The Behavior Analyst 1982, 5, 95-99 No. 1 (Spring)

Book Review

Is This the Right Road? A Review of Kratochwill's *Single Subject Research: Strategies for Evaluating Change*

Stephen A. Graf
Youngstown State University

Has the day arrived when inferential statistics are needed to guide decisions in single subject research? Or is visual inspection of data sufficient as a guide for appropriate interpretative responding? These are the issues raised in *Single Subject Research: Strategies for Evaluating Change,* editing by Thomas Kratochwill. We are confronted with a fork in the road. Which direction should we go, or should we split up and try both? As we ponder our choice, we might well consider another possibility. Might we be on the wrong road? If we are, our best strategy might well be to turn around, rather than choosing either fork or both.

OVERVIEW

Single Subject Research is a book that deals with research design, visual analysis, and statistical analysis in situations where subject's behaviors are measured repeatedly across time.

Both visual and statistical analyses have been used to evaluate change in time-dependent data, and a controversy exists between users of the two types of procedures. *Single Subject Research* addresses this controvesy from several viewpoints. Advocates of statistical analysis often suggest such analyses in addition to a visual analysis to facilitate the discovery of small but significant effects which a visual analysis probably would not uncover (e.g., Elashoff & Thorensen, p. 291). Advocates of visual analysis generally contend that statistical analyses are more harmful than helpful (e.g., Michael, 1974) because only large, clearly visible effects should be considered when one is concerned with building a science of behavior.

A second theme of the book addresses the more specific questions that arise from its general focus. If statistical analysis is appropriate, what type of statistic is most appropriate when? If visual analysis is appropriate, what type of graph is most appropriate when? In the foreword, Alan Kazdin commends the volume because no previous single source before 1978 had covered the examination and statistical evaluation of single-subject data so comprehensively. Kratochwill comments in the preface that this is a strategy book. "The essential tone . . . is that of encouraging selection of an appropriate research design and data analysis procedure from one of many available" (p. xvii).

The book progresses from an opening chapter on basic research design authored by Kratochwill to four data analysis procedures presented by other authors. It concludes with an overview on the selection of an appropriate analysis procedure. The candidates, in order of presentation, are visual analysis, randomization tests, time-series analysis, and Markov chain analysis.

After tracing the historical development of time-series analysis, Kratochwill offers a notational system to represent the various designs and then describes a host of such designs. He covers the literature of the visual analysis tradition and highlights three considerations that make visual analysis of data a complex endeavor: trend, variability, and autocorrelation. He concludes the chapter by tracing the emergence of statistical tests in single subject designs and sets the stage for the resulting controversy.

Address reprint requests to Stephen Graf, Department of Psychology, Youngstown State University, Youngstown, Ohio 44555.

Barry Parsonson and Donald Baer cover the analysis and presentation of graphic data. They begin by discussing the differences between group and single-subject designs. In their exploration of graphics as a form of data analysis, they underscore that statistical analysis is post hoc and thus not useful during the treatment or investigation, while visual analysis will show what has happened session by session, if the investigator plots the data properly. Their opinion that graphics are relatively insensitive is held to be an advantage, since only powerful variables are likely to be scrutinized. They suggest esoteric variables discoverable only with sophisticated measurement or statistical analysis are not likely to have much useful application. As an additional feature they mention the ease of independent replication when data are graphically analyzed. They conclude that there exists a strong rationale for using graphic analysis. They also cover in detail how visual analysis of graphic data is done, including forms of graphic representation and the construction and presentation of graphs.

Joel Levin, Leonard Marascuilo, and Lawrence Hubert describe nonparametric randomization tests. A randomization test ascertains the likelihood that the differences between treatment groups could have, by chance, been at least as large as the outcome observed. The crucial question to be asked in order to determine the appropriateness of a randomization test involves whether there is random assignment of treatments to phases. This requirement prohibits post hoc data analysis of designs which were not set up under these guidelines.

While other possibilities also exist for randomization analyses, the authors show examples using a "basic" randomization test. Levin, et al., themselves pinpoint some problems. For example, the computations can sometimes be overwhelming and may require aids such as transformation to ranks, computer approximations, or normal distribution approximations. They also explain that a summary statistic is needed from every phase. They contrast their approach with previously used data analysis procedures for small-sample behavior change experiments, specifically an analysis of variance approach.

John Gottman and Gene Glass begin their chapter by quoting an anonymous reviewer commenting on an article they had submitted. The reviewer typifies the position of those who have said that applied behavior analysis does not need statistics to salvage small effects because the field should stick with large effects. Gottman and Glass use an example to illustrate the insufficiency of eyeballing graphs. They consider the problem of autocorrelation and go on to mention analysis of variance and regression analysis alternatives. They outline interrupted time-series analysis, emphasizing aspects of the analysis they believe have been overlooked. The analysis proceeds by attempting to identify a model, which is represented as a formula. While 50 or more observations are required to identify a model with any confidence, the intervention effects can be tested with less observations. Once a model is identified, predictions can be made, and intervention effects can be tested for significance.

John Gottman and Cliff Notarius address themselves to the sequential analysis of behavior over time. They say that when one focuses on response rates, the dimension of time and the importance of sequence is ignored. Attending to sequence, they claim, allows one to use statistics in the study of interacting organisms one day at a time. They present Markoff chain analysis with examples and follow with matrix algebra and examples. They use the language of information theory to describe dyadic interaction patterns.

Janet Elashoff and Carl Thoresen discuss the choosing of an appropriate analysis procedure, although they do not include a visual analysis of the study used as an example for the statistical tests. They suggest asking questions concerning the nature of the response measure and the nature of the experiment to cue the choice of an appropriate measure. They also make a distinction between exploratory and confirmatory analysis. In exploration, one should attend to the data

in detail and not test for statistical significance. Confirmatory analysis involves statistical testing for significant effects and represents a follow-up experiment. They conclude that the time has arrived when appropriate inferential statistics are needed in single-subject research.

COMMENTS

The most important issue raised by the publication of this volume is whether the suggestions made by the various authors of *Single Subject Research* are conducive to the continued development of a science of behavior. My own conclusion is that in general they are not. I have based my conclusions on consideration of the question: How should we be operating if we are attempting to generate a parsimonious, objective, and empirical science? From this context the areas of concern include design strategies, definitions of behavior, measurements of behavior, data displays, and interpretative behaviors, in that order.[1]

Design Strategies.

The focus of my concern here involves the design strategies that guide the many particular design decisions that follow. An empirical approach to design implies a dynamic rather than static perspective, meaning that flexibility is required in scientific activity to take advantage of what one finds. Except for Parsonson and Baer (Chapter 2) and Elashoff and Thoresen (Chapter 6), the authors of this volume tend toward a static approach. Kratochwill's careful description of some 17 specific designs along with his recommendations on the selection of a design promote a static, cookbook approach, and the requirements of the statistical models in Chapters 3 through 5 further advance this viewpoint. The a priori requirements of the statistical design must be met in order to apply the particular significant test, and the investigator is thereby obligated not only to that design but to a variety of other associated practices. For example, between-subject statistical designs encourage measurement decisions such as collapsing data across individuals. The resulting variability introduced is used as an error term in the statistical inference. Within-subject statistical designs similarly encourage large numbers of observations within each phase and the collapsing of data within phases.

Definition of Behavior

A natural science must adhere to a definition of behavior which is consistent with nature rather than merely convenient. Inspection of the dependent variables found throughout the volume leaves one unclear as to how behavior is being defined. Kratochwill states that both the conceptualization and operational definition of dependent variables are important (p. 28) but does not specify other characteristics of such variables. Avoiding reifications of behavior as well as other abuses of the subject matter requires careful definition of what constitutes behavior. One does not find such definitions within *Single Subject Research* and some of the subsequent problems with measurements of behavior, data displays of behavior, and interpretations of behavior are probably the result of this deficiency.

Measurement of Behavior

Johnston and Pennypacker (1980) suggest an approach to the measurement of behavior which uses absolute and standard units of measurement to quantify the amount of selected dimensional quantities. Such "idemnotic measurement" requires units of measurement (e.g., the cycle and units of time) that are established before and apart from the actual measurement operation. The alternative strategy, "vaganotic measurement," creates scales and units of measurement based on the same data that are being assessed. For example, standard deviation is created by a statistical formula for a particular set of data but is then used to measure the statistical significance of that data. Vaganotic measurement is embodied in the inferential hypothesis testing approach in a number of ways and constitutes an integral part of time series

[1] Elaborations of each of these topics may be found in Johnston and Pennypacker (1980).

analyses and randomization tests found in Chapters 3 and 4.

Data Displays

A further complication of inferential statistics is that they tend to discourage plotting data graphically. Data displays ideally are done while an experiment is in progress, as well as afterwards. Such online displays characterize a dynamic approach in which the features of the data can influence the investigator's design decisions.

Of course the most parsimonious data display would be one with a standard ordinate and a standard abscissa. In their chapter on the analysis and presentation of graphic data Parsonson and Baer mention such a display (p. 153), the "Standard Celeration Chart," but do not include a representation of it. The requirements for an appropriate data display set forth in Chapter 2 are often not met in displays found in the other chapters.

Interpretative Behaviors

Gottman and Glass (p. 199) discuss the insufficiency of eyeballing graphs compared to statistical analysis. The problem they see may simply be that inappropriate data displays of highly variable data produce inappropriate interpretative responses. The example used by Gottman and Glass involves what Parsonson and Baer call a "change in level between phases" (p. 126). In other words, what happened to the behavior when an intervention occurred? Did the frequency of the behavior jump up, jump down, or stay the same? In Gottman and Glass's example graduate students in a seminar on time-series analysis inspected graphs showing very unstable patterns of responding and made a visual judgment as to whether the frequency jumps for each of two individuals were "significant."

The data were displayed on graphs using a 0-100 equal/interval scale, and only 7 of the 13 students judged the shift of Person A to be "statistically reliable." For Person B, 11 of 13 students judged the jump to be a significant shift. Statistical analysis showed the shift to be significant for Person A and not significant for Person B, and Gottman and Glass argue that the eyeballs were fooled in 18 out of the 26 judgments.

I replotted the data onto a Standard Celeration Chart developed and used by Ogden Lindsley and his associates (Sepler, 1981) and saw a much different and clearer picture. Inclusion of both displays would have demonstrated this point, but would still not have done justice to the viewer, who would have been attempting a visual analysis without access to information the experimenter had on how the data was collected, etc.

Displaying data involves transforming factual output into visual stimuli which should influence interpretative behavior, and insofar as possible the display should compel an interpretative response that is reliable and general. However, pitting visual against statistical analysis distracts attention from other, prerequisite issues. These considerations include the design strategy, the definition of behavior, and the measurement of behavior, as well as the data displays and the interpretative behaviors.

As I pointed out earlier, the central theme of *Single Subject Research* revolves around whether statistical analysis facilitates more appropriate interpretation than visual analysis alone. *What I have tried to suggest is that within the context of trying to develop a natural science of behavior, this theme is really only one of a number of critical themes. Each of the other issues I have touched upon are, in fact, more important than whether numbers versus lines and dots exert more veridical control over the interpreters responses.*

We must be aware of the full range of consequences of inferential statistics in the study of behavior because their effects permeate all aspects of research method. My brief criticisms are not directed at the statistical procedures themselves. They are perfectly legitimate for their proper uses. Discovering the facts of behavior just does not happen to be one of those uses because statistical procedures repeatedly conflict with the fundamental nature of behavior. Even when the data base comes from observations of in-

dividual behavior, it does not insure that the question being asked is really behavioral. For example, asking which of two treatment procedures works better is not a purely behavioral question, and inferential statistics may be quite appropriate assuming that the question is worth answering. In other words, the distinction is between actuarial questions at the population level and analytical behavioral questions at the individual level. Statistics are more appropriate for the former than for the latter.

In the analogy with which I began, when we come to a fork in the road but we are not even on the right road, it would be better to turn around and go back rather than to take either fork. *Single Subject Research* is not a volume that has mapped the future course for a science of behavior that I think will be the most productive.

REFERENCES

Johnston, J. M., & Pennypacker, H. S. *Strategies and tactics of human behavioral research.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Kratochwill, T. R. (Ed.) *Single subject research: Strategies for evaluating change.* New York: Academic Press, 1978.

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

Sepler, H. J. Measuring the effects of no-fault divorce laws across fifty states: Quantifying the zeitgeist. *Family Law Quarterly,* 1981, *15,* 65-102.