DOCUMENT RESUME

ED 322 151 TM 014 876

AUTHOR Smoot, Sharene; And Others
TITLE Meta-Analysis of Single Subject Research in Special Education: A Common Metric and a Computerized Method.
PUB DATE Apr 90
NOTE 15p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavior Disorders; Comparative Analysis; Computer Uses in Education; *Educational Research; *Effect Size; Elementary Secondary Education; Graphs; *Meta Analysis; Outcomes of Treatment; Regression (Statistics); *Research Methodology; Sample Size; *Special Education; Student Problems
IDENTIFIERS *Single Subject Research Design

ABSTRACT

This study was designed to find: a suitable effect size (ES) measure or common metric (CM) for comparing the results of a set of single subject research (SSR) studies; and an easy way to convert the published graphs back into raw data from which ESs could be calculated. To meet the first objective, three possible formulas for measuring treatment effect were evaluated and then compared using identical data sets. To meet the second objective, a computer method was developed. Data for the study were taken from two previously published meta-analyses of SSR studies. The larger of the two (n=23 articles) which used piece-wise regression (PR) to calculate an ES, was the Skiba, Casey, and Center (1985-86) study concerning the use of non-aversive procedures in the treatment of classroom behavior problems. The smaller study (n=13 articles) which used the percent of non-overlapping data (PND) as a CM, was conducted by Scruggs, Mastropieri, Cook, and Escobar (1986) and concerned early intervention for children with conduct disorders. The CM study included a total of 195 graphed experiments. A simple comparison of the baseline and treatment phases from each graph was used; and a third possible CM was developed from an adaptation of Glass' formula (AGF) for group studies. The computerized method study used a scanner and a "mouse" device with a microcomputer. Data points were estimated by finding the screen coordinates of each point and using the scale on the ordinate. The PR model calculated an ES based on the combined effect of a change in level and slope between the two phases; the PND model grossly measured the differences in level between the two phases; and the AGF ES measured a standardized difference between the means of the phases. In meta-analyses for SSR, when different ESs (based on different sets of assumptions) are calculated yielding similar results, the validity of the conclusions increases. Three graphs are included. (RLC)

META-ANALYSIS OF SINGLE SUBJECT RESEARCH IN SPECIAL EDUCATION: A COMMON METRIC & A COMPUTERIZED METHOD

Sharene Smoot, Georgia State University
William Curlette, Georgia State University
Samuel Deitz, Georgia State University

Sharene Smoot, Georgia State University
William Curlette, Georgia State University
Samuel Deitz, Georgia State University

Dr. Sharene Smoot
4242 Panola Rd
Lithonia, GA
30038

META-ANALYSIS OF SINGLE SUBJECT RESEARCH IN SPECIAL EDUCATION: A COMMON METRIC AND A COMPUTERIZED METHOD

Introduction

Many valuable single subject research (SSR) studies in special education have been excluded from meta-analyses because of the format incompatibility with group research studies. One solution is to do a meta-analysis of exclusively SSR studies, but the major drawback to this approach has been the historical practice of presenting the results in a graphed form for visual analysis. This creates two problems for the meta-analyst. The first is how to mathematically measure the effect size as is done in meta-analysis of group studies, and the second is how to recapture or estimate the data from a published graph about two by four inches in size in a journal article. The two objectives of this study were to find a suitable effect size measure or common metric for comparing the results of a set of SSR studies and to find an easy way to convert the published graphs back into raw data from which effect sizes could be calculated. To meet the first objective, three possible formulas for measuring treatment effect were evaluated and then compared using identical data sets. To meet the second objective, a computer method was developed.

Three previous attempts at meta-analysis of SSR studies in education were found in the literature. These were done by two sets of authors who used two different formulas for calculating a common metric, each justifying their choice. This generated much controversy regarding statistical assumptions which might affect the validity of the conclusions of these meta-analyses (Dunst & Snyder, 1986). In these studies the raw data was estimated manually by using drafting equipment. (Fortunately the practice today is for researchers to use a computer program to draw the graph from the raw data and to include statistical data such as phase means and standard deviations.) However a meta-analysis of a particular educational topic using SSR methodology might search back decades of published research and include many graphs plotted by hand with no raw data or statistical summaries.

Data Sources

Two of the three published meta-analyses furnished the data used for this study. Both were published in 1986. The larger study (N = 23 articles), which used piecewise regression (PR) to calculate an effect size, was the Skiba, Casey, and Center (1985-86) meta-analysis on the topic of using nonaversive procedures in the treatment of classroom behavior problems. The smaller study (N = 13 articles) used the percent of nonoverlapping data (PND) as a common metric. It was done by Scruggs, Mastropieri, Cook, and Escobar on the topic of early intervention for children with conduct disorders. To develop the computerized data recovery method, it was necessary to find graphs for which the actual data was available. Three doctoral dissertations from the Special Education Department at Georgia State University were found which met this requirement, and these were used to establish the validity and reliabiity of the microcomputer technique for estimating the data from published graphs. They were written by Drs. A. Troutman

(1978), M. Powell (1982), and T. Higgens (1982).

The Common Metric Study

Methodology

The substantive question of the choice of an appropriate quantitative method to summarize research in educational and clinical fields using SSR methodology was addressed by using both formulas previously used in the two meta-analyses on both sets of graphs. There were a total of 195 graphed experiments. A simple comparison of the baseline and treatment phases from each graph was used. In addition, a third possible common metric was developed from an adaptation of Glass's formula for group studies. Thus effect sizes were calculated in three different ways from three different theoretical bases on the larger set of graphs. These three sets of effect sizes were then correlated pairwise to see how they compared with each other. Then each set of effect sizes was grouped by a study characteristic (for example, the type of reinforcer that was used), the mean was calculated for each group, and the groups were ranked from most to least effective as a researcher would do to ascertain the results of his or her meta-analysis. Since the median is also a measure of central tendency, the groups were also ranked by this statistic to see if the results would be different. This process was repeated for a second study characteristic, the agent factor. Again the mean and median were calculated from the groupings. Finally all the calculations and processes were repeated on the smaller set of graphs to determine replicability of the results.

Formulas

Piecewise regression (PR) was the most time consuming formula used due to computational complexity. It is a regression discontinuity model that includes the change in frequency of the behavior as well as the rate of that change by adding a time factor (Center, Skiba, & Casey, 1985-86). It takes into account the fact that baseline data may not be stable, and there may be an underlying trend. This is a parametric statistic based on ANOVA via regression with the phases as a dummy coded variable. By adding the variable of time, this multiple regression model can yield three effect sizes based on level, trend, and the combined effects of level and trend. The full model is described as follows:

$$y = b0 + b1x + b2t + b3x(t - n) + e$$

where... y represents the value of the data point,
t represents the successive days,
n is the number of days in the first phase,
x represents a dummy coded variable (0 = baseline phase, 1 = treatment phase),
and..... e represents residual error.

Thus.... b0 represents the intercept,
b1 represents the change in level from the last day of baseline to the first day of treatment,
b2 represents the slope of the baseline phase,
and..... b3 is the the change in slope from the baseline phase to treatment phase.

Effect sizes computed were those for the combined effect of slope and level (full model). Rosenthal's (1978) formula which has been shown to be equivalent (Neter & Wasserman, 1974) to the effect size recommended by Glass (1978) follows:

$$ES = \frac{2\,t}{\sqrt{df}} = 2 \cdot \sqrt{\frac{MS\ (effect)}{MS\ (error)\ d.f.\ (error)}}$$

The percent of nonoverlapping data (PND) is an ordinal nonparametric statistic and is compatible with the precedent of looking for large effects and not using standard statistical techniques. It is a ratio based on the relationships between the graphed data points in the baseline and treatment phases (Tawney & Gast, 1984). The number of data points in the treatment phase that did not overlap with any of the points in the baseline phase were divided by the total number of data points in the treatment phase.

The adapted Glass formula (AGF) used the data in the two phases like the data in two groups. The data in the baseline phase is considered the control group data, likewise the treatment phase is analagous to the treatment group data. The mean of the baseline phase data was subtracted from the mean of the treatment phase data and the resulting effect size is standardized by dividing it by the standard deviation of the baseline phase.

Results and Conclusions

The highest pairwise correlation between the three sets of effect sizes was $r = .83$, $p < .001$ between the sizes generated by the adapted Glass formula and the piecewise regression formula in the larger data set ($N = 23$). In the other data set ($N = 13$), $r = .75$, $p = .001$ between the adapted Glass formula and the percent of nonoverlapping data measure. The highest correlation of the rankings of two of the study variables was $rho = .86$, $p < .025$ between the adapted Glass formula and the piecewise regression formula for the study characteristic of reinforcer, using the mean as a measure of central tendency. Since the median is a better measure of the typical score if the distribution is skewed, the distributions of each of the three sets of effect sizes were plotted. The distributions of the effect sizes were found to be distinctively different, but consistent, in the two data sets.

Insert Figures 1, 2, and 3 about here

Among the three formulas studied as possible common metrics for meta-analysis in SSR, the best one may be the adapted Glass effect size for the following reasons: (a) high correlation with both the piecewise regressions and the percent of nonoverlapping data measure, (b) the adapted Glass formula ranked study characteristics very similarly to the piecewise regression formula which includes behavior change over time as a factor in the equation, (c) the distribution of the effect sizes generated by the adapted Glass effect size were nearly normally distributed, (d) there is computational simplicity--it can be done on a hand held statistical calculator, and (e) similarity of the formula to the effect size measures now used with group designs. However interpretation of these effect sizes needs to consider that there is a diffference in scale between SSR (within-subject) and group studies (between subjects).

The Computerized Method Study

Methodology

The technical methodology for this study consisted of using a scanner and a "mouse" device with a microcomputer. Thus the data points were estimated by finding the screen coordinates of each point and using the scale on the ordinate. The equipment consisted of an IBM or compatible microcomputer, a Hewlett Packard Scan Jet scanner, and a Microsoft mouse. The software used was Publisher's Paintbrush (Zackman, White, & Albertine, 1987) and a BASIC program written to estimate the values of the data points and to do the simpler calculations of the effect sizes. A mainframe computer was used to do the multiple regression needed for the PR formula.

The procedure used the scanner to put photocopied graphs from published studies on a computer screen. (Care needs to be taken in photocopying by centering the graph on the photocopier.) Then the data was estimated by using the screen coordinates. Rules were developed to handle the variety of shapes and sizes of the "dots" which became a problem in the transition from photocopied graph to enlargement on the screen of the computer monitor.

Reliability and validity studies were also conducted on the computer method. The validity of this method was established by comparing 80 data points from five randomly selected graphs with the corresponding raw data from the three doctoral dissertations. The graphs were given a preliminary visual analysis to ascertain that the data had been graphed correctly, any with obvious errors were discarded from the sample. The graphs were then photocopied, scanned, saved to a file, retrieved, and the data estimated using the mouse pointer with the paintbrush software. Trial and error was used to develop the set of rules resulting in the most accurate estimation.

The reliability of the method was calculated by having three persons use the rules to estimate the data from 103 data points in five baseline and treatment phases randomly selected from the 195 files. Hoyt's method was used to calculate the reliability using the estimated value of each data point as the unit of analysis.

Results and Conclusions

The validity and reliability coefficients of the microcomputer methodology were both .999. The procedures are fairly simple and within the skills of most educational researchers. In addition, the cost of the system is reasonable. Probably most persons in educational research already have access to IBM compatible computers and photocopiers. The costs of the mouse (about $200) and software (about $400) are reasonable, only the scanner is a high priced item at about $2,000. In summary, the reliability and validity of the scanner and microcomputer approach is extremely high yielding a cost effective incentive to conduct meta-analyses in SSR.

Summary

In the perspective of the interdependence of research and practice, experimental SSR has much to offer. A practitioner can readily conduct an experimental study with one subject in just a few months (such as a school quarter) in a classroom or clinical situation. Such field research, if carefully planned, can be quite valuable without requiring a large time commitment or groups of subjects. The results of these field studies can then be combined via meta-analyses. This could be especially helpful in the study of low incidence disabilities which often present serious challenges to professionals in special education and rehabilitation.

Additionally, this study suggests that more than one way can be used to quantitatively rank treatment effectiveness in SSR. To review, the PR model calculates an effect size based on the combined effect of a change in level and slope between the two phases; the AGF effect size measures a standardized difference between the means of the phases; and the PND model grossly measures the difference in level between the two phases. The advantages and disadvantages of each effect size metric or a combination of the metrics can be considered in view of the various assumptions which previously caused considerable controversy. In meta-analysis for SSR, when different effect sizes (based on different sets of assumptions) are calculated yielding similar results, the validity of the conclusions increases. The findings in this study should reduce objections to the quantitative techniques of meta-analysis in comparing SSR studies.

References

Center, B.. A., Skiba, R. J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. Journal of Special Education, 19, 483-492.

Dunst, C. J., & Snyder, S. W. (1986). A critique of the Utah State University early intervention meta-analysis research. Exceptional Children, 53, 269-276.

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. Review of Research in Education, 5, 351-379.

Higgens, T. S. Jr. (1982). A comparison of two methods of practice on the spelling performance of learning disabled children

(Doctoral dissertation, Georgia State University, 1982). Dissertation Abstracts International, 42, DE082-26143.

Neter, J., & Wasserman, W. (1974). Applied linear statistical models. Homewood, IL: Richard D. Irwin.

Powell, M. D. (1982). Using motoric cues to occasion compliance responses in emotionally handicapped boys. (Doctoral dissertation, Georgia State University, 1982). Dissertation Abstracts International, 42, DE082-26148.

Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. Behavioral Disorders, 11, 260-271.

Skiba, R., Casey, A., & Center, B. (1985-86). Nonaversive procedures in the classsroom behavior problems. The Journal of Special Education , 19, 459-481.

Troutman, A. C. (1978). Stimulus control: A procedure to facilitate generalization from resource room to regular classrooms. (Doctoral dissertation, Gerogia State University, 1977). Dissertation Abstracts International, 38, DDK77-23238.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Tawney, J. & Gast, D. (1984). Single subject research in special education. Columbus, OH: Charles E. Merrill.

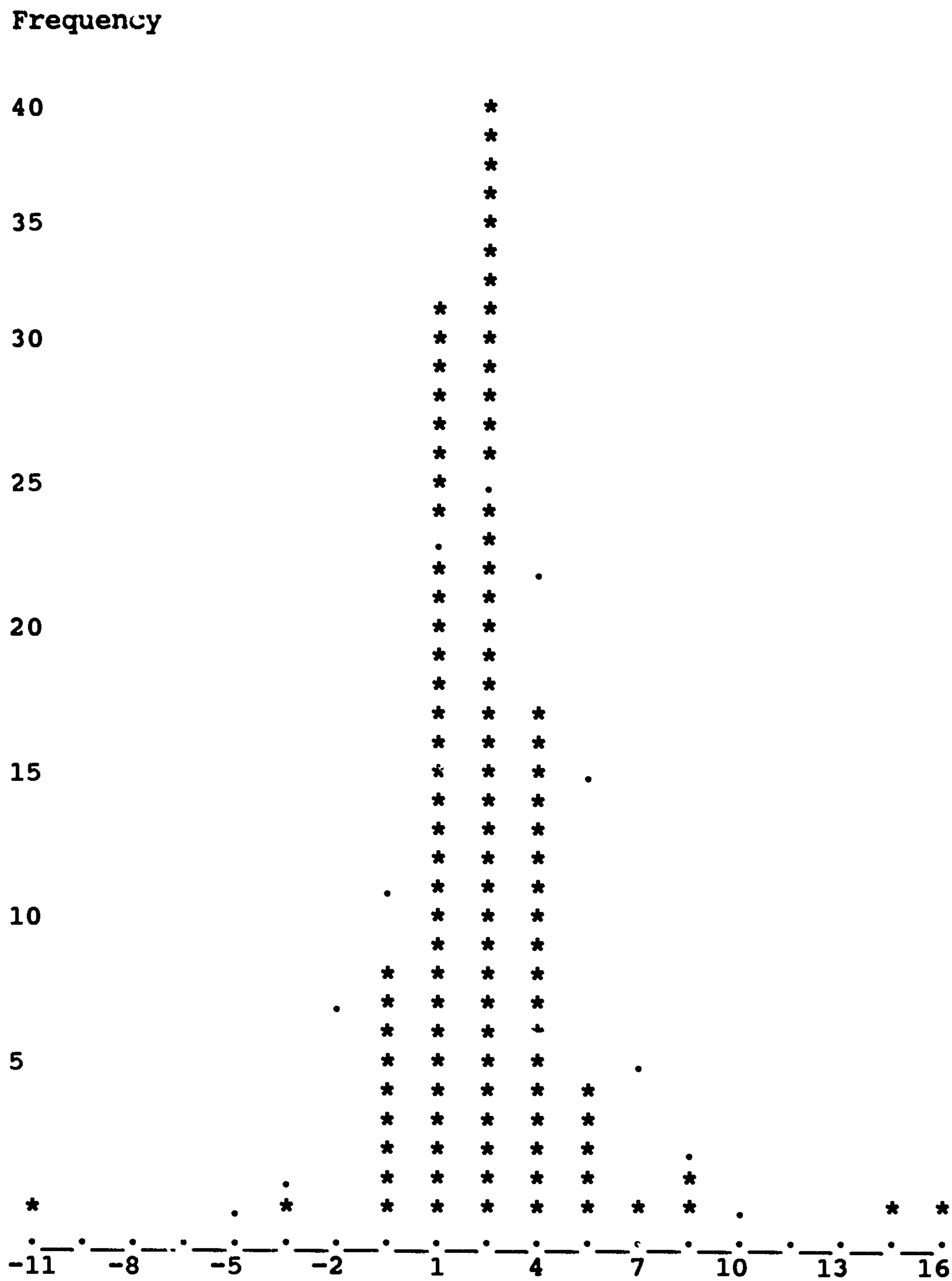

Figure 1. Distribution of AGF effect sizes. One * equals one occurance. The dots superimpose the standard normal curve.

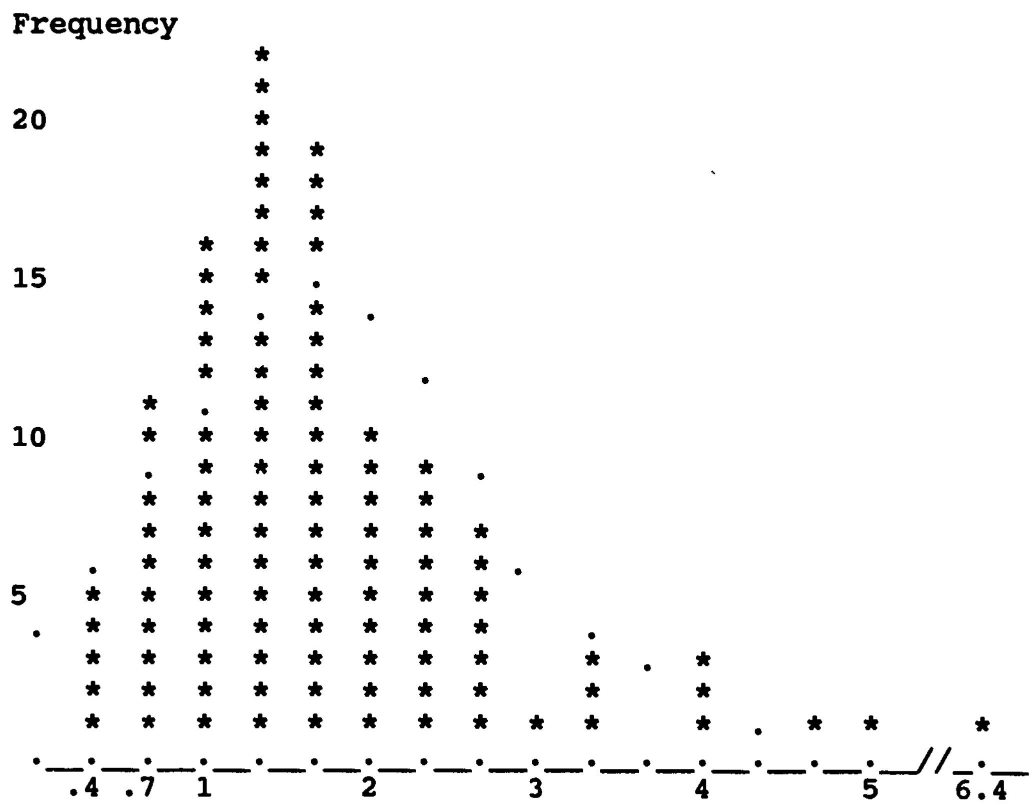

Figure 2. Distribution of PW effect sizes. One * equals one occurance. The dots superimpose the standard normal curve.

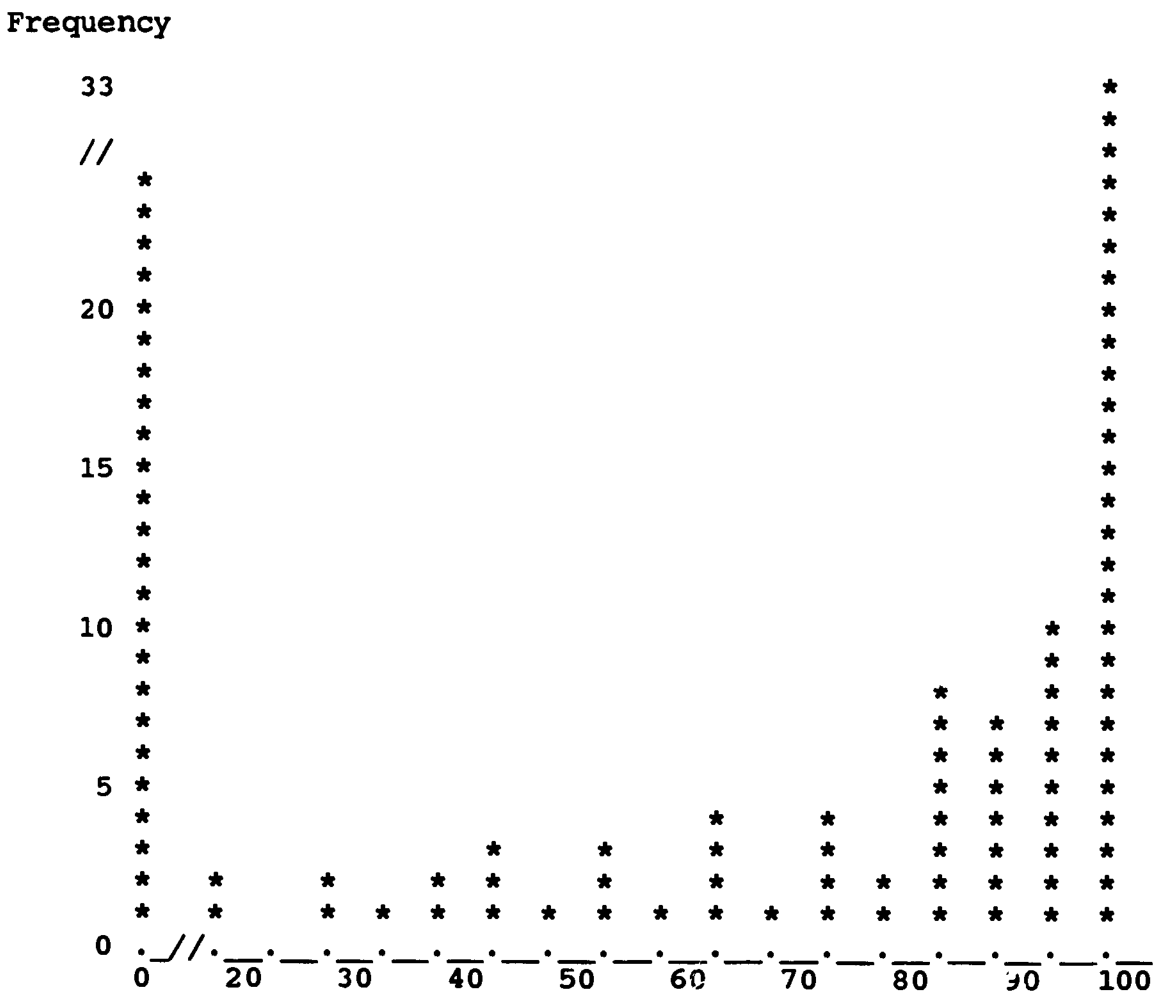

Figure 3. Distribution of PND values. One * equals one occurance. The tall columns at the values of 0 and 100 are due to floor and ceiling effects.