EFFECTS OF LINES OF PROGRESS AND SEMILOGARITHMIC CHARTS ON RATINGS OF CHARTED DATA

DONALD B. BAILEY, JR.

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

The extent to which interrater agreement and ratings of significance on both changes in level and trend are affected by lines of progress and semilogarithmic charts was investigated. Thirteen graduate students rated four sets of charts, each set containing 19 phase changes. Set I data were plotted on equal interval charts. In Set II a line of progress was drawn through each phase on each chart. In Set III data points were replotted on semilogarithmic charts. In Set IV a line of progress was drawn through each phase of each Set III chart. A significant main effect on interrater agreement was found for lines of progress as well as a significant 2-way interaction between lines of progress and change type. Three main effects (chart type, lines of progress, and type of change) and a significant 3-way interaction were found for ratings of significance. Implications of these data for visual analysis of charted data are discussed.

DESCRIPTORS: applied behavior analysis, data analysis, experimental effects, judgmental aids, visual inference

Folklore tells us that "seeing is believing." That belief has also been a basic tenet of applied behavior analysis. Charted data are generally accepted as sufficient evidence to determine whether experimental control has been demonstrated. Recent publications, however, have explored factors that affect the use of visual analysis to determine experimental effects. DeProspero and Cohen (1979) and Jones, Weinrott, and Vaught (1978) demonstrated that raters of charted data often differ in their opinion as to whether or not experimental control has been demonstrated. Wampold and Furlong (1981) found that a group of graduate students who had completed a single-subject research seminar failed to differentiate charts containing changes in both level and trend from those containing a change in trend only. Studies by Furlong and Wampold (1981, 1982) suggest that raters typically attend to large phase changes but do not attend to relative variation in data.

Problems in interrater agreement such as those

described have led many to suggest that visual analysis should be supplemented by statistical techniques to verify the effectiveness of interventions (e.g., Edgington, 1967; Gorsuch, 1983; Jones, Vaught, & Weinrott, 1977; Kazdin, 1976; Wolery & Billingsley, 1982). Although statistical analysis may be important in some cases, it is possible that the use of judgmental aids or alternate charting techniques could serve to improve ratings based on visual analysis. Unfortunately, little research has addressed this issue. One exception is a study by Knapp (1983) which was an investigation of the relative effects of three graphing techniques (cumulative plot, semilog, and frequency-polygon) and three styles of presenting data from different intervention phases (baseline data separated from intervention data either by a space, a vertical line, or not separated). Data plotted on semilogarithmic and frequency-polygon charts were rated significantly different from each other, with frequency-polygon charts producing the greatest amount of interrater agreement on "no change," whereas data plotted on semilog charts produced the least consensus on "no change."

The purpose of this study was to investigate the effects of two judgmental aids—lines of progress and semilogarithmic charts—on raters' judgments regarding charted data.

I express my appreciation to William B. Ware for his assistance in the statistical analyses and to Owen R. White for his insightful reviews and suggestions. Requests for reprints should be sent to Don Bailey, Frank Porter Graham Child Development Center, Highway 54 Bypass West 071A, University of North Carolina, Chapel Hill, North Carolina 27514.

Lines of Progress

Traditional analytic strategies summarize the central tendency of data sets using a mean or median. Such an analysis, however, may obscure any trends in behavioral performance. Because time-series data may be increasing or decreasing, it is important to include an estimate of trend in the analytic process.

Charted data may be summarized visually by a line describing the level and direction of growth. For example, a regression equation may be used to estimate the "least-squares" line for a set of data. The advantage of such a trend line, or line of progress, is that it provides a parsimonious summary of a phase against which performance in other phases may be compared. White (1974) has suggested a simple method, the "quickie split-middle technique," for estimating a line of progress. The procedure involves determining the intersections of the median level of performance and the median day (or session) for each half of a phase. Those two points are connected, resulting in a trend line, and the line is then shifted up or down (keeping the slope or angle of the line constant) until half the data points fall on or above the line and half fall on or below the line. The split-middle technique and examples of its application are described in detail by Kazdin (1976, 1982) and White and Haring (1980).

Lines of progress are rarely provided in published applied behavior analysis research. When a line is used to summarize a phase it typically summarizes only the central tendency of a phase. A line of progress may more accurately reflect changes in performance over time and provide a basis for predicting the future course of progress. The central tendency line can be misleading, thus affecting the use of visual inference to ascertain the nature and extent of changes in level, trend, and variability.

Semilogarithmic Charts

On an equal interval chart the distance between any two consecutive points on each axis is always the same. For example, on the abscissa, the distance between Day 1 and Day 2 is equal to the distance between Day 13 and Day 14; on the

ordinate, the distance between rates of 5 and 10 per minute is equal to the distance between rates of 20 and 25 per minute. Equal *absolute* changes are always represented by equal distances on the chart.

On a semilogarithmic chart, the abscissa (time or sessions) is also represented by equal intervals. However, the ordinate (behavioral performance) is drawn in logarithmic units. Equal *relative* changes in performance are represented by equal distances on the chart. Any proportionate change (e.g., doubling, halving) is represented by the same distance on the chart, regardless of the absolute values of those changes.

The representation of behavior in logarithmic units may more accurately reflect the effort involved in learning (and teaching). Although going from 5% correct to 10% correct is the same absolute increase as going from 80% correct to 85% correct, the first increase is proportionately much greater than the second, and thus should require more effort. Semilog charts may also be advantageous when making predictions because data that would usually appear as a normal learning curve on an equal interval chart (gradual rise at the beginning with a steeper increase over time) more closely approximate a straight line on the semilog chart. Semilogarithmic charts have proved to be useful in planning and evaluating instructional programs (Lindsley, 1964; White & Haring, 1980), in the application of decision-making rules for instructional changes (Haring, Liberty, & White, 1980), and in making predictions about the future course of instructional progress (White, Billingsley, & Munson, 1980).

Assuming that lines of progress and semilog charts could affect visual analysis, the individual and combined effects of the two procedures on interrater agreement and on ratings of significance were investigated in this study.

METHOD

Students

Thirteen female special education graduate students participated in the study. Each had just completed a course in applied behavior analysis, half

of which was devoted to an understanding of data collection procedures, charting procedures, and applied behavior analysis research designs. All were master's degree students in special education. None had conducted applied behavior analysis research, but all had experience in reading and interpreting charted data. Although the use of graduate students may be questioned due to lack of experience, Knapp (1983) found no differences in ratings obtained from editors or reviewers of applied behavior analysis journals, graduate students, and undergraduate students with at least one course in behavior analysis.

Materials

The charts used were selected from those used in the Jones et al. (1977) investigation of time-series analysis in operant research. Five of the seven charts presented in that study were used because they were difficult to analyze visually and there was a discrepancy between the visual interpretation and statistical interpretation. The five charts selected were: (1) Figure 3 in Boren and Colman (1970); (2) Figure 1 in Ingham and Andrews (1973); (3) Figure 5 (Hannah) in Baer, Rowbury, and Baer (1973); (4) Figure 5 (Charlotte) in Baer et al. (1973); and (5) Figure 7 in Wincze, Leitenberg, and Agras (1972).

Four sets of charts were prepared. Set I was comprised of each of the five charts as published by the authors. Set II was comprised of the same charts with a line of progress drawn through each experimental phase by using the quickie split-middle technique (White, 1974). In Set III, all data points were plotted on semilogarithmic percent charts. Set IV was comprised of the same charts as Set III with a line of progress drawn through each phase.

Procedure

Each student was given a one-page introduction to the study. This introduction contained a brief review of the concepts of level and trend. Figure 1 from Jones et al. (1977) was attached to illustrate combinations of level and trend changes.

Following the introductory page, each student received a packet that included each of the 20

charts (5 data sets $\times$ 4 charting procedures). The charts were randomly ordered. Each student then was asked to look at each phase change and determine if the change resulted in a significant change in level or trend of the data, or both. Students were instructed that in determining significance they should ask whether the changes in behavioral scores following interruptions (or phase changes) warranted the conclusion that experimental control over behavior had been obtained. Each student wrote *yes* or *no* to the questions "Was there a change in level?" and "Was there a change in slope?" for each phase change in each chart.

RESULTS

Across the five charts, 19 phase changes occurred. Thus, within each set of charts, each respondent made 19 decisions about changes in level and 19 decisions about changes in trend. The effects of visual aid and type of chart on ratings of changes in level and trend were analyzed in two ways: effects on interrater agreement (a measure of cross-rater consistency) and effects on ratings of significance (a measure of whether, in fact, a change actually occurred).

Interrater Agreement

Interrater agreement for each phase change on each chart was determined according to the following formula: [$\#$ of raters stating a change did or did not occur (whichever was larger)/total $\#$ of raters $\times$ 100]. A 3-factor analysis of variance was conducted as a part of a randomized block factorial design to determine the effects of chart type (equal interval or semilogarithmic), lines of progress (present or absent), and change type (level or trend) on interrater agreement. Because the unit of analysis was percentage of agreement rather than individual data, the design was a repeated measures model with different phase changes rated by each observer used as the blocking variable. The results are displayed in Table 1. A significant main effect was found for lines of progress, and a significant 2-way interaction was found between lines of progress and change type. The significant effect for phase changes (blocking variable) indicates that

Table 1

3-Factor Analysis of Variance for Interrater Agreement

Source	df	Mean square	F
Main effects			
Chart type	1	407.9	2.53
Line of progress	1	6697.9	41.48**
Change type	1	96.3	.59
Two-way interactions			
Chart type $\times$ line of progress	1	59.38	.37
Chart type $\times$ change type	1	192.38	1.19
Line of progress $\times$ change type	1	847.9	5.25*
Three-way interaction			
Chart type $\times$ line of progress $\times$ change type	1	142.2	.88
Phase changes	18	431.39	2.67**
Residual	126	161.48	

* $p < .05$.** $p < .01$.

different phase changes generated different levels of interrater agreement.

The data are displayed graphically in Figure 1. Equal interval charts with no lines of progress resulted in interrater agreement of 73% and 66% on level and trend changes, respectively. Interrater agreement on level and trend increased to 85% and 84%, respectively, when lines of progress were added to the equal interval charts. Semilog charts with no lines of progress resulted in interrater agreement of 77% on changes in level and 71% on changes in trend. When lines of progress were added to the semilog charts, interrater agreement increased to 83% for changes in level and 90% for changes in trend.

Ratings of Significance

A 3-factor analysis of variance was also conducted as a part of a randomized block factorial design to determine the effects of chart type, lines of progress, and change type on ratings of significance of effects after phase changes. Ratings of significance was defined for each phase change as

Table 2

3-Factor Analysis of Variance for Ratings of Significance

Source	df	Mean square	F
Main effects			
Chart type	1	6672	42.31**
Line of progress	1	1836.2	11.65**
Change type	1	6963.5	44.16**
Two-way interactions			
Chart type $\times$ line of progress	1	3290.6	20.86**
Chart type $\times$ change type	1	918.1	5.82*
Line of progress $\times$ change type	1	14.6	.09
Three-way interaction			
Chart type $\times$ line of progress $\times$ change type	1	1455	9.23**
Students	12	1233.9	7.83**
Residual	84	157.7	

* $p < .05$.** $p < .01$.

the percentage of raters who indicated that a significant change in performance had occurred. The design was a repeated measures model with students used as the blocking variable. The results are displayed in Table 2. Significant main effects were found for chart type, lines of progress, and change type. A significant 3-way interaction also was found. The significant effect for students (blocking variable) indicates that students varied among themselves in rating the significance of phase changes.

The data are displayed visually in Figure 2. Equal interval charts with no lines of progress resulted in an average of 68% of the phase changes rated as having a significant change in level and 51% of the phase changes rated as having a significant change in trend. Those figures increased to 81% rated as having significant level changes and 77% significant trend changes when lines of progress were drawn. Semilog charts with no lines of progress resulted in an average of 62% of the phases rated as having a significant level change and 48% a significant trend change. When lines of progress were added to the semilog charts, 67% of the phases were rated as having significant level

Figure 1. Percentage of interrater agreement on changes in level and trend as a function of chart type and use of lines of progress.

changes and 31% were rated as having significant trend changes.

DISCUSSION

The data demonstrate a complex interrelationship between chart type, lines of progress, and change type, and their differing effects on interrater agreement and ratings of significance. The application of a line of progress to charted data clearly had the most dramatic effect on interrater agreement. Although lines of progress increased agreement on both level and trend effects, the 2-way interaction suggests that lines of progress may improve agreement of ratings on trend more than agreement of ratings on level. The type of chart used did not appear to have a marked effect on interrater agreement.

Lines of progress also had a significant effect on ratings of significance. However, the effect varied as a function of both chart type and change type. When added to an equal interval chart, lines of progress seemed to have the effect of accentuating possible changes in trend and level, as indicated by the significant rise in both. When added to a semilog chart, lines of progress again seemed to have the effect of accentuating possible changes in level. However, lines of progress had the effect of obliterating any possible changes in trend, as indicated by the dramatic reduction in percentage of trend changes rated as significant.

The type of chart used did have an effect on ratings of significance, particularly in ratings of trend. The most common effect of replotting data on a semilog chart was a "flattening" of the data so that changes in trend were not so apparent. This

Figure 2. Percentage of phase changes rated as having a significant change in level and trend as a function of chart type and use of lines of progress.

had little effect on observers' ratings of trend, however, until a line of progress was added.

The findings generally correspond to expectations based on assumptions regarding lines of progress and semilog charts. For example, the assumption that a line of progress highlights the presence or absence of trend in time-series data was supported by the finding that lines of progress significantly improved interrater agreement on changes in trend. Likewise, the assumption that a semilog chart more accurately reflects linearity of progress was supported by the finding that raters tended to judge fewer phase changes as significant when data were presented on semilog charts.

At least two potential limitations of this study should be addressed. First, the charts evaluated may not be representative of the "population" of actual charts published. Although the study has demonstrated that chart type and lines of progress

can affect rater judgments in at least some real cases, the findings should not necessarily be generalized to all data sets. Second, the study addresses only the issue of agreement in opinion across raters and the general direction of rater opinions. These are reliability issues and are separate from the critical validity issue of correspondence of raters' opinions with some absolute standard of significance. We do not know if the judgmental aids examined in this study bring us any closer to the "truth" concerning performance change in the charts examined. Although agreement is both desirable and necessary for establishing confidence in the interpretability of research results, it is not sufficient to guarantee that results have been appropriately interpreted.

The implications of these findings need to be examined more carefully in future research. Studies to date have documented low rates of agreement

among judges of charted data (DeProspero & Cohen, 1979; Jones et al., 1978) and lack of agreement between visual and statistical inference (Jones et al., 1977). This study has documented that two judgmental aids can affect both interrater agreement and raters' judgments of significance of changes across experimental phases. Still to be resolved is whether the same analytic techniques should be used with all data sets or whether a technology is needed for matching analytic strategies to basic characteristics of performance change and variability displayed in the data. For example, White and Haring (1976) suggested that meaningful changes in rate of progress can occur *within* a given experimental phase, and that in such cases the use of a single, linear line of progress may be misleading, may reduce interrater agreement, or may increase the probability of inappropriate interpretations of the data. In other instances procedures such as drawing "smoothed curves" through data as described by Tukey (1977) may be required.

This study replicates the general finding of previous studies that "seeing is believing" is not such a simple proposition. It is clear that interrater agreement and ratings of significance of charted data are affected by certain characteristics of the charts themselves. This study extends previous research by demonstrating the specific effects of lines of progress and semilogarithmic charts, by documenting their relative impact on changes in level and changes in trend, and by emphasizing that using the two judgmental aids in combination to improve interrater agreement may result in more conservative ratings of significance.

REFERENCES

- Baer, A. M., Rowbury, T., & Baer, D. M. (1973). The development of instructional control over classroom activities of deviant preschool children. *Journal of Applied Behavior Analysis*, 6, 289-298.
- Boren, J. J., & Colman, A. D. (1970). Some experiments on reinforcement principles within a psychiatric ward for delinquent soldiers. *Journal of Applied Behavior Analysis*, 3, 29-37.
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, 12, 573-579.
- Edgington, E. (1967). Statistical inference from $N = 1$ experiments. *Journal of Psychology*, 65, 195-199.
- Furlong, M. J., & Wampold, B. E. (1982). Intervention effects and relative variation as dimensions in experts' use of visual inference. *Journal of Applied Behavior Analysis*, 15, 415-421.
- Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, 5, 141-154.
- Haring, N. G., Liberty, K. A., & White, O. R. (1980). Rules for data-based strategy decisions in instructional programs. In W. Sailor, B. Wilcox, & L. Brown (Eds.), *Methods of instruction for severely handicapped students* (pp. 159-192). Baltimore: Paul H. Brooks.
- Ingham, R. J., & Andrews, G. (1973). An analysis of a token economy in stuttering therapy. *Journal of Applied Behavior Analysis*, 6, 219-229.
- Jones, R. R., Vaught, R. S., & Weinrott, M. R. (1977). Time series analysis in operant research. *Journal of Applied Behavior Analysis*, 10, 151-166.
- Jones, R. R., Weinrott, M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis*, 11, 277-283.
- Kazdin, A. E. (1976). Statistical analyses for single-case experimental designs. In M. Hersen & D. H. Barlow (Eds.), *Single-case experimental designs* (pp. 265-316). New York: Pergamon Press.
- Kazdin, A. E. (1982). *Single-case research designs*. New York: Oxford University Press.
- Knapp, T. J. (1983). Behavior analysts' visual appraisal of behavior change in graphic display. *Behavioral Assessment*, 5, 155-164.
- Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education*, 147, 62-81.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Wampold, B. E., & Furlong, M. J. (1981). The heuristics of visual inference. *Behavioral Assessment*, 3, 79-82.
- White, O. R. (1974). *The "split-middle"—a "quickie" method of trend estimation*. Experimental Education Unit, Child Development and Mental Retardation Center, University of Washington, Seattle.
- White, O. R., Billingsley, F., & Munson, R. (1980, October). *Evaluation of the severely and profoundly handicapped*. Paper presented at the meeting of the Association for the Severely Handicapped, Los Angeles.
- White, O. R., & Haring, N. G. (1976). *Exceptional teaching*. Columbus, OH: Charles E. Merrill.
- White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus, OH: Charles E. Merrill.
- Wincze, J. P., Leitenberg, H., & Agram, W. S. (1972). The effects of token reinforcement and feedback on the delusional behavior of chronic paranoid schizophrenics. *Journal of Applied Behavior Analysis*, 5, 247-262.
- Wolery, M., & Billingsley, F. F. (1982). The application of Revusky's R_n test to slope and level changes. *Behavioral Assessment*, 4, 93-103.

Received October 13, 1983

Final acceptance April 2, 1984