DOCUMENT RESUME

ED 202 116 EA 013 565

AUTHOR Miskel, Cecil; Sandlin, Terry
TITLE Survey Research in Educational Administration: A Critical Analysis.
PUB DATE Apr 81
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Data Analysis; Data Collection; *Evaluation; *Evaluation Criteria; Evaluation Methods; Measurement Techniques; Reliability; Research Design; *Research Methodology; *Research Problems; Sampling; *Surveys; Validity
IDENTIFIERS Education Administration Quarterly; Journal of Educational Administration

ABSTRACT

In order to assess the methodological merit of the published research in educational administration based on surveys, the authors analyzed data from a random sample of 24 survey studies published in the "Education Administration Quarterly" and the "Journal of Educational Administration." Each article was evaluated according to six criteria governing sampling and instrumentation procedures. Raw scores, descriptive statistics, and graphical analysis were used to judge levels of quality and to test for systematic improvements in the scientific rigor of the published research. The study reveals that both journals suffer general shortcomings across a number of the quality criteria. Often methodology is treated so tersely that readers cannot evaluate the quality of procedures. The authors conclude that improvements in methodological rigor have been uneven and modest. (Author/WD)

SURVEY RESEARCH IN EDUCATIONAL ADMINISTRATION: A CRITICAL ANALYSIS

Cecil Miskel

Terry Sandlin

The University of Kansas

Paper Presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981

Survey Research in Educational Administration: A Critical Analysis

Cecil Miskel and Terry Sandlin

The University of Kansas

An examination of the research literature in educational administration, reveals a pervasive use of survey methods. A large portion of the published research includes a survey procedure. Given this pervasiveness, two important questions that need to be addressed are the following: What level of methodological rigor is evidenced by these studies? Did the rigor increase during the 1970s? The purpose of this paper is to assess the methodological merit of the published research in educational administration that has used survey procedures.

To answer the questions, data were taken from a random sample of 24 survey studies published in the Journal of Educational Administration and 23 in the Educational Administration Quarterly during the eight years of 1972-79. Each article was evaluated against six criteria for sampling and instrumentation procedures. Raw scores, descriptive statistics, and graphical analysis were used to assess the levels of quality and to test for systematic improvements in the scientific rigor of the published research. The deficiencies and strengths are presented and discussed. Qualitative differences between the journals and systematic trends were not evident. Finally, suggestions for future research are presented.

SURVEY RESEARCH IN EDUCATIONAL ADMINISTRATION: A CRITICAL ANALYSIS

Several critiques have been made of the literature in educational administration. In particular, five analyses standout. After making a comprehensive review of the Educational Administration Quarterly (EAQ), Campbell concluded that the announced purpose of the EAQ to publish conceptual, empirical, and analytic manuscripts has been largely achieved.[1] Moreover, most of the articles published in the EAQ are of good quality, some of them of superior quality and certainly as well done as many of those in other journals. In his judgment, the more recent articles seemed to be of better quality. Campbell also noted that the Journal of Educational Administration (JEA) represents the only journal that truly competes with the EAQ for publishing manuscripts focused primarily on the field of educational administration.

In a recent review of the research in educational administration, Boyan[2] agreed with Immegart and Boyd.[3] They believe that to an increasing extent internal specialization characterizes inquiry in the field. In other words, scholars tend to specialize in theories guiding inquiry and methods of doing research. Clearly, survey methods represent one area of specialization in educational administration research. A cursory examination of the published research in the educational administration literature yields the observation that survey procedures constitute the most frequently used empirical method. In contrast to Campbell's conclusion, Boyan believes that improvements in the states of inquiry in the field have been uneven and modest.

In a more directed critique, McNamara analyzed the statistical methodology employed in the articles published in the first twelve volumes of the EAQ.[4] McNamara concluded that univariate analysis techniques that test for differences between groups dominate the treatment

of data. Moreover, tests were lacking to estimate the significance of the variable relationships; that is, post hoc procedures to determine the proportion of variance explained in the dependent variable were not in evidence.

Apparently, some disagreement exists among the critiques regarding whether the scientific merit of the published research in educational administration is improving. Willson's evaluation of the research techniques used in the studies published in the American Educational Research Journal showed that over a ten year period (1969-1973) no broadening of the pool of research techniques had occurred, and deficiencies in randomization and unit of analysis remained.[5]

Given the pervasiveness of survey research, the limitations cited in earlier critiques and the possibility that the level of merit has changed, two questions that become important are: 1) What level of methodological rigor is evidenced by the published investigations that have used survey procedures? 2) Did the methodological rigor increase during the 1970s? The purpose of this paper is to respond to these questions with findings from a systematic evaluation of the articles reporting findings of survey research studies that are found in the Educational Administration Quarterly and the Journal of Educational Administration.

Survey Research

Survey research is a planned collection of data that consists of procedures used by investigators to enter a subject population and to measure a specific set of responses. As a branch of social scientific research, investigations comprised of sample survey methods examine large and small populations by selecting and studying samples chosen from the populations.[6] Although the purposes guiding specific studies vary, survey research usually provides descriptions, explanations, and

predictions of relationships among sociological and psychological variables--facts, opinions, attitudes, and behavior.

To gather data, sample surveys employ questionnaires and interviews, attitude scales, projective techniques, existent records such as census data, and various related methods.[7] Typically, independent variables are neither manipulated nor are control conditions employed.[8] Unfortunately, survey studies frequently fail to meet the common scientific criteria of quality and are too often conducted with insufficient planning.[9] The design and execution of a survey project not only requires technical expertise, but also arduous intellectual activities. Two areas of the design--sampling and measurement--are particularly susceptible to error and poor execution.

Criteria of Quality

Sampling

Population. A population for any investigation is the total number of units in which the researcher is interested. A sample is a subset of the population which is drawn because it is impossible or impractical to work with all of the units in the intended population.[10] By definition survey research links populations and samples. Survey researchers study samples drawn from populations and generalize the characteristics to the specified population.

The goal of sampling, therefore, is to select a smaller representative subset of elements from the entire population. Consequently, defining the population is an essential step before sampling procedures can be formulated. Sudman suggests the use of the following characteristics when defining populations: geography, personal variables of age, race, education, institutional affiliations and intentions, and organizational variables of size, school level, and private or public type.[11]

However, the generalizability of the findings can be jeopardized by several components of the sampling design including the unit of analysis, participation rate, and method of sampling.

Unit of analysis. According to Kish, the unit of analysis is the fundamental element of the population for which information is being sought.[12] Perhaps a more descriptive designation is the unit of investigation. In other words, unit of analysis refers to the fundamental elements of the population about which inferences are to be drawn.

As noted by Burstein, the hierarchical nature of educational organizations produces several levels that can serve as focal units.[13] For example, appropriate units of analysis for research in educational administration include individuals, classrooms, curriculum programs, school attendance centers, and districts. Moreover, the selection of the appropriate unit(s) of analysis should be based on the theoretical formulation of the study.[14] If the variables to be studied are concerned with administrative organization and processes in school buildings, for instance, the attendance center is a more appropriate unit of analysis than the individual.[15]

The unit of analysis or investigation directly affects sampling procedures by defining the appropriate sampling element and the number of subjects in the study. If, for example, elementary students are the focus of the study and hence the unit of analysis, a larger number of units exist and probably can be drawn more easily and in larger numbers than if the unit of analysis is the elementary school. Consequently, the specification by researchers of the unit of analysis allows a more accurate assessment of the adequacy of the sampling design.

Participation rates. The survey population usually differs somewhat from the target population. The primary difference frequently arises

from nonresponses.[16] Kerlinger, for example, concludes that the low return rates of mail questionnaires yield many studies that are worse than useless, except in highly sophisticated hands.[17] The problem is not in just the number of those participating, but the problem of potential bias in those returning questionnaires. When mail questionnaires are used, a general agreement seems to exist that the return rate should approach 80 percent or higher.[18] Even with such a high return rate, the representativeness of the responses may remain questionable.[19]

Methods of sampling. A major source of sampling error occurs because of the sampling design itself. Potential sampling designs include convenience, systematic, simple random, and probability selection techniques. In many cases, combinations of the procedures are employed.

Convenience sampling is selecting a particular subgroup within the population because that subgroup is easily accessible. This technique contains the assumption that the available respondents are representative of the total population, which sometimes is not true.[20] Nevertheless, convenience samples often are necessary and unavoidable. Kerlinger notes that the weaknesses of convenience samples can somewhat be reduced by the use of knowledge and care in selecting the samples, along with replicating studies with different samples.[21]

In systematic sampling, the first sample element is randomly chosen from a list, and subsequent units are chosen at regular intervals from the list. Two basic assumptions undergird this procedure: the list is arranged randomly, and the feature by which it is arranged is not related to the purpose of the survey. If these two assumptions are met, most systematic samples exhibit the same precision as random samples. Systematic samples sometimes are designated "pseudo-simple random samples" or "quasi-random samples."[22]

Random sampling occurs when each element in a population has an equal chance of being drawn, and all possible samples have an equal chance of being drawn.[23] The advantage of using random samples is that constant and independent probabilities are ensured. Therefore, random sampling forms the basis of adequate sampling procedures. Yet, in many cases, simple random sampling may not be sufficient to minimize error and, and more elaborate designs are required.

Probability samples use some form of random sampling in one or more of their stages. The most common form of probability sampling is stratified sampling. In stratified sampling the population is divided into strata such as administrators and teachers, females and males, or elementary and secondary schools. Random samples are then drawn from each stratum. Stratified sampling procedures are appropriate when specific strata are of interest or when prior information suggests differences among the strata. The purpose of stratified sampling is to reduce sampling error and to insure that the focal strata comprising the population are represented in the sample.[24]

Measurement

Oppenheim observed that, in general, great strides have been made in the improvement of sampling methods, but similar gains are not apparent in questionnaire and interview development.[25] Although many concerns exist about measurement techniques, Kerlinger agrees with Pfeiffer and Heslin who maintain that the most critical considerations are validity and reliability.[26]

Validity. The basic questions of validity center on what characteristics are being measured, what the scores mean, and how useful the data are. As complex as the subject is, the most common definition of validity is epitomized by the question: Are we measuring what we think we are measuring?[27]

In educational research, the three most important types of validity are content, criterion related, and construct.[28] Content validity refers to the initial impressions which the user or judge has of the instrument. More specifically, content validity is the judged representativeness of the items; that is, the items adequately sample the content tapped by the measuring device.

Criterion related validity is established by comparing test scores with one or more external variables, or criteria, that measure the attribute under study. Predictive, concurrent, convergent, and discriminant validity are common types of criterion validity.

Kerlinger asserts, however, that scientifically speaking, construct validity is the most important type.[29] Its significance resides in the fact that construct validity unites psychometric techniques with theoretic concepts. Three parts comprise construct validity: indicating what constructs potentially explain test variance, deriving hypotheses from the theory involving the construct, and testing the hypotheses empirically.[30]

Reliability. The basic questions of reliability pivot on concepts such as stability, dependability, accuracy, and unsystematic or error variance. The basic definition of reliability is illustrated by the question: How precise or accurate is the measuring instrument?[31] In other words, the greater the consistancy of responses to items measuring the same concept, the greater the reliability. The more reliable the measure, the less random error it generates.

Technically, the coefficient of reliability is the variance ratio of true scores to total scores on equivalent forms of a measure.[32] A number of different testing and statistical procedures have been proposed to provide coefficients of reliability. Stanley lists the follow-

ing as three major procedures.[33] (1) Administration of two parallel forms and correlating the resulting scores. (2) Administration of the same measure at a later time and correlating the resulting scores. This is sometimes referred to as an estimate of stability test-retest reliability. (3) Subdivision of a single measure into two presumably parallel groups of items and correlating the resulting two scores. Spearman-Brown's prophecy formula and Cronbach's alpha represent examples of this approach. Regardless of how the calculation is made, the reliability coefficient is only an estimate of the percentage of total variance that can be described as actual variance and not due to error.

In summary, four criteria of quality for sampling procedures and two for measurement techniques have been derived, defined, and discussed. These criteria were used systematically to evaluate a random sample of survey research studies.

Methods

Population and Sampling

The population for the study was the articles using a survey research procedure that had been published in 24 issues and eight volumes (8-15) of the Educational Administration Quarterly (EAQ) and in 16 issues and eight volumes (10-17) of the Journal of Educational Administration (JEA). These volumes were published from 1972 through 1979. The multi-level units of analysis were the issue and the volume comprising the journals.

During the eight year period, 141 and 161 articles were published in the EAQ and JEA respectively (see Table 1). Using the definition of survey research presented earlier in this paper, the articles were classified as either employing a survey research procedure or being of another type. As an estimate of reliability, both investigators had to

agree that an article contained a survey component. As shown in Table 1, both journals published a similar number (EAQ = 64 and JEA = 66) of survey studies. To test whether the frequency of survey investigations per volume exhibited systematic trends a chi square test of homogeneity was calculated. The chi square value was 4.67 ($p > .05$). Therefore, the number of survey articles per volume appears to be the same.

Table 1 about here

To ensure a representative sample, a stratified random sampling procedure was used. The population of survey articles was stratified by journal, volume, and issue. In the case of the EAQ, one article was randomly selected from each of the 24 issues or three per volume. The exception was volume 10, issue 3, 1974 because it did not contain a survey research article. Therefore, the sample from the EAQ included 23 articles in 23 issues and eight volumes.

Since a volume of the JEA is comprised of only two issues, one article was randomly selected from each issue and then a third article was chosen randomly from each volume. Thus, the sample from the JEA included 24 articles in sixteen issues and eight volumes.

Instrumentation

To operationalize the evaluation criteria of survey research methods, the nine item instrument shown in Table 2 was developed. As reviewed earlier in this paper, four sampling and two measurement criteria served as a conceptual guide in building the instrument. The sampling criteria items measured the specification of the population, the specification of the unit of analysis, participation rates, and the type of sampling method. The instrumentation criteria scale items assessed the specifications of

validity and reliability estimates. Similar to the measurement procedure used by Willson, instruments not subject to reliability analysis in the usual sense of psychometrics were not considered.[34] The most common examples of this type of measure is the use of published figures and demographic information such as sex, age, and job assignment. Three general items (5,8,9) were created by combining and averaging the values of two or more items.

Table 2 about here

To quantify the quality of each criterion three categories were developed for each item that indicated poor, adequate, and good survey methods. The three categories were assigned scale values of 0, 1, and 2 respectively. The definitions and category values for the items and the two scales are presented in Table 2. In addition, the four items comprising the sampling criteria scale were averaged to yield item 5, an indicator of overall adequacy of the sampling procedures. The possible range of scores was 0-2. Similarly, the two item scores for the instrumentation criteria were averaged to produce item 8, a summary statistic for measurement adequacy. The possible range of scores was 0-2. Finally, an overall assessment of the survey methodology was calculated by averaging item 5 (sampling criteria scale) and item 8 (measurement criteria scale). The range, therefore, was 0-2 with high scores indicating the use of better survey methods than low scores. When a study used more than one population, sample, or measure, the values were averaged to produce a single score for each item. The scores for items for the individual articles in each issue and volume were aggregated to create scores for the 23 issues and eight volumes of the EAQ and the 16 issues and eight volumes of the JEA.

Application of the measure given in Table 2 was accomplished by the two investigators independently rating each article on the eight items. Intercoder reliability was attained by calculating the percentage of agreement for each item. The average percentage of intercoder agreement across the eight items was 96% and the range was 95% to 98%. Intracoder reliability was estimated when each investigator recoded the data one week later. Intracoder agreements over all the variables were 97% and 98%.

A panel procedure was used to establish content validity for the instrument. Three research methodologists agreed that the six individual items represent important criteria for evaluating survey methods.

Analysis

To answer question one regarding the quality of research, raw scores and descriptive statistics were used. Specifically, means and standard deviations were calculated separately for each indicator of quality across the issues and volumes of both journals. To respond to question two about the possibility of a trend in the level of methodological rigor during the 1972-79 time period, the scores for the sampling criteria and measurement criteria. Moreover, these procedures were applied at two levels or units of analysis, that is, by issue and by volume for each journal.

Findings

The two questions guiding the research are addressed separately. In addition, the findings are presented for the two levels of analysis and for each journal. Comparisons across levels and journals complete the response to each question.

Question One: What level of methodologial rigor is evidenced by the published investigations that have used survey procedures?

Issue level of analysis. Table 3 presents a summary of data for the 23 issues of the EAQ. The scores for each criterion across the issues comprise the nine numbered columns, while the scores for each issue across the nine criteria form the rows. Means and standard deviations for each criterion across the issues are given at the bottom of Table 3.

Table 3 about here

Based on the values of means for the EAQ issues on the sampling criteria, only two of the four indicators had an average value that can be described as adequate to good (see Table 3). The population criterion (column 1) had a mean of 1.33. The population was specified in six of the 23 issues, and in all cases the samples were discussed adequately. Similarly, the methods of sampling (column 4) tended to be adequate with a mean of 1.36. Either a random sampling procedure or the entire population was typically used. However, in six articles it was impossible to determine the selection techniques. In contrast, the specification of the unit of analysis (column 2) was poorly detailed and had a mean of .20 on a .00-2.00 scale. The unit of analysis was mentioned in only five investigations, and multilevel analyses were conducted in only one study. Moreover, the unit of analysis in a dominate portion of the issues was the individual, student, teacher, or administrator. The 23 studies also demonstrated less than adequate quality on the criterion for the participation rates (column 3) with a mean of .75. Eight issues either

failed to provide adequate information to calculate the return rates or had less than a 60% level of participation. As indicated by the mean of .91 for the sampling criteria scale (column 5), the overall quality of the sampling procedures in the EAQ has been marginally adequate.

The quality of measurement in the EAQ studies varies with the criterion. Reliability (column 6) with a mean of 1.05 is addressed more systematically than validity (column 7) with a mean of .57. Moreover, many of the articles refer to reliability and validity by citing other studies or mentioning in a vague fashion that reliability and validity had been established. As indicated by the mean of .81 for the measurement criteria scale (column 8), the overall quality of the measurement procedures in the EAQ has been less than adequate.

Table 4 displays a summary of the data for the 16 issues of the JEA. Only two of the four sampling criteria can be considered adequate. Specification of the population has a mean of 1.19, and the sampling method criterion attained a mean of 1.30. The lowest possible score of zero was calculated for the unit of analysis criterion. No study explicitly stated what the focal unit was, and none used multilevel analyses. Similarly, the rates of participation were poor with a mean of .55. Overall, the sampling criteria scale mean of .71 indicates that the sampling procedures exhibited by the articles published in JEA tended to be inadequate.

The two means for the quality of measurement criteria in the JEA investigations attain similar levels. The reliability criterion with a mean of 1.07 reached adequacy, while the validity criterion approached the adequate level with a mean of .85. With one exception for reliability and two for validity, each issue addressed these criteria to some extent. As indicated by the mean of .96 for the measurement criteria

scale, the quality of the measurement procedures for the JEA was marginally adequate.

At the issue level of analysis, the EAQ and JEA exhibit similar strengths and weaknesses on the survey research criteria (see Tables 3 and 4). Both publish articles that show the highest quality on specifying the population, using random sampling procedures, and providing estimates of reliability. Conversely, both publish articles that are inadequate in specifying the unit(s) of analysis and giving or achieving sufficient participation rates. The JEA articles tend to explicate the validity of measures more adequately than those in the EAQ. Slightly better scores on the sampling criteria scale were attained for the EAQ ($\bar{X}$ = .91) than for the JEA ($\bar{X}$ = .71). On the measurement criteria score the reverse relationship holds with the JEA having a mean of .96, and the EAQ having a mean of .81. The overall criteria scale scores are essentially equal (EAQ = .86; JEA = .84).

Volume level of analysis. Tables 3 and 4 summarize the data by volumes for the EAQ and JEA respectively. Aggregating the issue level data to the volume level produced few variations in the results. The largest change for the EAQ was a decline in the mean from 1.05 to .93 for the reliability of measurement criterion. The largest change in the JEA was a decline in the mean from 1.30 to 1.03 for the sampling method criterion. All other changes were minimal.

Summary of findings for question one. The levels of methodological rigor are not high. With a score of 1.00 being defined as adequate, neither the EAQ nor JEA attained this level on the sampling criteria scale, measurement criteria scale, or overall quality scale. Therefore, the survey research rigor approaches, but does not attain adequacy in the two journals.

Question Two: Did the methodological rigor increase during the 1970s?

Issue level of analysis. Graph 1 pictorially displays the means of sampling and measurement criteria scales across the 23 issues of the EAQ. No trends are apparent in either scale. The data points bounce widely and show no tendancies to form a narrower band or to become higher over time.

Graph 1 about here

Graph 2 presents the means of the sampling and measurement criteria scales across the 16 issues of the JEA. No trends are evidenced for either scale.

Graph 2 about here

Volume level of Analysis. Graph 3 exhibits the data summaries for the sampling and measurement criteria scores for the eight volumes of EAQ. While no trends exist for improved rigor, volumes 11 through 15 show more stability than volumes 8 through 11. For the later volumes, the means show less variability but remain at or somewhat below 1.0. However, the means on the measurement criteria sc declined for volumes 14 and 15.

Graph 3 about here

Graph 4 plots the sampling and measurement criteria means across the eight volumes of JEA. No trends are evident for criteria, and the variability does not seem to lessen in the later volumes.

Graph 4 about here

Summary of the findings for question two. The results from graphing the data show no increased or decreased rigor during the 1970s in the quality of survey research methodology in the EAQ or JEA. In fact, the data plots for both journals suggest wide variations in quality from issue to issue and from volume to volume.

Discussion

Level of Methodological Rigor

Even a cursory examination of data discussed earlier and presented in Tables 3-6 reveals several important characteristics about the quality of the survey research published in the EAQ and JEA. Perhaps most alarming finding is the lack of consideration given to the unit of analysis. Only five EAQ articles and no JEA articles explicitly mentioned the focal unit of the study. With one exception, the investigations used only a single unit of analysis. Typically, the focus was on the individual with no evidence of the data being aggregated to the classroom, school building, or district. This result is similar to the finding of Willson that few studies published in the American Educational Research Journal recognize different aggregation levels.[35] Potentially, a large amount of information from different perspectives is being lost by the practice of using a single unit of analysis. Greater emphasis should be placed on organizational levels such as the school attendance center and district in research for the field of educational administration.

The reporting of research with either low participation rates or no mention of the rates occurs too frequently. In particular, over one-half of the JEA issues and over one-third of the EAQ issues failed to meet

minimal standards on this criterion. Inadequate participation rates certainly calls into question the generalizability of many of the findings that appear in the educational administration literature.

Hardly any of the investigations specifically defined the population of interest. Rather, emphasis was typically placed on describing the sample. After reviewing a large number of studies, a somewhat cynical impression emerges that the elaborate description of the sample is part of an effort to suggest that the findings are highly generalizable. Yet in many cases the population is not mentioned, and the participation rate is low or not identified. This condition is poor practice even when the intentions are positive, but it is an unethical practice when the intentions are negative.

As cited earlier, Oppenheim has observed that great strides have been made in the improvement of sampling methods.[36] The scores on the sampling criteria items and scale, however, suggest that the improvements have not been applied systematically in the field of educational administration. Either many researchers have not learned of the advances, or they have chosen to ignore them.

Inadequacies also exist in the quality of the survey measures. While references to the reliability estimates resulted in relatively large mean scores of about 1.0, the finding must be tempered. Almost 30% of the EAQ issues made no mention of reliability. A possible explanation for this omission is that investigators attached little importance to reporting the reliability of previously developed and used measures such as the OCDQ, LBDQ, PCII and so forth. On the other side of the issue, almost 40% of the EAQ meticulously provided empirical values for all of the instruments included in the study. In a sense, the reporting of reliability estimates is either a feast or famine.

Mentioning and describing validity was particularly weak in the EAQ. About 40% failed to acknowledge this measurement criterion. However, all but two of the JEA issues at least mentioned validity.

An interesting difference between the EAQ and JEA should be noted. The overall sampling criteria values were higher for the EAQ, while the overall measurement values were higher for the JEA. In fact, the largest discrepancies between the EAQ and JEA occurred between the sampling criterion for participation rates and the measurement criterion of validity. EAQ had the highest participation rate scores, while JEA had the highest validity scores. Evidently the editorial boards emphasize different specific criteria.

Trends in Methodological Rigor

The findings of the current study regarding survey research procedures do not support Campbell's conclusion that recent articles in the EAQ are improving.[37] Rather, the results reported in this paper suggest that Boyan's ssertion reflects the state of research.[38] Improvements have been uneven and modest. Perhaps the topics, problem definitions, and use of conceptual models are better in more recent volumes of the EAQ and JEA. But the sampling and measurement criteria of quality show wide fluctuations and demonstrate no discernable trends. Consequently, the present findings support the assertion that the levels of quality remain eratic for the survey research procedures used in the field of educational administration.

Related Observations

Although the concept of specialization of methods as mentioned by Immegart, Boyd and Immegart, and Boyan was not addressed directly, several names seemed to appear more frequently than would normally be expected.[39] This observation provides some support for the idea that some

researchers in educational administration specialize in using survey research methods. Since over 40% of the published articles in the EAQ and JEA are survey based, the contention could be made that the field itself specializes in survey research.

Comparing the scores of the two journals on survey research criteria reveals few substantial differences in quality. The EAQ seems to emphasize sampling adequacy, while the JEA tends to insist on at least mentioning validity in each study. Both suffer general shortcomings across a number of the quality criteria.

Conclusion

Two common approaches to the evaluation of research are to find fatal flaws or redeeming features. In a sense the present study has focused on fatal flaws. The blame for the observed weaknesses of the EAQ and JEA should not be placed on the respective editors and editorial boards, however. The two journals probably reflect the general sophistication level of the field and publish the best survey research articles that are available in educational administration. Many of the investigations had several redeeming characteristics. Indeed, the quality ranged from excellent to abysmally low.

The editors and editorial boards can improve the quality of presentation if not the quality of the methodology itself. They should demand that the researchers provide a clear exposition of the methods that were used. There is no reason for the methodology to be treated so tersely that the readers can not evaluate the quality of procedures.

Finally, the quality criteria for survey research are well developed and sources are available to guide the design of research projects. For most studies, slight increases in technical expertise, planning, and physical effort would improve the survey research procedures immensely. Scholars in educational administration should make the commitment to improve the methods and hence the quality of our knowledge base.

Notes

1. Roald Campbell, "A Critique of the Educational Administration Quarterly," Educational Administration Quarterly 15 (Fall 1979): 1-19.
2. Norman J. Boyan, "Follow the Leader: Commentary on Research in Educational Administration," Educational Researcher 10 (February 1981): 6-13, 21.
3. Glenn L. Immegart, "The Study of Educational Administration, 1954-1974," in L. L. Cunningham, W. G. Hack, and R. O. Nystrand (eds.), Educational Administration: The Developing Decades (Berkeley, Calif.: McCutchan, 1977); and William L. Boyd and Glenn L. Immegart, "Education's Turbulent Environment and Problem Finding: Lines of Convergence," in Glenn L. Immegart and William L. Boyd (eds.), Problem-Finding in Educational Administration: Trends in Research and Theory (Lexington, Mass.: Heath, 1979).
4. James F. McNamara, "Practical Significance and Statistical Models," Educational Administration Quarterly 14 (Winter 1978): 48-63.
5. Victor L. Willson, "Research Techniques in AERJ Articles: 1969 to 1978," Educational Researcher 9 (June 1980): 5-10.
6. Fred N. Kerlinger, Foundations of Behavioral Research, 2nd ed. (New York: Holt, Rinehart, and Winston, 1973), pp. 410-423.
7. Abraham N. Oppenheim, Questionnaire Design and Attitude Development (New York: Basic Books, 1966), p. 1.
8. Lawrence S. Meyers and Neal E. Grossen, Behavioral Research: Theory, Procedure, Design, 2nd. ed. (San Francisco: Freeman, 1978), p. 191.
9. Oppenheim, Questionnaire Design, op. cit., p. 3.

10. John L. Sullivan and Stanley Feldman, Multiple Indicators: An Introduction (Beverly Hills: Sage, 1979), p. 76.

11. Seymour Sudman, Applied Sampling (New York: Academic Press, 1976), pp. 12-14.

12. Leslie Kish, Survey Sampling (New York: Wiley, 1965), pp. 6-7.

13. Leigh Burnstein, "The Analysis of Multilevel Data in Educational Research and Evaluation," Review of Research in Education 8 (1980): 158-233.

14. Ibid., p. 161.

15. Jon Conklin, "Issues and Policy Implications of the Analysis of Multilevel Data," Paper Presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

16. Kish, Survey Sampling, op. cit., p. 7.

17. Kerlinger, Foundations of Behavioral Research, p. 414.

18. Ibid. and Sudman, Applied Sampling, op. cit., p. 30.

19. C. A. Mosher, Survey Methods in Social Investigation (London: Heineman, 1966), p. 181.

20. Meyers and Grossen, Behavioral Research, op. cit., p. 113-114.

21. Kerlinger, Foundations of Behavioral Research, op. cit., p. 129.

22. Sudman, Applied Sampling, op. cit., p. 57.

23. B. J. Winer, Statistical Principles in Experimental Design (New York: McGraw Hill, 1971), p. 5.

24. Sudman, Applied Sampling, op. cit., p. 111.

25. Oppenheim, Questionnaire Design, op. cit., p. 21.

26. Kerlinger, Foundations of Behavioral Research, op. cit., p. 442 and J. William Pfeiffer and Richard Heslin, Instrumentation in Human Relations Training (San Diego: University Associates, 1973), p. 27.

27. Kerlinger, Foundations of Behavioral Research, op. cit., pp. 456-459.

28. Ibid.

29. Ibid.

30. Lee J. Cronbach, Essentials of Psychological Testing, 3rd ed. (New York: Harper & Row, 1970), p. 143.

31. Kerlinger, Foundations of Behavioral Research, op. cit., pp. 442-455.

32. Edward E. Cureton, "The Definition and Estimation of Test Reliability," Educational and Psychological Measurement 18 (1958): 715-738.

33. Julian C. Stanley, "Reliability," in R. L. Thorndike (ed.), Educational Measurement, 2nd ed. (Washington, D.C.: American Council on Education, 1971), pp. 356-442.

34. Willson, "Research Techniques," op. cit., p. 8.

35. Ibid.

36. Oppenheim, Questionnaire Design, op. cit., p. 21.

37. Campbell, "A critique," op. cit.

38. Boyan, "Follow the Leader," op. cit.

39. Immegart, "The Study," op. cit.; Boyd and Immegart, "Education's Turbulent Environment," op. cit; and Boyan, "Follow the Leader," op. cit.

Table 1

Frequency and Percent of Survey Articles

Published in Each Journal by Year

	Educational Administration Quarterly			Journal of Educational Administration		
Year	Total Articles	Survey Based	Percent of Total	Total Articles	Survey Based	Percent of Total
1972	16	8	50.0	15	6	40.0
1973	16	10	62.5	24	7	29.2
1974	18	6	33.3	20	10	50.0
1975	18	6	33.3	21	10	47.6
1976	18	11	61.1	22	7	31.8
1977	18	9	50.0	21	9	42.8
1978	19	9	47.4	19	6	31.6
1979	18	5	27.8	19	11	57.9
Summary	141	64	45.4	161	66	41.0

Table 2

Scales and Items Comprising the Survey Research Evaluation Instrument

SAMPLING CRITERIA SCALE

1. Specification of the Population
   _____ 0. Poor. The population was neither identified nor was the sample described.
   _____ 1. Adequate. The population was not mentioned, but it could be inferred from the description of the sample.
   _____ 2. Good. The population was explicitly identified.

2. Unit of Analysis
   _____ 0. Poor. The unit of analysis was not mentioned, but an identifiable single level unit was used.
   _____ 1. Adequate. A single level unit of analysis was identified, or multilevel units of analysis were used.
   _____ 2. Good. Multilevel units of analysis were explicitly identified and used.

3. Participation Rates
   _____ 0. Poor. The participation rate was not specified or less han 60%.
   _____ 1. Adequate. The participation rate was between 60-80%.
   _____ 2. Good. More than 80% of the sample participated.

4. Sampling Method
   _____ 0. Poor. The method of sampling was either not specified or convenience procedures were used.
   _____ 1. Adequate. A mixture of convenience and random selection procedures were used.
   _____ 2. Good. Random selection procedures were used.

5. Sampling Criteria Scale (Sum of items 1-4).
   _____ Mean of items 1-4.

Table 2 (continued)

Scales and Items Comprising the Survey Research Evaluation Instrument

MEASUREMENT CRITERIA SCALE

6. Reliability
   _____ 0. Poor. Reliability was not mentioned.
   _____ 1. Adequate. Reference was made to standardized instruments, t previous studies or to the type(s) without specifying values
   _____ 2. Good. Empirical values were reported.

7. Validity
   _____ 0. Poor. Validity indicators were not mentioned.
   _____ 1. Adequate. Reference was made to standardized instruments, to previous studies, or to the type(s) without describing the procedures.
   _____ 2. Good. The type(s) and procedures used to establish validity were described.

8. Measurement Criteria Scale
   _____ Mean of items 6-7.

9. Overall Quality
   _____ Mean of items 5 and 8.

Table 3

Summary of the Data for the Evaluation Criteria for the Survey Studies Published in 23 Issues of the Educational Administration Quarterly

		Sampling					Measurement			Overall
Issue	Volume	1	2	3	4	5	6	7	8	9
1	8	2.00	.00	1.00	2.00	1.25	.63	.38	.51	.88
2	8	1.00	1.00	1.33	2.00	1.33	.00	.00	.00	.67
3	8	2.00	.00	1.00	2.00	1.25	.00	.00	.00	.63
4	9	1.00	.00	.00	.00	.25	1.00	1.00	1.00	.63
5	9	1.00	.00	1.00	1.00	.75	2.00	1.00	1.50	1.13
6	9	1.00	.00	.00	.00	.25	2.00	2.00	2.00	1.13
7	10	1.50	.00	2.00	2.00	1.38	.00	.00	.00	.69
8	10	2.00	.00	1.00	2.00	1.25	.33	.33	.33	.79
9	11	1.00	.00	1.00	2.00	1.00	2.00	.63	1.32	1.16
10	11	1.00	.00	1.00	1.00	.75	.50	.50	.50	.63
11	11	2.00	.00	1.00	2.00	1.25	2.00	.00	1.00	1.13
12	12	2.00	1.00	.00	2.00	1.25	.00	.00	.00	.63
13	12	1.00	.50	.00	1.33	.70	2.00	1.00	1.50	1.10
14	12	1.00	1.00	1.50	2.00	1.38	2.00	1.00	1.50	1.44
15	13	1.00	.00	.00	2.00	.75	2.00	2.00	2.00	1.38
16	13	2.00	.00	1.33	2.00	1.33	1.67	.83	1.25	1.29
17	13	1.00	.00	.00	.00	.25	.00	.00	.00	.13
18	14	1.00	.00	.00	2.00	.75	.00	.00	.00	.38
19	14	2.00	.00	1.00	.00	.75	2.00	1.00	1.50	1.13
20	14	1.00	.00	.50	.00	.38	1.00	1.00	1.00	.75
21	15	1.00	.00	.00	.00	.25	1.08	.33	.71	.48
22	15	1.00	1.00	.50	2.00	1.13	2.00	.00	1.00	1.07
23	15	1.00	.00	2.00	2.00	1.25	.00	.00	.00	.63
Mean		1.33	.20	.75	1.36	.91	1.05	.57	.81	.86
Standard Deviation		.47	.39	.66	.88	.41	.88	.62	.69	.34

Note: 1 = Specification of the Population; 2 = Specification of the Unit of Analysis; 3 = Participation Rates; 4 = Sampling Method; 5 = Sampling Criteria Scale; 6 = Reliability; 7 = Validity; 8 = Measurement Criteria Scale; 9 = Overall Quality Scale.

Table 4

Summary of the Data for the Evaluation Criteria for the Survey Studies Published in 16 Issues of the Journal of Educational Administration

		Sampling					Measurement			Overall
Issue	Volume	1	2	3	4	5	6	7	8	9
1	10	1.00	.00	.84	2.00	.96	1.88	1.50	1.69	1.33
2	10	1.00	.00	.00	.00	.25	1.00	2.00	1.50	.88
3	11	1.00	.00	.00	1.00	.50	.50	.50	.50	.51
4	11	2.00	.00	.00	2.00	1.00	.00	.00	.00	.50
5	12	1.00	.00	1.75	2.00	1.19	.25	.00	.13	.72
6	12	1.00	.00	.00	2.00	.75	1.00	1.00	1.00	.88
7	13	1.50	.00	.75	1.00	.81	1.38	.38	.88	.82
8	13	1.00	.00	.00	.00	.25	2.00	2.00	2.00	1.13
9	14	1.00	.00	.00	1.00	.50	1.00	1.09	1.05	1.78
10	14	1.00	.00	.00	.00	.25	2.00	1.00	1.50	.88
11	15	1.00	.00	2.00	2.00	1.25	1.00	1.00	1.0	1.13
12	15	1.50	.00	1.50	1.00	1.00	1.50	.84	1.17	1.09
13	16	1.00	.00	.00	.50	.63	.25	.25	.25	.44
14	16	1.00	.00	.00	.00	.25	.78	.33	.56	.41
15	17	2.00	.00	.00	.00	.50	1.00	1.00	1.00	.75
16	17	1.00	.00	2.00	2.00	1.25	1.50	.75	1.25	1.19
Mean		1.19	.00	.55	1.30	.71	1.07	.85	.96	.84
Standard Deviation		.36	.00	.80	.87	.37	.62	.61	.56	.28

Note: 1 = Specification of the Population; 2 = Specification of the Unit of Analysis; 3 = Participation Rates; 4 = Sampling Method; 5 = Sampling Criteria Scale; 6 = Reliability; 7 = Validity; 8 = Measurement Criteria Scale; 9 = Overall Quality Scale.

Table 5

Summary of the data for the Evaluation Criteria for the Survey Studies Published in the 8 Volumes of the Educational Administration Quarterly

		Sampling					Measurement			Overall
Volume	Year	1	2	3	4	5	6	7	8	9
8	1972	1.67	.33	1.11	2.00	1.28	.21	.13	.17	.73
9	1973	1.00	.00	.33	.33	.42	1.67	1.33	1.50	.96
10	1974	1.75	.00	1.50	2.00	1.32	.17	.17	.17	.74
11	1975	1.33	.00	1.00	1.67	1.00	.83	.38	.94	.97
12	1976	1.33	.83	.50	1.78	1.11	1.33	.67	1.00	1.06
13	1977	1.33	.00	.44	1.33	.78	1.22	.94	1.08	.93
14	1978	1.33	.00	.50	.67	.63	1.00	.67	.83	.75
15	1979	1.00	.33	.83	1.33	.88	1.03	.11	.57	.73
Mean		1.34	.19	.78	1.39	.83	.93	.55	.78	.86
Standard Deviation		.27	.30	.41	.61	.31	.52	.44	.45	.13

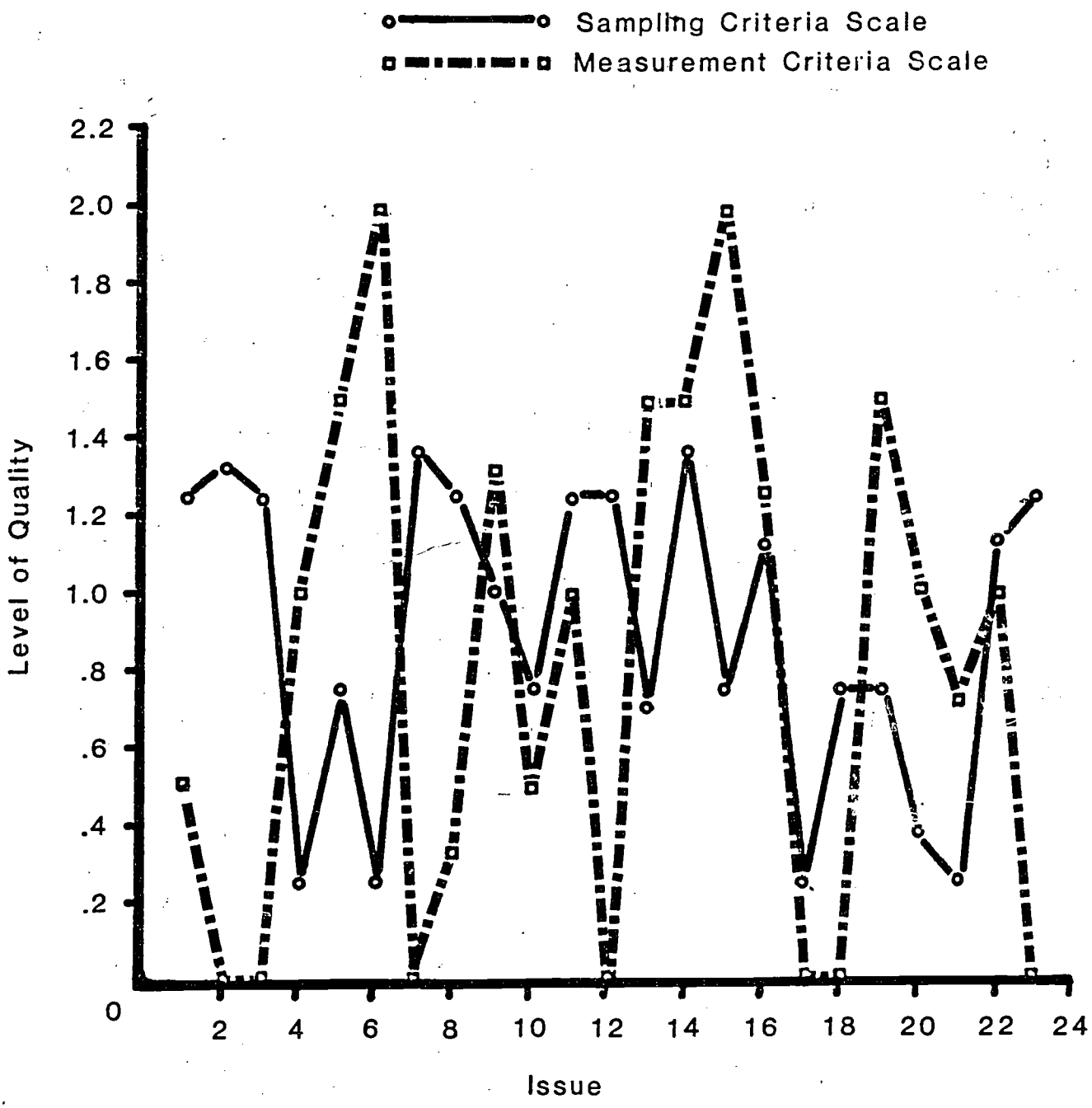
Note: 1 = Specification of the Population; 2 = Specification of the Unit of Analysis; 3 = Participation Rates; 4 = Sampling Method; 5 = Sampling Criteria Scale; 6 = Reliability; 7 = Validity; 8 = Measurement Criteria Scale; 9 = Overall Quality Scale.

Table 6

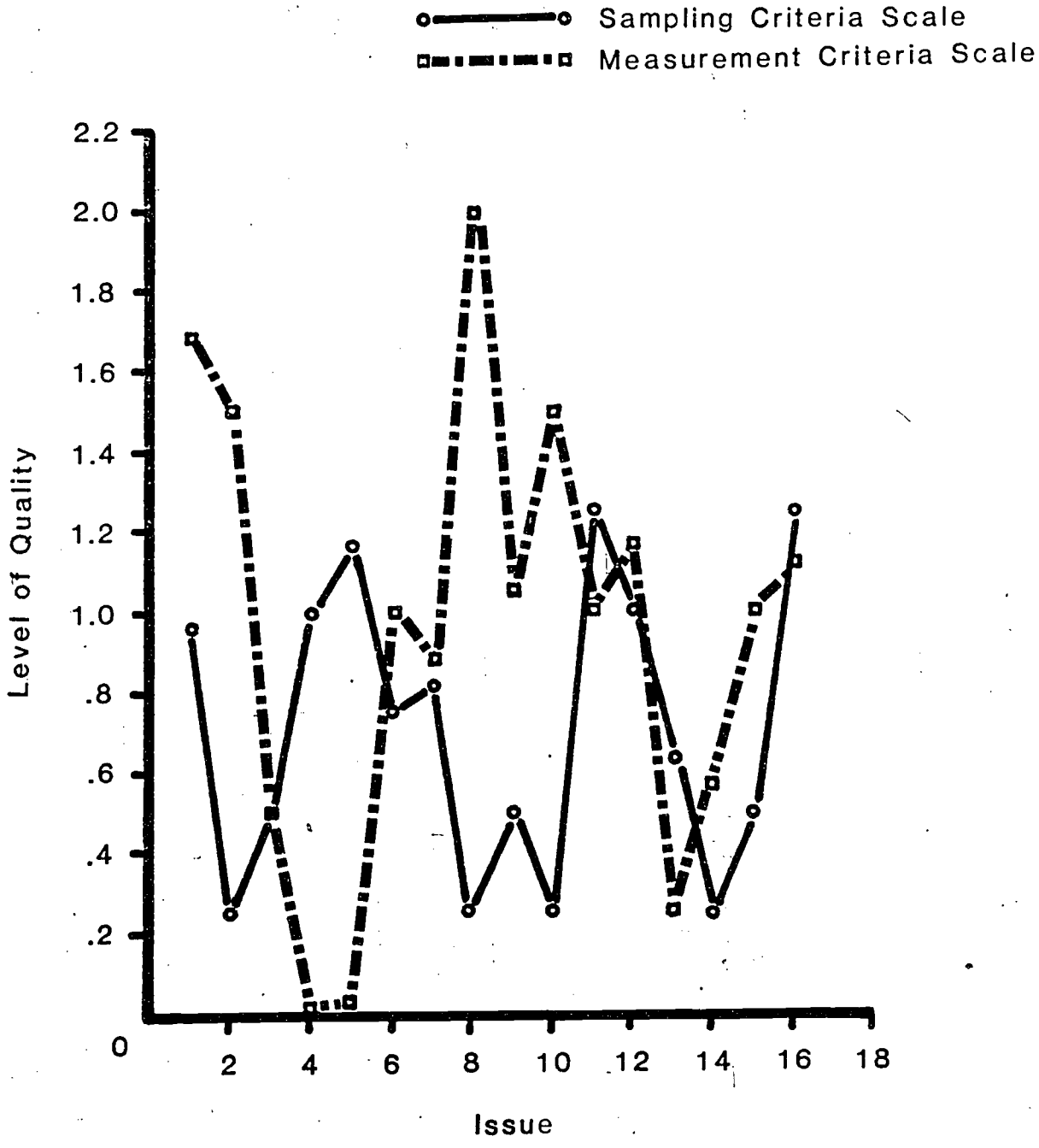
Summary of the Data for the Evaluation Criteria for the Survey Studies Published in the 8 Volumes of the Journal of Educational Administration

		Sampling					Measurement			Overall
Volume	Year	1	2	3	4	5	6	7	8	9
10	1972	1.00	.00	.42	1.00	.61	1.44	1.75	1.60	1.11
11	1973	1.50	.00	.00	1.50	.75	.25	.25	.25	.51
12	1974	1.00	.00	.88	2.00	.97	.63	.50	.57	.80
13	1975	1.25	.00	.38	.50	.53	1.69	1.19	1.44	.98
14	1976	1.00	.00	.00	.50	.38	1.50	1.05	1.28	.83
15	1977	1.25	.00	1.75	1.50	1.13	1.25	.92	1.09	1.11
16	1978	1.00	.00	.00	.25	.44	.52	.29	.41	.43
17	1979	1.50	.00	1.00	1.00	.88	1.25	.88	1.63	.97
Mean		1.18	.00	.55	1.03	.71	1.06	.85	.96	.84
Standard Deviation		.62	.00	.62	.60	.27	.52	.50	.50	.26

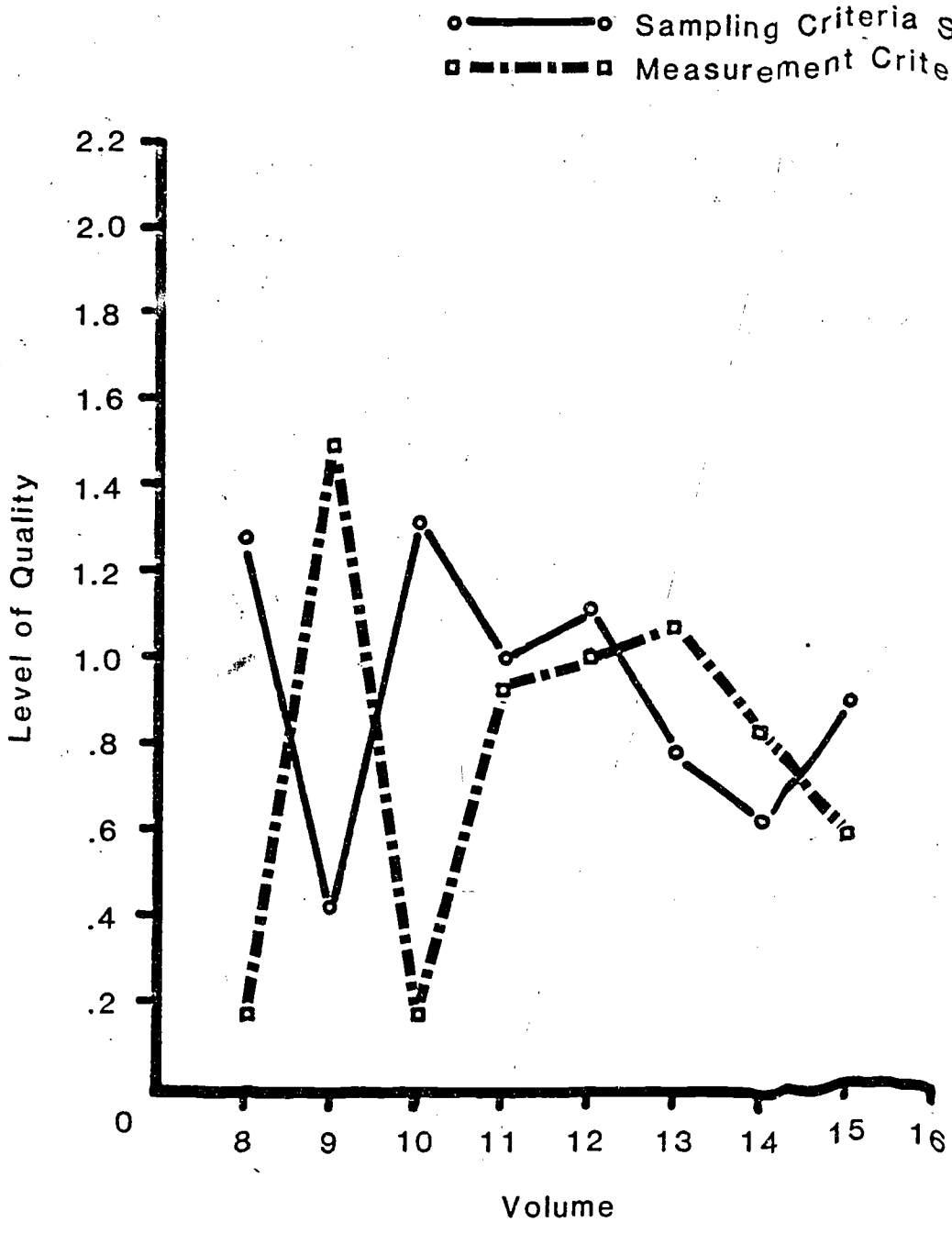
Note: 1 = Specification of the Population; 2 = Specification of the Unit of Analysis; 3 = Participation Rates; 4 = Sampling Method; 5 = Sampling Criteria Scale; 6 = Reliability; 7 = Validity; 8 = Measurement Criteria Scale; 9 = Overall Quality Scale.



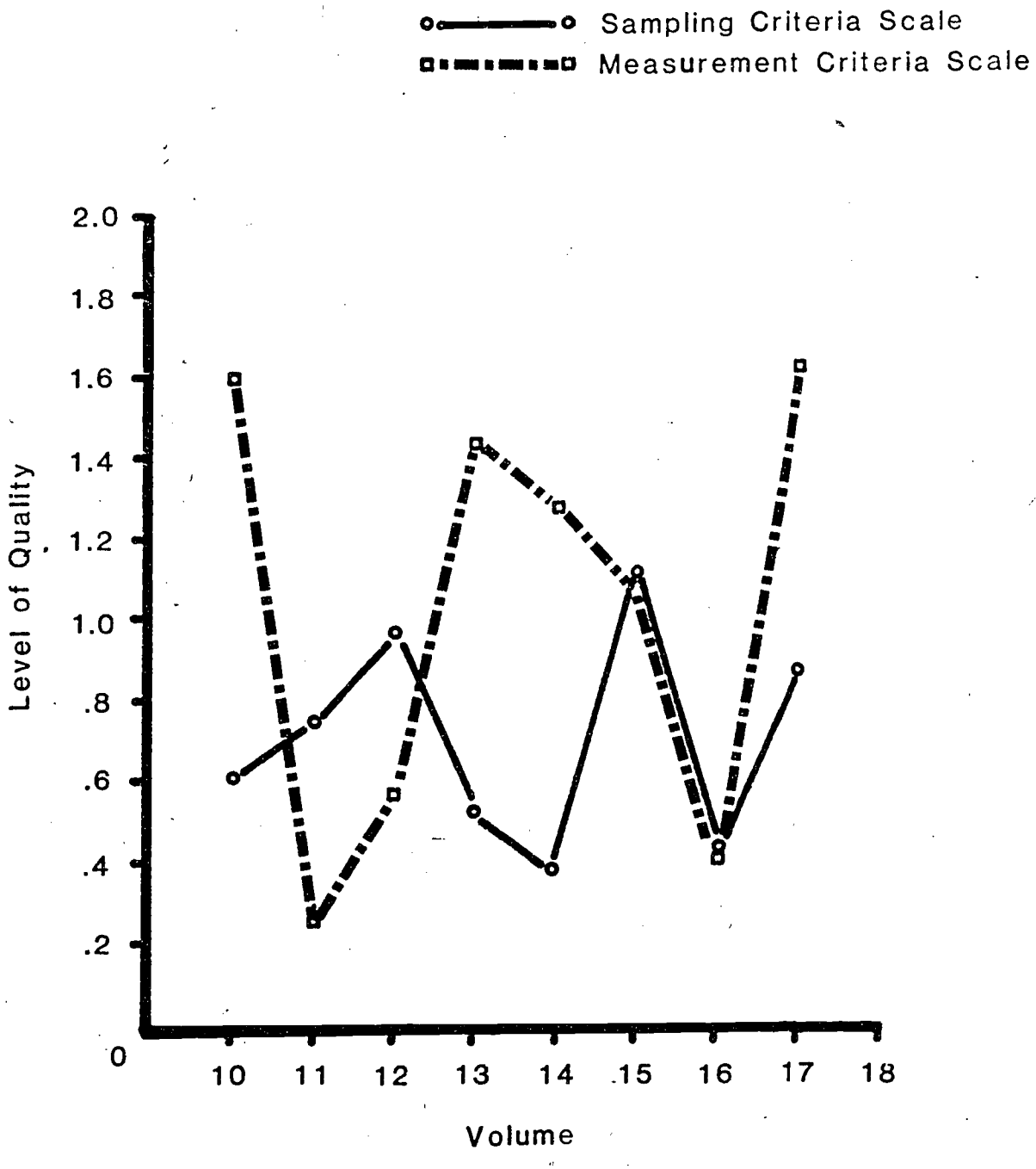
Graph 1. Level of Quality on the Sampling and Measurement Criteria Scales for Twenty-three Issues of EAQ



Graph 2. Level of Quality on the Sampling and Measurement Criteria Scales for Sixteen Issues of JEA



Graph 3. Level of Quality on the Sampling and Measurement Criteria Scales for Eight Volumes of EAQ



Grapgh 4. Level of Quality on the Sampling and Measurement Criteria Scales for Eight Volumes of JEA