other physical educators have also appealed for the increased use of multivariate and repeated measures analyses (Cox & Serfass, 1981; Gould, 1982; Karpman, 1981; Korell & Safrit, 1977; Levine, 1977). Apparently, such encouragement has been heeded by at least some of the current researchers in physical education. Some type of multivariate statistic was used in approximately 25% of the articles comprising the current sample. One may infer, then, that a reader with no knowledge of multivariate statistics might not be able to understand one out of four published research reports in physical education. It is the opinion of the principal investigator that this is an inordinately high proportion of the research literature to go unread by scholars and professionals because they lack knowledge of multivariate methods.

Repeated measures analyses. As for repeated measures analyses, it is unclear whether or not use has increased. Inasmuch as Burkhardt (1969) did not categorize repeated measures ANOVAs separately, a direct comparison with his research findings is not possible. However, it should be noted that two-way and three-way repeated measures analyses of variance ranked ninth and tenth overall in frequency of use in the current sample of research. Each was employed in approximately 10% of the total number of research reports. It was not the purpose of the present study to determine appropriate use of statistical techniques, yet it was readily discernible to the present investigator that a number of within-subjects designs did not apply repeated measures analytic techniques. Nearly half of all reports employed either a pure within-subjects or mixed design, yet only 61 studies employed repeated measures analyses. Perhaps too many physical education researchers are unaware of the error committed when failing to recognize the need for repeated-measures analyses.

Statistics Use According to Subspecialties

The comparison of frequencies of statistics used in the published research of physical education subspecialties revealed limited agreement in rank ordering. Rankings for Motor Learning & Development, Mechanical & Muscular Analysis of Motor Skills, and Program Development were positively associated with every other subspecialty ranking. However, four of the six Kendall's tau coefficients for agreement between pairs of rankings failed to reach the critical value for statistical significance at the .05 level, thus indicating differences in rankings of use of statistical analyses. Three of the four nonsignificant associations involved Functional Effects of Physical Activity studies. This suggests that further research be conducted to determine if there is something characteristically different about the research in this subspecialty. An inspection of Table 7 revealed (a) less use of proportions, chi square analysis, discriminant function analysis, and factor analysis, and (b) greater use of t-tests, standard error of the mean, and repeated measures analysis of variance by researchers of Functional Effects of Physical Activity. It would appear that researchers of Functional Effects often employ two-group pretest-posttest designs. The ranking association for studies classified as Measurement & Evaluation and Management Theory & Practice was also nonsignificant. Measurement & Evaluation researchers apparently use (a) correlation, t-tests, analyses of variance, and factor analysis more frequently, and (b) proportions and chi square analyses less frequently than do Management researchers.

Nonparametric, multivariate, and repeated measures analyses. The use of nonparametric, multivariate, and repeated measures analyses varied among the subspecialties. Nonparametric statistical techniques were employed most often in the following subspecialties: (a) Management Theory & Practice, (b) Program Development, and (c) Sociocultural & Behavioral Aspects. As expected, these were also the subspecialties that frequently employed proportions and survey research strategies. The reporting of multivariate statistics was most common in Measurement & Evaluation research. This occurred because of the frequent use of factor analyses in the evaluation of test structures. Repeated measures analyses (ANOVR, ANCOVR, and MANOVR) were employed most frequently by researchers of (a) Motor Learning & Development, (b) Functional Effects of Physical Activity, (c) Mechanical & Muscular Analysis of Motor Skills, and (d) Measurement & Evaluation. These were also expected findings when one considers the type of research typically conducted in these specialized areas of physical education. Each one is concerned with research that is "longitudinal" in the sense that the same subjects are typically measured at two or more points in time to discover changes due to maturation, treatment, and the like. Such studies are characteristically different from research that quantitatively compares groups at a single point in time.

There were frequent observations of data from a within-subjects design analyzed as if they were from a randomized design. This was especially problematic for reports categorized as Program Development. Nearly 50% of the articles in this subspecialty used a within-subjects or mixed design, yet only 14% of the articles reported use of a repeated-measures analysis. This may be a chance finding peculiar to the present sample. Or, it may indicate that the Program Development researchers are particularly negligent with regard to repeated measures analyses.

Number of Variables Studied

Physical education researchers design studies that are "complex" in number of variables, i.e., a median of 10 variables per study was found for the present sample. Consistent with the findings of Schutz, Smoll, and Gessaroli (1983), physical education research seems also to be characterized by use of multiple dependent variables, i.e., a median of 4 dependent variables per study. Schutz, Smoll, and Gessaroli (1983) found, however, that only about 40% of the studies that reported multiple dependent variables employed multivariate analyses. Findings from the present investigation that 32% of the studies reporting multiple dependent variables used multivariate analyses are similar to the findings of Schutz, Smoll, and Gessaroli.

It was also found in the present study that typically only 1 of the 4 dependent variables studied was simultaneously analyzed by use of a multivariate technique. Schutz, Smoll, and Gessaroli (1983) lamented the relative neglect of multivariate statistics by many researchers; the present investigator similarly bemoans such a condition. Unless there is logic in conducting research that specifies independent treatment of dependent variables, multivariate data should be analyzed by using multivariate statistical techniques. Failure to do so results in an increase in Type I errors, and/or a loss of information regarding interrelationships among the dependent variables. The potential value of the research is unfortunately reduced for both the producers and consumers of published reports.

Number of Subjects

Several scholars concluded that much physical education research has been conducted with inadequate sample sizes (Baumgartner, 1974; Christensen & Christensen, 1977; Dotson, 1980; Jones & Brewer, 1972; Schutz, 1973). Jones and Brewer (1972) found from study of a sample of reports published in the Research Quarterly between 1969 and 1971 that sample sizes ranged from 3 to 1200. Mean and median sample sizes were not reported. However, it was determined that the value of the median was between 51 and 75. In the present study, sample sizes ranged from 1 to 9433, with a median value of 64 and a semi-interquartile deviation of 65.5. This is certainly comparable to the median sample size found by Jones and Brewer 15 years ago. Current physical education researchers appear to be guilty of continuing to employ inadequate sample sizes. The consequence of using samples that are too small is inadequate statistical power thus making it harder to find statistical significance for a stated alpha level. Again, the worth of the reported research is reduced when researchers ignore the relationship between sample size and statistical power. Ideally, researchers should be employing samples sufficient in size to detect real relationships among variables.

Statistical power should not, however, be the only guide to determination of adequate sample size. Practical significance or "meaningfulness" is even more important. Roundy (1968) recommended that physical education researchers first determine the smallest possible relationship that would be of practical importance, then use appropriate formulas to determine the needed sample size. Tolson (1980) and Thomas and Nelson (1985) added to this idea by recommending that physical education researchers employ the omega-squared statistic to test for meaningfulness of an effect, given the number of subjects tested. In the present sample of 233 reports, only two calculated omega-squared. They were studies classified as Sociocultural & Behavioral Aspects and Measurement & Evaluation.

Another of the factors upon which sample size depends is the basic design of the research study. It was found in the present study that almost half of all the research reports employed either a pure repeated measures or mixed design. The required sample size for these designs is smaller than the size for a comparable randomized design. Consistent with the researcher's expectations, mean sample sizes for subspecialties were inversely related to frequency of use of within-subjects designs.

Type of statistic used also influences the required sample size. In general, multivariate statistics require large sample sizes (Schutz, Smoll, & Gessaroli, 1983) and nonparametric statistics are used in conjunction with smaller sample sizes. If physical education researchers are now increasingly analyzing data by using multivariate techniques and using relatively few nonparametric statistics, then sample sizes should be larger than they were a decade or two ago. Yet this was not observed in the present sample. The nature of the relationships between sample size and use of nonparametric and multivariate statistics for the seven physical eduation subspecialties is unclear from the present investigation.

Significance Testing

Physical education researchers have been criticized for overconcern with statistical significance (Nelson & Hurst 1963; Schutz, 1973). Schutz (1973) accused researchers and journal editors of harboring a "prejudice against the null hypothesis." In a small scale study limited to two issues of the Research Quarterly, he found that 89% of the studies that used significance testing reported significant findings. Schutz worried that studies reporting significance were somehow believed to be more valuable than studies that report nonsignificance. In the present investigation, approximately 88% of the research articles tested for significance, and 98% of them reported at least one of the analyses as statistically significant. This is higher than

Schutz's finding of approximately 12 years ago. It should be noted, however, that in the present investigation no attempt was made to identify which of several analyses conducted within a study was considered to be the "primary" statistical test. This may have inflated the present count of articles reporting significance.

Nelson and Hurst (1963) focused criticism on interpretations of p values. They suggested that physical education researchers seemed to believe that significance found at the .01 level was "better" than significance found at the .05 level. The inference was that they seemingly ignore the rationale behind a researcher's a priori setting of an alpha level which he or she believed to be consistent with the nature of the investigation. Support for Nelson and Hurst's criticism was found in the present study. In 59.3% of the inferential research reports, an alpha level was neither stated nor clearly implied. Among the reported alpha levels, a p value of .05 was predominantly used. But, never was there any rationale given for its selection.

A relatively new phenomenon in the reporting of statistical significance was observed in the present sample of research. Exact p values were reported in 15.7% of the inferential studies. This is considered to be a positive practice. However, this could be a result of how signifcance levels are now recorded in the printouts of computer analyses rather than the researcher's reasoned decision. If such were the case, little is said for the meanings of p values among physical education researchers. Nelson and Hurst's concern may be as valid today as it was in 1963.

Testing of Statistical Assumptions

A researcher who applies an incorrect statistical model for the evaluation of data has committed a "vulgar error" (Slater-Hammel, 1969). Dotson (1980) stated that the majority of Research Quarterly contributors failed to attend to statistical assumptions that underlie particular statistical models. Evidence of assumption testing was found in only 6.9% of the present sample of research articles. Physical education researchers either are not concerned with parametric assumptions or they simply failed to include reports of their tests within their published reports.

Transformations. One of the parametric assumptions that can be tested rather easily is the assumption of normality of distribution. If a distribution is highly skewed, platykurtic or leptokurtic, it can sometimes be normalized by means of a nonlinear transformation. Berenson and Wolf (1977) recommended that more physical education researchers consider use of data transformations or use of appropriate nonparametric techniques when data are not normally distributed. In the present sample of research reports, only 3 instances were discovered in which nonlinear transformations were performed on raw scores for the purpose of normalizing the distribution. Evidently, the practice is not common among physical education researchers.

The result of using an inappropriate statistic is loss of precision in probabilistic conclusions because actual alpha levels differ from nominal alpha levels. While many of the parametric statistics are robust with respect to violations in normality, in some cases an erroneous conclusion could be made.

Reporting of Statistical Analyses

It is considered good practice for researchers to (a) report the rationale for use of particular statistics, (b) cite helpful statistical references, (c) identify data processing procedures, and (d) report any other potentially important information to explain the nature of an effect (American Psychological Association, 1983; Isaac & Michael, 1982; Leedy, 1985; Slater-Hammel,1965a; Teraslinna, 1967). The results of the present investigation revealed limited reporting of this nature. Why? Perhaps the ultimate fault can be ascribed to journal editors rather than to the researchers themselves. Such information is not difficult to report in an article nor is it necessarily space consuming. If journal editors required such information from their contributors, undoubtedly, research writers would include the information in their reports. Continued omission of details concerning statistical analyses limits the meaningfulness of the reported projects for the consumers of physical education research.

Implications for the Field of Physical Education

Improving the Quality of Published Research

It is commonly acknowledged that the published research of a field of study serves as an "index" of scholarship (Crase, 1978). It also serves as a model that novice researchers can emulate (Baumgartner, 1974; Slater-Hammel, 1965b). The results of the present study, although concerned with selected statistical issues, suggest that published research is not an index of good scholarship. Neither does it serve as an exemplary model. Following are some suggestions that may have the potential to improve the quality of physical education research. They are not all-encompassing but, rather, are limited to just a few of the many decisions a researcher must consider in the design and conduct of a study. The ideas were suggested by the findings of the present study.

The first suggestion for physical education research is to take more care in the selection of appropriate statistics. Before using any statistical technique, the researcher must make sure that the data fit the model. If the statistic requires interval level of measurement, independence, normality, homogeneity of variance, etc., one should not cavalierly proceed in analyzing data without investigating these assumptions. This initial effort should be made despite the fact that many parametric statistics are robust with respect to assumption violations.

If the data do not accurately fit the model, then available options should be explored. In some cases, nonnormal distributions can be transformed to approach normality. In other cases, there may be an analogous nonparametric statistic that can be used to answer the statistical question of concern. In still other cases, the researcher may decide to use a parametric statistic but employ a more conservative alpha value for testing. Or simply, conclusions may be stated more tentatively than usual.

Whatever the decision, the results of investigating assumptions and subsequent decision-making should be shared with the reader. In other words, it is the researcher's responsibility to convince the reader that the selected statistical methods are appropriate. Only then can the consumer trust the statistical conclusions resulting from the analyses. As Leedy (1985) posited, ". . . failure to substantiate what one has done with a solid rationale as to _why_ one has done it" may be one of the "weakest links" in the research process (p. 231). Far too many physical education researchers seem to neglect the statistical methods and statistical results sections of the research report.

The citation of a good statistical reference to which the reader could turn for better understanding of an unfamiliar statistic is a minimally acceptable remedy for the present practice.

The selection of an appropriate statistic also calls for understanding of the research design. The present study revealed relatively frequent use of within-subjects designs and multiple-dependent-variable designs. However, there was evidence of only limited use of repeated measures and multivariate analyses. Treating such designs as randomized and univariate is an incorrect practice. This is not to say that studies that have made this "error" should be dismissed as totally invalid. Rather, their conclusions should be taken taken more tentatively. Ideally, the data should be subjected to appropriate re-analyses.

Another suggestion for physical education researchers is to divorce themselves from their "marriage" to statistical significance. The criterion for judging research quality should not be the level at which significance can be reported. An alpha level should be selected in light of the nature of the investigation and the consequences of an erroneous conclusion. It should also be realized that there is nothing magical about the .05 level. For many educational and psychological studies in which control of extraneous variables is very difficult, researchers should consider use of $\underline{p}$ values of .10 or even .20.

Physical education researchers also need to distinguish between statistical significance and practical meaningfulness. The present investigator applauds the approach taken in a recently-published research text (Thomas & Nelson, 1985) in which the authors repeatedly differentiate between the question of reliability of an effect or relationship and the question of strength, i.e., meaningfulness of an effect or relationship. Physical education researchers should habitually calculate the omega squared value as a follow-up to every significant ANOVA finding. Using examples from reports published in the Research Quarterly, Tolson (1980) clearly demonstrated that statistical significance does not guarantee practical meaningfulness. Statistical significance merely means that a similar result is likely to be found in a replication study.

Statistical power analyses of reports published in the Research Quarterly (Christensen & Christensen, 1977; Jones & Brewer, 1972) showed that despite preoccupation with statistical significance many physical education researchers effectively reduce the power of the statistics they employ by using small numbers of subjects. That is to say, they make it harder to detect effects or relationships that may really exist. It is not possible to generalize as to how many subjects are required for different types of studies and designs. However, there are statistical guidelines and aids to help the investigator make this decision. Researchers should be reminded of Roundy's (1968) suggestion that one work backward from the smallest difference or relationship which is meaningful to discover the sample size needed to detect such a difference or relationship. Survey researchers can also refer to King's (1978) published nomogram to determine the size of sample needed for a stated confidence level and acceptable error.

Improving Academic Preparation in Statistics

Perhaps the most clearcut finding of the present investigation is that many different statistical techniques are being used by physical education researchers today. It is naive to believe that one can acquire sufficient statistical competence to read and conduct research within an interest area in the traditional series of graduate level research and statistics courses. The goal of academic preparation in statistics coursework should be to develop an "independent learner" of statistics, i.e., one who can continue to learn new statistical techniques from texts, journal articles, workshops, oral presentations at professional meetings, and the like. In statistics courses, concepts that can be generalized to categories of statistics should be taught. Concepts of appropriate use and interpretation of statistical techniques should take precedence over the "how to do" aspects of the techniques. For example, rather than teaching separate formulas for testing group differences in means, variances, proportions, and correlation coefficients, statistics students should be taught the concept that a test statistic is computed by dividing the difference between the sample statistic and the hypothesized parameter by the standard error of the statistic (Hinkle, Wiersma, & Jurs, 1982).

Cause and effect relationships are always difficult to substantiate in research. Likewise, the relationship between statistics learned in graduate coursework and statistics subsequently used in independent research is unclear. It has been suggested that too many researchers, not just in physical education, tend to ask questions and design studies that are appropriate for the statistical techniques they know and use comfortably. How much better it would be if all researchers let their questions evolve naturally from observations of the world and from previous research and, thereafter, sought to determine the best methodology by which to discover an answer. Researchers should not be tied to a particular research or statistical methodology. Such a notion is consistent with the call for "independent" learners. If one does not know a statistical technique that is needed, then he or she must seek out a way to learn it! And, preferably, the new technique should be learned prior to writing a research proposal that calls for the use of the technique.

Finally, the results of the present study can perhaps provide some valuable guidance in the design of courses for statistical preparation for subdiscipline specialization within the field of physical eduation. The frequency of statistical analyses used within different subspecialties seems to suggest that academic preparation in statistics should differ somewhat across subspecialties. This implies a type of needs assessment approach to curriculum where emphasis is given to statistical methods used most frequently within one's own subspecialty. For example, research consumers and producers in Functional Effects of Physical Activity and Motor Learning & Development must definitely learn repeated measures techniques. A Measurement & Evaluation specialist must know factor analytic techniques. Consumers and producers in Sociocultural & Behavioral Aspects especially need the tools of multivariate analysis of variance and discriminant function analysis. While this is an enormous challenge to graduate curriculum planners, it appears worthy of consideration.

Summary

Findings of the present study were discussed in relationship to pertinent references cited in the review of literature. The use of particular statistical analyses and categories of analyses, variables influencing the numerical complexity of the data, significance, assumption testing, and reporting of statistical analyses were discussed. The latter portion of the chapter presented implications of the present findings for the field of physical education. It was suggested that the quality of published research could be improved, and that effective statistical preparation of physical educators may demand strengthening experiences in present programs of study.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter briefly outlines the methods used in this study and reports the major findings and conclusions in light of the data collected and the broad framing questions specified in Chapter I. In addition, as is customary, recommendations for further study are proposed.

Summary

The present investigation was designed to determine the types and frequencies of statistical analyses reported in recently-published physical education research. It also examined several statistical issues. Most variables were analyzed both for the population of physical education research and for physical education subspecialty research.

Despite obvious interest in research among physical educators and prevalent use of statistical analyses, the review of literature identified few research investigations that focused on statistical usage in physical education. Brady (1968) determined that the statistical techniques reported most frequently in doctoral dissertations in physical education were the mean, standard deviation, Pearson's product moment correlation, and analysis of variance. Burkhardt (1969) reported that 42% of the studies published in the Research Quarterly between 1962 and 1966 employed techniques of hypothesis testing about means, variances, and proportions, 24% used correlation and/or regression, 15% used reliability and/or validity techniques, 13% used descriptive statistics, 5% used nonparametric statistics, and 1% employed factor analysis. More recent studies revealed that physical education researchers frequently designed investigations with multiple dependent measures, but employed appropriate multivariate and/or repeated-measures analyses less frequently (Morrow & Frankiewicz, 1979; Schutz, Smoll, & Gesarolli, 1983).

Physical education researchers have been criticized for their overconcern with statistical significance. Schutz (1973) found that 89% of the research reports in two issues of Research Quarterly reported statistically significant findings. Yet, surprisingly, few physical education researchers employed sample sizes that would enable them to detect anything less than a large effect size (Christensen & Christensen, 1977; Jones & Brewer, 1972).

Physical education researchers have also been criticized for some of their reporting practices. Slater-Hammel (1965a) expressed the concern that researchers seemed to be unfamiliar with the most current statistical references. Teraslinna (1967) believed that the use of sophisticated statistical procedures required more explanation than was usually given. According to the American Psychological Association

(1983), the authors of research reports are responsible for selection of statistical method and presentation of all supporting data. This may include citations of references and formulas for less common statistics. It may also include reporting of statistic values, degrees of freedom, exact probability levels, and other such supporting information.

Research questions were specified for the present investigation to reflect concerns suggested by the review of literature. Using stratified random sampling with proportional allocation from seven different physical education journals, 233 studies were selected and content analyzed. All articles were published between July 1, 1977 and June 30, 1984. Reliability was determined by procedures specified by Krippendorff (1980) and Winer (1971). Classification of articles by subspecialty focus followed the Zeigler (1983) scheme. Descriptive statistics were calculated to summarize findings regarding statistical usage and the related issues of interest for the sample data. Inferential statistics were employed in order to estimate population parameters.

Major Findings and Conclusions

1. What statistical techniques were used in a sample of recently-published physical education research reports?

Eighty-seven different statistical techniques were reported in the sampled research studies. Included were many well-known statistics and also several relatively less-known techniques. It was concluded that a consumer of today's physical education research needs a very broad statistical background.

The estimated population proportion for use of the arithmetic mean, the most commonly reported statistical analysis, was 78.5% with a 95% confidence interval of 74.7 to 82.3%. Other commonly reported statistical techniques and the estimated population proportions were (a) standard deviation, 48.1%, (b) Pearson's product moment correlation, 34.3%, (c) proportion, 30.0%, and (d) range, 18.0%. There were only 13 statistical techniques with confidence intervals that spanned the 10% level of usage, and only 26 techniques with confidence intervals that spanned 5% usage.

Analysis of statistical usage according to subspecialty focus revealed that the mean and standard deviation were employed frequently in the research of all subspecialties, and variance and two-way ANCOVA were used infrequently. However, many differences in use were observed.

In order of estimated use in the population of published physical education research, the 26 top-ranked statistical analyses were: mean, standard deviation, Pearson's product moment correlation coefficient, range, two-way ANOVA, one-way ANOVA, independent t-test, two-way repeated measures ANOVA, three-way repeated measures ANOVA, discriminant function analysis, Newman-Keuls multiple comparison, multiple regression, dependent t-test, Scheffé multiple comparison, principal components factor analysis, one-sample chi square, two-sample chi square, median, standard error of the mean, two-way MANOVA, Tukey's HSD test, three-way ANOVA, variance, linear regression, and two-way ANCOVA. When analyzed by subspecialty, several differences in rank ordering of the 26 most frequently reported statistical analyses were noted. For example, the standard error of the mean ranked fourth in use by reports classified as Functional Effects of Physical Activity but ranked nearly last in use among other subspecialty reports.

Kendall's tau values for the associations in rank order of statistical techniques used among subspecialties ranged from +.62 to +.07. Four Kendall's tau values were found to be nonsignificant at the .05 level for a one-tailed positive test, thus suggesting that some of the subspecialties differed in their use of statistics. It was hypothesized that researchers and consumers from different subspecialties may need different statistical training.

When individual statistical techniques were classified by similar purpose, the following ordering and estimated population proportions of occurrences in physical education research resulted: () dispersion, 24.5%, (b) central tendency, 19.8%, (c) ANOVA, 14.9%, (d) correlation, 9.5%, (e) multiple comparison, 6.5%, (f) regression, 6.1%, (g) t-test, 4.6%, (h) MANOVA, 3.4%, (i) association, 2.6%,

(j) factor analysis, 2.6%, (k) goodness-of-fit, 2.1%, (l) ANCOVA, 1.9%, (m) reliability, 1.1%, and (n) MANCOVA, 0.4%. There were many similarities in usage across the seven subspecialties. However, there were again a few observed differences in statistical usage. For example, the category of factor analysis ranked 3.5 in Measurement & Evaluation reports and ranked 13.5 in Motor Learning & Development reports.

2. What was the estimated population proportion of physical education research reports that used as their most complex analysis a 1-variable statistical technique? ...a 2-variable technique? ...a multiple-variable technique? ...a multivariate technique?

The most complex analysis was a 1-variable technique in 9.0% of the published research reports, a 2-variable technique in 25.8% of the reports, and a multiple-variable technique in 39.9% of the reports. A multivariate technique was employed in 25.3% of the physical education research reports. It was suggested that a consumer who lacked knowledge of multivariate statistical techniques might not be able to read a meaningful portion of the published physical education research.

Twenty-one different multivariate statistics were reported, contributing 9.4% to the total number of statistical techniques reported in the sampled research. Most commonly reported were discriminant function analysis, principal components factor analysis, and two-way MANOVA. Discriminant function analysis was employed in 10.3% of the physical education research reports, factor analysis was used in 6.4% of the reports, and two-way MANOVA was used in 6.0% of the reports.

Multivariate statistics, especially factor analytic techniques, were reported most frequently in Measurement & Evaluation research studies. They were employed least frequently in Functional Effects of Physical Activity research. However, at least one multivariate statistic was reported in every subspecialty category of research reports.

3. What was the estimated population proportion of physical education research reports that used inferential statistical techniques? ...only descriptive techniques?

The vast majority, 87.6% of the published reports, employed one or more inferential statistics to test for significance. The remaining 12.4% of the reports employed only descriptive statistical procedures. Significance was reported in 98.0% of the articles in which an inferential procedure was used.

The greatest use of inferential techniques, in 100% of the sampled reports, was in Motor Learning & Development research. The greatest use of descriptive techniques, in 28.6% of the sampled reports, was in the research classified as Program Development.

4. What was the estimated population proportion of physical education research reports that employed a p value of .05 for tests of significance? ...reported exact p values?

Alpha values were not stated or clearly implied in 59.3% of the reports. In the reports that did report p values, 86.7% employed a nominal alpha of .05. Exact p values were reported in 15.7% of the inferential research studies.

Considering only the inferential research reports that reported a p value, the .05 level was used in 100% of the reports classified as Management Theory & Practice, Socio-cultural & Behavioral Aspets, Motor Learning & Development, and Functional Effects of Physical Activity. The greatest use of exact p values, in 35.0% of the inferential reports, was in Program Development research.

5. What was the estimated population median number of variables studied in physical education research? ...of dependent variables studied?

On the average, physical education research studies investigated 10 variables per study. However, there was considerable variability in the number of variables studied. The median number of dependent variables studied was 4.

The largest median number of variables and dependent variables studied in a subspecialty area, 12 variables and 9 dependent variables, was in Program Development research. The smallest medians, 7 and 3, were found for Motor Learning & Development research.

6. What was the estimated population median for the largest number of variables simultaneously analyzed in a single statistical analysis? ...of dependent variables simultaneously analyzed?

The median number of variables simultaneously analyzed by a single statistical procedure was 3, while the median number of dependent variables simultaneously analyzed was 1. It was concluded that physical education research designs may be more sophisticated than the statistical analyses being applied to them. In terms of the number of variables simultaneously analyzed, Measurement & Evaluation research employed the most complex analyses and Mechanical & Muscular Analysis of Motor Skills employed the least complex analyses.

7. What was the estimated population median number of subjects employed in physical education research?

There was great variability in sample sizes employed in physical eduation research. The median number of subjects was 64, but the semi-interquartile deviation was 65.5. The median number of subjects ranged from 272 for the subspecialty research of Management Theory & Practice to 20 subjects for Functional Effects of Physical Activity research.

8. What was the estimated population proportion of physical education research reports that used a within-subjects or mixed design?

Almost exactly half, 49.8%, of the physical eduation research reports employed designs classified as within-subjects or mixed. The greatest use of within-subjects and/or mixed designs was in the research on Functional Effects of Physical Activity and Motor Learning & Development. It was concluded that consumers and researchers in these subspecialties especially must know repeated measures statistical techniques.

9. What was the estimated population proportion of physical education research that tested for a statistical assumption?

Only 6.9% of physical education research studies reported testing for a statistical assumption. It was concluded that assumption testing is not a common practice among physical education researchers. An assumption was tested in 12.5% of the sampled reports classified as Motor Learning & Development. This was the greatest use observed among the subspecialties.

10. What was the estimated population proportion of physical education research reports that employed data transformation?

Only 12.0% of the published research employed data transformation, and a mere 1.3% performed a nonlinear transformation. Nonlinear data transformation was observed in one research report from each of the subspecialties of Sociocultural & Behavioral Aspects, Functional Effects of Physical Activity, and Mechanical & Muscular Analysis of Motor Skills.

11. What was the estimated population proportion of physical education research reports that used a nonparametric statistical technique?

Nonparametric statistical techniques were employed relatively infrequently, contributing to only 4.6% of the total number of analyses reported. Most commonly used of these techniques were chi square analyses. One or more nonparametric techniques were reported in 15.9% of physical education research reports.

Management Theory & Practice researchers employed nonparametric techniques more frequently, in 11.9% of the total statistical techniques used, than did researchers representing any other subspecialty. The sampled research classified as Functional Effects of Physical Activity and Mechanical & Muscular Analysis of Motor Skills employed no nonparametric techniques.

12. What was the estimated population proportion of physical eduation research reports that provided justification for use of a particular statistical technique? ...reported data analysis methods? ...cited a statistical reference?

Explanations of statistical analysis selection were provided in 26.2% of physical education research reports. Statistical references were cited in 24.0% of the reports. Data analysis methods were specified in 6.0% of the research reports. It was concluded that these reporting practices were not extremely common among physical education researchers.

Measurement & Evaluation research provided justification for statistical analyses in 69.2% of the reports and cited statistical references in 53.8% of the reports. The greatest identification of data analysis was the 22.2% observed for research reports classified as Mechanical & Muscular Analysis of Motor Skills.

Recommendations for Further Study

The principal investigator recommends continued investigation into the nature of statistics usage in physical education research. The following suggestions may be considered for future studies.

1. Sampling procedures should be revised and expanded so that population parameters can be estimated for the physical education subspecialties. The findings of the present investigation suggest that there may be some real differences in the use of statistical techniques by various subspecialty researchers. However, one must be tentative regarding generalizations to the subspecialties from the

present sample. The present sample size of 233 articles was determined to yield precise estimates only for the population of physical education research papers, not for the individual populations of subspecialty research.

2. Different criteria for selection of research journals to be studied might be considered. There is a wealth of research literature that physical educators read and to which they contribute that is not published in the research journals identified for the present sutdy. For example, many exercise physiologists read and contribute research to Medicine and Science in Sports and Exercise; motor learning specialists read and contribute to Perceptual and Motor Skills. Generalizations from the present study must be limited to the research published in the journals included in the sample.

3. Graphical techniques should be included as a category of statistical analysis procedures. For example, graphical techniques such as scattergrams and box-and-whisker plots should be included in future analyses.

4. Replication of the present investigation after passage of a few years might be informative. For example, it could be interesting to compare the use of multivariate statistical techniques in 5 years with the observations in the present study. Several other variables of usage would also be interesting to compare.

5. Although it would be a very challenging research endeavor, the assumption that one needs to know a statistical technique in order to fully appreciate and understand its use in a research study should be tested. What is the nature of the relationship between general and specific knowledge of statistics and comprehension of quantitative research? A similarly challenging research endeavor would be to attempt to determine the relationship between the statistical techniques a researcher knows and the research questions that he or she asks. The answers to the above questions are essential to the content and design of learning experiences intended to establish competence for producing and consuming high quality research.

BIBLIOGRAPHY

American Academy of Physical Education. (1982). Synthesizing and transmitting knowledge: Research and its application. The Academy Papers, No. 16. Reston, VA: Author.

American Alliance for Health, Physical Education, Recreation and Dance. (1982). Physical education gold book: Directory of physical educators in higher education. Champaign, IL: Human Kinetics.

American Psychological Association. (1983). Publication manual (3rd ed.). Washington, DC: Author.

Anderson, S. B., Ball, S., Murphy, R. T., & Associates. (1975). Encyclopedia of educational evaluation. San Francisco: Jossey-Bass.

Arnold, R. K. (1982). Research on sport skills: Is it applicable to teaching and coaching? Motor Skills: Theory into Practice, 6, 93-102.

Baumgartner, T. A. (1969a). A study of intact classes. Research Quarterly, 40, 632-634.

Baumgartner, T. A. (1969b). Critical comments on Rosemier's exaggerated alpha test. Research Quarterly, 40, 863-864.

Baumgartner, T. A. (1974). Remarks concerning sampling used in the Research Quarterly. Research Quarterly, 45, 215-216.

Berelson, B. (1952). Content analysis in communications research. Glencoe, IL: Free Press.

Berenson, M. L., & Wolf, E. H. (1977). Aligned rank tests for randomized block experiments. Research Quarterly for Exercise and Sport, 48, 623-627.

Berlin, P. (1973). Research in physical education. The North Carolina Journal, 9(3), 3-7.

Brady, G. M. (1968). An evaluation of content in statistics and experimental design courses for doctoral programs in physical education. Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Bryant, J., Comisky, P., & Zillman, D. (1977). Drama in sports commentary. Journal of Communication, 27, 140-149.

Burkhardt, J. E. (1969). Statistical comprehension for graduate students in physical education: Test and norms. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Callahan, E. J., & Ziegler, S. G. (1980). The application of within-subject design methodology to sports psychology. Journal of Sport Behavior, 3, 174-183.

Carney, T. F. (1972). Content analysis: A technique for systematic inference from communications. London: Batsford.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). Graphical methods for data analysis. Boston: Duxbury Press.

Christensen, J. E., & Christensen, C. E. (1977). Statistical power analysis of health, physical education, and recreation research. Research Quarterly for Exercise an Sport, 48, 204-208.

Clarke, H. H. (1938). History of the research section of the American Association for Health and Physical Education. Research Quarterly, 9, 25-36.

Clarke, D. H., & Clarke, H. H. (1984). Research processes in physical education. Englewood Clifs, NJ: Prentice-Hall.

College and University Council of NASPE. (1984). Standards for graduate programs in physical education. Journal of Physical Education, Recreation, and Dance, 55(2), 54-62.

Condor, R., & Anderson, D. F. (1984). Longitudinal analysis of coverage accorded black and white athletes in feature articles of Sports Illustrated (1960-1980). Journal of Sports Behavior, 7, 39-43.

Cowell, C. C. (1937). Diary analysis: A suggested technique for the study of children's activities and interests. Research Quarterly, 8, 158-172.

Cox, D. R. (1979, May). Statistical analysis and the computer. Proceedings of the Computer Science and Statistics: 12th Annual Symposium on the Interface, J. T. Gentleman (Ed.). University of Waterloo.

Cox, D. R., & Snell, E. J. (1981). Applied statistics: Principles and examples. New York: Chapman and Hall.

Cox, R. H., & Serfass, R. C. (Eds.). (1981). AAHPERD research consortium symposium papers: Analysis of repeated measures and identifying causal relationships in physical eduation research. Reston, VA: AAHPERD Publications.

Crase, D. (1978). Has physical education achieved a scholarly dimension? Journal of Physical Education and Recreation, 49(8), 21-24.

Crase, D. (1979). National/international resources in the sport sciences. Journal of Sport Behavior, 2(1), 63-67.

Cureton, T. K. (1944). Guide for tracing research in the health, physical education, and recreation fields. Research Quarterly, 15, 150-180.

Cureton, T. K. (1949). Doctorate theses reported by graduate departments of health, physical education and recreation 1930-1946, inclusively. Research Quarterly, 20, 21-59.

Daniels, W. D. (1978). Applied nonparametric statistics. Boston: Houghton Mifflin.

Dotson, C. O. (1980). Logic of questionable destiny. Resarch Quarterly for Exercise and Sport, 51, 23-36.

Earls, N. F. (1981). Distinctive teachers' personal qualities, perceptions of teacher education and the realities of teaching. Journal of Teaching in Physical Education, 1, 59-70.

Everitt, B. S., & Dunn, G. (1983). Advanced methods of data exploration and modelling. London: Heinemann Educational Books.

Fellingham, G. W., Bryce, G. R., & Carter, M. W. (1978). Latin square changeover design in physical education research. Research Quarterly for Exercise and Sport, 49, 125-134.

Gabert, T. (1976). Responsibility in research communication. The Physical Educator, 33(3), 149-151.

Gensemer, R. E. (1985). Physical education: Perspectives, inquiry, applications. Philadelphia: Saunders.

Gephart, W. J. (1969). Profiling completed research. Phi Delta Kappan, Occasional Papers #3.

Glass, G. V., & Hopkins, K. D. (1984). Statistical methods in education and psychology (2nd ed.). Englewood Clifs, NJ: Prentice-Hall.

Goodman, L. A. (1984). The analysis of cross-classified data having ordered categories. Cambridge: Harvard University Press.

Gould, D. (1982). Sport psychology in the 1980s: Status, direction, and challenge in youth sports research. Journal of Sport Psychology, 4, 203-218.

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

Groves, D. L., Heekin, R., & Banks, C. (1978). Content analysis: International Journal of Sport Psychology. International Journal of Sport Psychology, 9, 1-6.

Hair, J. E., Jr., Anderson, R. E., Tatham, R. L., & Grabowsky, B. J. (1979). Multivariate data analysis with readings. Tulsa: Petroleum.

Henry, F. M. (1959). On the use of a rubber yardstick in evaluating a difference. Research Quarterly, 30, 241.

Henry, F. M. (1960). A plea for better scholarship in research reports. Research Quarterly, 31, 234-236.

Henry, F. M. (1964). Physical eduation: An academic discipline. Journal of Health, Physical Education, and Recreation, 35, 32-33.

Henry, F. M. (1977). Blocking (pairing by ranks within groups) is statistically correct. Research Quarterly for Exercise and Sport, 48, 56.

Hildreth, K. (1980). Sexism in elementary physical education literature: A content analysis (Doctoral dissertation, University of North Carolina at Greensboro, 1979). Dissertation Abstracts International, 40, 3472A.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1982). Basic behavioral statistics. Boston: Houghton Mifflin.

Hirsch, A. A. (1981). An analysis of professional conceptions of modern dance education as reflected in Journal of Physical Education and Recreation and in Dance Magazine articles from January 1965 to December 1979 (Doctoral dissertation, The State University of New Jersey at New Brunswick, 1980). Dissertation Abstracts International, 41, 3943A.

Holsti, O. R. (1969). Content analysis for the social sciences and humanities. Reading, MA: Addison-Wesley.

Hooke, R. (1983). How to tell the liars from the statisticians. New York: Marcel Dekker.

Huff, D. (1954). How to lie with statistics. New York: W. W. Norton.

Hunter, M. D. (1966). A dictionary for physical educators. Ann Arbor, MI: University Microfilms, Inc.

Isaac, S., & Michael, W. B. (1981). Handbook in research and evaluation (2nd ed.). San Diego: EdITS.

Jaeger, R. M. (1983). Statistics: A spectator sport. Beverly Hills: Sage.

Jaeger, R. M. (1984). Sampling in education and the social sciences. New York: Longman.

Jones, B. J., & Brewer, J. K. (1972). An analysis of the power of statistical tests reported in the Research Quarterly, 43, 23-30.

Karpman, M. B. (1981). Redundancy in canonical analysis. Research Quarterly for Exercise and Sport, 52, 291-292.

Kendall, M. G., & Buckland, W. R. (1971). A dictionary of statistical terms. New York: Hafner Publishing.

Kenyon, G. S. (1965). Multiple comparisons and the analysis of variance: An empirical illustration. Research Quarterly, 36, 413-419.

Keogh, J. (1960). Comments on the selection of data for presentation. Research Quarterly, 31, 240.

Keppel, G. (1982). Design & analysis: A researcher's handbook (2nd ed.). Englewood Clifs, NJ: Prentice-Hall.

Kerlinger, F. (1973). Foundations of behavioral research. New York: Holt, Rinehart & Winston.

Kerlinger, F. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart & Winston.

Kimble, G. A. (1978). How to use (and misuse) statistics. Englewood Clifs, NJ: Prentice-Hall.

King, H. A. (1978). A nomogram to assist in planning surveys of small (n = 2,000) populations. Research Quarterly for Exercise and Sport, 49, 552-557.

King, H. A., & Baker, J. A. (1982). Conceptualization and bibliography of research in teaching physical education based on theses and dissertations, 1969-1979. Journal of Teaching in Physical Education, 2, 63-102.

Korell, D. M., & Safrit, M. J. (1977). Comparison of seriation and multidimensional scaling: Two techniques for validating constructs in physical education. Research Quarterly for Exercise and Sport, 48, 333-340.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Beverly Hills: Sage.

Kroll, W. P. (1982). Graduate study and research in physical education. Champaign, IL: Human Kinetics.

Landers, D. M. (1973). The deception paradigm and debriefing. Quest, 20, 20-25.

Lau, R. R., & Russell, D. (1980). Attributions in the sports pages. Journal of Personality and Social Psychology, 39, 29-38.

Leedy, P. D. (1985). Practical research: Planning and design (3rd ed.). New York: Macmillan.

Levine, D. M. (1977). Nonmetric multidimensional scaling and hierarchical clustering: Procedures for the investigation of the perception of sports. Research Quarterly for Exercise and Sport, 48, 341-348.

Lloyd, F. S. (1938). The research specialist: His role in teacher education. Research Quarterly, 9, 33-37.

Lock, J. C. (1976). A study of the objectives for teaching physical education in secondary schools and colleges from 1930-73, inclusive, as reported in the Journal of Health, Physical Education, Recreation (Doctoral dissertation, East Texas State University, 1975). Dissertation Abstracts International, 36, 5869A.

Locke, L. F. (1969). Research in physical education. New York: Teachers College Press.

Locke, L. F. (1972). Implications for physical education. Research Quarterly, 43, 374-386.

Loucks, D. (1952). An analytical frequency study of the content of the Research Quarterly, 1930-1949. Research Quarterly, 23, 209-220.

Loy, J. W., & Segrave, J. O. (1973). Research methodology in the sociology of sport. In J. H. Wilmore (Ed.), Exercise and sport science review: Vol. 1 (pp. 289-333). New York: Academic Press.

Martens, R. (1973). People errors in people experiments. Quest, 20, 16-19.

Martin, T. P. (1975). Why research? Journal of Health, Physical Education, and Recreation, 46(2), 56, 58.

Massey, B. H. (1966). The physical educator as researcher. Quest, 7, 46-52.

McCloy, C. H. (1930). Professional progress through research. Research Quarterly, 1, 63-73.

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

Mihevic, P. A., & Spray, J. A. (1979). An alternative multivariate one- and two-sample post hoc procedure. Research Quarterly for Exercise and Sport, 50, 649-660.

Montoye, H. J. (1955). Probability as applied to the design of experiments in health, physical education, and recreation. Research Quarterly, 26, 486-487.

Montoye, H. J., & Washburn, R. (1980). Research Quarterly contributors: An academic genealogy. Research Quarterly, for Exercise and Sport, 51, 261-266.

Morrow, J. R., & Frankiewicz, R. G. (1979). Strategies for the analysis of repeated and multiple measures designs. Research Quarterly for Exercise and Sport, 50, 297-304.

Naisbitt, J. (1982). Megatrends: Ten new directions transforming our lives. New York: Basic Books.

Nelson, D. O., & Hurst, R. L. (1963). Significant or not significant. Research Quarterly, 34, 239-242.

Nelson, R. C., & Morehouse, C. A. (1966). Statistical procedures used in multiple-group experiments. Research Quarterly, 37, 441-445.

Novak, B. J. (1942). Health material in a representative newspaper. Research Quarterly, 13, 347-350.

Park, R. J. (1980). The Research Quarterly and its antecedents. Research Quarterly for Exercise and Sport, 51, 1-22.

Pearman, W. A. (1978). Race on the sports page. Review of Sport and Leisure, 3, 54-68.

Pelton, B. C. (1976). Research redirected for the twenty-first century. The Physical Educator, 33(1), 34-38.

Petrie, B. M. (1969). Statistical analysis of attitude scores. Research Quarterly, 40, 434-436.

Pierson, W. R. (1960). Meaning of a significant t. Research Quarterly, 31, 667.

Pruzek, R. M. (1974). Data analysis: Decompositions, reformulations, and iterative processing. In J. H. Wilmore (Ed.), Exercise and sport sciences review: Vol. 1 (pp. 357-375). New York: Academic Press.

Puhl, J. (1982). The gap between research and practice in exercise physiology. Motor Skills: Theory into Practice, 6(2), 103-107.

Rand Corporation. (1955). A million random digits with normal deviates. Glencoe, IL: Free Press.

Reid, L. N., & Soley, L. C. (1979). Sports Illustrated's coverage of women in sports. Journalism Quarterly, 56, 861-863.

Rosemier, R. A. (1968). The use of an exaggerated alpha in a test for the initial equality of groups. Research Quarterly, 39, 829-830.

Roundy, E. S. (1968). Notes on problems associated with accepting the null hypothesis. Research Quarterly, 39, 831-832.

Russell, R. D. (1962). An analysis of the health and health education research in the Research Quarterly, 1951-1960. Research Quarterly, 33, 137-140.

Safrit, M. J. (1979). Women in research in physical education. Quest, 31, 158-171.

SAS Institute Inc. (1982). SAS User's Guide: Basics. Cary, NC: SAS Institute Inc.

Schutz, R. W. (1973). Sampling theory and procedures. In J. H. Wilmore (Ed.), Exercise and sport sciences review: Vol. 1 (pp. 365-392). New York: Academic Press.

Schutz, R. W., Smoll, F. L., & Gessaroli, M. E. (1983). Multivariate statistics: A self-test and guide to their utilization. Research Quarterly for Exercise and Sport, 54, 255-263.

Sharp, R. G. (1976). An investigation of professional periodical reading habits of physical educators in selected Tennessee colleges and universities (Doctoral dissertation, Middle Tennessee State University, 1976). Dissertation Abstracts International, 37, 1456A.

Shearer, F. M. (1936). A study of the amounts and types of physical education material found in education magazines. Research Quarterly, 7, 131-135.

Silva, J. M., III, & Parkhouse, B. L. (1982). On answering questions worth asking: Alternative designs for sport and exercise research. Quest, 34(1), 43-52.

Singer, R. N. (1966). Individual t tests versus multiple range tests. Research Quarterly, 37, 284-286.

Singer, R. N. (1973). Methodological controls for social psychological problems in experimentation. Quest, 20, 32-38.

Slater-Hammel, A. T. (1959). The meaning of a nonsignificant t. Research Quarterly, 30, 242-243.

Slater-Hammel, A. T. (1965a). Computational design for evaluating the significance of a difference between means. Research Quarterly, 36, 212-215.

Slater-Hammel, A. T. (1965b). Evaluation of matched groups and repeated measurement data. Research Quarterly, 36, 499-503.

Slater-Hammel, A. T. (1968). Statistical model and experimental procedures. Research Quarterly, 39, 414-416.

Slater-Hammel, A. T. (1969). Vulgar errors. Research Quarterly, 40, 437-442.

Stamm, C. L., & Safrit, M. J. (1975). Comparison of significance tests for repeated measures ANOVA design. Research Quarterly, 46, 403-409.

Stamm, C. L., & Safrit, M. J. (1977). Comparison of two nonparametric methods for estimating the reliability of two motor performance tests. Research Quarterly for Exercise and Sport, 48, 169-176.

Steinhaus, A. H. (1949). Why this research? In A. W. Hubbard (Ed.), Research methods in health, physical education, and recreation (3rd ed.), pp. 3-19. Washington, DC: AAHPER.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. The general inquirer: A computer approach to content analysis. Cambridge, MA: The M.I.T. Press.

Sudman, S. (1972). Applied sampling. New York: Academic Press.

Teraslinna, P. (1967). On publishable research articles. Research Quarterly, 38, 154-156.

Thomas, J. R., & Nelson, J. K. (1985). Introduction to research in health, physical education, recreation, and dance. Champaign, IL: Human Kinetics.

Tolson, H. (1980). An adjunct to statistical significance ω^2. Research Quarterly for Exercise and Sport, 51, 580-584.

Tuckman, B. W. (1978). Conducting educational research (2nd ed.). New York: Harcourt, Brace, and Jovanovich.

Van Dalen, D. B. (1962). Hypotheses and deduced consequences. Research Quarterly, 33, 316-317.

Van Doren, C. S., & Heit, M. J. (1973). Where it's at: A content analysis and appraisal of the Journal of Leisure Research. Journal of Leisure Research, 5, 67-73.

Vanderzwaag, H. J. (1973). Sport studies and exercise science: Philosophical accommodations. Quest, 20, 73-78.

Weber, J. C., & Lamb, D. R. (1970). Statistics and research in physical education. St. Louis: Mosby.

Weiss, M. R., & Bredemeier, B. J. (1983). Developmental sport psychology: A theoretical perspective for studying children in sport. Journal of Sport Psychology, 5, 216-230.

Williams, H. G. (1973). Volunteerism, the beneficient subject and ecological validity. Quest, 20, 26-31.

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

Zeigler, E. (1982). Physical education and sport: An introduction. Philadelphia: Lea & Febiger.

Zeigler, E. F. (1983). Relating a proposed taxonomy of sport and developmental physical activity to a planned inventory of scientific findings. Quest, 35, 54-65.

APPENDIX A

RESEARCH ARTICLES ANALYZED

Dance Research Journal

*Lord, M. (1980). A characterization of dance teacher behaviors in choreography and technique classes. Vol. 14, 15-24.

Manley, M. E , & Wilson, V. E. (1981-82). Anxiety, creativity, and dance performance. Vol. 12, 11-22.

Journal of Sport and Social Issues

Evans, A. S. (1979). Differences in the recruitment of black and white football players at a Big Eight university. Vol. 3(2), 1-10.

Marple, D. (1983). Tournament earnings and performance differentials between the sexes in professional golf and tennis. Vol. 7(1), 1-14.

*McCollum, R. H., & McCollum, D. F. (1980). Analysis of ABC-TV coverage of the 21st Olympiad Games, Montreal. Vol. 4(1), 25-33.

Nixon, H. L. (1982). Idealized functions of sport: Religion and political socialization through sport. Vol. 6(1), 1-11.

Phillips, J. C. (1983). Race and career opportunities in major league baseball: 1960-1980. Vol. 7(2), 1-17.

Journal of Sport Behavior

Anderson, D. F., & Gill, K. S. (1983). Occupational socialization patterns of men's and women's interscholastic basketball coaches. Vol. 6, 105-116.

*Anderson, D. F., & Pease, D. G. (1981). Children's motor and social skill and attitudes toward sport team involvement. Vol. 4, 128-136.

Brown, J. M., & Davies, N. (1978). Attitude towards violence among college athletes. Vol. 1, 61-70.

Buhrmann, H. G., & Zaugg, M. K. (1983). Religion and superstition in the sport of basketball. Vol. 6, 146-157.

---

*Article in reliability sample

Condor, R., & Anderson, D. F. (1984). Longitudinal analysis of coverage accorded black and white athletes in feature articles of Sports Illustrated (1960-1980). Vol. 7, 39-43.

*Crabbe, J. M., & Johnson, G. O. (1980). Male and female coaction in a competitive environment. Vol. 3, 86-95.

DuBois, P. E. (1979). Participation in sport and occupational attainment: An investigation of selected athlete categories. Vol. 2, 103-114.

DuBois, P. E. (1981). The youth sport coach as an agent of socialization: An exploratory study. Vol. 4, 95-107.

Gilliland, K., & Tutko, T. A. (1978). Differences in parent-child relations between athletes and non-athletes. Vol. 1, 51-60.

Gould, D., Feltz, D., Horn, T., & Weiss, M. (1982). Reasons for attrition in competitive youth swimming. Vol. 5, 155-165.

Gray, J. J., Haring, M. J., & Banks, N. M. (1984). Mental rehearsal for sport performance: Exploring the relaxation-imagery paradigm. Vol. 7, 68-78.

Gundersheim, J. (1982) A comparison of male and female athletes and nonathletes on measures of self-actualization. Vol. 5, 186-201.

Hailey, B. J., & Bailey, L. A. (1982). Negative addiction in runners: A quantitative approach. Vol. 5, 150-154.

Krotee, M. L. (1980). The effects of various physical activity situational settings on the anxiety level of children. Vol. 3, 158-164.

McCutcheon, L. E. (1982). Does running make people more creative? Vol. 5, 202-206.

Lee, M. J., Coburn, T., & Partridge, R. (1983). The influence of team structure in determining leadership function in Association football. Vol. 6, 59-66.

Lennon, J. X. (1980). The effects of crowding and observation of athletic events on spectator tendency toward aggressive behavior. Vol. 3, 61-68.

*Leonard, W. M., II. (1983). Mortality ratios of professional baseball players and managers. Vol. 6, 117-135.

Nation, J. R., & LeUnes, A. D. (1983). Personality characteristics of intercollegiate football players as determined by position, classification, and redshirt status. Vol. 6, 92-102.

Nixon, H. L., II. (1981). Birth order and preferences for risky sport among college students. Vol. 4, 13-24.

Pilz, G. A. (1979). Attitudes toward different forms of aggressive and violent behavior in competitive sports: Two empirical studies. Vol. 2, 3-26.

Purdy, D. A., Eitzen, D. S., & Haufler, S. E. (1982). Age-group swimming: Contributing factors and consequences. Vol. 5, 28-43.

Rees, C. R. (1983). Instrumental and expressive leadership in team sports: A test of leadership role differentiation theory. Vol. 6, 17-27.

Seabourne, T., Weinberg, R., & Jackson, A. (1984). Effect of individualized practice and training of visuo-motor behavior rehearsal in enhancing karate performance. Vol. 7, 58-67.

Skipper, J. K., Jr. (1984). The sociological significance of nicknames: The case of baseball players. Vol. 7, 28-38.

Snyder, E. E., & Spreitzer, E. (1983). Change and variation in the social acceptance of female participation in sports. Vol. 6, 3-8.

Thirer, J. (1978). The effect of observing filmed violence on the aggressive attitudes of female athletes and non-athletes. Vol. 1, 28-36.

Thirer, J., & Greer, D. L. (1981). Personality characteristics associated with beginning, intermediate, and competitive bodybuilders. Vol. 4, 3-12.

Thirer, J., Knowlton, R., Sawka, M., & Chang, T. J. (1978). Relationship of psycho-physiological characteristics to perceived exertion and levels of anxiety in competitive swimmers. Vol. 1, 169-173.

Warren, J., & Lanning, W. (1982). Biorhythm: Its relationship to football injuries. Vol. 5, 132-138.

Journal of Sport Psychology

Allard, F., Graham, S., & Paarsalu, M. T. (1980). Perception in sport: Basketball. Vol. 2, 14-21.

Allard, F., & Starkes, J. L. (1980). Perception in sport: Volleyball. Vol. 2, 22-33.

Bird, A. M., Foster, C. D., & Maruyama, G. (1980). Convergent and incremental effects of cohesion on attributions for self and team. Vol. 2, 181-194.

Brawley, L. R., Powers, R. C., & Phillips, K. A. (1980). Sex bias in evaluating motor activity: General or task-specific performance expectancy? Vol. 2, 279-287.

Carmack, M. A., & Martens, R. (1979). Measuring commitment to running: A survey of runners' attitudes and mental states. Vol. 1, 25-42.

Carron, A. V., & Chelladurai, P. (1981). The dynamics of group cohesion in sport. Vol. 3, 123-139.

Carver, C. S., DeGregorio, E., & Gillis, R. (1981). Challenge and Type A behavior among intercollegiate football players. Vol. 3, 140-148.

Chelladurai, P. (1984). Discrepancy between preferences and perceptions of leadership behavior and satisfaction of athletes in varying sports. Vol. 6, 27-41.

Chelladurai, P., & Carron, A. V. (1983). Athletic maturity and preferred leadership. Vol. 5, 371-380.

Corbin, C. C. (1981). Sex of subject, sex of opponent, and opponent ability as factors affecting self-confidence in a competitive situation. Vol. 3, 265-270.

Corbin, C. C., Stewart, M. J., & Blair, W. O. (1981). Self-confidence and motor performance of preadolescent boys and girls studied in different feedback situations. Vol. 3, 30-34.

Epstein, M. L. (1980). The relationship of mental imagery and mental rehearsal to performance of a motor task. Vol. 2, 211-220.

Fisher, A. C., Mancini, V. H., Hirsch, R. L., Proulx, T. J., & Staurowsky, E. J. (1982). Coach-athlete interactions and team climate. Vol. 4, 388-404.

Gill, D. L. (1980). Success-failure attributions in competitive groups: An exception to egocentricism. Vol. 2, 106-114.

Gill, D. L., Ruder, M. K., & Gross, J. B. (1982). Open-ended attributions in team competition. Vol. 4, 159-169.

Gould, D., Horn, T., & Spreemann, J. (1983). Sources of stress in junior elite wrestlers. Vol. 5, 159-171.

Gould, D., Weiss, M., & Weinberg, R. (1981). Psychological characteristics of successful and nonsuccessful Big Ten wrestlers. Vol. 3, 69-81.

Hale, B. D. (1982). The effects of internal and external imagery on muscular and ocular concomitants. Vol. 4, 379-387.

Jagacinski, R. J., Newell, K. M., & Isaac, P. D. (1979). Predicting the success of a basketball shot at various stages of execution. Vol. 1, 301-310.

Kleiber, D. A., & Roberts, G. C. (1981). The effects of sport experience in the development of social character: An exploratory investigation. Vol. 3, 114-122.

Landers, D. M., Wilkinson, M. O., Hatfield, B. D., & Barber, H. (1982). Causality and the cohesion-performance relationship. Vol. 4, 170-183.

Mark, M. M., Mutrie, N., Brooks, D. R., & Harris, D. V. (1984). Causal attributions of winners and losers in individual competitive sports: Toward a reformulation of the self-serving bias. Vol. 6, 184-196.

McAuley, E., & Gill, D. (1983). Reliability and validity of the physical self-efficacy scale in a competitive sport setting. Vol. 5, 410-418.

*Noel, R. C. (1980). The effect of visuo-motor behavior rehearsal on tennis performance. Vol. 2, 221-226.

Passer, M. W. (1983). Fear of failure, fear of evaluation, perceived competence, and self-esteem in competitive-trait-anxious children. Vol. 5, 172-188.

Reis, J., & Bird, A. M. (1982). Cue processing as a function of breadth of attention. Vol. 4, 64-72.

Rejeski, W., Darracott, C., & Hutslar, S. (1979). Pygmalion in youth sport: A field study. Vol. 1, 311-319.

Rejeski, W. J., & Ribisl, P. M. (1980). Expected task duration and perceived effort: An attributional analysis. Vol. 2, 227-236.

Roberts, G. C., Kleiber, D. A., & Duda, J. L. (1981). An analysis of motivation in children's sport: The role of perceived competence in participation. Vol. 3, 206-216.

Roberts, G. C., & Pascuzzi, D. (1979). Causal attributions in sport: Some theoretical implications. Vol. 1, 203-211.

Ruder, M. K., & Gill, D. K. (1982). Immediate effects of win-loss on perceptions of cohesion in intramural and intercollegiate volleyball teams. Vol. 4, 227-234.

Rushall, B. S., & Smith, K. C. (1979). The modification of the quality and quantity of behavior categories in a swimming coach. Vol. 1, 138-150.

Ryan, E. D., & Simons, J. (1981). Cognitive demand, imagery, and frequency of mental rehearsal as factors influencing acquisition of motor skills. Vol. 3, 35-45.

Ryan, E. D., & Simons, J. (1982). Efficacy of mental imagery in enhancing mental rehearsal of motor skills. Vol. 4, 41-51.

Scanlon, T. K., & Lewthwaite, R. (1984). Social psychological aspects of competition for male youth sport participants: I. Predictors of competitive stress. Vol. 6, 208-226.

*Scanlon, T. K., & Passer, M. W. (1980). Self-serving biases in the competitive sport setting: An attributional dilemma. Vol. 2, 124-136.

Simon, J. A., & Martens, R. (1979). Children's anxiety in sport and nonsport evaluative activities. Vol. 1, 160-169.

Spink, K. S., & Roberts, G. C. (1980). Ambiguity of outcome and causal attributions. Vol. 2, 237-244.

Starkes, J. L., & Allard, F. (1983). Perception in volleyball: The effects of competitive stress. Vol. 5, 189-196.

Summers, J. J., Machin, V. J., & Sargent, G. I. (1983). Psychosocial factors related to marathon running. Vol. 5, 314-331.

Vallerand, R. J., & Reid, G. (1984). On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. Vol. 6, 94-102.

*Van Schoyck, S. R., & Grasha, A. F. (1981). Attentional style variations and athletic ability: The advantages of a sport-specific test. Vol. 3, 149-165.

Weiss, M. R., & Knoppers, A. (1982). The influence of socializing agents on female collegiate volleyball players. Vol. 4, 267-279.

*Willis, J. D. (1982). Three scales to measure competition-related motives in sport. Vol. 4, 338-353.

Yukelson, D., Weinberg, R. S., & Jackson, A. (1984). A multidimensional sport cohesion instrument for inter-collegiate basketball teams. Vol. 6, 103-117.

*Yukelson, D., Weinberg, R. S., West, S., & Jackson, A. (1981). Attributions and performance: An empirical test of Kukla's theory. Vol. 3, 46-57.

Journal of Teaching in Physical Education

Aufderheidi, S. K., McKenzie, T. L., & Knowles, C. J. (1982). Effect of individualized instruction on handicapped and nonhandicapped students in elementary physical education classes. Vol. 1(3), 51-57.

*Bischoff, J. (1982). Equal opportunity, satisfaction and success: An exploratory study on coeducational volley-ball. Vol. 2(1), 3-12.

Devlin, G., Mancini, V., & Frye, P. (1981). Teaching con-tingency management skills to disruptive elementary students: Effects on student self-concept and physical educators' behaviors. Vol. 1(1), 47-58.

Martinek, T., & Karper, W. (1982). Canonical relationships among motor ability, expression of effort, teacher expectations and dyadic interactions in elementary age children. Vol. 1(2), 26-39.

McBride, R. E. (1984). Perceived teaching and program concerns among preservice teachers, university super-visors, and cooperating teachers. Vol. 3(3), 36-43.

McKenzie, T. L. (1981). Modification, transfer and maintenance of the verbal behavior of an experienced physical education teacher: A single-subject analysis. Introductory issue, 48-56.

McKenzie, T. L., Clark, E. K., & McKenzie, R. (1984). Instructional strategies: Influence on teacher and student behavior. Vol. 3(2), 20-28.

*Paese, P. C. (1982). Effects of interdependent group contingencies in a secondary physical education setting. Vol. 2(1), 29-37.

Phillips, D. A., & Carlisle, C. (1983). A comparison of physical education teachers categorized as most and least effective. Vol. 2(3), 55-67.

Shute, S., Dodds, P., Placek, J., Rife, F., & Silverman, S. (1982). Academic learning time in elementary school movement education: A descriptive analytic study. Vol. 1(2), 3-14.

van der Mars, H., Mancini, V., & Frye, P. (1981). Effects of interaction analysis training on perceived and observed teaching behaviors. Introductory issue, 57-65.

Williams, L. T. R. (1983). Professional self-perception of physical education teachers. Vol. 2(2), 77-87.

Yerg, B. J. (1981). The impact of selected presage and process behaviors on the refinement of a motor skill. Vol. 1(1), 38-46.

Research Quarterly for Exercise and Sport

Arnett, B., & Rikli, R. (1981). Effects of method of subject selection (volunteer vs. random) and treatment variable on motor performance. Vol. 52, 433-440.

Ashby, A. A., Shea, C. H., & Tolson, H. (1982). Memory characteristics of two dimensional movement information. Vol. 53, 93-100.

Baker, J. A. W., & Wilson, B. D. (1978). The effect of tennis racket stiffness and string tension on ball velocity after impact. Vol. 49, 255-259.

Barnett, M. L., & Stanicek, J. A. (1979). Effects of goal setting on achievement in archery. Vol. 50, 328-332.

Belka, D. E., & Williams, H. G. (1980). Canonical relationships among perceptual-motor, perceptual, and cognitive behaviors in children. Vol. 51, 463-477.

Bird, A. M., & Rikli, R. (1983). Observational learning and practice variability. Vol. 54, 1-4.

Bryant, F. O., Burkett, L. N., Chen, S. S., Krahenbuhl, G. S., & Lu, Ping. (1977). Dynamic and performance characteristics of baseball bats. Vol. 48, 505-509.

Budney, D. R., & Bellow, D. G. (1982). On the swing mechanics of a matched set of golf clubs. Vol. 53, 185-192.

Burke, E. J. (1977). Physiological effects of similar training programs in males and females. Vol. 48, 510-517.

Burke, E. J., & Brush, F. C. (1979). Physiological and anthropometric assessment of successful teenage female distance runners. Vol. 50, 180-187.

Butts, N. K. (1982). Physiological profiles of high school female cross country runners. Vol. 53, 8-14.

Carron, A. V., & Bennett, B. B. (1977). Compatibility in the coach-athlete dyad. Vol. 48, 671-679.

Chai, D. X. (1977). Future of leisure: A Delphi application. Vol. 48, 518-524.

Christensen, C. E., Milner, K., & Christensen, J. E. (1978). An analysis of faculty perceptions of leadership qualities of male and female physical education departments. Vol. 49, 269-277.

Clarke, D. H., & Vaccaro, P. (1979). The effect of swimming training on muscular performance and body composition in children. Vol. 50, 9-17.

*Clarkson, P. M., Johnson, J., Dextradeur, D., Leszcznski, W., Wai, J., & Melchionda, A. (1982). The relationships among isokinetic endurance, initial strength level, and fiber type. Vol. 53, 15-19.

Conley, D. L., Krahenbuhl, G. S., Burkett, L. N., & Millar, A. L. (1981). Physiological correlates of female road racing performance. Vol. 52, 441-448.

Cooter, G. R., & Mowbray, K. W. (1978). Effects of iron supplementation and activity on serum iron depletion and hemoglobin levels in female athletes. Vol. 49, 114-118.

Coyle, E. F., Bell, S., Costill, D. L., & Fink, W. J. (1978). Skeletal muscle fiber characteristics of world class shot-putters. Vol. 49, 278-284.

*Coyle, E. F., Costill, D. L., Fink, W. J., & Hoopes, D. G. (1978). Gastric emptying rates for selected athletic drinks. Vol. 49, 119-124.

Crompton, J. L., Lamb, C. W., Jr., & Vedlitz, A. (1979). Age and sex differences among adolescent participants in nine outdoor recreation activities. Vol. 50, 589-598.

Dainis, A. (1979). Cinematographic analysis of the handspring vault. Vol. 50, 341-349.

Del Ray, P., Wughalter, E. H., & Whitehurst, M. (1982). The effects of contextual interference on females with varied experience in open sport skills. Vol. 53, 108-115.

DiNucci, J. M., & Shows, D. A. (1977). A comparison of the motor performance of Black and Caucasian girls age 6-8. Vol. 48, 680-684.

*Dubois, P. E. (1978). Participation in sports and occupational attainment: A comparative study. Vol. 49, 28-37.

Ewing, J. L., Jr., & Stull, G. A. (1984). Rate of force development in the handgripping muscles by females as a function of fatigue level. Vol. 55, 17-23.

Falls, H. B., & Humphrey, L. D. (1978). Body type and composition differences between placers and nonplacers in an AIAW gymnastics meet. Vol. 49, 38-43.

Feltz, D. L., & Landers, D. M. (1977). Informational-motivational components of a model's demonstration. Vol. 48, 525-534.

Foster, C., Costill, D. L., & Fink, W. J. (1980). Gastric emptying characteristics of glucose and glucose polymer solutions. Vol. 51, 299-305.

Freedson, P. S., Gilliam, T. B., Sady, S. P., & Katch, V. L. (1981). Transient VO_2 characteristics in children at the onset of steady-rate exercise. Vol. 52, 167-173.

Frekany, G. A. (1978). Retention characteristics of location and distance cues in motor short-term memory. Vol. 49, 293-300.

*Gabbard, C., Patterson, P., & Elledge, J. (1981). Grip and forearm position effects on tests of static and dynamic upper body endurance. Vol. 52, 174-179.

Gagnon, M., & Rodigue, D. (1979). Determination of physical properties of the forearm by anthropometry, immersion, and photography methods. Vol. 50, 188-198.

Gallemore, S. L. (1981). Perceptions about the objectives of student teaching. Vol. 52, 180-190.

Genuario (Baltrusaitis), S. E., & Dolgener, F. A. (1980). The relationship of isokinetic torque at two speeds to the vertical jump. Vol. 51, 593-598.

Gill, D. L. (1977). Influence of group success-failure and relative ability on intrapersonal variables. Vol. 48, 685-694.

Gilliam, T. B., Villanacci, J. F., Freedson, P. S., & Sady, S. P. (1979). Isokinetic torque in boys and girls ages 7 to 13: Effect of age, height, and weight. Vol. 50, 599-609.

Gillis, J. H. (1979). Effects of achieving tendency, gender, and outcome on causal attributions following motor performance. Vol. 50, 610-619.

Goldberger, M., Gerney, P., & Chamberlain, J. (1982). The effects of three styles of teaching on the psychomotor performance and social skill development on fifth grade children. Vol. 53, 116-124.

Greendorfer, S. L., & Ewing, M. E. (1981). Race and gender differences in children's socialization into sport. Vol. 52, 301-310.

Greendorfer, S. L., & Lewko, J. H. (1978). Role of family members in sport socialization of children. Vol. 49, 146-152.

Hall, E. G. (1980). Comparison of postperformance state anxiety of internals and externals following failure or success on a simple motor task. Vol. 51, 306-314.

Hall, S. J., & DePauw, K. P. (1982). A photogrammetrically based model for predicting total body mass centroid location. Vol. 53, 37-45.

Heyward, V., & McCreary, L. (1977). Analysis of the static strength and relative endurance of women athletes. Vol. 48, 703-710.

Higgs, C. (1983). An analysis of racing wheelchairs used at the 1980 Olympic games for the disabled. Vol. 54, 229-233.

Hoffman, S. J., Imwold, C. H., & Koller, J. A. (1983). Accuracy and prediction in throwing: A taxonomic analysis of children's performance. Vol. 54, 33-40.

Holmen, M. G., & Parkhouse, B. L. (1981). Trends in the selection of coaches for female athletes: A demographic inquiry. Vol. 52, 9-18.

Hopkins, D. R. (1977). Factor analysis of selected basketball skill tests. Vol. 48, 535-540.

Hopkins, D. R. (1979). Using skill tests to identify successful and unsuccessful basketball performers. Vol. 50, 381-387.

Horgan, J. S. (1977). Stabilometer performance of educable mentally retarded children under differential feedback conditions. Vol. 48, 711-716.

Housner, L. D. (1984). The role of imaginal processing in the retention of visually-presented sequential motoric stimuli. Vol. 55, 24-31.

Hovell, M. F., Bursick, J. H., Sharkey, R., & McClure, J. (1978). An evaluation of elementary students' voluntary physical activity during recess. Vol. 49, 460-474.

Hunt, S. L. (1979). Work and leisure in an academic environment: Relationships between selected meanings. Vol. 50, 388-395.

Iacobellis, J. (1979). Cardiovascular responses to sustained isometric work in a hot environment. Vol. 50, 620-629.

*Imwold, C. H., & Hoffman, S. J. (1983). Visual recognition of a gymnastics skill by experienced and inexperienced instructors. Vol. 54, 149-155.

*Iso-Ahola, S. (1978). Perceiving the causes of objective and subjective outcomes following motor performance. Vol. 49, 62-70.

Israel, R. G., & Brown, R. L. (1981). Response time comparisons among four base running starting techniques in slow pitch softball. Vol. 52, 324-329.

Johnson, R. T., Bjorkland, R., & Krotee, M. L. (1984). The effects of cooperative, competitive and individualistic student interaction patterns on the achievement and attitudes of students learning the golf skill of putting. Vol. 55, 129-134.

*Kelly, J. M., Serfass, R. C., & Stull, G. A. (1980). Elicitation of maximal oxygen uptake from standing bicycle ergometry. Vol. 51, 315-322.

Keogh, J. F., Griffin, N. S., & Spector, R. (1981). Observer perceptions of movement confidence. Vol. 52, 465-473.

Kermond, J., & Konz, S. (1978). Support leg loading in punt kicking. Vol. 49, 71-79.

Kingsley, J. L., Brown, F. L., & Seibert, M. E. (1977). Social acceptance of female athletes by college women. Vol. 48, 727-733.

Krahenbuhl, G. S., Pangrazi, R. P., & Chomokos, E. A. (1979). Aerobic responses of young boys to submaximal running. Vol. 50, 413-421.

Leger, L. A. (1982). Energy cost of disco dancing. Vol. 53, 46-49.

Linden, M. (1977). Factor analytical study of Olympic decathlon data. Vol. 48, 562-568.

Locke, L. F., & Massengale, J. D. (1978). Role conflict in teachers/coaches. Vol. 49, 162-174.

Lydon, M. C., & Cheffers, J. T. F. (1984). Decision-making in elementary school-age children: Effects upon motor learning and self-concept development. Vol. 55, 135-140.

Magill, R. A., & Parks, P. F. (1983). The psychophysics of kinesthesis for positioning responses: The physical stimulus-psychological response relationship. Vol. 54, 346-351.

Mann, R., & Sprague, P. (1980). A kinetic analysis of the ground leg during sprint running. Vol. 51, 334-348.

Margolis, J. F., & Christina, R. W. (1981). A test of Schmidt's schema theory of discrete motor skill learning. Vol. 52, 474-483.

*Marino, G. W. (1979). Acceleration-time relationships in an ice skating start. Vol. 50, 55-59.

Matthews, P. R. (1979). The frequency with which the mentally retarded participate in recreation activities. Vol. 50, 71-79.

McCaughan, L. R., & McKinlay, S. (1981). Effects of success/failure and extrinsic rewards on intrinsic motivation using a competitive motor task. Vol. 52, 208-215.

McElroy, M. A., & Kirkendall, D. R. (1980). Significant others and professionalized sport attitudes. Vol. 51, 645-653.

McGrain, P. (1980). Trends in selected kinematic and myoelectric variables associated with learning a novel motor task. Vol. 51, 509-520.

McIntyre, D. R., & Pfautsch, E. W. (1982). A kinematic analysis of the baseball batting swings involved in opposite-field and same-field hitting. Vol. 53, 206-213.

Mihevic, P. M. (1983). Cardiovascular fitness and the psychophysics of perceived exertion. Vol. 54, 239-246.

Miyashita, M., & Kanehisa, H. (1979). Dynamic peak torque related to age, sex, and performance. Vol. 50, 249-255.

Morris, A. M., Williams, J. M., Atwater, A. E., & Wilmore, J. H. (1982). Age and sex differences in motor performance of 3 through 6 year old children. Vol. 53, 214-221.

Murray, M. J. (1979). Matching preferred cognitive mode with teaching methodology in learning a novel motor skill. Vol. 50, 80-87.

Nelson, J. K. (1978). Motivating effects of the use of norms and goals with endurance tests. Vol. 49, 317-321.

Noble, B. J., & Maresh, C. M. (1979). Acute exposure of college basketball players to moderate altitude: Selected physiological responses. Vol. 50, 668-678.

Nolan, G. N., & Bates, B. T. (1982). A biomechanical analysis of the effects of two paddle types on performance in North American canoe racing. Vol. 53, 50-57.

Noland, M., & Kearney, J. T. (1978). Anthropometric and densitometric responses of women to specific and general exercise. Vol. 49, 322-328.

Ordway, G. A., Kearney, J. T., & Stull, G. A. (1977). Rhythmic isometric fatigue patterns of the elbow flexors and knee extensors. Vol. 48, 734-740.

Pigott, R. E., & Shapiro D. C. (1984). Motor schema: The structure of the variability session. Vol. 55, 41-45.

Plagenhoef, S. (1979). Biomechanical analysis of Olympic flatwater kayaking and canoeing. Vol. 50, 443-459.

Pollock, M. L., Foster, C., Anholm, J., Hare, J., Farrell, P., Maksud, M., & Jackson, A. S. (1982). Body composition of Olympic speed skating candidates. Vol. 53, 150-155.

Purcell, K. (1981). A tennis forehand-backhard drive skill test which measures ball control and stroke firmness. Vol. 52, 238-245.

Reeve, T. G., & Cone, S. L. (1980). Coding of learned kinesthetic location information. Vol. 51, 349-358.

Reeve, T. G., & Magill, R. A. (1981). The role of the components of knowledge of results information in error correction. Vol. 52, 80-85.

Reeve, T. G., & Mainor, R., Jr. (1983). Effects of movement context on the encoding of kinesthetic spatial information. Vol. 54, 352-363.

*Rikli, R., & Arnett, B. (1978). Motor performance scores of volunteer versus randomly selected subjects. Vol. 49, 329-334.

Riley, W. W., Jr., Powers, S. K., & Welch, H. G. (1981). The effect of two levels of muscular work on urinary creatinine excretion. Vol. 52, 339-347.

Sady, S. P., & Katch, V. L. (1981). Relative endurance and physiological responses: A study of individual differences in prepubescent boys and adult men. Vol. 52, 246-255.

Sady, S. P., Katch, V. L., Villanacci, J. F., & Gilliam, T. B. (1983). Children-adult comparisons of VO_2 and HR kinetics during submaximum exercise. Vol. 54, 55-59.

Sage, G. H., & Hornak, J. E. (1978). Progressive speed practice in learning a continuous motor skill. Vol. 49, 190-196.

* Sage, G. H., & Loudermilk, S. (1979). The female athlete and role conflict. Vol. 50, 88-96.

Serfass, R. C., Stull, G. A., Alexander, J. F., & Ewing, J. L., Jr. (1984). The effects of rapid weight loss and attempted rehydration on strength and endurance of the handgripping muscles in college wrestlers. Vol. 55, 46-52.

Shapiro, R. (1978). Direct linear transformation method for three-dimensional cinematography. Vol. 49, 197-205.

* Siegel, D., Johnson, J., & Kline, G. (1984). Attentional load and the reproduction of physical work. Vol. 55, 146-152.

Silva, J. M., III, Shultz, B. B., Haslam, R. W., & Murray, D. (1981). A psychophysiological assessment of elite wrestlers. Vol. 52, 348-358.

* Slaughter, M. H., Lohman, T. G., & Boileau, R. A. (1977). Relationship of Heath and Carter's second component to lean body mass and height in college women. Vol. 48, 759-768.

Slava, S., Laurie, D. R., & Corbin, C. B. (1984). Long-term effects of a conceptual physical education program. Vol. 55, 161-168.

Slovia, P. (1977). Empirical study of training and performance in the marathon. Vol. 48, 769-777.

Southard, D. (1983). Importance of selected competencies and relationship to corresponding coursework in programs of teacher preparation. Vol. 54, 383-388.

Stamford, B. A., Weltman, A., & Fulco, C. (1978). Anaerobic threshold and cardiovascular responses during one- versus two-legged cycling. Vol. 49, 351-362.

Stamford, B. A., Weltman, A., Moffatt, R. J., & Fulco, C. (1978). Effects of severe prior exercise on assessment of maximal oxygen uptake during one- versus two-legged cycling. Vol. 49, 363-371.

*Strache, C. V. (1979). Players' perceptions of leadership qualities for coaches. Vol. 50, 679-686.

Templin, T. T. (1979). Occupational socialization and the physical education student teacher. Vol. 50, 482-493.

Thomas, J. R., Gallagher, J. D., & Purvis, G. J. (1981). Reaction time and anticipation time: Effects of development. Vol. 52, 359-367.

Tu, J., & Rothstein, A. L. (1979). Improvement of jogging performance through application of personality specific motivational techniques. Vol. 50, 97-103.

Tucker, L. A. (1983). Effect of weight training on self-concept: A profile of those influenced most. Vol. 54, 389-397.

Vickers, J., Lashuk, M., & Taerum, T. (1980). Differences in attitude toward the concepts "male," "female," "male athlete," and "female athlete." Vol. 51, 407-416.

Washburn, R., Chin, M. K., & Montoye, H. J. (1980). Accuracy of pedometer in walking and running. Vol. 51, 695-702.

Weinberg, R. S., & Jackson, A. (1979). Competition and extrinsic rewards: Effect on intrinsic motivation and attribution. Vol. 50, 494-502.

Weiss, M. R. (1983). Modeling and motor performance: A developmental perspective. Vol. 54, 190-197.

Weltman, A., Katch, V., Sady, S., & Freedson, P. (1978). Onset of metabolic acidosis (anaerobic threshold) as a criterion measure of submaximal fitness. Vol. 49, 218-227.

Weltman, A., Stamford, B. A., Moffatt, R. J., & Katch, V. L. (1977). Exercise recovery, lactate removal, and subsequent high intensity exercise performance. Vol. 48, 786-796.

Wrisberg, C. A., & Mead, B. J. (1983). Developing coincident timing skill in children: A comparison of training methods. Vol. 54, 67-74.

Yu, J., & Mendell, R. (1980). The development and utility of a leisure behavior index. Vol. 51, 553-558.

Review of Sport and Leisure

Anderson, D. F., Rupnow, A. A., & Ludwig, D. A. (1981). Importance of reward systems in youth sports: A replication of Watson's study using soccer participants. Vol. 6(1), 37-47.

Dewar, C. K. (1979). Spectator fights at professional baseball games. Vol. 4(1), 12-25.

*Frey, J. H. (1981). The place of athletics in the educational priorities of university alumni. Vol. 6(1), 48-64.

*Gillis, J. H. (1980). Olympic success and national religious orientation. Vol. 5(2), 1-20.

Godbey, G., & Robinson, H. (1979). The American sports fan: 'Spectatoritis' revisited. Vol. 4(1), 1-11.

Greenstein, T. N., & Marcum, J. P. (1981). Factors affecting attendance of major league baseball: Team performance. Vol. 6(2), 21-34.

Jackson, J. J. (1979). Some consequences of Participaction's campaign in Victoria. Vol. 4(2), 89-100.

Jackson, J. J., Heron, R. P., & McLachlin, H. J. (1978). Examining the diffusion of innovations in physical education, recreation and sport settings: A methodology. Vol. 3(2), 93-111.

Jensen, T. M., Leonard, W. M., II, & Liverman, R. D. (1981). College students' attitudes toward intercollegiate athletics. Vol. 6(1), 65-96.

Kjeldsen, E. K. M. (1981). Centrality and leadership recruitment: A study of their linkage. Vol. 6(2), 1-20.

Leuch, M. R., Krahenbuhl, G. S., & Odenkirk, J. E. (1979). Assessment of spectator aggression at intercollegiate basketball contests. Vol. 4(1), 40-52.

Lowe, B., Jackson, J. J., & Roberts, J. (1981). The cross-cultural analysis of Commonwealth athletes: Phase II. Vol. 6(1), 114-123.

Morgan, M. J. (1978). Composite personality profiles of female athletes: Some theoretical and practical possibilities. Vol. 3(1), 47-63.

Nabil, P. A. (1980). The present-day Afro-American major-league baseball player and socioeconomic mobility in American society. Vol. 5(2), 49-68.

Purdy, D. A. (1980). The relationship among self-esteem, competition anxiety and competitiveness. Vol. 5(1), 114-123.

APPENDIX B

CODE SHEET FOR ARTICLE ANALYSIS

JOURNAL____________________ #_____

A. What is the primary subspecialty focus of the article?
   1 = Background, Meaning, & Significance
   2 = Functional Effects of Physical Activity
   3 = Sociocultural & Behavioral Aspects
   4 = Motor Learning & Development
   5 = Mechanical & Muscular Analysis of Motor Skills
   6 = Management Theory & Practice
   7 = Program Development
   8 = Measurement & Evaluation

B. Number of subjects (or similar unit of analysis)? ______
   B1. What is the design of the study?
   1 = Completely Randomized 2 = Pure Within-S 3 = Mixed

C. Number of variables (IV, DV, classification)? ______
   C1. Number of dependent variables? ______
   C2. What is the largest number of variables which were entered into a single statistical analysis? ______
   C2a. How many were dependent variables? ______

D. What statistical analyses were used?

E. Were any significance tests performed? 1 = No (Go to F) 2 = Yes
   E1. Was significance reported for any of the significance tests? 1 = No 2 = Yes

   E2. Was a single criterion p-value reported/implied?
   1 = No 2 = .10 3 = .05 4 = .01 5 = Other

   E3. Were exact p-values reported? 1 = No 2 = Yes

F. Was an explanation and/or a rationale given for any of the statistical analyses employed?
1 = No 2 = Yes (specify)________________________________

G. Were any statistical references cited?
1 = No 2 = Yes (specify)________________________________

H. Were any statistical assumptions tested for?
1 = No 2 = Yes (specify)________________________________

I. Does the article report data transformation of any kind?
1 = No 2 = Yes (specify)________________________________

J. Does the article report data analysis methods (i.e , computer programs used)?
1 = No 2 = Yes (specify)________________________________

APPENDIX C

COMPUTATION OF RELIABILITY COEFFICIENTS

Coding of Question A by Judge 1 and Judge 2

		Judge 1 Category	Judge 2 Category
RQES	#27	2	2
	#34	3	3
	#38	2	2
	#42	3	3
	#62	8	8
	#86	5	5
	#90	3	3
	#141	6	6
	#149	2	2
	#195	5	5
	#222	2	2
	#266	7	7
	#303	3	2
JOSP	#26	3	3
	#32	4	4
	#44	3	3
	#50	8	8
	#80	8	8
JOSB	#18	3	3
	#32	3	3
	#64	3	3
JTPE	#13	3	3
	#16	3	3
ROSL	#23	3	3
	#30	3	6
JSSI	#3	3	3
DNRJ	#3	7	7

Computation of Reliability for Question A

$$\text{Agreement Coefficient} = 1 - \left[\frac{rm - 1}{m - 1}\right] \left[\frac{\sum_i \sum_b \sum_{c>b} n_{bi} \; n_{ci} \; d_{bc}}{\sum_b \sum_{c>b} n_b \; n_c \; d_{bc}}\right]$$

(Krippendorff, 1980, p. 138)

$n_1 = 0$ $n_2 = 9$ $n_3 = 26$ $n_4 = 2$ $n_5 = 4$ $n_6 = 3$ $n_7 = 4$ $n_8 = 6$

r = # of articles = 27 m = # of independent judges = 2

$$\therefore 1 - \left[\frac{(27)(2) - 1}{2 - 1}\right] \left[\frac{2}{1039}\right] = .90 = \text{90\% better than chance}$$

Coding of Question C by Judge 1 and Judge 2

		Judge 1 Category	Judge 2 Category	TOTAL (Pi)
RQES	#27	17	19	36
	#34	10	10	20
	#38	5	5	10
	#42	7	10	17
	#62	8	8	16
	#86	6	8	14
	#90	5	4	9
	#141	27	18	45
	#149	11	9	20
	#195	4	4	8
	#222	11	12	23
	#266	4	4	8
	#303	10	11	21
JOSP	#26	11	10	21
	#32	6	10	16
	#44	5	13	18
	#50	16	15	31
	#80	5	6	11
JOSB	#18	3	5	8
	#32	6	15	21
	#64	5	7	12
JTPE	#13	8	10	18
	#16	3	4	7
ROSL	#23	5	6	11
	#30	4	5	9
JSSI	#3	8	8	16
DNRJ	#3	4	7	11
				G = 457

Computation of Reliability for Question C

$$SS_{\text{b/t articles}} = \frac{\Sigma Pi^2}{m} - \frac{G^2}{mr} = \frac{9845}{2} - \frac{(457)^2}{2(27)} = 1054.9259$$

$$MS_{\text{b/t articles}} = \frac{1054.9259}{27 - 1} = 40.57$$

$$SS_{\text{w/in articles}} = \Sigma(\Sigma X^2) - \frac{\Sigma Pi^2}{m} = 5069 - 4922.5 = 146.5$$

$$MS_{\text{w/in articles}} = \frac{146.5}{27} = 5.43$$

(Winer, 1971, pp. 286-288)

$$\therefore 1 - \frac{5.43}{40.57} = .87 = \underline{87\% \text{ reliable}}$$

Coding of Question F by Judge 1 and Judge 2

		JUDGE 2	
		Yes	No
JUDGE 1	Yes	6	1
	No	1	19

Computation of Reliability for Question F

Yes/Yes = (14/54)(13/53) X 27 = 1.72

Yes/No = (14/54)(40/53) X 27 = 5.28

No/Yes = (40/54)(14/53) X 27 = 5.28

No/No = (40/54)(39/53) X 27 = 14.72

$\therefore 1 - \frac{\text{Observed Disagreements}}{\text{Expected Disagreements}} =$

$1 - \frac{1 + 1}{10.56} = .81 =$ 81% better than chance

(Krippendorf, 1980, p. 134)

APPENDIX D

DEFINITIONS OF ABBREVIATIONS

USED IN CHAPTER IV

Physical Education Subspecialties

Func Effects = Functional Effects of Physical Activity
Socio & Beh = Sociocultural & Behavioral Aspects
M Lng & Dev = Motor Learning & Development
M & M Analy = Mechanical & Muscular Analysis of Motor Skills
Mgmt T & P = Management Theory & Practice
Prgm Dev = Program Development
Meas & Eval = Measurement & Evaluation

Physical Education Research Journals

RQES = Research Quarterly for Exercise and Sport
JOSP = Journal of Sport Psychology
JOSB = Journal of Sport Behavior
ROSL = Review of Sport and Leisure
JTPE = Journal of Teaching in Physical Education
JSSI = Journal of Sport and Social Issues
DNRJ = Dance Research Journal

Statistical Techniques

ACV1 = One-way Analysis of Covariance (ANCOVA)
ACV2 = Two-way ANCOVA
ACV3 = Three-way (or greater) ANCOVA

ACR1 = One-way Repeated Measures Analysis of Covariance (ANCOVR)
ACR2 = Two-way ANCOVR
ACR3 = Three-way (or greater) ANCOVR

ALFA = Alpha Factor Analysis

ANV1 = One-way Analysis of Variance (ANOVA)
ANV2 = Two-way ANOVA
ANV3 = Three-way (or greater) ANOVA

AVR1 = One-way Repeated Measures Analysis of Variance (ANOVR)
AVR2 = Two-way ANOVR
AVR3 = Three-way (or greater) ANOVR

BISC = Biserial Correlation
BFSH = Behrens-Fisher Statistic
BTPH = Bonferroni t Multiple Comparison Test

CAFA = Canonical Factor Analysis
CANC = Canonical Correlation
CCAL = Cronbach's Coefficient Alpha
CLFA = Cluster Analysis

CONM = Confusion Matrix
CSQ1 = One-Sample Chi Square Analysis
CSQ2 = Two-Sample Chi Square Analysis
CVAR = Coefficient of Variation

DBPH = Duncan-Bonner Post Hoc Test
DETT = Dependent t-Test
DFNA = Discriminant Function Analysis
DIFP = Difference in Proportions
DNMR = Duncan's New Multiple Range Test

ESQC = Eta-Squared Correlation

FLSD = Fisher's Least Significant Difference Test

GENC = Generalizability Coefficient
GKGM = Goodman-Kruskal's Gamma

HTSQ = Hotelling's T-Square

ICCC = Intraclass Correlation Coefficient
IMFA = Image Factor Analysis
INTT = Independent t-Test

KCCW = Kendall's Coefficient of Concordance
KDRR = Kuder-Richardson Reliability
KTAB = Kendall's tau beta
KTAC = Kendall's tau C
KWAR = Kruskal-Wallis ANOVA by Ranks

LKGF = Likelihood Goodness-of-Fit Test
LLAN = Log Linear Analysis
LREG = Linear Regression

MAN1 = One-way Multivariate Analysis of Variance (MANOVA)
MAN2 = Two-way MANOVA
MAN3 = Three-way (or greater) MANOVA

MCLA = Multiple Classification Analysis

MCV1 = One-way Multivariate Analysis of Covariance (MANCOVA)
MCV2 = Two-way MANCOVA
MCV3 = Three-way (or greater) MANCOVA

MEAN = Arithmetic Mean
MEDI = Median
MLFA = Confirmatory Maximum Likelihood Factor Analysis
MODE = Mode
MREG = Multiple Regression

MVR1 = One-way Repeated Measures Multivariate Analysis of Variance (MANOVR)
MVR3 = Three-way (or greater) MANOVR

MWUT = Mann-Whitney U Test

NKPH = Newman-Keuls Post Hoc

OMSQ = Omega Squared

PARC = Partial Correlation
PATH = Path Analysis
PCFA = Principal Components Factor Analysis
PFLT = Pearson-Filon Test

PPMC = Pearson's Product Moment Correlation
PRAN = Profile Analysis
PROP = Proportion
PTBC = Point Biserial Correlation

RANG = Range
RTFA = R-Type Factor Analysis

SCHE = Scheffé Test
SDEV = Standard Deviation
SERM = Standard Error of the Mean
SIAM = Scored-Interval Agreement Method
SROC = Spearman's Rank Order Correlation

TAPH = Tukey's Alpha Test
TBPH = Tukey's Beta Procedure
TETC = Tetrachoric Correlation
THSD = Tukey's Honestly Significant Difference Test
TNDA = Trend Analysis
TOPH = Tukey's Omega Test
TWSD = Tukey's Wholely Significant Difference Test

VARI = Variance

WILT = Wilcoxon t-Test

ZTST = z Critical Ratio Test

APPENDIX E

CODING FOR CLASSIFICATION OF STATISTICS

Category of Statistics		Individual Statistical Analyses						
Central Tendency	=	MEAN	MEDI	MODE				
Dispersion	=	RANG	SDEV	VARI	CVAR	SERM	PROP	
Correlaton	=	PPMC	SROC	BISC	PTBC	TETC	ESQC	PARC
Association	=	GKGM	KTAB	KTAC	KCCW	CSQ2	OMSQ	
Reliability	=	SIAM	CCAL	ICCC	KDRR	GENC		
Goodness-of-Fit	=	CSQ1	LKGF	LLAN*	ZTST*			
t-Test	=	DETT	INTT	BFSH	WILT	MWUT	DIFP	
ANOVA	=	ANV1-3	AVR1-3	TNDA	KWAR			
ANCOVA	=	ACV1-3	ACR1-3					
MANOVA	=	MAN1-3	MVR1-3	HTSQ	PRAN			
MANCOVA	=	MCV1-3	MCR1-3					
Regression	=	LREG	MREG	DFNA	CANC	PATH	MCLA	CONM
Factor Analysis	=	PCFA	ALFA	CAFA	IMFA	CLFA	MLFA	RTFA
Multiple Comparison	=	THSD	SCHE	DNMR	NKPH	BTPH	DBPH	FLSD
		TAPH	TBPH	TWSD	TOPH	PFLT		

* Log linear analysis and z-test have been classified as "goodness-of-fit" tests because of their particular use in the sampled research reports. Log linear analysis was used to test a model. The z-tests were used to test the differences between a sample estimate and a known population value.