Accuracy and Efficiency of Data Interpretation: A Comparison of Data Display Methods

Elizabeth Lefebre, Michael Fabrizio, and Charles Merbitz

The Chicago School of Professional Psychology

Although behavior analysis relies primarily on visual inspection for interpreting data, previous research shows that the method of display can influence the judgments. In the current study, 26 Board-Certified Behavior Analysts reviewed two data sets displayed on each of three methods—equal-interval graphs, tables, and Standard Celeration Charts—and determined which data set showed more change per unit of time and which showed less variability. Regardless of their preference for one method of display, participants handled data displayed on the Standard Celeration Chart most efficiently. However, the accuracy of their judgments across all types of data display was low and may indicate the need for additional training, even for persons holding the Board-Certified Behavior Analyst credential.

DESCRIPTORS: data analysis, graphic displays, Standard Celeration Chart

The Importance of Data-Based Decision Making and Prediction

The experimental analysis of behavior rests on baseline logic, a form of inductive reasoning (Cooper, Heron, & Heward, 1987)[1] that rests on an assumption that if a participant responds at a steady state under a set of conditions, changes in the dependent variable that occur as a result of changes in the independent variable become easier to detect. As more and more baseline data are collected, predictive power increases until a stable state of responding occurs (Baer, Wolf, & Risley, 1968). As research progresses, the experimenter makes predictions about the participant's responding, and future data either verify or refute those earlier predictions. When the conditions change and the participant's responses change accordingly, the experimenter compares changes in the participant's responding under the new conditions with responding predicted had the old conditions stayed in effect.

Just as prediction plays an important role in research, it plays an equally important role in clinical intervention. Behavior analysts must know how their clients respond to changes in their environment (Behavior Analysis Certification Board®, 2004). Behavior analysts must also be familiar with the characteristics of the client's behavioral data so they can both easily predict future performance and quickly detect potential problems with programming or other environmental factors. Further, they must easily recognize features of their client's data that may hold important clinical implications, such as the presence of normal or abnormal levels of variability, and also the speed with which a client acquires new skills. For example, if a client's responding during a new intervention becomes extremely variable, the behavior analyst should be able to look back at previous response patterns, detect any difference in variability, and make appropriate changes to the program. Prediction based on data also allows changes in programming so a client will meet intervention goals in an efficient manner.

Common Methods of Prediction

Experimenters use two types of tools to make predictions: line graphs and statistics (Cooper et al., 1987). In a statistical approach, an experimenter or behavior analyst derives an equation for a line that best fits the data set and then uses that linear equation to predict where the next data point(s) should fall. From these predictions, the behavior analyst then makes ongoing decisions based on the

This study was completed in partial fulfillment of the requirements for the degree of Master of Arts in Applied Behavior Analysis by the first author. The authors thank Behavior Research Company for granting its permission to reproduce the Daily per Minute (Dpmin-12EC) Standard Celeration Chart. Correspondence about this study should be sent to: Michael Fabrizio, 1110 24th Avenue South, Seattle, WA 98144-3037.

[1]Although there is a newer edition of Cooper, Heron, and Heward's book (2008), the text cited in this manuscript comes from the earlier edition (1987) of the book.

degree of alignment between prediction and actual observations. When using a graphic analysis, an experimenter or behavior analyst does not rely on mathematically calculated values for predictions, but instead plots data on a graph and makes predictions from the current data about the future data by relying primarily on visual inspection of such features as the data's level, trend, and variability.

The Importance of Visual Analysis Tools and Procedures

In applied behavior analysis, visual analysis of graphic displays is the most common method used to evaluate the effects of changes in independent variables (DeProspero & Cohen, 1979). Visual records and displays of data validate whether clinical interventions are associated with the intended, positive differences in a client's responding. Researchers and practitioners can also see whether they have isolated and controlled important variables affecting the behavior of interest; they can also respond to changes in environmental variables as those changes occur, and in cost- and time-efficient ways (Parsonson & Baer, 1978).

In applied settings, data should drive a feedback loop in which changes in client performance occasion changes in the intervention or reinforce maintenance of the current conditions. In such a feedback loop, data should serve as stimuli that occasion changes in the behavior analyst's behavior to make changes in procedures when warranted and to predict future outcomes more accurately. Given this important role in clinical intervention, data should be collected and displayed in a way that allows behavior analysts to make fast, easy, and accurate decisions (Johnston & Pennypacker, 1993).

To facilitate fast, easy, and accurate decision making based on graphed data, important dimensions of the data such as trend, level, and variability should be readily distinguishable. Data displayed graphically should enable behavior analysts and researchers to see patterns in the data and to make comparisons across changes in treatment conditions, since such comparisons have proven more useful in influencing interpretive reactions than statistical analyses (Johnston & Pennypacker, 1993). Given this reliance on visual rather than mathematical properties of the data, the type of graphic display selected may play an important role in the analysis. Because the type of graph chosen plays an important role in the decisions made for a client, behavior analysts and researchers must choose a display that most clearly represents the behavior of interest as it occurs naturally in the environment.

Commonly Used Types of Graphic Displays

Tufte (1983) described the purpose of graphic displays as showing data and inducing the viewer to pay attention to the data themselves rather than to other features of the display. According to Tufte (1983), displays should avoid distorting what the data say and should encourage the viewer's eye to compare different parts of the data with one another. Behavior analysts have tended to use three main types of data displays: (1) line graphs, such as equal-interval graphs or semilogarithmic charts; (2) data tables; and (3) cumulative records (Cooper et al., 1987). Of these three types of displays, behavior analysts rarely use cumulative records, and instead most commonly use equal-interval line graphs (Cooper et al., 1987).

With equal-interval graphs, also called add-subtract graphs, both the vertical and horizontal axes are divided into equal distances between marks to denote the addition or subtraction of constant amounts. These graphs display some dimension of the dependent variable on the vertical axis (typically, percent correct or frequency of responding) and some dimension of time (typically, days or weeks) or an analogue of time (sessions) on the horizontal axis (Cooper et al., 1987). Because of their mathematical properties, equal-interval graphs tend to produce curved rather than straight lines to display human learning data. Because equal-interval graphs tend to produce curved lines when they display learning data, those curved lines likely negatively affect the behavior analyst's ability to accurately and quickly predict future performance or to easily compare variability across the full range of responding.

Semilogarithmic graphs, such as the Standard Celeration Chart (SCC) developed by Lindsley in the 1960s, display frequency measures along the vertical axis. Celeration refers to change in frequency of behavior across time and is depicted on the SCC as the trend in the data. In contrast to equal-interval graphs, the vertical axis of the SCC has a multiplicative scale; that is, equal distances on the graph are equal ratios instead of additive distances. A doubling of the response frequency would cover the same vertical distance on the SCC regardless of the base frequency. Thus, data that grow from a frequency of two per minute to four per minute would cover the same vertical distance as data that grow from a frequency of 500 per minute to 1,000 per minute. On the horizontal axis, the SCC displays units of real calendar time such as days, weeks, or months. When data are plotted on the SCC, the multiplicative scale of the ordinate shows proportional change in responding. Because it reliably produces a straight line with human learning data (Lindsley, 1991), standard trend (or celeration) measures can be calculated and drawn easily and quantified visually. This is important because changes in trend are often a major factor in an analysis of data (DeProspero & Cohen, 1979). With the SCC, behavior analysts may be able to detect important changes in a client's performance more easily than with other graphic displays that do not consistently produce straight lines.

Previous Research on Visual Analysis of Graphed Data

Stevens and Savin (1962) replotted data from a selection of cumulative learning experiments onto graphs that had logarithmic scales on both the ordinate and the abscissa. They found that all 8 data sets created a straight line when plotted on the log-log graph, and a line of best fit could easily be fitted to the data by eye without the need for a statistical procedure. Although log-log graphs can produce very straight lines and can be very helpful for predicting future behavior, such graphs are rarely used in clinical practice.

Furlong and Wampold (1982) assessed the ability of experts trained in single-subject design to sort graphed data into groups showing similar experimental effects. The researchers transformed the graphs in one of three ways. One group of graphed data underwent scaling transformations in which the vertical axis was stretched. The second group of graphed data underwent transformations in variability in which the standard deviation of random deviates was multiplied by the same constant. Such a transformation produced data with larger, more random levels of variability. The third group of graphed data underwent a standard transformation in which a randomly selected deviate was added to each value in the data pattern. Furlong and Wampold (1982) found that the raters attended to large changes in trend and level, but they did not focus on comparable changes in variability across phases. The researchers suggested that variability should be defined clearly in the literature to clarify the guidelines for its use as an interpretive dimension of visual analysis. The scaling transformations also altered participants' decisions about experimental control demonstrated on the graphs, since the transformations seemed to be correlated with a change in response. This prompted a suggestion that experimenters should pay closer attention to scale when analyzing their data.

Mawhinney and Austin (1999) examined the speed and accuracy with which an expert in each of three types of visual displays identified the onset of intervention: equal-interval graphs, SCCs, and Statistical Process Control (SPC) Charts. These experts reviewed data sets and plotted the data on their preferred chart, point by point. Accuracy, as measured by the expert's ability to identify the location of the onset of an intervention, was higher for the data plotted on traditional equal-interval graphs (63% correct), followed by SCCs (32% correct), and then SPC charts (25% correct). However, the SCC proved the most *efficient* method of data analysis because the SCC expert spent significantly less time completing the tasks than either the equal-interval or the SPC expert. The SCC expert took less than half the time to plot the data than did the equal-interval or the SPC expert, although the SCC expert's correct response frequency was lower. These findings along with the others suggest that the different display methods used to analyze data may have an effect on the viewer's ability to accurately interpret data.

The Current Study and Its Goals

Similar to the study of Mawhinney et al. (1999), in the current study, the efficiency with which behavior analysts make accurate data-based decisions was assessed with equal-interval graphs, semilogarithmic charts, and data tables. The previous research points to the importance of studying the variables that affect interpretive behavior. This is critical because behavior analysts make most of their decisions via visual analyses of data. The current study sought to determine the accuracy and efficiency of making data-based decisions across different display types and the influence of preferences for and experience with certain types of graphs on participants' ability to make data-based decisions.

METHODS

Participants and Setting

A total of 26 Board-Certified Behavior Analysts who held a minimum of a master's degree participated in the study. Participants were selected through the certificant registry on the Behavior Analyst Certification Board™ web site using a selection procedure in which every ninth certificant in each state within the United States received an invitation to participate. Because each certificant had to be individually invited to participate, it was not feasible to invite all certificants (greater than 3,000 at the time this research occurred) in the registry. These 26 participants responded from a pool of 446 who were invited. For those potential participants who did not respond to initial invitations, a maximum of two follow-up e-mails were sent to them reminding them of the opportunity to participate. This pool of professionals was selected because they all had previously met the minimum requirements for certification by the Behavior Analyst Certification Board™ and, as such, likely had both training and experience in analyzing data derived from behavior analytic interventions. Participants received invitations via e-mail, and they logged into the Web site to participate at their convenience.

Demographic Information About the Participants

Demographic data collected from participants included information about their experience in the field, the highest educational degree they held, the number of years they had practiced behavior analysis, and the type of data display they most commonly used. Participants also indicated which data display (equal-interval graphs, SCCs, or data tables) they thought easiest to interpret and which display they preferred.

All participants in the study held certificates as Board-Certified Behavior Analysts for a median of 3 years, with a range of 2 months to 7 years. The median number of years participants had experience in interpreting data was 8 years, with a range of 3 to 30 years. Of the participants, 16 held a master's degree and 10 held a doctoral degree as their highest level of education. Of the 26 participants, 16 reported that they used equal-interval graphs as their primary method of interpretation; 2 participants reported that they used data tables; 5 participants reported that they used SCCs; and 3 participants reported that they used all three display methods for interpretation. This information allowed participants to be placed into groups based on their affinity toward one type of display.

After completing the survey, participants indicated which display method they preferred and which display method they found easiest to interpret. As to their preference, 12 participants reported that they preferred equal-interval graphs, 1 participant preferred tables, and 13 participants preferred SCCs. As to ease of interpretation, 10 participants reported that SCCs were the easiest method, 15 participants reported that equal-interval graphs were the easiest, and 1 participant reported that tables were the easiest.

Materials

A Web site designed by the first author served as the stimulus presentation and data-collection vehicle for this study. The Web site collected data both on responses selected and on response times for each participant. This Web site led participants through a number of pages, including a letter of agreement, an informational questionnaire, several pages of instructions, and specific tips to remember when answering questions. The questionnaire included questions about the participants' experience in the field, the degree held, their number of years practicing behavior analysis, and the type of data display they most commonly used. The survey began after both the informational questionnaire and the instructions, and it consisted of graphic displays or data tables depicting 15 sets of data, each graphed on an equal-interval graph and an SCC, as well as in tabular form. The SCCs were constructed with an SCC template for Microsoft Excel (Harder, 2008); equal-interval graphs were developed with Microsoft Excel's scatter plot graphing function; and data tables were created in Microsoft Excel.

The data sets used in the study consisted of both real data sets and modified data sets, all depicting AB designs. Of the 15 data sets, 6 were real data sets collected within early intervention applied behavior analysis programs for children with autism and overseen by behavior analysts, and 3 of the data sets were generated by college students as they performed flashcard timings designed to teach them verbal relations important to graduate coursework they completed. The remaining 6data sets were modified versions of 9 additional real data sets. The data sets were modified by a change either in the length of the observation or in the frequency of certain data points in the data set to increase or decrease the variability or to modify the slope.

These data sets were modified to provide a range of stimuli that best represented how the different types of displays show the same data sets differently. The modified data sets were changed so that the slopes and variability of the two data paths differed by a ratio of at least X1.25 on an SCC. A difference in celeration and variability of at least X1.25 appears to be a readily detectable difference between two sets of data plotted on the SCC. Each display consisted of two data paths, displayed in two panels (Panel A and Panel B) on the graph. Each data path differed from the other by the slope of the trend line, the variability, or a combination of both.

Dependent Measures

The dependent measures for the study were the participants' frequency of correct and incorrect identification of differences in trend and variability per minute, along with interresponse time. Interresponse time was calculated to determine the time each participant took to respond to each stimulus. A linear regression equation was used to calculate the slope of each line created by the data in each panel. The display showing the most success in increasing the behavior of interest (the steepest trend) determined the accuracy, and the panel showing a greater slope became the correct answer. For detecting amounts of variability, the range that the data deviated from the slope was calculated. The panel showing a smaller variability value became the correct answer.

Procedure

The Web site led participants through a series of pages before they accessed the data sets. The first page included the introduction, explanation, and purpose of the study as well as a letter of agreement and experimenter contact information. After reading the introduction and agreeing to participate, participants proceeded to the informational questionnaire. After completing the questionnaire, participants then read a vignette explaining that they were to act as a behavior analyst in a clinical setting. The vignette asked participants to make data-based decisions about the trend and variability of the data they would see on the following pages. At the bottom of each page of instructions, a "next" button appeared that enabled the participant to advance to the next page. The vignette read:

> *You have been hired at a central office to consult on several different behavioral programs running at different sites. The staff use different data collection systems and display their data for interpretation on different types of visual displays. You will be looking at pairs of data sets from these programs. On each page, panel A and panel B show the results of two behavior improvement procedures for a client. Your decision will be to recommend Procedure A, Procedure B, or either one for future work with that client, based on which procedure has been more effective/efficient overall. Staff members have also complained about client inconsistency, so you will also select the panel showing the least variability, or indicate that the variability is the same. Finally, in some cases you may need more information to make these decisions, so for either question you may indicate that there is not enough information to make the choice.*

The next page following the vignette read:

> *It is important to remember the following when making your decisions: Both procedures took the same amount of staff time/resources per session. In each procedure, the client reached the goal that was set. In some pairs, the client may have started at different levels with the two different procedures. With some data sets, clients may have had unusually good or bad days that are probably not associated with the procedure. Now staff members are working to increase or decrease other behaviors, so you are making important recommendations for future work.*

Following the instructions, the first page of the survey appeared. The participants saw the equal-interval graphs first, with data sets presented in random orders, followed by the data tables, and then SCCs also in random order. Each display appeared on two pages, making the survey a total of 92 pages, including the 15 data sets that each appeared two times across the three different display types.

The first page (Figure 1) included a graphic display and the following question: "*If the goal was met for each procedure, which procedure was most successful in increasing the behavior in the quickest, most economical way?"*

Each page had four options for answering the question: "Panel A," "Panel B," "Either One," and "Not Enough Information," as Figure 1 shows. When the participant clicked on an answer, the Web site recorded the time and advanced to the next page.

This next page included the same graphic display and the question, *"Which program shows less variability?"* The same options as on the previous page were offered for the participant to answer the question. Each page also included a button entitled "Review Instructions" that opened up a separate page with the instructions previously presented. Following the page with the second question, a new display appeared and the questions repeated.

Accuracy Calibration

Accuracy was calibrated between the first author's assessments and the Web site's server for both the correctness of answers selected and the time taken to answer each question. Accuracy of recording the selection of the same answer was calculated from the total number of agreed-upon responses divided by the total number of possible responses. Accuracy scores for recording the selection of the same answer were 100%, and 92% of time intervals were within 3 seconds of each other. Here, differences in the Internet connection speeds may have accounted for the differences in frequencies collected. Because this study sought primarily to determine which of three types of data displays allowed more accurate and efficient data-based decision making in clinical situations, a time difference of less than 3 seconds would likely not translate into clinically significant differences in ease of interpretation.

RESULTS

Response Frequencies-Rate of Change-All participants

With all participants considered, the correct

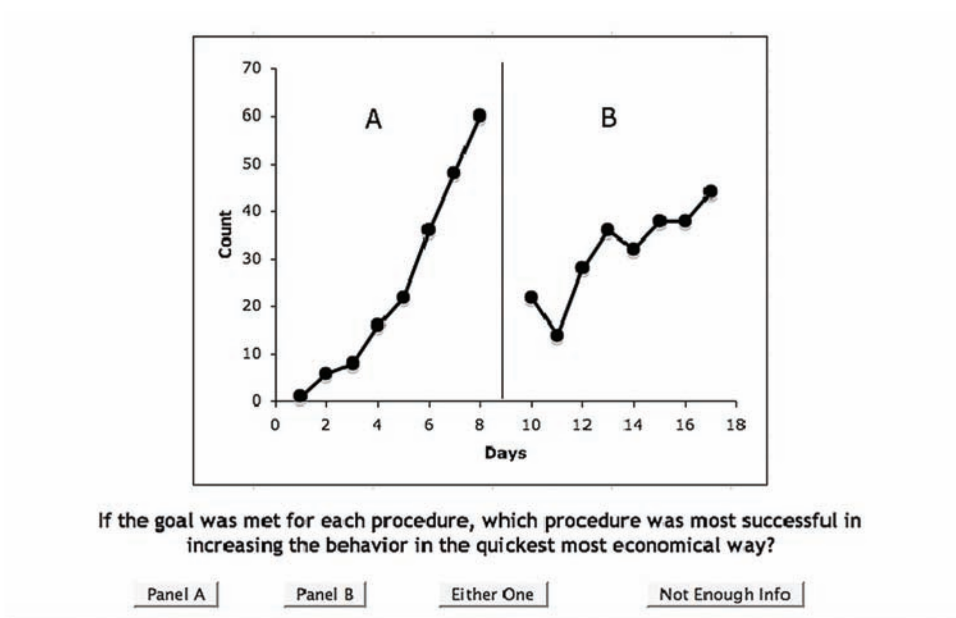


Figure 1. The display participants viewed as they moved through the survey.

Table 1. Mean, median, range, and median percentage of correct and incorrect responses across display methods for questions about both rate of change and variability.

Correct and Incorrect Response Rates Across Display Methods

	Equal Interval Graph		Table		Standard Celeration Chart	
Rate of Change	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Mean	1.45	1.78	1.44	1.52	2.54	2.09
Median	1.39	1.78	1.25	1.22	2.44	1.99
Range	0.41 – 3.00	0.35 - 3.16	0.29 – 3.11	0.37 - 3.53	0.62 – 6.28	0.64 – 4.19
Median Percentage	45%	53%	47%	48%	57%	42%
Range of Percentages	20 - 67%	33 – 73%	0 – 67%	0 – 80%	20 - 80%	20 – 67%
Variability						
Mean	2.71	3.50	2.04	4.21	3.72	7.87
Median	2.48	3.50	1.78	3.75	3.33	7.40
Range	0.30 – 6.10	0.16 – 8.14	0.49 – 5.71	0.79 – 15.71	0.81 – 9.79	2.69 – 20.00
Median Percentage	40%	53%	33%	63%	33%	66%
Range of Percentages	0 – 73%	0 – 73%	0 – 67%	0 – 93%	7 – 60%	40 - 93%

Figure 2. Overall correct and incorrect response frequencies for rate of change and variability across equal-interval graphs, tables, and SCCs. Each dot represents one person's correct frequency and each x one person's error frequency. →

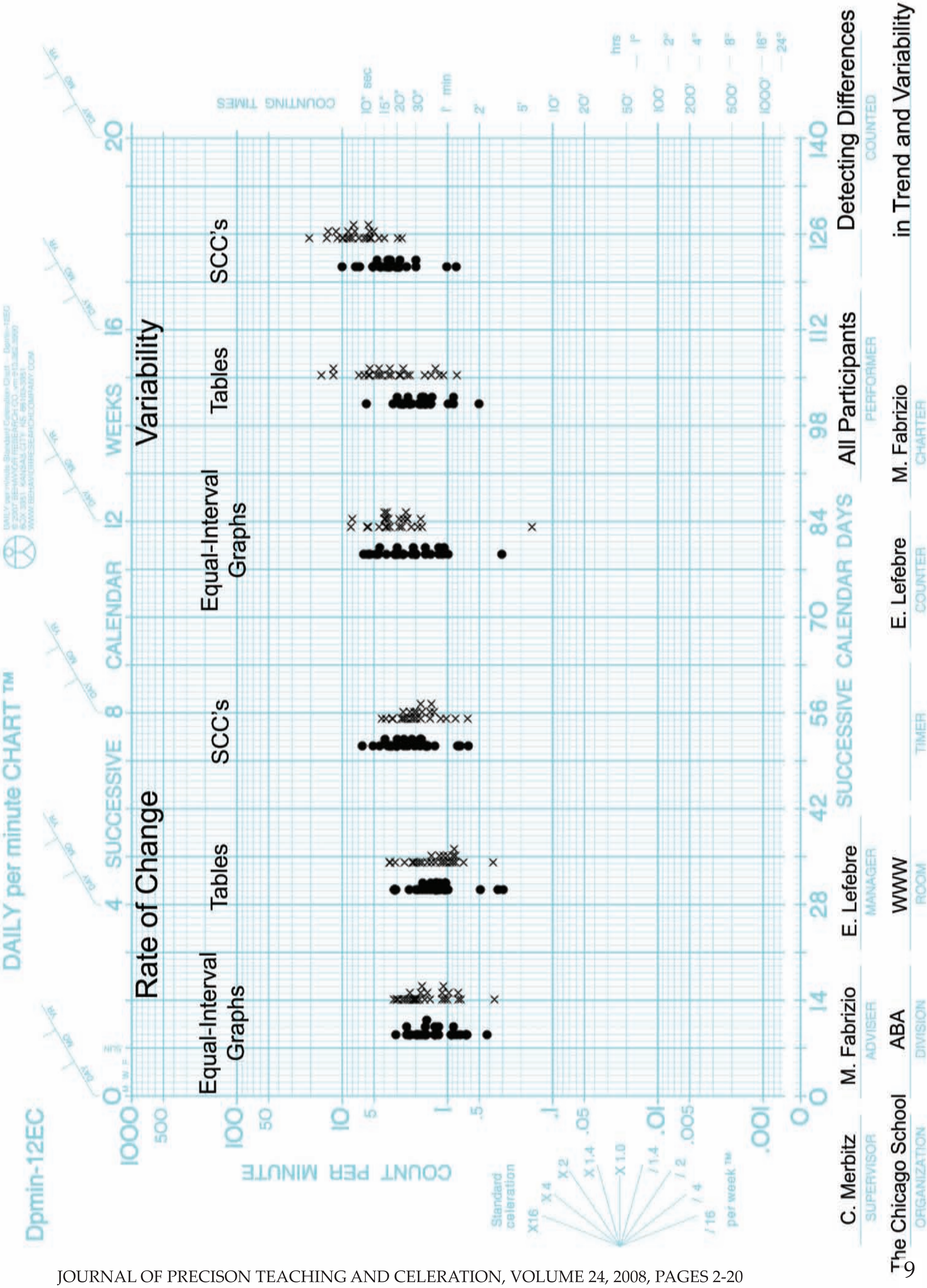

Dpmin-12EC
DAILY per minute CHART ™
SUCCESSIVE CALENDAR WEEKS
COUNT PER MINUTE
COUNTING TIMES
SUCCESSIVE CALENDAR DAYS
Rate of Change
Variability
Equal-Interval Graphs
Tables
SCC's
Equal-Interval Graphs
Tables
SCC's
C. Merbitz
SUPERVISOR
M. Fabrizio
ADVISER
E. Lefebre
MANAGER
TIMER
E. Lefebre
COUNTER
All Participants
PERFORMER
Detecting Differences
COUNTED
The Chicago School
ORGANIZATION
ABA
DIVISION
WWW
ROOM
M. Fabrizio
CHARTER
in Trend and Variability
Standard celeration
per week ™

Table 2. Mean, median, range, and median percentage of correct and incorrect responses across display methods for the equal-interval affinity group for questions about both rate of change and variability.

Correct and Incorrect Response Rates Across Display Method for the Equal Interval Affinity Group

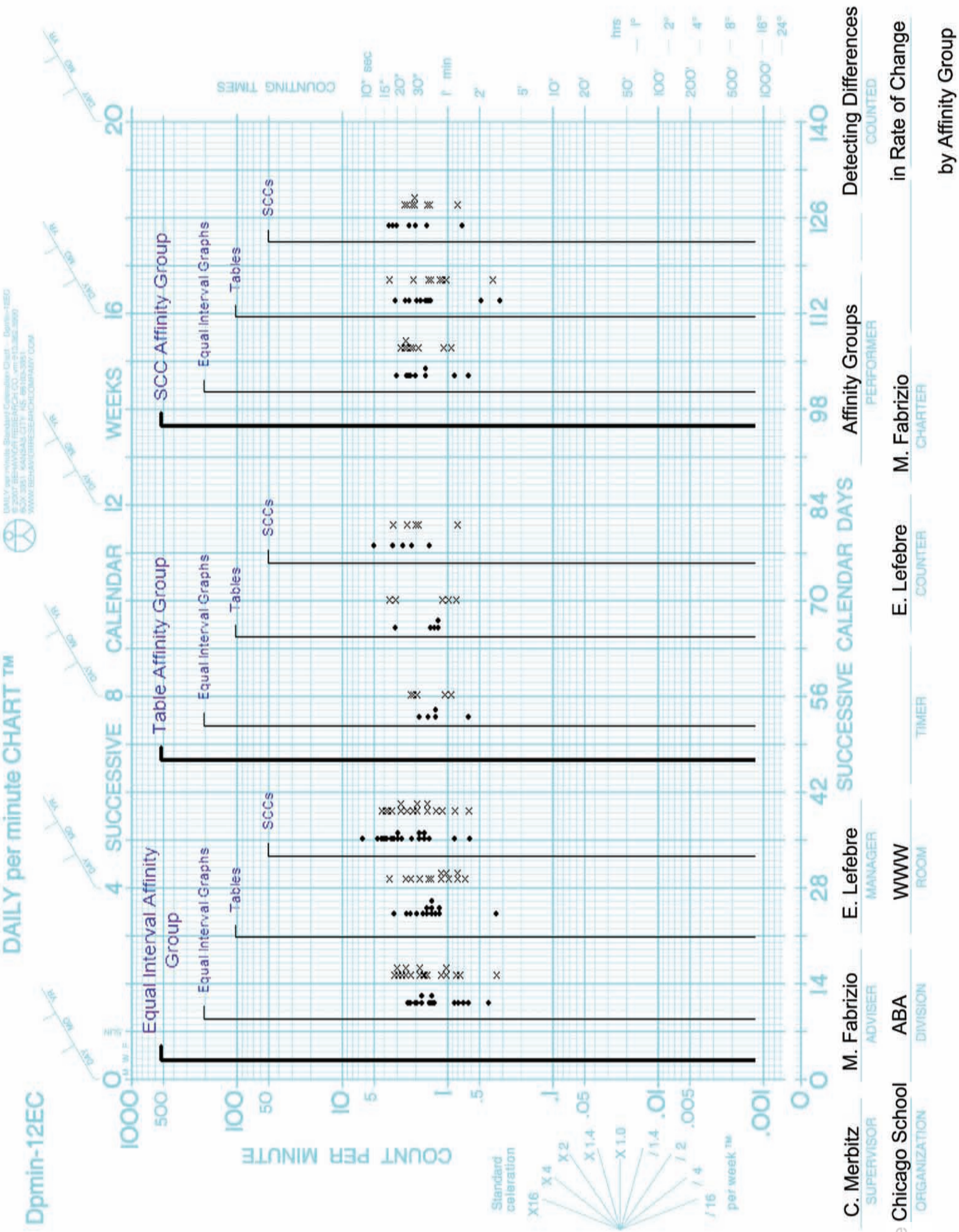
	Equal Interval Affinity Group					
	Equal Interval Graphs		Tables		Standard Celeration Charts	
Rate of Change	correct	incorrect	correct	incorrect	correct	incorrect
Mean	1.32	1.82	1.45	1.61	2.44	2.08
Median	1.28	1.75	1.24	1.38	2.10	1.96
Range	0.41 - 2.35	0.35 - 3.16	0.35 - 3.16	0.69 - 3.53	0.62 - 6.28	0.64 - 4.19
Median Percentage	40%	57%	47%	53%	53%	47%
Range of Percentages	20 - 67%	33 - 73%	0 - 60%	0 - 80%	13 - 80%	20 - 60%
Variability						
Mean	2.41	3.5	2.10	4.58	3.31	8.24
Median	2.11	3.19	1.78	4.20	3.18	7.07
Range	0.97 - 6.10	0.16 - 8.14	0.86 - 5.71	1.18 - 15.71	0.81 - 7.35	2.69 - 20.00
Median Percentage	37%	60%	27%	67%	33%	67%
Range of Percentages	27 - 73%	27 - 73%	0 - 60%	0 - 80%	7 - 40%	40 - 93%

median response frequencies on the rate of change were higher on SCCs, followed by equal-interval graphs, and then data tables. Figure 2 and Table 1 show these results. The SCCs also had a higher percentage of correct responses, followed by tables, and then equal-interval graphs for questions about rate of change. Table 1 shows these results. Error frequencies were high across all display methods, with SCCs having the highest, followed closely by equal-interval graphs, and then data tables.

Response Frequencies-Variability-All participants

Correct response frequencies on variability were also higher on SCCs (as Figure 2 and Table 1 show), followed by equal-interval graphs, and then data tables. Error frequencies were also high across all display methods, with SCCs having the highest error frequencies, followed by tables, and then

Figure 3. Correct and incorrect responses to questions about rate of change on each display method, → by each affinity group.


Dpmin-12EC
DAILY per minute CHART ™
COUNT PER MINUTE
SUCCESSIVE CALENDAR WEEKS
SUCCESSIVE CALENDAR DAYS
COUNTING TIMES
Equal Interval Affinity Group
Table Affinity Group
SCC Affinity Group
Equal Interval Graphs
Tables
SCCs
C. Merbitz
SUPERVISOR
M. Fabrizio
ADVISER
E. Lefebre
MANAGER
E. Lefebre
COUNTER
M. Fabrizio
CHARTER
Affinity Groups
PERFORMER
Detecting Differences
in Rate of Change
by Affinity Group
COUNTED
The Chicago School
ORGANIZATION
ABA
DIVISION
WWW
ROOM
TIMER
Standard celeration
per week

equal-interval graphs. Incorrect response frequencies were higher than correct response frequencies for all three display methods. Equal-interval graphs had a slightly higher percentage correct than both SCCs and data tables. Figure 2 shows each individual participant's correct and incorrect response frequencies per minute for each type of graph, displaying the differences in response frequencies across participants.

Affinity Groups Response Rates

Whereas in Figure 3 and Table 1 the participants are considered as a whole group, further analysis divided the participants into affinity groups based on the type of display they reported using most frequently. Participants who reported that they used equal-interval graphs most frequently were assigned to the equal-interval affinity group, and participants who reported that they used SCCs most frequently were assigned to the SCC affinity group. Affinity group assignment was made based on the participants' report of the type of display they used most frequently, assuming that a participant who used a given type of display frequently should be more skilled at interpreting data on that display than on other display methods. After this partitioning of the participants into affinity groups, the same analyses reported were conducted again.

Affinity Group Members' Detection of Changes in Trend

Equal-Interval Affinity Group--Trend

The highest median correct frequencies and percentage-correct scores for the equal-interval affinity group were for data displayed on SCCs (Figure 3) for questions about rate of change, followed by equal-interval graphs, and then tables (as Table 2 shows). Figure 3 displays correct and incorrect frequencies for questions about rate of change by each affinity group. Table 2 provides mean, median, range, and percentage scores for correct and incorrect response frequencies for the equal-interval affinity group. The equal-interval affinity group also had the greatest range of frequencies for correct and incorrect responses for data displayed on SCCs. The equal-interval affinity group had similar median correct frequencies on equal-interval graphs and tables. The highest median incorrect frequencies for questions about rate of change for the equal-interval affinity group were also for data displayed on SCCs, followed by equal-interval graphs, and then tables. The equal-interval affinity group had higher incorrect response frequencies than correct response frequencies for equal-interval graphs and tables.

Table Affinity-Trend

Although the table affinity group had fewer participants, it had higher correct median response frequencies and percentage-correct scores on SCCs than on equal-interval graphs and tables for questions about rate of change (Figure 3 and Table 3). Table 3 provides mean, median, range, and percentage scores for correct and incorrect response frequencies for the table affinity group. The table affinity group also had the greatest range for frequency of responding in correct responses on SCCs, as well as the greatest range for frequency of responding in incorrect responses on their preferred display method. Incorrect response frequencies were also higher on SCCs and equal-interval graphs than on tables.

SCC Affinity-Trend

The SCC affinity group had the highest correct response frequencies on SCCs, followed by equal-interval graphs, and then tables (Figure 3). The SCC affinity group had the highest percentage-correct score on SCCs, followed by tables, and then equal-interval graphs (Table 4). Table 4 provides mean, median, range, and percentage scores for correct and incorrect response frequencies for the SCC affinity group. The SCC affinity group had the greatest range of correct response frequencies on SCCs and the greatest range of incorrect response frequencies on tables. Incorrect response frequencies were also higher on SCCs, followed by equal-interval graphs, and then tables.

Affinity Group Members' Detection of Change in Variability

Equal-Interval Affinity-Variability

The highest median correct and incorrect frequencies across all affinity groups were for data displayed on SCCs (Figure 4) for questions about variability, followed by equal-interval graphs, and then data tables. Figure 4 displays correct and incorrect response frequencies for questions about variability by each affinity group. The equal-interval

Table 3. Mean, median, range, and median percentage of correct and incorrect responses across display methods for the table affinity group for questions about both rate of change and variability.

Correct and Incorrect Response Rates Across Display Method for the Table Affinity Group

	Table Affinity Group					
	Equal Interval Graphs		Tables		Standard Celeration Charts	
Rate of Change	correct	incorrect	correct	incorrect	correct	incorrect
Mean	1.31	1.67	2.02	1.90	3.04	2.09
Median	1.28	1.91	1.39	1.09	2.92	2.04
Range	0.63 - 1.86	0.94 - 2.27	1.24 - 3.11	0.84 - 3.53	1.48 - 4.91	0.82 - 3.27
Median Percentage	40%	60%	53%	47%	60%	40%
Range of Percentages	40 - 47%	40 - 60%	47 - 60%	40 - 53%	53 - 60%	33 - 47%
Variability						
Mean	2.44	2.69	1.81	1.72	3.21	5.59
Median	2.13	2.49	1.64	1.65	3.22	5.27
Range	1.09 - 4.40	1.90 - 3.85	1.39 - 2.55	0.79 - 2.78	2.86 - 3.56	3.97 - 7.86
Median Percentage	43%	50%	53%	40%	40%	60%
Range of Percentages	27 - 60%	40 - 67%	33 - 67%	33 - 67%	27 - 40%	53 - 73%

affinity group had the greatest range of correct and incorrect response frequencies with data plotted on SCCs, as shown in Table 2. The equal-interval affinity group had the highest median incorrect response frequencies with data plotted on SCCs, followed by tables, and then equal-interval graphs. Median incorrect response frequencies were higher than median correct response frequencies across all display methods.

Table Affinity-Variability

The table affinity group also had higher median correct response frequencies for data plotted on SCCs (Figure 4), with questions about variability, followed by equal-interval graphs, and then their preferred display method. The table affinity group had the greatest range of correct response frequencies on equal-interval graphs, and the greatest range of incorrect response frequencies for data plotted on SCCs. The table affinity group also had the highest median incorrect response frequencies for data plotted on SCCs, followed by equal-interval graphs, and then data tables. Median incorrect response frequencies were also higher than median correct response frequencies across all display methods.

Table 4. Mean, median, range, and median percentage of correct and incorrect responses across display methods for the SCC affinity group for questions about both rate of change and variability.

Correct and Incorrect Response Rates Across Display Method for the SCC Affinity Group

	Standard Celeration Affinity Group					
	Equal Interval Graphs		Tables		Standard Celeration Charts	
Rate of Change	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Mean	1.74	1.74	1.41	1.34	2.41	1.85
Median	1.72	1.90	1.32	1.04	2.63	2.03
Range	0.63 - 3.00	0.94 - 2.27	0.32 -3.09	0.37 - 3.53	0.73-3.58	0.82 - 2.56
Median Percentage	0.47	50%	53%	47%	60%	40%
Range of Percentages	20 - 67%	33 - 60%	27 - 67%	33 - 73%	33 - 60%	33 - 67%
Variability						
Mean	3.27	3.13	2.21	3.38	4.8	6.37
Median	3.7	2.69	2.49	3.33	3.85	5.35
Range	0.30 - 5.53	1.90 - 5.53	0.95 - 3.16	2.23 - 5.41	2.40 - 9.80	3.97 - 9.60
Median Percentage	53%	47%	40%	60%	40%	53%
Range of Percentages	7 - 60%	40 - 60%	20 - 53%	43 - 67%	20 - 60%	40 - 80%

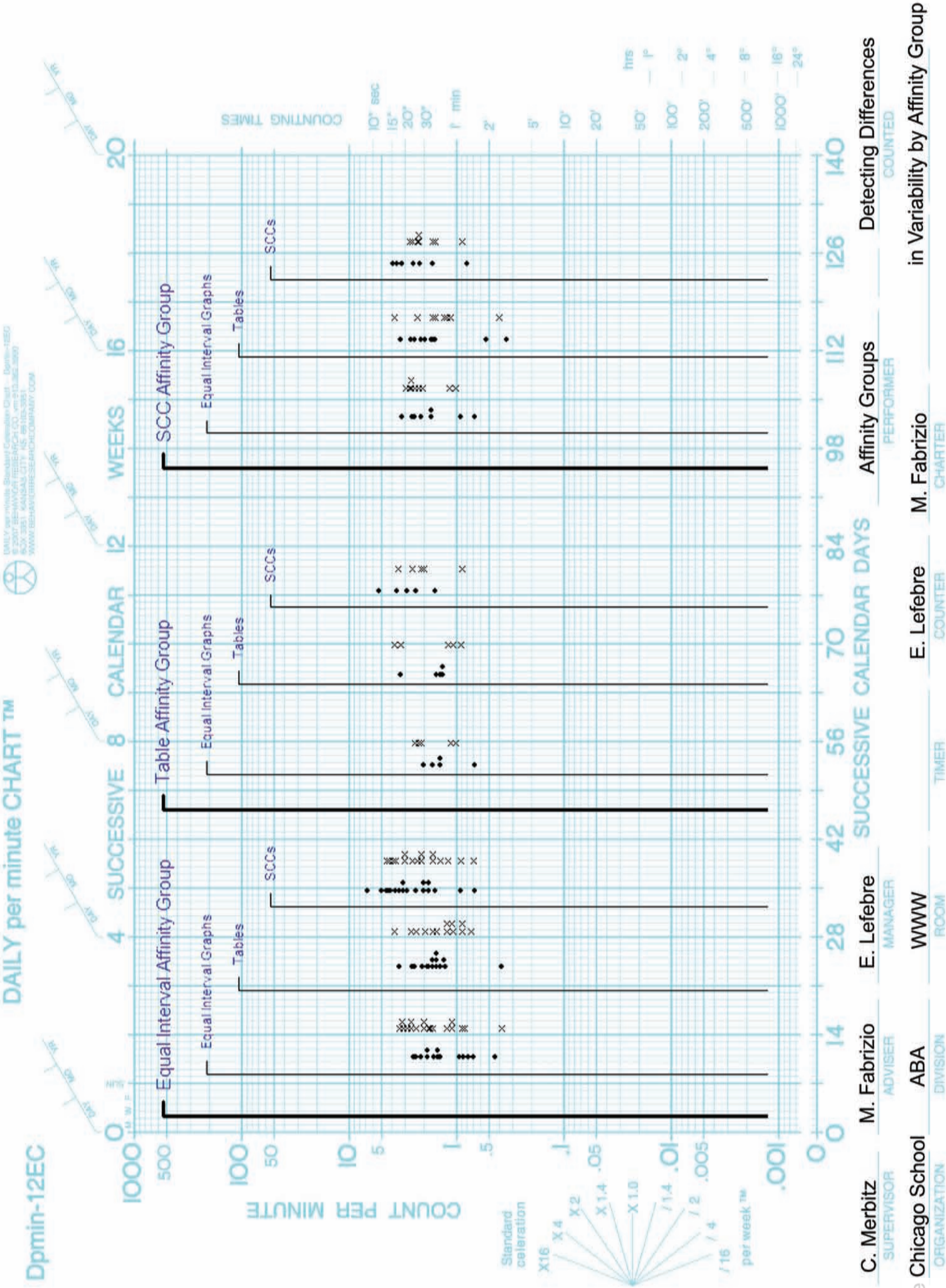
SCC Affinity-Variability

The SCC affinity group also had higher median correct response frequencies for data plotted on SCCs (Figure 4), followed by equal-interval graphs, and then tables. The SCC affinity group also had the greatest range of correct and incorrect response frequencies for data displayed on SCCs. The highest median incorrect frequencies also came from data displayed on SCCs.

Not Enough Information

On each page of the survey, participants had the option to select "Not Enough Information" as a possible response to both questions about rate of change and variability. This option was never the correct answer to any question throughout the survey. Participants across all affinity groups

Figure 4. Correct and incorrect responses to questions about variability on each display method, by each affinity group. →



Dpmin-12EC
DAILY per minute CHART ™
SUCCESSIVE CALENDAR WEEKS
SUCCESSIVE CALENDAR DAYS
COUNT PER MINUTE
COUNTING TIMES
Equal Interval Affinity Group
Table Affinity Group
SCC Affinity Group
Equal Interval Graphs
Tables
SCCs
C. Merbitz
SUPERVISOR
M. Fabrizio
ADVISER
E. Lefebre
MANAGER
E. Lefebre
COUNTER
M. Fabrizio
CHARTER
Affinity Groups
PERFORMER
Detecting Differences in Variability by Affinity Group
COUNTED
The Chicago School
ORGANIZATION
ABA
DIVISION
WWW
ROOM
TIMER

Table 5. Percentages of responses selected as "Either One."

"Either One" Percentages Selected across Display Methods by Affinity Group

Rate of Change	Equal Interval Affinity Group			Table Affinity Group			SCC Affini	
	Equal Interval Graph	Table	SCC	Equal Interval Graph	Table	SCC	Equal Interval Graph	Tab
Percentage of Responses	17%	9%	13%	13%	3%	8%	10%	6%
Variability								
Percentage of Responses	22%	25%	17%	12%	8%	12%	8%	16%

seldom selected "Not Enough Information" as the correct response. Participants selected "Not Enough Information" for between zero and one percent of responses on all display methods for both rate of change and variability regardless of affinity group.

Participants also had an option to select "Either One" as a possible response to both questions about rate of change and variability. This option was the correct answer to two of the questions in the survey. Participants within all groups selected "Either One" more often on equal-interval graphs than on tables and SCCs for questions about rate of change (Table 5). For questions about variability, participants within the SCC affinity group responded "Either One" to more questions about data plotted on SCCs than on equal-interval graphs; participants within the equal-interval graph affinity group responded "Either One" to more questions about data plotted on equal-interval graphs; and participants within the table affinity group responded "Either One" to more questions about data plotted on both equal-interval graphs and SCCs than to data plotted on tables.

Interresponse Time

Interresponse time decreased as participants progressed through stimulus sets for each type of display. Figure 5 is a scatter plot with a logarithmic scale that displays response time decreasing as participants progressed through the stimulus sets for each type of display. As a new type of display was introduced, interresponse time temporarily increased. Each datum on Figure 5 represents a participant's response time on each stimulus set.

DISCUSSION

This study sought to (1) determine the accuracy and efficiency of making data-based decisions across different display types, and (2) determine the influence of preferences and experience with certain types of graphs on participants' ability to make data-based decisions. Overall the results suggested that the highest correct response frequencies occurred when participants made decisions by looking at data displayed on an SCC for both changes in variability and rate of change.

Best Performance of Affinity Groups

All affinity groups had higher correct response frequencies and percentages of correct responses when they looked at rate of change and variability of data displayed on SCCs regardless of a participant's preferred method of analysis. DeProspero and Cohen (1979) found that changes in trend often serve as a major factor that behavior analysts consider when analyzing data, and therefore SCCs may be our best method for interpretation according to the findings of the current study because of the higher correct response frequencies participants emitted when looking at rate of change on an SCC. Although the SCC proved to have the highest correct response frequencies, further training is critical to decrease the high incorrect response frequencies on all methods, including the SCC.

High Error Frequencies Across All Display Methods

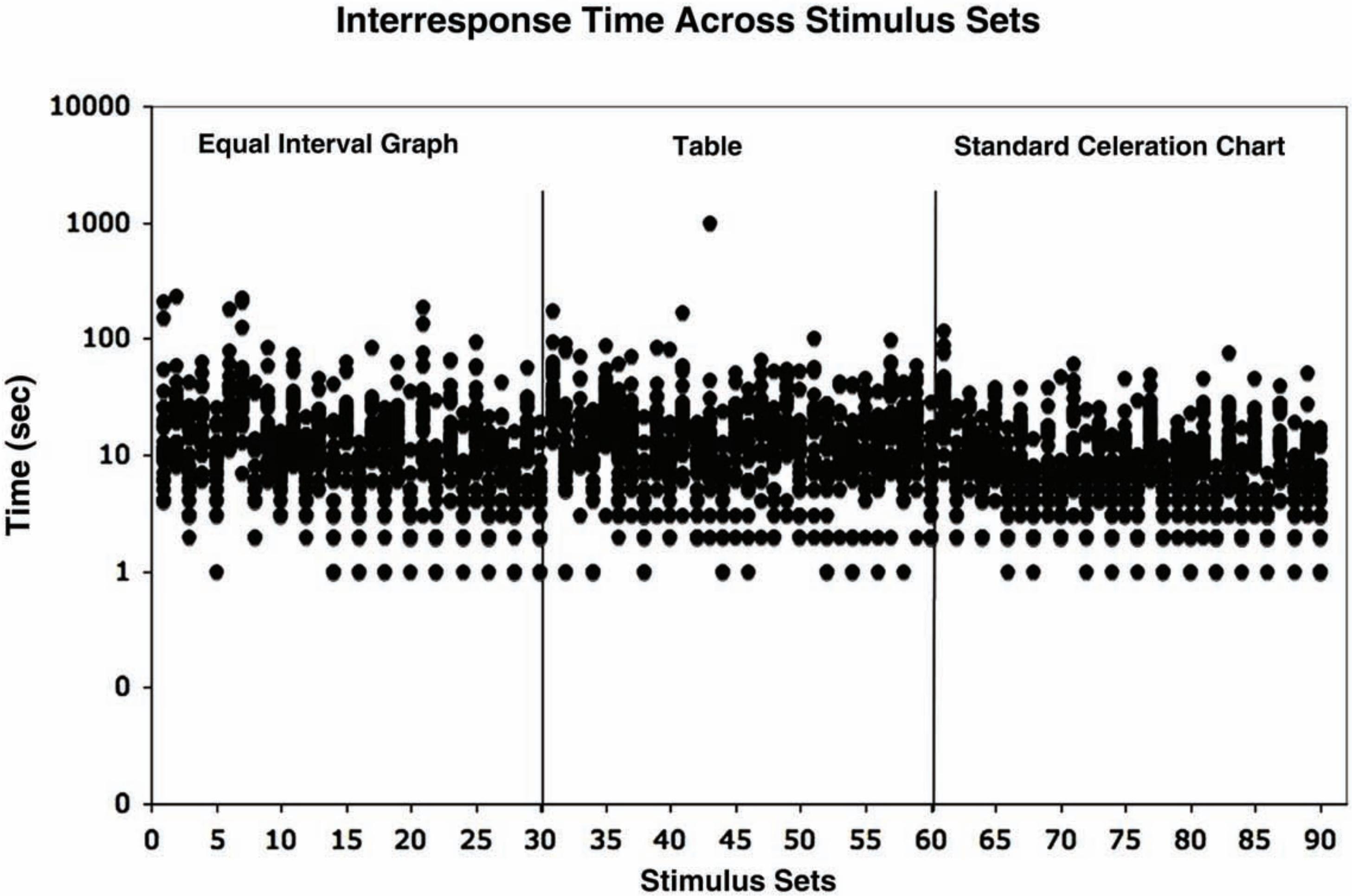


Figure 5. Interresponse time displayed across stimulus sets and display methods.

Across all display methods, incorrect response frequencies were relatively high. Incorrect response frequencies were higher for equal-interval and SCC affinity groups on their preferred display methods than correct response frequencies; and for the table affinity group, the correct and incorrect frequencies were very similar. The high incorrect response frequencies on commonly used displays have some significant clinical implications. If the interpretation (and subsequent programming changes made based on that interpretation) in clinical practice are incorrect, behavior analysts may fail to provide clients with the services they are ethically mandated to provide. For a science that distinguishes itself from others by its use of graphic display of data for interpretation, the present results are disturbing and suggest at least that closer attention must be paid to the methods used to analyze data. Seemingly, when it comes to data collection, data display, data analysis, and data-based decision making, more emphasis may be needed in the training and continuing education of behavior analysts working in applied settings. Many clients come to behavior analysts because they have lower skill levels in specific areas. For behavior analysts to provide effective services reasonably calculated to help improve the already often impoverished repertoires of their clients and thus enhance their clients' lives, more accurate and efficient methods may be needed to use data to make sound, ongoing decisions on behalf of those clients.

Real vs. Modified Data Sets

The stimuli selected for the current study were chosen to illustrate the effects of different display methods on the interpretation of data. The stimuli included a selection of real data from clinical settings so that findings from the study could be generalized more easily to data collected in applied settings. Although they were taken from real clinical settings, some of the data sets were modified to represent the characteristics thought to influence interpretation of data when displayed visually (DeProspero & Cohen, 1979). Modified data sets were created so that specific characteristics of the data would be either attenuated or accentuated as a result of the type of display used. For example, out-

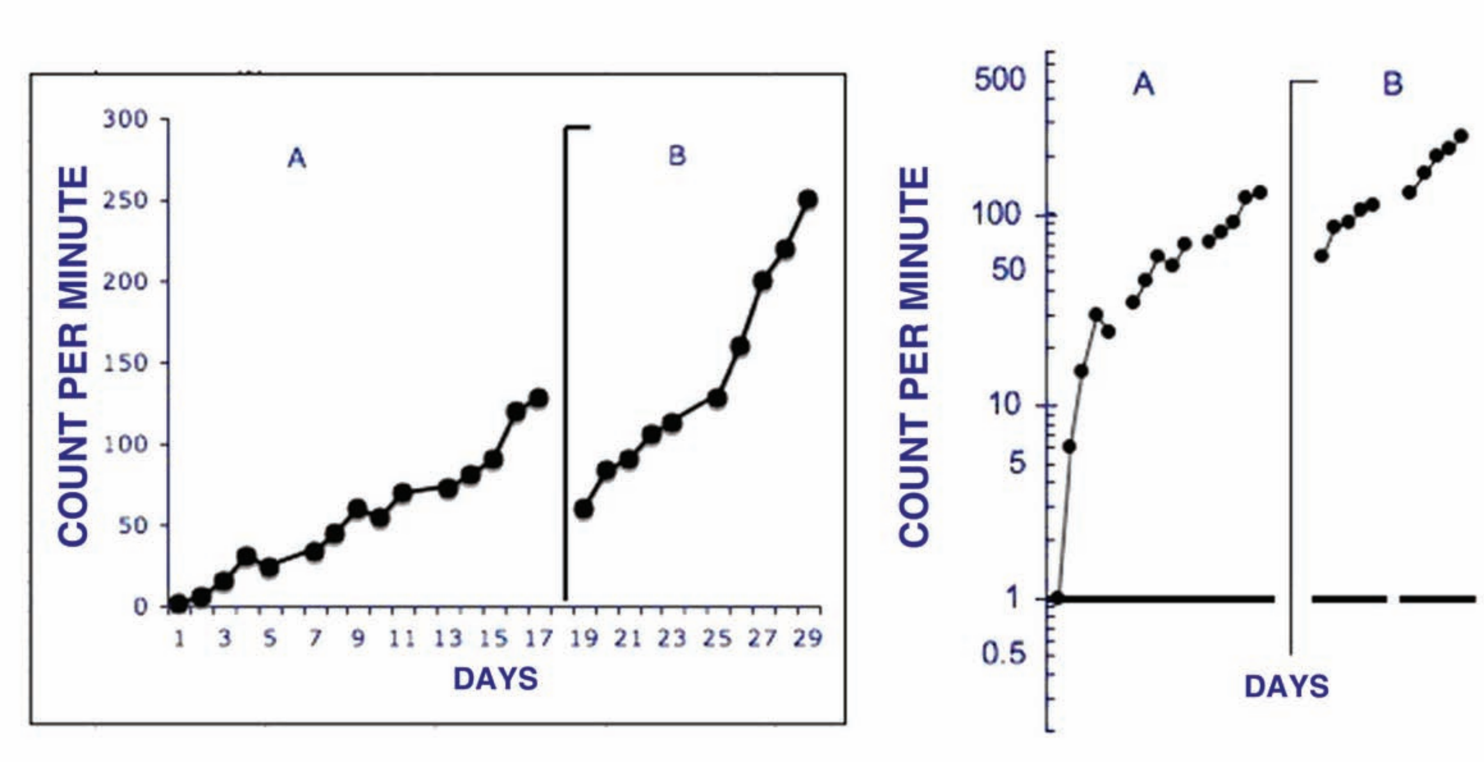


Figure 6. An illustration of how the same sets of data may look different when plotted on an equal-interval graph (left) and when plotted on a Standard Celeration Chart (right).

liers were multiplied by a constant in the data from one of the panels to accentuate its variability.

Such differences in how data appear on both types of graphs used here arise from inherent differences in how those two graphs are constructed. By design, the SCC uses a multiplicative scale along its ordinate. As a result, data displayed on the SCC move in equal distances either up or down the SCC's ordinate by factors rather than by the addition or subtraction of equal amounts. Data plotted on the SCC may then look different from data plotted on an equal-interval scale because vertical change is proportional on the SCC and not on an equal-interval scale. Figure 6 shows an example of how different display methods can show the same data sets quite differently.

Differences in how graphed data appear on a display may (and often should) lead behavior analysts to make different decisions about treatment. These differences may cause behavior analysts to overlook important information about client progress. A client may be progressing at a rate that is too slow, and the behavior analyst might miss opportunities to provide more efficient intervention. For example, in the equal-interval graph in Figure 6, Panel B shows more change per unit of time. The same data plotted on the SCC appear to show more change per unit of time on Panel A. But if the behavior displayed in Figure 6 is in fact changing at a faster pace in Panel A than in Panel B, a behavior analyst using an equal-interval graph may make unnecessary changes to the intervention when in fact the client's performance is progressing quite rapidly. Such changes may inhibit client progress, and behavior analysts may avoid such errors in judgment simply by selecting a display type that more accurately represents the data.

Limitations

Although a concerted effort went into selecting data from real-world clinical settings to promote extension of the current findings to such situations, the method of display of the graphs in the Web panel deviated from standard clinical practice in an important way that may have affected how participants interpreted the data they saw. On each type of graph, days were plotted on the abscissa, and the tables provided temporal information about these dates when the measures were recorded. Panels A and B were presented on the same timeline, as if Panel B occurred after Panel A in time. Panel A started with day 1, and Panel B started on the day following the last data point in

Panel A. Although the data were presented on the same timeline, the directions stated that the two procedures were independent of one another, as were the target behaviors.

In actual clinical settings, data for two different behaviors may not be plotted on the same graph. Thus, the presentation of two different behaviors on the same graphic display may have affected the participants' responses on assumptions they may have made about the data presented in Panel B because of the location of the data relative to Panel A. For example, if Panel A ended on day 9, and Panel B started on day 10, participants may have assumed that the data in Panel B represented a change in the independent variable used for the same behavior. Such an assumption may have led participants to select an incorrect response because they assumed that the behavior had already been acquired in the previous phase.

Additionally, practice in interpreting data on each type of graph may have had a facilitating effect on accuracy and efficiency of responding as participants moved through the survey. Because the equal-interval graphs, SCCs, and tables were presented consecutively rather than randomly interspersed, performance on earlier stimuli—all equal-interval graphs—may have been slower than performance on later stimuli because of, at least in part, the presentation order for the stimuli. For example, the accuracy of responses to questions presented toward the end of each section may have been higher because of practice and the similarity of stimuli on the previous pages. Further, fewer accurate responses toward the beginning of each section may be correlated with the presentation of a new display type that the participant had no recent practice with. The decreasing trend of inter-response time within the equal-interval condition, as in Figure 5, and no trend seen in the table and the SCC conditions suggests that possible practice effects may have influenced the participants' performance. Considerations for future research may include having an initial set of practice stimuli for participants before the test stimuli are presented, and random interspersing of the display methods throughout the survey.

The current study used a nonrandom sample. The degree to which any sample, regardless of details about how it was selected, can provide meaningful information about a larger population depends on the degree to which it matches its referent population along important parameters. In the current study, participants received invitations to participate through a nonrandom selection procedure in which every ninth certificant listed in alphabetical order in each state within the U.S. was selected and invited to participate.

Differences in training programs and procedures across the country may be correlated with the behavior analyst's ability to interpret data when those data are shown on varying graphic displays. Different data-based decision rules may also affect how a behavior analyst interprets data. Therefore, geographical location may have been correlated with a participant's ability to interpret data. For example, this ability may be due to a behavior analyst's proximity to training centers or other behavior analysts.

The sampling procedure used here may have caused participants whose last names begin with certain letters to be selected depending on the number of Board-Certified Behavior Analysts in each state. Although this may have been the case, the first initial of the participant's last name is not likely to have any correlation with the participant's ability to interpret data. Another implication resulting from the sampling procedure is that the sample used did not match the referent sample in terms of the type of display the participants typically used. The current study had a higher percentage of behavior analysts who typically use SCCs in clinical practice, and those participants may have more experience with interpreting data that is displayed on SCCs. This may be responsible for the higher correct response frequencies for data plotted on the SCC. Future researchers may wish to include a stratified random-sampling procedure in selecting participants who proportionally represent the distribution of behavior analysts across the U.S. and are likely to regularly use equal-interval graphs and SCCs.

Future directions for research in this area include addressing the previous limitations, expanding the sample of participants and the size of the sample stimulus sets, and potentially more closely examining the variables that are controlling participant responding, and how these types of displays may effect responding.

Importance of Efficient and Accurate Analysis

Visual inspection of graphed data remains the most commonly used method of data interpretation in applied behavior analysis (Cooper et al., 1987). The field is distinguished from other behavioral sciences because of its reliance on graphic analysis and its ability to deliver effective and efficient services to clients who depend on these data analyses. Therefore, the methods used for interpretation of data must permit the most accurate and efficient interpretation reasonably possible in most circumstances. Such accurate and efficient analyses will help ensure that clients receive the best treatment in the most efficient manner possible. With efficient methods of analysis, behavior analysts can spend more time with clients rather than spending unnecessary time on data analysis. Inaccurate analysis interferes with client progress and wastes time and resources. Choosing a display method that assists with accurate analysis may prevent unnecessary or counterproductive effort on a client's behalf and may keep clients happy and progressing quickly toward their goals.

REFERENCES

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions in applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Barrett, B. H., Beck R., Binder, C., Cook, D. A., Engelmann, S., Greer, R. D., et al. (1990). *Statement on students' right to effective education*. The Association for Behavior Analysis Task Force. Retrieved November 2, 2009, from http://www.abainternational.org/ABA/statements/education.asp

Behavior Analysis Certification Board™. (2004). *Guidelines for responsible conduct for behavior analysts.* Retrieved March 18, 2008, from http://www.bacb.com/maint_frame.html

Cooper, J. O., Heron, T.E., & Heward, W. L. (1987). *Applied behavior analysis.* Columbus, OH: Merrill Publishing.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis, 12,* 573–579.

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.

Harder, S. R. (2008). Standard Celeration Chart: Daily Per-Minute, Version 6-0C(20). Freeware charting template, archived at: http://lists.psu.edu/cgi-bin/wa?A2=ind0805&L=SCLISTSERV&T=0&F=&S=&P=1580

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of human behavioral research.* Hillsdale, NJ: Erlbaum.

Kazdin, A. E. (1976) Statistical analysis for single case experimental designs. In M. Hersen & D. H. Barlow (Eds.), *Single case experimental designs. Strategies for studying behavior change* (pp. ???–???). New York: Pergamon.

Lindsley, O. R. (1971). Precision teaching in perspective: An interview. *Teaching Exceptional Children, 2*, 114–119.

Lindsley, O. R. (1991, Spring). Precision teaching by teachers for children. *Teaching Exceptional Children*, pp. 10–15.

Mawhinney, T. C., & Austin, J. (1999). Speed and accuracy of data analysts' behavior using methods of equal interval graphic data charts, standard celeration charts, and statistical process control charts. *Journal of Organizational Behavior Management, 18*(4) 5–45.

Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T.R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 101–165). New York: Academic Press.

Stevens, J. C., & Savin, H. B. (1962). On the form of learning curves. *Journal of the Experimental Analysis of Behavior, 5*(1), 15–18.

Tufte, E. R. (1993). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Van Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foxx, R. M., Iwata, B. A., & Lovaas, O. I. (1989). *Statement on the right to effective behavioral treatment*. The Association for Behavior Analysis Task Force. Retrieved DATE, from http://www.abainternational.org//ABA/statements/treatment.asp