

THE QUALITY OF EXPERIMENTAL
METHODOLOGY IN COUNSELING AND
COUNSELOR EDUCATION

Dissertation for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
CONSTANCE C. RIPSTRA
1974



24841781



This is to certify that the
thesis entitled
THE QUALITY OF
EXPERIMENTAL METHODOLOGY IN
COUNSELING AND COUNSELOR EDUCATION
presented by

Constance C. Ripstra

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling

Bob B. Winborn

Major professor

Date July 19, 1974

0-7639



ABSTRACT

THE QUALITY OF EXPERIMENTAL METHODOLOGY IN COUNSELING AND COUNSELOR EDUCATION

By

Constance C. Ripstra

The purpose of this study was to systematically evaluate the quality of experimental research which has been published in the fields of counseling and counselor education from 1962 through 1973. Attention was directed at the methodology and reporting of studies rather than at the subject matter or variables being examined. The specific independent variable was time, in order to determine whether there has been an improvement since 1962 in the quality of published research. Four three-year spans were chosen as levels of the independent variable: 1962 - 1964; 1965 - 1967; 1968 - 1970; 1971 - 1973.

Following a survey of three journals, Journal of Counseling Psychology, Personnel and Guidance Journal, and Counselor Education and Supervision, to specify the population of pre-, true-, and quasi-experimental studies, a sample of 38 studies was randomly chosen for each year

690208

Constance C. Ripstra

span. Each study was evaluated by a trained rater on the Evaluation Instrument for Experimental Methodology, which produced six measures of the quality of reporting and methodology. Three raters independently rated the studies. Fifteen randomly chosen studies were commonly rated to establish the average interrater reliability estimate of .78.

A 1 x 4 design with equal cell sizes was utilized to examine for differences between the four year spans. A multivariate analysis of variance using orthogonal polynomials was used to test the hypotheses of the trend of the quality over time. A slight linear trend was distinguished across the four year spans. Graphic illustration demonstrated a very slight positive increasing trend over time. Examination of the means derived from the EIEM for the last year span revealed that the quality of reporting and the introduction was "clearly adequate." However, quality of the method, results, and discussion sections was generally "barely adequate." In total the quality of experimental research in counseling and counselor education was characterized as "barely adequate."

THE QUALITY OF EXPERIMENTAL METHODOLOGY IN
COUNSELING AND COUNSELOR EDUCATION

By

Constance C. Ripstra

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services
and Educational Psychology

1974

TABLE OF CONTENTS

Chapter	Page
I. THE PROBLEM, RATIONALE, AND RELATED RESEARCH	1
Rationale.	1
Purpose	5
Review of the Literature.	6
Reporting	6
Sampling and Generalization.	8
Designs and Controls	10
Measurement and Criteria.	12
Analysis	13
Replication	15
Hypothesis	16
Summary	16
II. EXPERIMENTAL DESIGN AND METHODOLOGY	18
Sample.	18
Instrument	23
Procedures	25
Design and Statistical Analysis	29
III. ANALYSIS OF RESULTS	31
Preliminary Data	31
Test of Hypotheses.	33
Observations.	41
Summary	42
IV. SUMMARY AND DISCUSSION	44
Summary	44
Discussion	45
Recommendations.	51
Conclusion	53

APPENDICES

Appendix

A. Frequency Count of the Number of Experimental, Correlational and Miscellaneous Studies for the Three Journals for the Four Year Spans	55
B. Evaluation Instrument for Experimental Methodology	56
C. Relevant Definitions	60
D. Notes on Rating.	61
E. Means and Standard Deviations for Items of the Evaluation Instrument for Experimental Methodology	62
F. Univariate Tests of the Six Dependent Variables for a Linear Trend.	64
G. Principle Components of the Correlation Matrix for the Six Dependent Variables of the ELEM	65
H. Ninety-five Percent Confidence Interval for Estimated Means of the Six Dependent Measures of the ELEM	66
REFERENCES.	67

LIST OF TABLES

Table	Page
2.1. Frequency Count of the Number of Experimental Studies for the Three Journals and Four Year Spans in the <u>Population</u> . . .	21
2.2. Frequency Count of the Number of Experimental Studies for the Three Journals and Four Year Spans in the <u>Sample</u> . . .	22
2.3. Hoyt Reliability Estimates for the Fifteen Commonly Rated Studies and the Five Studies Rated for a Validity Estimate in the Order of Rating.	28
3.1. Mean Scores and Corrected Standard Deviations for the Scales of the EIEM . . .	32
3.2. Sample Intercorrelation Matrix for Scales of the EIEM	33
3.3. Multivariate Test for Orthogonal Polynomials for Six Scales of EIEM . . .	34
3.4. Multivariate Test for Orthogonal Polynomials for Two Overall Items of the EIEM	35
3.5. Slopes of Estimated Means for Year Spans .	35

LIST OF FIGURES

Figure	Page
3.1. Observed Means on the Six Measures of the EIEM	36
3.2. Estimated Means on the Six Measures of the EIEM	37
3.3. Graphic Description of Univariate Confidence Intervals for Six Scales of EIEM. . . .	40

CHAPTER I

THE PROBLEM, RATIONALE, AND
RELATED RESEARCH

Rationale

The aim of science above all else is to discover new and useful information in the form of verifiable data, that is, of data obtained under conditions such that other qualified people can make similar observations and obtain the same results. This calls for orderliness and precision in uncovering relationships and communicating them to others. (Hilgard, 1962, p. 9)

Counseling psychology is usually considered an applied science, and presumably the aim of science given in the above quotation is also a goal for this branch of psychology. Some counseling psychologists (Hansen & Warner, 1971; Thoresen, 1969; Whiteley, 1967) have questioned whether the profession is making significant progress toward this goal. The quality of research studies has been questioned and calls for improvement have been made (Kelley, et al., 1970; Pawlicki, 1970; Schmidt & Pepinsky, 1965; Thoresen, 1969). There are several considerations which make it imperative to pay attention to these professional needs. One is that the profession may be building a research base on a

foundation of sand. If an initial study in a particular area shows statistically significant results, the tendency is to take the results as truth and continue investigation of the problem in an attempt to further define the construct of interest. Due to the reticence of most professionals to replicate studies (Smith, 1970), the significance is never tested. Consequently, further research or conclusions may build upon a faulty base.

The probability of a faulty base increases considerably when the methodology of the study is examined. "Research which is not well formulated is more than worthless since it becomes deceptive as well (Whiteley, 1967, p. 281)." It is probable that a majority of research has significant errors that confuse or invalidate the results entirely or restrict conclusions to the sample. Glass and Robbins (1967) expertly demonstrated this in an evaluation of studies by Delacato and his associates in the field of reading theory. All of the empirical studies cited by Delacato as supporting his theory of the role of neurological organization in reading were shown to contain major faults. Thus, Glass and Robbins illustrated how research can build on a faulty base of prior research and seemingly validate a theory without legitimate evidence.

A second reason for examining the quality of research in counseling is that most research does not

have a likely chance of rejecting the null hypothesis unless the treatment effect is powerful (Cohen, 1962). The choices of design, sample size and analysis are often not appropriate, and, therefore, the study does not have sufficient precision to correctly reject the null hypothesis of no treatment effects. For the profession this may mean much effort and time expended for little more than a researcher's personal experience in the experimental process. Consequently, the progress of counseling to establish a research base can be inhibited by design and methodological errors.

The research progress of the profession is also slighted when researchers do not closely examine their data for findings not directly associated with the stated hypotheses. As Tukey (1969) stated, "Data analysis needs to be both exploratory and confirmatory [p. 90]." Therefore, when an experimenter stops after his data analysis, the scientific endeavor is halted at the beginning of the process (Eastwood, 1967).

While the consensus of the literature is that there is a lack of well-planned and executed research in counseling, the authors of such conclusions base their comments on varying types of data. Some are communicating intuitive feelings about the state of counseling research (Coleman, 1957; Dressel, 1953; Fisher & Roth, 1961; Holland, 1974). Others reach the same conclusion

following a systematic review of the literature on group counseling (Gazda & Larsen, 1968), practicum supervision (Hansen & Warner, 1971), behavior therapy with children (Pawlicki, 1970), and research published in 1963 (Schmidt & Pepinsky, 1965). An occasional astute reader has written a critical review of a published research study which has reporting or methodological flaws (Crittenden, 1973; Marks, Conry, & Foster, 1973; Mills & Mencke, 1967; Sieka, Taylor, Thomason, & Muthard, 1971).

While several have reviewed counseling journals to examine such variables as types of statistics used (Edgington, 1964), the institutional sources of published research (Goodstein, 1963), common errors in manuscripts submitted for publication (Smith, Smith, Schefers, & Steinmann, 1971), and publication trends, empirical versus theoretical papers (Foreman, 1966), only one study has been published which systematically evaluated the methodological quality of counseling research.

Kelley, Smits, Leventhal, and Rhodes (1970) critiqued the designs of all empirical studies published in the Journal of Counseling Psychology from 1964 through 1968. Using Campbell and Stanley's system (1963), they labeled the designs as pre-experimental, true-experimental, or quasi-experimental and rated the studies according to the internal and external validity criteria. Their evaluation, however, scrutinized only one aspect of research methodology, that of design.

Purpose

Several authors recommend that a qualitative analysis of published counseling research be pursued (Foreman, 1966; Samler, 1958; Stone & Shertzer, 1964). The purpose of the present study was to accomplish such an evaluation. It empirically determined those aspects of methodology which are consistently weak in published research related to counseling and counselor education. Specifically, the intent of the investigator was to systematically evaluate the quality of experimental research which has been published in the fields of counseling and counselor education from 1962 through 1973. Such information can be used in several ways: as a baseline of the status of published counseling research at a given point in time; as an attentional device directed at the need for more carefully executed research; as an educational tool for those who read and evaluate professional research; as an educational tool for those who teach research skills for improvement and/or reemphasis; and as feedback to editors of journals for improvement in review and acceptance criteria. By examining the quality of research across years, one may conclude, as some have postulated (Carkhuff, 1965; Myers, 1966; Patterson, 1963), whether in fact the quality of research is improving.

This investigator recognizes that this study has examined only one of the two issues of quality of counseling and counselor education research. The research methodology has been evaluated, while relevance of the results and studied phenomena to the profession has not been examined. While neither is a sufficient condition, both are necessary conditions for quality research in a profession.

The overall rationale of outcome of the study was to encourage what is implied by Lykken (1968):

The value of any research can be determined, not from the statistical results, but only by skilled, subjective evaluation of the coherence and reasonableness of the theory, the degree of experimental control employed, the sophistication of the measuring techniques, the scientific or practical importance of the phenomenon studied, and so on.
[pp. 158-159]

Review of the Literature

Many articles are devoted to examining the recurring methodological problems encountered in counseling research reports. Among these are problems with reporting, sampling and the accompanying difficulties in generalization, design, controls, measurement and criteria, analysis, and the lack of replication. These problems will be discussed in the next sections.

Reporting

The relevance of good reporting lies mainly with the issue of replication, although its benefits also

contribute to valid evaluation and reliable usage of results. Inadequate reporting is a commonly made criticism of counseling research. A recommendation made to the Division of Counseling Psychology, American Psychological Association, concerning modifications in scientific inquiry and reporting was to encourage " . . . a practice of reporting in greater detail the research methodology employed, the characteristics of the clients, the precise nature of the professional interventions, and the outcome measures (Whiteley & Allen, 1969, p. 84)." Others note specific deficiencies which commonly occur in counseling publications: lack of clear and concise definition of the problem of interest (Harrison, 1971; Smith, Smith, Scheffers, & Steinmann, 1971); inadequate statements regarding treatment process, counselors' theoretical orientation or qualifications in therapy research (Gazda & Larsen, 1968; Kiesler, 1966b; Patterson, 1966; Pawlicki, 1970; Whiteley & Allen, 1969); description of dependent variables (Kiesler, 1966b); and poor usage of grammar and style (Smith, Smith, Scheffers, & Steinmann, 1971). Other authors make general comments about the importance of careful and complete reporting of disciplined inquiry (Fisher & Roth, 1961; Kelley, Smits, Levanthal, & Rhodes, 1970; Orne, 1962; Spithill, 1973; Thoresen, 1969). Kelley, et al. (1970), suggest that authors not only

specify all details of procedures but also include statements of inadequacies in their studies.

Sampling and Generalization

Sampling refers to the process of defining a population of interest and then, assuming it is too large to use in its entirety, to choose a sample from which inferences can be generalized to that population. Orne (1962) feels that "ecological validity," generalization, is one of the two requirements for meaningful experimentation. The ideal procedure of sampling from a population is random selection of a sample sufficiently large to satisfy statistical considerations.

The fields of counseling and counselor education must contend with the usual problems encountered by those professions interested in human beings. The population of interest is often spread across the nation, if not world, and therefore, too often the sampling procedure is dictated by proximity or convenience. The consequences of such sampling procedures are usually seen in inaccurate and illegitimate generalizations beyond the sample. In counseling research one must be aware of many populations of interest possible even in a single study: counselors, counselees, counselor educators, methods and techniques, environmental-situational variables, and measuring variables (Meltzoff & Kornreich, 1970; Patterson, 1960). The use of volunteers poses a common problem (Orne, 1962;

Patterson, 1956), as does the frequent use of counselor trainees when the population of interest is counselors (Herr, 1964; Patterson, 1966). In an evaluative survey of counseling process and outcome studies, Kelley, et al. (1970), found that 61.6% of the studies reviewed had an interaction between subject selection and treatments, which is a source of external invalidity (Campbell & Stanley, 1963). Each of these problems results in limited generalization.

An additional problem often encountered in counseling research using group design is small sample size. Individual differences of humans create a problem for sampling. To assume a representative sample on all variables which contribute to the problem of interest, a large sample size is required (Cohen, 1962; Tukey, 1969). Reviews of group counseling research (Gazda & Larsen, 1968), abnormal-social psychology (Cohen, 1962), behavior therapy with children (Pawlicki, 1970) and psychotherapy research (Meltzoff & Kornreich, 1970) point out the consistent use of small sample size. Although there are a number of problems created with small n (Tversky & Kahneman, 1971), the post hoc solution is replication of the study (Patterson, 1956). Unfortunately, replication studies are not valued as professional activity (Barker & Gurman, 1972).

Though sampling procedures are recognized as important aspects of research (Coleman, 1957; Patterson, 1963), much counseling research cannot legitimately be generalized beyond the sample because of restrictions due to error (Dressel, 1954; Krause, 1972). However, the argument has been made that the data from a nonrandomly selected sample may be generalized to the type of population which the sample characterizes (Cornfield & Tukey, 1956). Implicit is the requirement that the sample be very carefully described so that the reader can infer beyond the sample. Unfortunately, as was noted in a previous section, the general quality of reporting in counseling research is inadequate. Thus, many studies cannot use the Cornfield-Tukey Argument to allow generalization beyond the nonrandom sample.

Designs and Controls

Kelley, et al. (1970), evaluated studies published in the Journal of Counseling Psychology from 1964 through 1968 by using Campbell and Stanley's (1963) criteria for design analysis. The majority of studies were found to have sources of invalidity that were not controlled in the design. They concluded that this group of studies "has little relevance beyond that of generating testable hypotheses [p. 340]." Dressel (1953) came to the same conclusion following an evaluation

similar to Kelley's. Of twelve studies reviewed in detail, ten had errors in design.

In their survey of counseling research Kelley, et al. (1970), found that the designs of a majority of the published studies they reviewed were classified as pre-experimental. Such designs have a treatment group but no adequate comparison or control group. Results from such designs should be considered tentative, and no causal inferences can be made legitimately. However, Gazda and Larsen (1968) found that 70% of the group counseling studies reviewed had "true-experimental" designs. These designs have adequate controls for evaluating treatment effects and allow causal statements to be made.

In experimental studies control of all contributing variables is desirable in order to say with some degree of confidence that the change in the dependent variable is due to the manipulated variable. Problems of improper or absent consideration of control seem to be a major criticism of counseling research (Calvin, 1954; Coleman, 1957; Dressel, 1953; Harrison, 1971; Hobbs & Seeman, 1955; Kiesler, 1966b; Patterson, 1963, 1966). Pawlicki (1970) evaluated behavior therapy research with children and found that 85% did not provide a control group. This is misleading, however, as many of the reviewed studies were single-subject designs.

Gazda and Larsen (1968) report that 15% of the group and multiple counseling research studies published prior to 1967 did not report use of control groups or statistical controls.

The use of statistical control through use of analysis of covariance does not seem widespread, though its use is recommended in counseling literature (Feldman & Hass, 1970; Herr, 1964; Patterson, 1956, 1963). Matching seems to remain a favorite technique of counselor researchers (Patterson, 1956), despite warnings of loss of power and difficulties in obtaining truly matched groups (Campbell & Stanley, 1963; Feldman & Hass, 1970). Recommended changes in design include utilization of factorial designs to simultaneously investigate and control the many variables which are thought to contribute to human interaction and learning (Ford, 1959; Kiesler, 1966b; Whiteley & Allen, 1969).

Measurement and Criteria

While it is generally recognized that instrumentation is a major aspect of any scientific endeavor (Coleman, 1957; Thoresen, 1969), inadequate measuring devices continue to contribute to the problems in counseling research. The measurement of process and outcome variables seems to be a major stumbling block for counseling research (Herr, 1964). Jensen, Coles, and Nestor (1955) specify that the necessary characteristics

of a criterion variable include definability, stability, and relevance. These are apparently difficult to attain. Many researchers choose as dependent variables standardized educational or psychological instruments or "home-made" rating instruments (Kiesler, 1966a). Independent raters are often employed (Bordin et al., 1954). These introduce error of measurement which contribute to a reduction of the needed power to correctly reject null hypotheses. Poor choice of dependent measures also contributes to the preponderance of irrelevant research.

An additional consideration arises because of the subject of counseling research. Many variables typically contribute to a concept, and, therefore, to prevent imposing unidimensionality on it, multivariate models must be encouraged (Borden et al., 1954; Edwards & Cronbach, 1952; Fisher & Roth, 1961; Gazda & Larsen, 1968; Lachenmeyer, 1970; Thoresen, 1969). This means inclusion of those dependent variables thought to be affected by the independent variables.

Analysis

When compared to other aspects of research, analysis is infrequently pointed to as a source of methodological error in counseling research. Criticisms center not on the inappropriateness of those statistics used, but on insufficient use of available techniques or procedures which add to the analysis

process. Thoresen (1969) supports Tukey's (1969) arguments for going beyond the statistical significance test; Nunnally (1960) concurs. "Such 'detective work' facilitates serendipity . . . (and) careful analysis and re-analysis may suggest new hypotheses and provide the basis for speculation . . . (Thoresen, 1969, p. 268)."

Likewise, Nunnally (1960) and Tversky and Kahneman (1971) advocate the use of confidence intervals in addition to the traditionally used hypothesis tests. Such reporting gives more information than a statement of significance. Kiesler (1966b) and Nunnally (1960) suggest that variances of groups be examined for differences in addition to the traditional analysis of means. Nunnally (1960) and Thoresen (1969) stress the use of statements of meaningful significance in preference to the use of the .05 statistical level of significance. Nunnally (1960) also questions the wide use of significance tests of limited meaning in correlational studies, where a significant finding usually specifies only that the correlation is not zero. Cohen (1962) and Tversky and Kahneman (1971) offer the criticism that most research does not consider or report a value for beta, the probability of a Type II error, which is a decision to not reject a false null hypothesis. Cohen (1962) reviewed all the articles published in 1960 in the Journal of Abnormal and Social Psychology. By an

analysis of beta, he concluded that none of the articles had a chance of rejecting the null hypothesis unless the treatment effect were large. The implication is that under present conditions the probability of correctly rejecting the null hypothesis is small. Suggestions relating to sample size, control of error variance, alpha level and size of treatment effects are made.

Replication

Two aspects of replication are important: the quality of reporting and choice of procedures which allows for replication (Cronbach & Suppes, 1969; Orne, 1962); and the frequency with which it occurs in professional literature. The first has been commented on in a previous section, while the second has been alluded to in a number of sections. That replication is a necessary component in any research plan is recognized often in professional articles (Herr, 1964; Kiesler, 1966b; Krause, 1972; Lykken, 1968; Nunnally, 1960; Smith, 1970; Stanley, 1967; Thoresen, 1969). "In studies where random sampling from a defined population is difficult or impossible, it is of crucial importance that a number of replications be planned as part of the original design or be carried out by other workers (Patterson, 1955, p. 255)." "Confirmation comes from repetition (Tukey, 1969, p. 84)." However, Smith (1970) concludes that replication is rarely done for either experimental

or correlational studies; Gazda and Larsen (1968) found no replication studies in their comprehensive review of group and multiple counseling research.

Hypothesis

The following hypothesis was the primary focus of this investigation:

Differences exist between the scores on the Evaluation Instrument for Experimental Methodology for the four groups of years of published counseling research, such that there is an increasing linear trend, indicating an increase in the quality of the research across the year spans.

A statistical significance level of .05 was used. It was deemed a reasonable value when considering both Type I and II decision errors. Meaningful significance was especially relevant for examination of specific items of the Evaluation Instrument for Experimental Methodology. To establish a summary of research weaknesses found across the articles, the means of individual items were examined. An item whose mean was less than four would indicate an aspect of research that was rated less than adequate across the sample of experimental studies.

Summary

The purpose of this study was to systematically evaluate the quality of experimental research which has been published in the fields of counseling and counselor

education from 1962 through 1973. Attention was directed at the methodology and reporting of studies rather than at the subject matter or variables being examined. The variable of specific interest was time. Has there been an improvement over time from 1962 through 1973 in the quality of published research? The results will be most pertinent to counselor researchers and educators, for the journals from which research studies were selected are those journals regularly read by these members of the profession and which publish empirical studies. The data consist of scores on the Evaluation Instrument for Experimental Methodology, which examines the quality of design, procedures, analysis, and reporting.

CHAPTER II

EXPERIMENTAL DESIGN AND METHODOLOGY

Sample

The population of interest was group experimental studies published from 1962 through 1973 in the three major counseling and counselor education journals: Journal of Counseling Psychology, Personnel and Guidance Journal, and Counselor Education and Supervision. The three journals were chosen as the major publication outlets for experimental studies for counselor researchers and educators. The choice of two of the journals is supported by the empirical evidence that the Journal of Counseling Psychology and Personnel and Guidance Journal were cited most often in a survey of the references of published articles (Cotton & Anderson, 1973). Counselor Education and Supervision, as the publication of the Association of Counselor Education and Supervision, is the official journal for professional counselor educators.

The term "experimental study" was operationally defined as a study in which at least one variable was manipulated and the effects on another variable were

observed (Campbell & Stanley, 1963). In other words the experimenter systematically introduced a treatment and recorded results of that treatment on some variable(s). Three types of experimental studies are described by Campbell and Stanley (1963): pre-experimental, true-experimental, and quasi-experimental designs. All were considered part of the population of interest.

The complete population of experimental studies was specified by the following process. Two individuals competent in research design and statistics labeled each empirical study published from 1962 through 1973 in the three journals as either experimental, correlational or miscellaneous (see Appendix A). One of the experts had a Ph.D. in research and statistics and at the time was employed as a research associate. She had taught three statistics classes and during her degree program had worked as a research consultant for three years. The investigator of the present study, serving as the second consultant, had completed five of seven courses of a cognate in research methodology in a Ph.D. program. She had earned grades of 4.0 in the completed research and statistics classes and for four terms had been a graduate assistant for the research methodology series offered by the College of Education, Michigan State University.

An empirical study was considered to be any study which contained a report of a systematic collection of data. The definition of "experimental" was given in the paragraph above. A correlational study was defined as a study that compared existing groups of individuals on some dependent measure. Studies not classifiable as either experimental or correlational were labeled as miscellaneous. Surveys and factor analytic studies comprised the majority of these. The studies designated as correlational or miscellaneous were not included in the population of interest.

A total sample size of 152 was decided upon because it was the largest possible sample size if equal cell sizes were to be maintained. The population size of the first year span, 1962 - 1964, was 38, thereby setting 38 as the largest possible cell size. The sample of 152 studies to be evaluated were randomly selected from the total population of 363 experimental studies (Table 2.1). Specifically, the complete population for the year span 1962 - 1964, 38 studies, was included in the sample. The decision to use the population resulted in a reduction of the error variance. The samples of 38 studies for each of the remaining three year spans were randomly selected from the respective populations. Table 2.2 describes the sample according to year span and journal. The sample size was 41.87% of the population size.

Table 2.1
Frequency Count of the Number of Experimental Studies for the
Three Journals and Four Year Spans in the Population

	1962- 1964	1965- 1967	1968- 1970	1971- 1973	Totals	% of Total
<u>Journal of Counseling Psychology</u>	24	53	75	93	245	67%
<u>Personnel and Guidance Journal</u>	12	19	26	4	61	17%
<u>Counselor Education and Supervision</u>	2	4	23	28	57	16%
Totals	38	76	124	125	363	100%

Table 2.2
Frequency Count of the Number of Experimental Studies for the
Three Journals and Four Year Spans in the Sample

	1962- 1964	1965- 1967	1968- 1970	1971- 1973	Totals	% Of Total
<u>Journal of Counseling Psychology</u>	24	26	22	27	99	65%
<u>Personnel and Guidance Journal</u>	12	10	8	2	32	21%
<u>Counselor Education and Supervision</u>	2	2	8	9	21	14%
Totals	38	38	38	38	152	100%

Of the 152 studies in the sample, 40 studies or 26% were pre-experimental, 79 or 52% were true-experimental, and 33 studies, 22% of the sample, were quasi-experimental designs. Sixty-one percent, 93 studies, were applied research studies with outcome measures, and 31% of 47 studies were applied research with process measures. Twelve studies, 8%, were considered basic research. Six studies, 4% of the sample, were master degree theses, and 30 studies or 20% of the studies were doctoral dissertations. Thirty studies or 20% were at least partially supported by a grant. As the sample was randomly selected, these can be considered estimates of the specified population's characteristics.

Instrument

Assessment of the reporting and methodology of the studies was done using the Evaluation Instrument for Experimental Methodology (EIEM) (Appendix B), a rating form developed by the investigator. It has 37 Likert-scaled items, each item having six response options. Thirty-five items are divided into five sections, four of which correspond to the traditional sections of an experimental report: reporting (9 items), introduction (5 items), methods (8 items), results (7 items), and discussion (6 items). Two additional items provide an overall rating of the reporting and methodology. The reporting section evaluates the clarity of writing and description throughout the study. The introduction

section covers the literature review, purpose and hypothesis statements, and definition of the independent variables. The methods section includes items on the appropriateness of the dependent variables, sampling, subject assignment and design. The results section evaluates the statistical analysis. The discussion section includes assessment of the conclusions, generalizations, and qualifications of the study. A mean score is reported for each section, and a mean rating for the entire instrument is given as a total score.

The instrument was constructed by a compilation of the recurring problems in experimental counseling research cited previously in Chapter I. Special attention was also given to Smith, Smith, Scheffers, and Steinmann's (1971) survey of the common errors which occur in psychological studies. Other guides to the evaluation of research (Burck, Cottingham, & Reardon, 1973; Borg, 1963; Davitz & Davitz, 1967; Farquhar & Krumboltz, 1959; Isaac & Michael, 1971; Roberts, 1969), as well as experts in research methodology in the Department of Educational Psychology, Michigan State University, were consulted during the initial and trial stages of instrument development.

The interrater reliability of the instrument for three raters prior to the data collection was calculated as .79 using Hoyt's Analysis of Variance (1941). During the data collection an average reliability estimate

was also calculated for the three independent raters as .78 on fifteen studies evenly distributed throughout the evaluation process. This estimate was considered high enough to substantiate having one rater evaluate the quality of a study. An attempt was made to estimate the validity of the instrument. The two consultants described earlier as having detailed the population of research studies and considered qualified in the field of research methodology evaluated five of the same studies on which the interrater reliability was calculated. The average Hoyt's ANOVA value for this form of concurrent validity estimate was .85. This was considered high enough to conclude that the instrument was reasonably valid for the intended use of evaluation of experimental methodology.

Procedures

Random selection of the sample was accomplished by use of a random numbers table. Fifteen of the total of 152 studies were randomly chosen to be independently rated by all of the raters in order to establish interrater reliability estimates. The remaining 137 were randomly ordered and then assigned to the three raters. The fifteen studies designated for reliability checks were evenly placed throughout the sequence of the other studies for each rater. The random sequence of the studies was intended to avoid a time or fatigue bias,

and the random assignment to rater was done to avoid a rater bias. Prior to the rating process each study was photocopied and blinded for journal name, author's name and affiliation, and dates.

Three individuals were paid to rate the studies. They were recommended as superior students in the research design and statistics classes at Michigan State University by the professor who taught those classes. Each had successfully completed the three basic research courses offered by the Department of Counseling, Personnel Services and Educational Psychology: Quantitative Methods in Education, Advanced Quantitative Methods in Education, and Experimental Design in Education. Two raters had also completed a nonparametric statistics course. Rater A was a doctoral student in counselor education and had completed the three-course statistics series immediately prior to the rating process with a 4.0 or "A" grade in each course. Rater B was a doctoral student in statistics and had worked as an assistant to a research consultant. She also had finished the three-course statistics series, as well as an advanced course in nonparametric statistics, with a 4.0 grade in each. Rater C was a doctoral student in rehabilitation counseling and had completed the same four courses as Rater B with 4.0 grades. He had taught experimental psychology, which included research design and statistics,

for four years at the college level. The high reliability with the two consultants tends to support the above evidence of the raters' competence for the rating task.

Training with the EIEM took place immediately prior to the rating process. It consisted of independent evaluations of randomly chosen studies from the remaining population of interest after the sampling process. Group discussion of the rating of each item was conducted in order for the three raters to agree on the meaning of a particular item. In several instances this discussion resulted in a revision of the instrument. In addition, each rater was provided with definitions of relevant terms (Appendix C) and an instruction sheet for the rating process (Appendix D).

During the two-week rating process each rater worked independently. Checks were made at fifteen points throughout the process to establish that a reliability of at least .70 was maintained. Table 2.3 contains the Hoyt's reliability coefficient for each of the fifteen studies in the order they were rated. If the reliability had gone below .70 for two successive studies, retraining sessions would have been held to reestablish the inter-rater reliability beyond the criterion of .70. In addition to rating each article the rater was asked to define the type of design (pre-, true-, or quasi-experimental), type of experiment (applied-outcome,

Table 2.3

Hoyt Reliability Estimates for the Fifteen Commonly
Rated Studies and the Five Studies Rated for a
Validity Estimate in the Order of Rating

Order	Study Number	Reliability ^a	Validity
1	83	.89	
2	108	.84	
3	54	.82	
4	42	.79	
5	130	.83	.85
6	143	.87	.89
7	18	.82	
8	110	.79	.84
9	76	.74	.83
10	75	.51	
11	117	.79	.82
12	90	.63	
13	45	.88	
14	63	.84	
15	61	.61	

^aStandard deviation of reliability estimates
equals .107.

applied-process, or basic research) and statistical tests used in each study (see Appendix B).

Design and Statistical Analysis

The independent variable of interest was years of published research in counseling and counselor education. The total time span of 1962 through 1973 was considered. This was divided into four levels, each level containing three years: 1962 - 1964; 1965 - 1967; 1968 - 1970; 1971 - 1973. Thus, the design of this correlational study is a 1 x 4 matrix with an equal number of observations per cell:

Years				
Y_1	Y_2	Y_3	Y_4	$Y_1 = 1962 - 1964$
$n_1 = 38$	$n_2 = 38$	$n_3 = 38$	$n_4 = 38$	$Y_2 = 1965 - 1967$
				$Y_3 = 1968 - 1970$
				$Y_4 = 1971 - 1973$

The statistical treatment was a multivariate analysis of variance using the six scores derived from the EIEM: reporting (REP), introduction (I), methods (M), results (R), discussion (D), and total (T). This analysis would specifically answer the research hypotheses. An analysis for orthogonal polynomials, linear, quadratic, and residual trends across the groups of years, was performed. It was done to establish whether

there has been a trend in the quality of methodology for published research across the year spans. Since the population number of published experimental studies in each year span was known, the analyses included a correction of the variance-covariance matrix for having a known finite population.

The hypotheses tested were:

Hypothesis 1:

There is a significant linear trend for the dependent measures across the four year spans.

Hypothesis 2:

There is a significant quadratic trend for the dependent measures across the four year spans.

Hypothesis 3:

There is a significant residual trend for the dependent measures across the four year spans.

CHAPTER III

ANALYSIS OF RESULTS

Statistical analyses were calculated at the Michigan State University Computer Center on a Control Data 6500 computer system. Use of the Michigan State University computer facilities was made possible through support, in part, from the National Science Foundation. Data analyses were generated by a multivariate analysis of variance program developed by Finn (1967) and a program for computing a corrected variance-covariance matrix by Scheifley (1973).

Preliminary Data

Mean scores and standard deviations for the four groups on the five subscales and one total score on the Evaluation Instrument for Experimental Methodology are shown in Table 3.1. The standard deviations reported are those used in the data analysis following a correction for having a finite population.

The mean ratings and standard deviations for all items in the EIEM are reported by group in

Appendix E. An item might not have been appropriate for a particular study and, therefore, was omitted by the rater. This is reflected in the differing number of studies included in the calculation of the mean for an item. Unless otherwise noted, the number of studies equals 38 for each group.

Table 3.1

Mean Scores and Corrected Standard Deviations
for the Scales of the EIEM

	Y ₁		Y ₂		Y ₃		Y ₄	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Reporting	4.53	.74	4.73	.65	4.63	.60	4.90	.57
Intro- duction	4.42	.84	4.73	.80	4.67	.75	4.90	.60
Method	3.19	.86	3.56	1.00	3.56	.85	3.77	1.06
Results	3.31	.95	3.71	.90	3.52	1.01	3.84	.91
Discussion	3.33	1.19	3.62	1.02	3.36	1.10	3.70	.92
Total	3.76	.70	4.09	.69	3.94	.65	4.22	.64

The sample within cell intercorrelation matrix for the scales of the EIEM are reported in Table 3.2. Using Fisher's r to z transformation (Glass & Stanley, 1970) with an alpha level of .05, the minimum sample correlation to be statistically significant from zero is .16. Therefore, each reported correlation is statistically greater than zero.

Table 3.2
Sample Intercorrelation Matrix for
Scales of the EDEM

	Rep	I	M	R	D	T
Reporting	1.00					
Introduction	.72	1.00				
Method	.57	.49	1.00			
Results	.50	.48	.45	1.00		
Discussion	.39	.37	.41	.46	1.00	
Total	.82	.74	.79	.75	.66	1.00

Test of Hypotheses

An analysis of orthogonal polynomials was accomplished to determine the form of relationship between the year spans for the six dependent variables. The purpose of this analysis, commonly called a trend analysis, was to determine whether the means of the dependent variables were influenced by changes in the independent variable. For this investigation the question was whether a trend over time existed for the quality of published experimental research. The results of the test for orthogonal polynomials are found in Table 3.3. The univariate F-tests for the test of a linear trend are shown in Appendix F.

A separate multivariate analysis of orthogonal polynomials was performed for the two overall items of

the EIEM (Table 3.4). The results were consistent with the analysis of the six EIEM scales.

Table 3.3
Multivariate Test for Orthogonal Polynomials
for Six Scales of EIEM

Test	F-ratio	df	p
Linear	2.135	1,148	< .05
Quadratic	.320	1,148	< .93
Residual	1.043	1,148	< .40

A significant linear relationship with a nonzero slope was found to exist across time. After graphing the observed and estimated means for each dependent measure (Figures 3.1 and 3.2), a slightly positive significant linear trend was evident. Therefore, the quality of methodology in counseling and counselor education has improved over the twelve years. However, as can be seen from the graphs of estimated means, the degree of increase is slight. The estimated slopes (Table 3.5), each defined as the increase in the mean of the dependent variable from one year span to the next, vary from .08 to .17 on the 1 - 6 scale used for the EIEM. For example, for the measure Reporting there will be a predicted increase of .10 on the criterion scale every three years.

Table 3.4

Multivariate Test for Orthogonal Polynomials for
Two Overall Items of the EIEM

Test	F-ratio	df	p
Linear	3.042	1,148	< .05
Quadratic	1.088	1,148	< .34
Residual	1.472	1,148	< .23

Table 3.5

Slopes of Estimated Means for Year Spans

Scale/Item	Slope
Reporting	.10
Introduction	.14
Method	.17
Results	.14
Discussion	.08
Total	.12
Overall Reporting Item	.12
Overall Methodology Item	.15

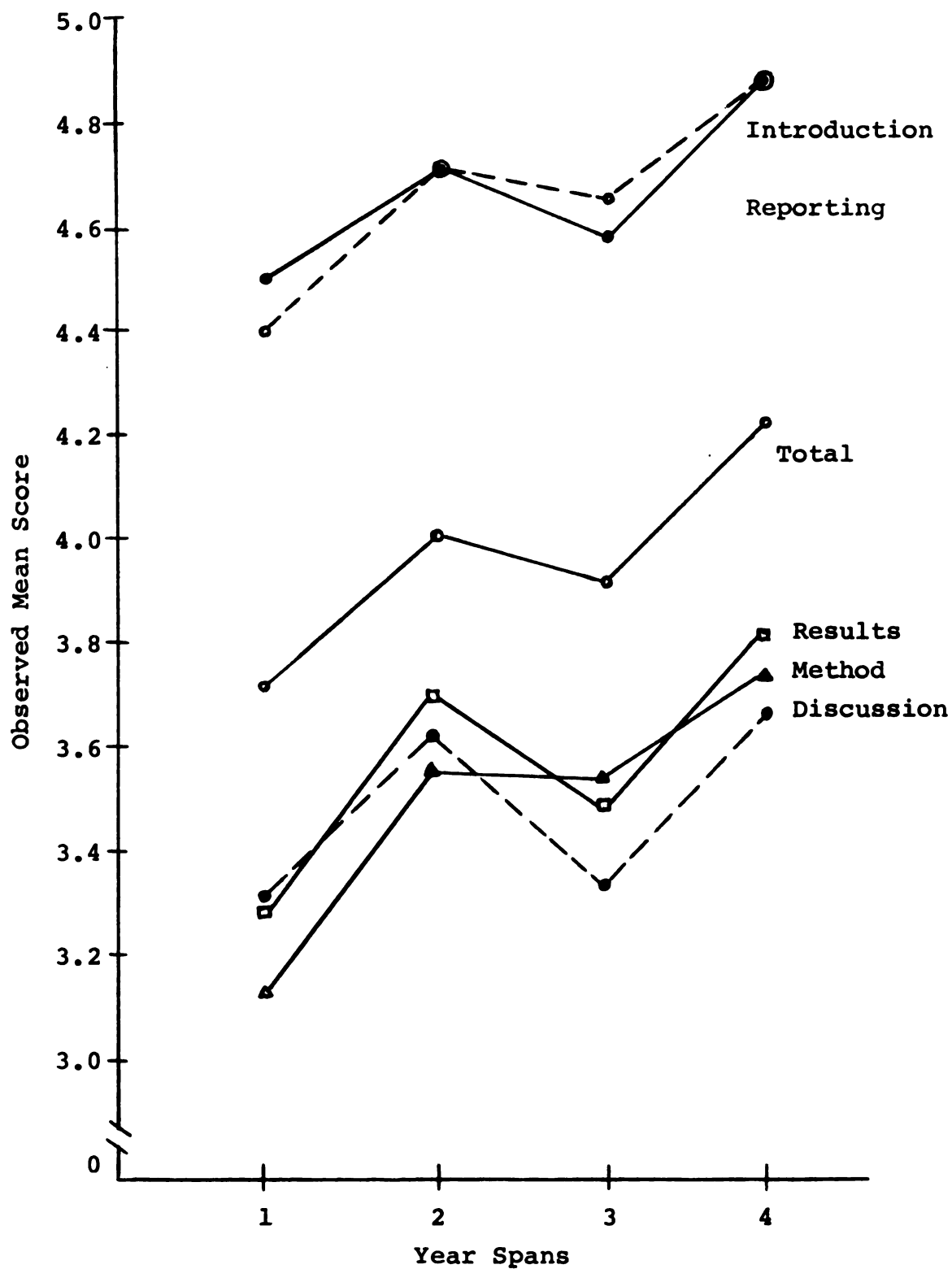


Fig. 3.1. Observed Means on the Six Measures of the EIEM.

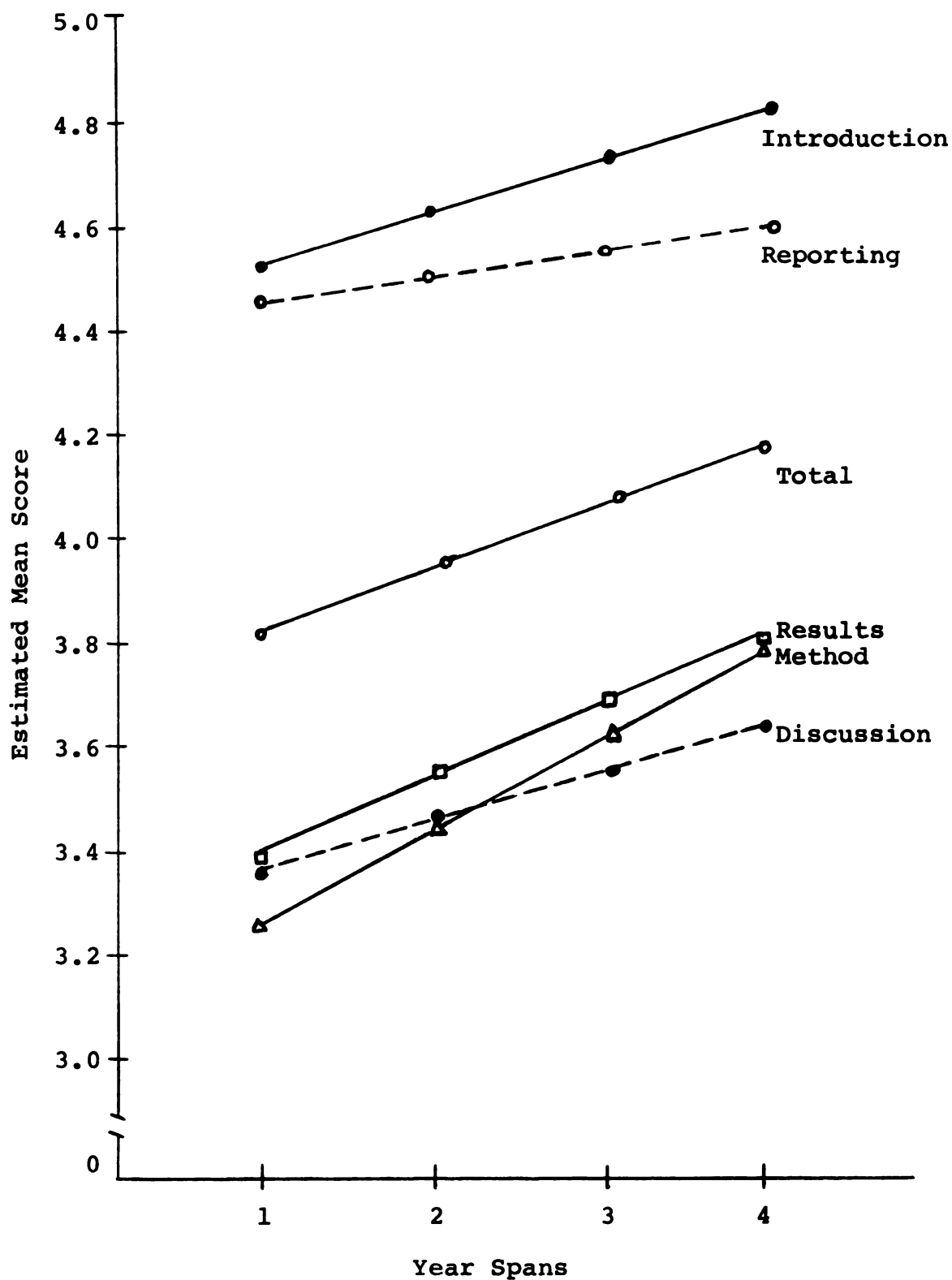


Fig. 3.2. Estimated Means on the Six Measures of the EIEM.

Although prediction into time should be made with caution, the trend based on experimental research for 1962 through 1973, if maintained at the same rate, will predict a mean rating for the total score for quality of methodology and reporting of 4.46 by 1979, 4.94 by 1991, and 5.30 by 2000. By 1994 the mean will indicate that in an overall evaluation, experimental research in counseling and counselor education will be clearly adequate. The graphs of means (Figures 3.1 and 3.2) reveal two interesting points. The results for the reporting and introduction scales cluster together, and the methods, results, and discussion scales cluster below the first two scales. This seems reasonable, in that the elements in the latter cluster are more concrete, and seem to be dependent on each other, in that they evaluate knowledge of research methodology and statistics. However, the reporting and introduction scales evaluate the description of what was done in the study and are both based on writing skill. The second point of interest is a consistent slight decrease in the means for the third year span compared to the second. Although this was not a significant decrease, as tested by the quadratic trend, the consistency for each dependent measure, excepting the method measure, should be noted. An examination of the residuals, observed mean minus the estimated mean for each dependent variable, agreed

with this visual observation. The estimated mean consistently overestimated the means for year span three, while consistently underestimating the means for year span two. This lack of fit, however, was not statistically significant.

A principle components analysis of the correlation matrix was computed (Appendix G). It indicated that there was an overall and general factor of quality which explained 65% of the variation of the measures.

Univariate and multivariate confidence intervals were generated around the estimated means of the dependent variables to consider the present state of experimental research in counseling and counselor education. For this evaluation only the most recent year span, 1971 - 1973, was considered, since this time span is contiguous with the year of this investigation, 1974. Appendix H details the upper and lower limits of a 95% confidence interval for each variable. The conclusion can be formulated that with 95% certainty the true value of the estimated mean for each variable lies within these bounds. Figure 3.3 contains the graphic representation of the univariate intervals compared to the scale of the dependent measures derived from the EIEM.

Two subscales, reporting and introduction, are clearly on the adequate end of the scale, while the three other subscales, method, results, and discussion,

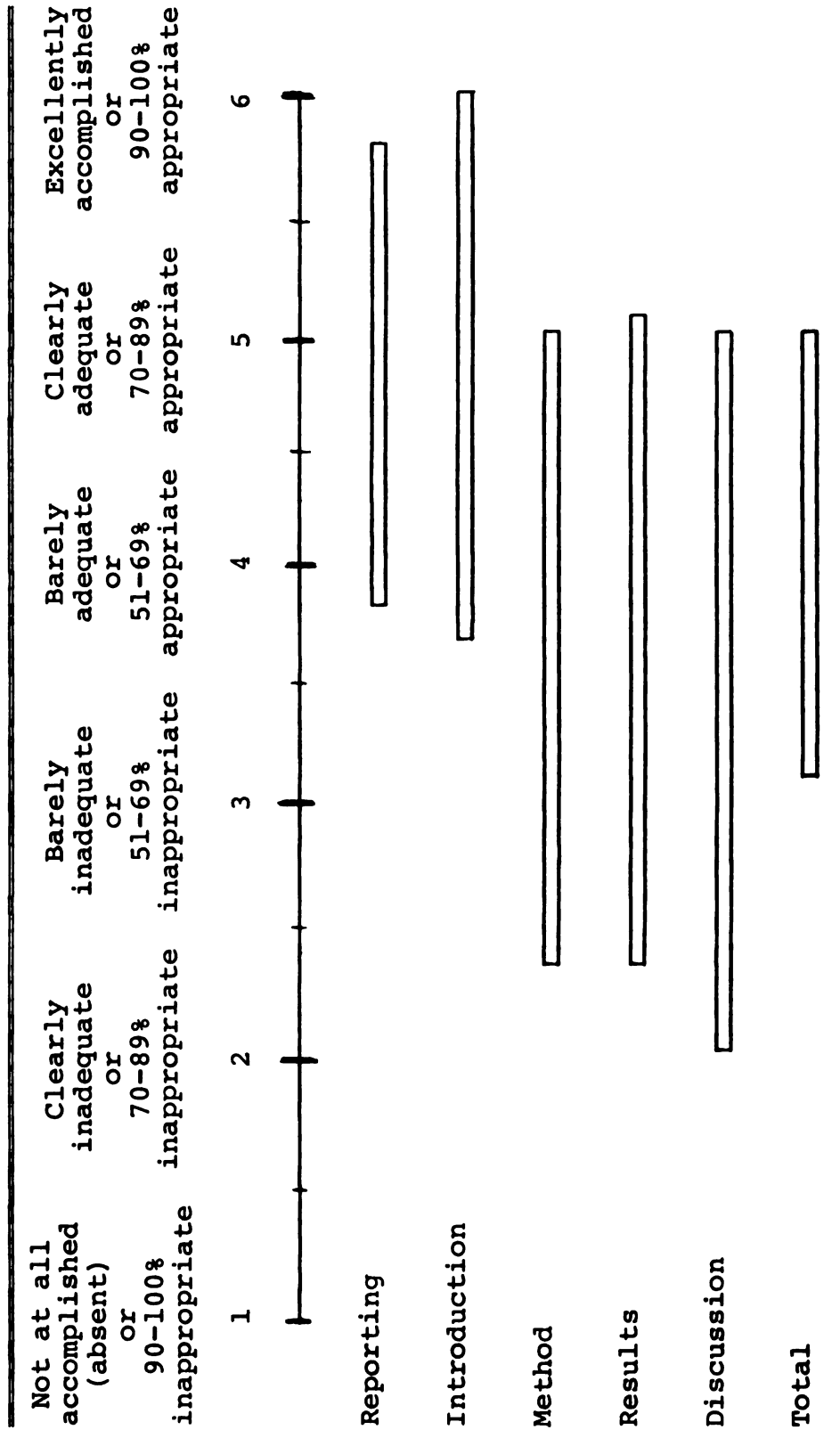


Fig. 3.3. Graphic Description of Univariate Confidence Intervals for Six Scales of FIEM.

span the middle area of the scale. The quality of the reporting and the introduction section was "clearly adequate" for the last year span. However, the measures which evaluated the essence of the experimental research were considerably lower. These aspects of the evaluated research studies were in the gray area, neither "clearly inadequate" nor "clearly adequate." The interval for the total score was predictably between the two groupings and could be characterized as "barely adequate."

Observations

The comments to follow have not been examined statistically, but have been deemed of worth in the attempt to delineate errors which occur in recently published experimental research. The means for individual items for year span four, 1971 - 1973 (see Appendix E), were compared to the scale used in the rating process. A criterion for meaningful significance of 3.51 was established. Any item whose mean was less than 3.51 would indicate an aspect of the research for 1971 - 1973 which was less than adequate.

The means for items 16, 19, 25, 31, 32, and 34 were below the criterion. The evaluation for item 16 seemed to suggest that reports of experimental studies do not include adequate information, such as reliability and validity estimates, for measurement instruments used as dependent measures. Rated as "clearly inadequate"

was the degree of random selection from the population of interest. The conclusion is that few of the studies for this year span indicated random selection from even a limited population. This affects the generalizability of the results.

The rating for item 25 indicated that authors do not give evidence that the assumptions necessary for legitimate hypothesis tests are satisfied. This could reflect that the authors do not mention the assumptions or that the assumptions have been violated. In the discussion section authors failed to generalize appropriately to populations, treatments, or settings allowable by the design and sampling procedure. They also tended not to indicate limitations or weaknesses of their studies when these were evident. The final item to be rated below the criterion referred to the author making suggestions for further investigation which follow from his results. Apparently, few authors made such comments.

Summary

A multivariate trend analysis revealed a linear relationship between the four year spans for all dependent measures, as well as for the two overall evaluation items of the EIEM. Graphic representation of observed and estimated means illustrated a slightly positive increasing slope for each measure. The largest slope

would predict only a .17 increase in the mean rating of quality for one three-year span. Although prediction into time should be made with appropriate caution, the trend based on experimental research for 1962 through 1973, if maintained at the same rate, will predict a mean rating for the total score for quality of methodology and reporting of 4.46 by 1979, 4.94 by 1991, and 5.30 by 2000. By 1994 the mean will indicate that in an overall evaluation experimental research in counseling and counselor education had become clearly adequate.

Confidence intervals were generated for the six scales of the EIEM for the fourth year span, 1971 - 1973, to provide evidence of the level of quality of experimental research in the fields of counseling and counselor education. The measures for reporting and introduction indicated that the quality for these two related aspects of an experimental study was "clearly adequate," though the band extended from "barely adequate" to "excellently accomplished." The measures for method, results, and discussion indicated a lower quality estimate for these three aspects of research. While each had a confidence span from "clearly inadequate" to "clearly adequate," the conclusion was offered that the quality of methodology of the counseling research published from 1971 to 1974 was mediocre.

CHAPTER IV

SUMMARY AND DISCUSSION

Summary

The purpose of this study was to systematically evaluate the quality of experimental research which has been published in the fields of counseling and counselor education from 1962 through 1973. Attention was directed at the methodology and reporting of studies rather than at the subject matter or variables being examined. The specific independent variable was time, in order to determine whether there has been an improvement since 1962 in the quality of published research. Four three-year spans were chosen as levels of the independent variable: 1962 - 1964; 1965 - 1967; 1968 - 1970; 1971 - 1973.

Following a survey of three journals, Journal of Counseling Psychology, Personnel and Guidance Journal, and Counselor Education and Supervision, to specify the population of pre-, true-, and quasi-experimental studies, a sample of 38 studies was randomly chosen for each year span. Each study was evaluated by a trained rater on

the Evaluation Instrument for Experimental Methodology, which produced six measures of the quality of reporting and methodology. Three raters independently rated the studies. Fifteen randomly chosen studies were commonly rated to establish the average interrater reliability estimate of .78.

A 1 x 4 design with equal cell sizes was utilized to examine for differences between the four year spans. A multivariate analysis of variance using orthogonal polynomials was used to test the hypotheses of the trend of the quality over time. A slight linear trend was distinguished across the four year spans. Graphic illustration demonstrated a very slight positive increasing trend over time. Examination of the means derived from the EITEM for the last year span revealed that the quality of reporting and the introduction was "clearly adequate." However, quality of the method, results and discussion sections was generally "barely adequate." In total the quality of experimental research in counseling and counselor education was characterized as "barely adequate."

Discussion

The evaluation of experimental studies in counseling and counselor education resulted in both good and bad news for the profession. The results indicate that there are slight differences in the form of a linear trend which were discriminated by the trend analysis.

The linear trend is the major finding of this investigation. Caution in interpretation is advisable, however, because the amount of increase in quality for succeeding year spans is minimal. Prediction over time is also risky. Contributing factors to the quality of published research are complex and probably do not act uniformly over time.

Speculation about factors contributing to the gradual increase of quality of research is relevant. Obviously the effect of the computer on the expansion of knowledge of statistics and research methodology has been great. The ability to analyze data from complex designs has been of direct benefit to the counseling profession. The problem of adequate controls for studies with human subjects has been somewhat relieved by the readily available alternatives provided by computer data analysis for statistical controls or complex designs with blocking variables. The improvement of the instruction of research methodology or a change in the requirements for a professional certificate or degree to include research methodology could be contributing to the gradual improving trend. With the increasing number of submitted manuscripts to professional publications, the criteria for acceptance could be changing to require better quality research now than in the past. Hopefully, investigations such as this will have impact

on researchers, members of the profession, and editors toward improving the quality of research literature.

Possible factors which have inhibited the development of higher quality research should be considered. The fields of counseling and counselor education have not received as much financial support for development and research as some of the other applied sciences. This may mean that the motivation to accomplish sophisticated research is affected. The fields are also quite young with research holding a lower priority than in more mature professions. The trend of training counselors as practitioners versus researchers has surely affected the quality of published research. As the profession matures, research should be established as a respectable priority among its members.

The postulation by several counselor educators (Carkhuff, 1965; Myers, 1966; Patterson, 1963) that the quality of research in counseling and counselor education has been improving over time is supported by this empirical investigation of methodology, but with the cautions previously stated. The significant linear trend demonstrated that there is a slightly positive trend in the quality of research in counseling and counselor education. The results of this study, however, are applicable only to the population of experimental studies of counseling and counselor education research. The

conclusions of others (Gazda & Larsen, 1968; Hansen & Warner, 1971; Pawlicki, 1970; Thoresen, 1969) that there is a lack of well-planned and executed research is also partially supported, as demonstrated by the examination of the confidence intervals for the dependent measures for the last year span. The quality of reporting for recent experimental publications is relatively high, as evidenced by a mean of 4.89 for the overall rating of reporting for the year span 1971 - 1973. However, the quality of methodology was rated less than "barely adequate." The mean for the overall methodology rating was 3.87. While the quality of the reporting of an experimental project is important for replication and communication within the profession, the impact of poor quality methodology is greater than that for poor quality reporting. Misleading or false results can be costly, especially in fields that deal with human beings.

In an effort to provide evaluation for specific aspects of experimental methodology, the means for the items for the last year span, 1971 - 1973, were examined (see Appendix E). Aspects of methodology that were rated below the meaningful significance criterion of 3.51 included descriptive statements about the reliability and validity of dependent measures, random selection of the sample, consideration of hypothesis testing assumptions, and three items in the discussion

section. Low ratings for the two items covering dependent measures and statistical assumptions have less impact on the overall quality of counseling research than the others. Of significant impact are the low ratings of the random sample selection item and discussion items. It is probable that many readers, especially those with inadequate knowledge of research methodology, look to the discussion section for conclusions without careful consideration of the previous sections of the experimental report. Thus, considering the rated inadequacy of generalization statements (item 31), too many researchers are misrepresenting the applicability of their results, and too many consumers are possibly not perceiving the illegitimate generalizations. Such occurrences are potentially harmful to the profession and to clients. The continued growth of counseling research is also hampered by such practices.

Examination of items that had marginal ratings, means of less than 4.0, for the last year span might be useful. When judges evaluated the subjects on the dependent variable, interrater reliabilities were not consistently reported (item 17). The absence of such statements partially inhibits the reader from evaluating the precision of the analysis.

The designs in this sample of experimental studies were rated as only "barely adequate" in providing

the maximum precision possible, given the data the researcher had (item 21). This result concurs with Cohen's (1962) conclusions that a majority of published research provides only a minimum degree of precision. The evaluated research also only marginally controlled for unbiased treatment effects (item 22). Adequate controls continue to be a problem in counseling research, as has been studied by Kelley, et al. (1970) and commented on by Calvin (1954), Harrison (1971), and Patterson (1966).

In the results section the means of two items fall in the 3.51 to 4.00 span. Ratings for items 28 and 29 can be interpreted that there is marginal consistency in reporting the descriptive statistics of dependent measures and pertinent information of the hypothesis tests performed. These are essential components of a results section, especially for the professional who carefully examines the correctness of data analyses.

In the discussion section items 33 and 35 had means between 3.51 and 4.00. Apparently researchers were marginally consistent in comparing their findings to theory or previous research. Of more importance was the marginal appropriateness of causal statements made in conclusions. This could refer to causal statements made when the design does not allow such conclusions or

when the results do not warrant such conclusions. Such incorrect statements are misleading.

In summary for this discussion of meaningful results, the reporting and introduction sections of the EIEM have no items with means of less than 4.00. This is consistent with previously reported results. However, the method section contains five of eight items with means less than 4.00, a criterion indicating at best a marginally adequate rating. The results section has three of seven items so rated, and the discussion section has five of six items with means below 4.00. The method and discussion sections of experimental studies should particularly be noted for inadequacies. These conclusions agree with the previous analyses of the dependent measures derived from the EIEM.

Recommendations

For subsequent investigations of the quality of experimental methodology, continuing refinement of the evaluation instrument is recommended. One revision could be the construction of a scale with greater detail or larger span to prevent a ceiling or floor effect, the effect which results from the frequent use of maximum or minimum scale values. Such revision would add clarity to the results derived from the instrument and might contribute to increasing interrater reliabilities.

Longer training sessions for the raters would also probably increase reliability estimates.

Recommendations for subsequent investigations include evaluation of other types of research in counseling and counselor education, most notably correlational research. Such an investigation would round out the evaluation of the quality of research in these fields. Evaluation of future years of counseling research would also be beneficial and could build on the present investigation to establish more firmly the trend of improving research.

As has been emphasized in Chapter I this investigation strenuously avoided evaluation of the content and relevance of counseling research. An evaluation of this essential aspect of the profession's research is strongly recommended. It would require noted professionals as evaluators and would be an extremely difficult task to operationalize. However, for a complete estimation of the state of research in the profession such an evaluation is essential.

Many recommendations to researchers have been covered in Chapter IV, namely those aspects of research reports to avoid which contribute to questionable and deceptive experimental studies. Those points of importance to experimental research were operationalized in the EIEM. An additional recommendation is for more

researchers and editors to consider replication of previous research as valuable professional effort, which is necessary to build a reliable research base for the profession. The state of counseling and counselor education research would be significantly benefited. Currently few studies and results are challenged. The pace of improvement of quality could be speeded by such a tactic.

The recommendation for the research consumer as well as the counselor educator is to carefully consider all aspects of a research report. For experimental studies for the years studied, the methods, results, and discussion sections were shown to have the highest probability for error or misleading statements. These sections also have the biggest impact on the significance of the results of an experimental study. For those counselor educators who teach research skills, the examination of the ratings of individual items of the rating instrument points to those areas which should be stressed. The instrument itself could be used as a learning tool for the counselor.

Conclusion

The systematic evaluation of the methodology and reporting of research in counseling and counselor education revealed mixed results. The quality of reporting was quite good, while the quality of experimental

methodology was barely mediocre. Despite the trend of increasing quality, the research in these fields must be viewed critically. Whiteley's (1967) comment that poorly formulated research is not only worthless but deceptive should be heeded by counseling researchers, editors, and research consumers in an effort to upgrade the profession's research, protect future clients and trainees, and promote better counseling service and training.

APPENDICES

APPENDIX A

FREQUENCY COUNT OF THE NUMBER OF EXPERIMENTAL, CORRELATIONAL AND MISCELLANEOUS STUDIES FOR THE THREE JOURNALS FOR THE FOUR YEAR SPANS

		1962- 1964	1965- 1967	1968- 1970	1971- 1973	TOTALS
<u>Journal of Counseling Psychology</u>	Exp	24	53	75	93	245
	Corr	67	96	135	132	430
	Misc	10	17	21	17	65
<u>Personnel & Guidance Journal</u>	Exp	12	19	26	4	61
	Corr	113	138	89	1	341
	Misc	44	41	16	0	101
<u>Counselor Education & Supervision</u>	Exp	2	4	23	28	57
	Corr	5	32	26	36	99
	Misc	15	6	11	10	42
TOTALS	Exp	38	76	124	125	363
	Corr	195	266	250	169	870
	Misc	69	64	48	27	208

APPENDIX B

EVALUATION INSTRUMENT FOR EXPERIMENTAL METHODOLOGY

ARTICLE NUMBER _____

RATER _____

TITLE OF ARTICLE _____

TYPE OF DESIGN: _____ Pre-experimental
_____ True-experimental
_____ Quasi-experimental

STATISTIC USED:
(to test main hypotheses)

_____ ANOVA (type: _____)

_____ ANCOVA

_____ MANOVA

_____ t or z Tests

_____ Nonparametric; name _____

_____ Correlation

_____ Factor Analysis

_____ Other; name _____

TYPE OF RESEARCH: _____ Applied: Process

_____ Applied: Outcome

_____ Basic Research

COMMENTS:

EVALUATION INSTRUMENT FOR EXPERIMENTAL METHODOLOGY

Note: The items are grouped according to convention, but the content of an individual item may be found anywhere in the study.

Rate each item using the following rating scale:

1	2	3	4	5	6
strongly disagree	clearly disagree	barely disagree	barely agree	clearly agree	strongly agree
OR	OR	OR	OR	OR	OR
not at all accomplished (absent)	clearly inadequate	barely inadequate	barely adequate	clearly adequate	excellently accomplished
OR	OR	OR	OR	OR	OR
90-100% inappropriate	70-89% inappropriate	51-69% inappropriate	51-69% appropriate	70-89% appropriate	90-100% appropriate

REPORTING (attend to the quality of reporting, not to the content of the item)

- | | | | | | | |
|---|---|---|---|---|---|---|
| 1. The review of the literature is <u>concise</u> , <u>understandable</u> , and <u>logical</u> . | 1 | 2 | 3 | 4 | 5 | 6 |
| 2. The research hypothesis is <u>clearly stated</u> . | 1 | 2 | 3 | 4 | 5 | 6 |
| 3. The population of interest is <u>clearly specified</u> . | 1 | 2 | 3 | 4 | 5 | 6 |
| 4. The procedure for selection of subjects is <u>clearly specified</u> . | 1 | 2 | 3 | 4 | 5 | 6 |
| 5. The subjects are <u>completely described</u> on relevant variables. | 1 | 2 | 3 | 4 | 5 | 6 |
| 6. The treatment procedures are <u>clearly enough defined</u> to allow for replication. | 1 | 2 | 3 | 4 | 5 | 6 |
| 7. <u>All</u> statistics used in the analysis are named. | 1 | 2 | 3 | 4 | 5 | 6 |
| 8. The results are <u>clearly</u> and <u>concisely</u> reported (no unnecessary data are included). | 1 | 2 | 3 | 4 | 5 | 6 |
| 9. The discussion is <u>understandable</u> and <u>concisely</u> written. | 1 | 2 | 3 | 4 | 5 | 6 |
| Give an overall rating of the quality of reporting of this study. | 1 | 2 | 3 | 4 | 5 | 6 |

INTRODUCTION

- | | | | | | | |
|--|---|---|---|---|---|---|
| 10. The purpose of the study is clearly stated. | 1 | 2 | 3 | 4 | 5 | 6 |
| 11. The review of the literature is relevant to the problem and independent variables of interest. | 1 | 2 | 3 | 4 | 5 | 6 |
| 12. Research hypotheses are stated for <u>all</u> variables (if exploratory, this is stated clearly). | 1 | 2 | 3 | 4 | 5 | 6 |
| 13. Each independent variable and its levels are clearly described; the design is clearly enough described to allow you to diagram it. | 1 | 2 | 3 | 4 | 5 | 6 |
| 14. An excellent rationale is given for the use of the particular dependent variables chosen. | 1 | 2 | 3 | 4 | 5 | 6 |

METHODS

- | | | | | | | |
|---|---|---|---|---|---|---|
| 15. The dependent measures are the most appropriate for the purpose of the study. | 1 | 2 | 3 | 4 | 5 | 6 |
| 16. The reliability and validity data are given for each instrument used as a dependent measure. | 1 | 2 | 3 | 4 | 5 | 6 |
| 17. The interrater reliabilities are given if raters are used. | 1 | 2 | 3 | 4 | 5 | 6 |
| 18. The <u>stated population</u> (not sample) is the relevant one in terms of the nature of the problem and hypotheses. | 1 | 2 | 3 | 4 | 5 | 6 |
| 19. Subjects were randomly selected from the population. | 1 | 2 | 3 | 4 | 5 | 6 |
| 20. Subjects were randomly assigned to treatment groups. | 1 | 2 | 3 | 4 | 5 | 6 |
| 21. Given the data collected by the researcher, the design is such that it provides the maximum precision possible. | 1 | 2 | 3 | 4 | 5 | 6 |
| 22. The design allows for <u>unbiased treatment effects</u> ; there are no confounding or uncontrolled irrelevant variables which confuse the results; necessary controls for <u>internal validity</u> are either built into the design or statistically managed. | 1 | 2 | 3 | 4 | 5 | 6 |

_____	regression
_____	subject mortality
_____	instrumentation
_____	history
_____	maturation
_____	testing
_____	selection bias
_____	selection-maturation interaction

RESULTS

- | | | | | | | |
|---|---|---|---|---|---|---|
| 23. The best statistical analysis for the design, data, and hypotheses was used. | 1 | 2 | 3 | 4 | 5 | 6 |
| 24. The data analysis is consistent with the design. | 1 | 2 | 3 | 4 | 5 | 6 |
| 25. The authors gave evidence that the assumptions necessary for the hypothesis test statistic(s) were met (normality, independence, equality of variance, additivity, etc.). | 1 | 2 | 3 | 4 | 5 | 6 |
| 26. The unit of analysis is equal to the experimental unit. | 1 | 2 | 3 | 4 | 5 | 6 |
| 27. Specific answers to the hypotheses are given. | 1 | 2 | 3 | 4 | 5 | 6 |
| 28. Means and variances or standard deviations are given for each dependent variable according to groups. | 1 | 2 | 3 | 4 | 5 | 6 |
| 29. The results section includes values of the test statistic, df, and p-value (for ANOVA the MSs are given). | 1 | 2 | 3 | 4 | 5 | 6 |

DISCUSSION

- | | | | | | | |
|--|---|---|---|---|---|---|
| 30. The conclusions drawn are consistent with the data results and hypotheses. | 1 | 2 | 3 | 4 | 5 | 6 |
| 31. The author generalizes to the population, treatments, or settings allowable by the design. | 1 | 2 | 3 | 4 | 5 | 6 |
| 32. If there were limitations of design, sampling, data collection or data analysis, the author indicates the qualifications to his study which <u>limit inference</u> . | 1 | 2 | 3 | 4 | 5 | 6 |
| 33. The author compares his findings to previous research findings or to a theory. | 1 | 2 | 3 | 4 | 5 | 6 |
| 34. The author makes suggestions for further investigation which logically follow from his study. | 1 | 2 | 3 | 4 | 5 | 6 |
| 35. The causal inferences made were entirely appropriate according to the design, sampling and analysis. | 1 | 2 | 3 | 4 | 5 | 6 |

Make an overall rating of the quality of the methodology of this study (considering items 10 - 35 and not those items in the reporting section).	1	2	3	4	5	6
--	---	---	---	---	---	---

Comments or any errors which you found not covered in the preceding items:

APPENDIX C

RELEVANT DEFINITIONS

Pre-experimental Design: Any design which has a treatment group but no reasonable comparison group. Causal statements cannot be made.

Examples:	1. One-shot case study	X	0
	2. One-group pretest posttest design	0	X 0
	3. Static-group compar- ison	$\begin{array}{r} X \quad 0 \\ \hline 0 \end{array}$	

True-experimental Design: Random assignment occurs to at least one treatment and one control group or several treatment groups. Causal statements are appropriate.

Examples:	1. Pretest-posttest control group design	R	O	X	O
		R	O		O
	2. Solomon four-group design	R	O	X	O
		R	O		O
		R		X	O
		R			O
	3. Posttest only control group design	R	X	O	
		R			O

Quasi-experimental Design: For field settings where complete control of experimental stimuli is impossible; the "when" and "to whom" of measurement is controllable, while the "when" and "to whom" of stimuli exposure and ability to randomize exposures are not controllable. Causal inferences cannot be made.

Examples:	1. Time series	O	O	O	O	X	O	O	O	O
	2. Equivalent Time Samples Design	X ₁ O	X ₂ O	X ₃ O	X ₄ O					
	3. Nonequivalent Control Group Design	$\frac{O \quad X \quad O}{O \quad \text{-----} \quad O}$								
	4. Counterbalanced Design	X ₁ O	X ₂ O	X ₃ O	X ₄ O					
		X ₂ O	X ₄ O	X ₁ O	X ₃ O					
		X ₃ O	X ₁ O	X ₄ O	X ₂ O					
		X ₄ O	X ₃ O	X ₂ O	X ₁ O					
	5. Separate-sample pretest-posttest design	R	O	(X)						
		R		X	O					

Applied Research-Process: A study whose purpose is the investigation of variables directly related to the practice and process of counseling (e.g. interaction variables, counselee variables, technique variables, counselor variables, etc.) None of the dependent variables are measures of the success of a counseling contact.

Applied Research-Outcome: A study whose purpose is the investigation of variables directly related to the end result of counseling -- successful treatment of a problem. At least one of the dependent variables is related to the end objective of counseling (successful information seeking behavior, a decision made, higher grades, more self-actualized, etc.).

Basic Research: Laboratory research whose purpose is to define and refine constructs of theories which though ultimately applicable are not directly applicable to counseling or counselor education.

APPENDIX D

NOTES ON RATING

BEFORE YOU BEGIN TO RATE THE FIRST ARTICLE, READ THE RATING FORM TO ACQUAINT YOURSELF WITH THE MINOR CHANGES THAT HAVE BEEN MADE.

1. Of utmost importance is the accuracy of your ratings. Therefore, I suggest that you rate only several articles at any one sitting. This is to avoid any interaction between articles, as well as to avoid a fatigue effect.
2. Frequently consult the notes that you took during the training session. The objective is to maintain the same set of criteria for all raters across all articles.
3. Freely consult any relevant sources, such as notes from statistics classes, stat texts, experts, and especially Campbell and Stanley.
4. Rate the studies in the order given to you -- alphabetically A to HHH.
5. Remember that the first section on "Reporting" is evaluation of the clarity of the reporting and not evaluation of the appropriateness or adequate nature of the content of the particular item.
6. Leave out any question which clearly does not apply to a particular article. However, this should occur very infrequently.
7. Comment freely on a particular article, noting especially any weaknesses which were not picked up in the standard items.
8. The information to answer an item may be found anywhere in the study.
9. Keep an accurate accounting of the time you spend rating.
10. If you have questions or problems call me at 517-337-0545 or leave a message at 517-353-9242 (Department of Psychiatry).
11. GOOD LUCK -- and I hope that this is as much a learning experience as a money-earning one for you. I appreciate the effort that you are contributing to my project.

APPENDIX E

MEANS AND STANDARD DEVIATIONS FOR ITEMS OF THE EVALUATION INSTRUMENT FOR EXPERIMENTAL METHODOLOGY

		Y ₁		Y ₂		Y ₃		Y ₄	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Reporting	1	4.50	1.45	5.13	.88	4.76	1.13	5.37	.59
	2	4.68	1.14	4.79	1.09	4.66	1.21	4.84	.97
	3	4.82	.95	5.05	1.18	5.03	.91	4.95	.90
	4	4.76	1.36	4.87	1.28	4.89	.98	4.89	1.03
	5	3.92	1.17	3.82	1.35	3.82	1.01	4.11	1.29
	6	4.34	1.48	4.61	1.17	4.42	1.06	4.97	.75
	7	4.59 ^a	1.48	4.92	1.38	4.97 ^a	1.44	5.24	1.13
	8	4.26	1.20	4.53	1.06	4.42	1.20	4.76	.97
	9	4.84	.79	4.89	.80	4.71	.84	4.95	.96
Introduction	10	5.21	.91	5.34	.63	5.34	.67	5.37	.63
	11	4.13	1.58	4.95	.90	4.50	1.13	5.00	.93
	12	4.32	1.19	4.53	1.29	4.26	1.43	4.61	1.24
	13	4.87	1.49	5.11	1.18	5.16	.97	5.32 ^a	.85
	14	3.63	1.36	4.13	1.32	4.11	1.23	4.24 ^a	1.19
Method	15	4.71	1.04	4.76	1.08	4.66	1.17	4.92	.94
	16	1.94 ^f	1.55	2.31 ^c	1.66	1.53 ^e	.98	2.82 ^d	1.91
	17	3.45 ^j	2.11	3.37 ^h	2.41	4.53 ⁱ	2.10	3.92 ^g	2.34
	18	5.29	.84	4.97	1.42	5.34	.75	5.30 ^a	.74
	19	1.42	1.31	1.79	1.61	1.47	1.35	1.53	1.37
	20	2.63 ^c	2.34	3.79	2.46	3.74	2.41	4.06 ^c	2.33
	21	3.16 ^a	1.38	3.50	1.45	3.70 ^a	1.29	3.73 ^a	1.57
	22	2.79	1.73	3.55	1.70	3.79	1.71	3.84	1.87
Results	23	3.21	1.49	3.92	1.26	3.68	1.45	4.03	1.40
	24	3.89	1.57	4.36 ^b	1.22	4.24	1.38	4.42	1.37
	25	1.42	.95	1.87	1.49	1.81 ^a	1.33	1.87	1.51
	26	3.89	2.12	3.58	2.14	3.53	2.32	4.16 ^a	2.24
	27	4.61	1.03	4.95	1.11	4.62 ^a	1.14	4.76	.91
	28	2.79	1.88	3.43 ^a	2.08	3.13	2.09	3.92 ^a	2.06
	29	3.41 ^a	1.54	3.87	2.09	3.78 ^a	1.80	3.68	1.88

Item	Y_1		Y_2		Y_3		Y_4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
30	4.34	1.05	4.82	.83	4.32	1.16	4.55	1.18
31	3.29	1.58	3.61	1.67	3.37	1.70	3.49 ^a	1.48
32	2.79	1.70	2.97	1.70	3.03	1.70	3.11	1.71
33	2.84	1.76	3.45	1.75	3.03	1.87	3.66	1.74
34	3.00	1.96	3.32	1.88	3.34	1.65	3.50	1.69
35	3.22 ^b	1.85	3.53 ^b	1.78	3.11	1.90	3.62 ^a	1.60
Overall-Reporting	4.45	.95	4.68	.81	4.53	.76	4.89	.89
Overall-Methodology	3.34	1.15	3.76	.94	3.68	1.12	3.87	1.12

^a_n = 37^d_n = 34^g_n = 25^j_n = 11^b_n = 36^e_n = 32^h_n = 19^c_n = 35^f_n = 31ⁱ_n = 15

APPENDIX F

UNIVARIATE TESTS OF THE SIX DEPENDENT VARIABLES FOR A LINEAR TREND

	F-ratio	df	p
Reporting	6.224	1,148	.014
Introduction	8.808	1,148	.004
Method	8.867	1,148	.003
Results	5.486	1,148	.021
Discussion	1.575	1,148	.211
Total	8.746	1,148	.004

APPENDIX G

PRINCIPLE COMPONENTS OF THE CORRELATION MATRIX FOR THE SIX DEPENDENT VARIABLES OF THE EIM

Variable	Component 1	Component 2
Reporting	-.8413	-.3509
Introduction	-.7950	-.3848
Method	-.7764	-.0574
Results	-.7518	+.2424
Discussion	-.6612	+.6421
Total	-.9898	+.0393

Percent of Variation Explained by Component 1 = 65.42%

Percent of Variation Explained by Component 2 = 12.45

APPENDIX H

NINETY-FIVE PERCENT CONFIDENCE INTERVAL FOR ESTIMATED MEANS OF THE SIX DEPENDENT MEASURES OF THE EIEM

Measure	Univariate		Multivariate	
	Lower limit	Upper limit	Lower limit	Upper limit
Reporting	3.93	5.76	2.73	6.00+
Introduction	3.83	5.96	2.42	6.00+
Method	2.47	5.08	.75	6.00+
Results	2.47	5.13	.72	6.00+
Discussion	2.12	5.13	.14	6.00+
Total	3.24	5.13	2.00	6.00+

REFERENCES

REFERENCES

- Barker, H. R., & Gurman, E. B. Replication versus tests of equivalence. Perceptual and Motor Skills, 1972, 35, 807-815.
- Bordin, E. S., Cutler, R. L., Dittman, A. T., Harway, N. I., Raush, H. L., & Rigler, D. Measurement problems in process research on psychotherapy. Journal of Consulting Psychology, 1954, 18, 79-82.
- Borg, W. R. Educational research: An introduction. New York: David McKay Co., 1963.
- Burck, H. D., Cottingham, H. F., & Reardon, R. C. Counseling and accountability: Methods and critique. New York: Pergamon Press, Inc., 1973.
- Calvin, A. D. Some misuses of the experimental method in evaluating the effect of client-centered counseling. Journal of Counseling Psychology, 1954, 1, 249-251.
- Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.
- Carkhuff, R. R. Counseling research, theory, and practice--1965. Journal of Counseling Psychology, 1966, 13, 467-480.
- Cohen, J. The statistical power of abnormal-social psychological research: A review. Journal of Abnormal and Social Psychology, 1962, 65, 145-153.
- Coleman, W. The role of evaluation in improving guidance and counseling services. Personnel and Guidance Journal, 1957, 35, 441-444.
- Cornfield, J., & Tukey, J. W. Average values of mean squares in factorials. Annals of Mathematical Statistics, 1956, 27, 907-949.

- Cotton, M. C., & Anderson, W. P. Citation changes in the Journal of Counseling Psychology. Journal of Counseling Psychology, 1973, 20, 272-274.
- Cronbach, L. J., & Suppes, P. Research for tomorrow's schools: Disciplined inquiry for education. Toronto, Ontario: The MacMillan Co., 1969.
- Crittenden, R. L. Comment on "Group reactive inhibition and reciprocal inhibition therapies with anxious college students." Journal of Counseling Psychology, 1973, 20, 353-354.
- Davitz, J. R., & Davitz, L. J. A guide for evaluating research plans in psychology and education. New York: Teachers College Press, 1967.
- Dressel, P. L. Implications of recent research for counseling. Journal of Counseling Psychology, 1954, 1, 100-105.
- Dressel, P. L. Some approaches to evaluation. Personnel and Guidance Journal, 1953, 31, 284-287.
- Eastwood, G. R. A note on hypothesis testing. Alberta Journal of Educational Research, 1967, 13, 265-273.
- Edgington, E. S. A tabulation of inferential statistics used in psychology journals. American Psychologist, 1964, 19, 202-203.
- Edwards, A. L., & Cronbach, L. J. Experimental design for research in psychotherapy. Journal of Clinical Psychology, 1952, 8, 51-59.
- Farquhar, W. W., & Krumboltz, J. D. A checklist for evaluating experimental research in psychology and education. Journal of Educational Research, 1959, 52, 354-355.
- Feldman, C. F., & Hass, W. A. Controls, conceptualization, and the interrelation between experimental and correlational research. American Psychologist, 1970, 25, 633-635.
- Finn, J. Multivariate: Fortran program for univariate and multivariate analysis of variance and covariance. State University of New York of Buffalo, 1967.

- Fisher, M. B., & Roth, R. M. Structure: An essential framework for research. Personnel and Guidance Journal, 1961, 39, 639-644.
- Ford, D. H. Research approaches to psychotherapy. Journal of Counseling Psychology, 1959, 6, 55-60.
- Foreman, M. E. Publication trends in counseling journals. Journal of Counseling Psychology, 1966, 13, 481-485.
- Gazda, G. M., & Larsen, M. J. A comprehensive appraisal of group and multiple counseling research. Journal of Research and Development in Education, 1968, 1 (2), 57-66.
- Glass, G. V., & Robbins, M. P. A critique of experiments on the role of neurological organization in reading performance. Reading Research Quarterly, 1967, 3 (1), 5-51.
- Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood, N.J.: Prentice-Hall, Inc., 1970.
- Goodstein, L. D. The institutional sources of articles in the Journal of Counseling Psychology. Journal of Counseling Psychology, 1963, 10, 94-95.
- Hansen, J. C., & Warner, R. W. Review of research on practicum supervision. Counselor Education and Supervision, 1971, 10, 261-272.
- Harrison, R. Research on human relations training: Design and interpretation. Journal of Applied Behavioral Science, 1971, 7, 71-85.
- Herr, E. L. Basic issues in research and evaluation of guidance services. Counselor Education and Supervision, 1964, 4, 9-16.
- Hilgard, E. Introduction to psychology. New York: Harcourt, Brace, & World, 1962.
- Hobbs, N., & Seeman, J. Counseling. In C. P. Stone & Q. McNemar, Annual Review of Psychology Volume 6, Stanford, California: Annual Reviews, Inc., 1955, pp. 379-404.
- Holland, J. L. Vocational guidance for everyone. Educational Researcher, 1974, 3, 9-15.

- Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.
- Isaac, S., & Michael, W. B. Handbook in research evaluation. San Diego: Robert R. Knapp, 1971.
- Jensen, B. T., Coles, G., & Nestor, B. The criterion problem in guidance research. Journal of Counseling Psychology, 1955, 2, 58-61.
- Kelley, J., Smits, S. L., Leventhal, R., & Rhodes, R. Critique of the designs of process and outcome research. Journal of Counseling Psychology, 1970, 17, 337-341.
- Kiesler, D. J. Basic methodological issues implicit in psychotherapy process research. American Journal of Psychotherapy, 1966a, 20, 135-155.
- Kiesler, D. J. Some myths of psychotherapy research and the search for a paradigm. Psychological Bulletin, 1966b, 65, 110-136.
- Krause, M. S. Experimental control as sampling problem in counseling and therapy research. Journal of Counseling Psychology, 1972, 19, 340-346.
- Lachenmeyer, C. W. Experimentation--A misunderstood methodology in psychological and social-psychological research. American Psychologist, 1970, 25, 617-624.
- Lykken, D. T. Statistical significance in psychological research. Psychological Bulletin, 1968, 70, 151-159.
- Marks, S. E., Conry, R. F., & Foster, S. F. The marathon group hypothesis: An unanswered question. Journal of Counseling Psychology, 1973, 20, 185-187.
- Meltzoff, J., & Kornreich, M. Research in psychotherapy. New York: Alterton Press, Inc., 1970.
- Mills, D. H., & Mencke, R. Characteristics of effective counselors: A re-evaluation. Counselor Education and Supervision, 1967, 6, 332-334.
- Myers, R. A. Research in counseling psychology--1964. Journal of Counseling Psychology, 1966, 13, 371-379.

- Nunnally, J. The place of statistics in psychology. Educational and Psychological Measurement, 1960, 20, 641-650.
- Orne, M. T. One the social psychology of the psychological experiment. American Psychologist, 1962, 17, 776-783.
- Patterson, C. H. Counseling. Annual Review of Psychology, 1966, 17, 70-110.
- Patterson, C. H. Program evaluation. Review of Educational Research, 1963, 33, 214-224.
- Patterson, C. H. Methodological problems in evaluation. Personnel and Guidance Journal, 1960, 39, 270-274.
- Patterson, C. H. Matching versus randomization in studies of counseling. Journal of Counseling Psychology, 1956, 3, 262-271.
- Patterson, C. H. Comment. Journal of Counseling Psychology, 1955, 2, 154-155.
- Pawlicki, R. Behavior-therapy research with children: A critical review. Canadian Journal of Behavioural Science, 1970, 2, 163-173.
- Roberts, K. H. Understanding research: Some thoughts on evaluating completed educational projects. 1969, ERIC-Stanford, California, ED 032 759.
- Samler, J. Comments. Personnel and Guidance Journal, 1958, 36, 555-556.
- Scheifley, V. M. Program for correction factor for a finite population. Unpublished.
- Schmidt, L. D., & Pepinsky, H. B. Counseling research in 1963. Journal of Counseling Psychology, 1965, 12, 418-427.
- Sieka, F., Taylor, D., Thomason, B., & Muthard, J. A critique of "Effectiveness of counselors and counselor aids." Journal of Counseling Psychology, 1971, 18, 362-364.
- Smith, N. C. Replication studies: A neglected aspect of psychological research. American Psychologist, 1970, 25, 970-975.

- Smith, O. W., Smith, P. C., Scheffers, J., & Steinmann, D. Common errors in reports of psychological studies. Perceptual and Motor Skills, 1971, 32, 3-7.
- Spithill, A. C. To leave a scratch on the wall: Getting published. Personnel and Guidance Journal, 1973, 52, 35-38.
- Stanley, J. C. Quasi-experimentation in educational settings. The School Review, 1967, 75, 343-352.
- Stone, S. C., & Shertzer, B. Ten years of the Personnel and Guidance Journal. Personnel and Guidance Journal, 1964, 42, 958-969.
- Thoresen, C. E. Relevance and research in counseling. Review of Educational Research, 1969, 39, 263-281.
- Tukey, J. W. Analyzing data: Sanctification or detective work? American Psychologist, 1969, 24, 83-91.
- Tversky, A., & Kahneman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.
- Whiteley, J. M. (Ed.) Research in counseling. Columbus, Ohio: Charles E. Merrill Publishing Co., 1967.
- Whiteley, J. M., & Allen, T. W. Suggested modifications in scientific inquiry and reporting of counseling research. Counseling Psychologist, 1969, 1 (2), 84-88.