Communicating Experimental Findings in Single Case Design Research: How to Use Celeration Values and Celeration Multipliers to Measure Direction, Magnitude, and Change of Slope

Shawn M. Datchuk and Richard M. Kubina, Jr.
The Pennsylvania State University - University Park

The accumulation of scientific knowledge greatly depends upon the critical review of experimental findings by ones peers. In single case design research, experimenters present findings with graphical displays of data and narrative description of a visual analysis. To aid in efficient and accurate description of experimental findings, the research community would benefit from use of the Standard Celeration Chart and two important metrics, celeration value and the celeration multiplier or celeration turn. The present article explains the importance of embracing standard graphic displays depicting proportional distance and provides step-by-step directions to calculate both metrics. Also the results section contrasts celeration value and celeration turn with the commonly used metric of percentage of nonoverlapping data.

Many scientific discoveries, such as the foundational concept of reinforcement (Skinner & Ferster, 1957) and elements of effective instruction (Christensen, Ysseldyke, & Thurlow, 1989), have stemmed from single case design (SCD) research. SCDs facilitate detection of a functional relationship between the introduction of an independent variable and changes in a dependent variable (Johnston & Pennypacker, 2009; Sidman, 1960). Using only a single or small number of participants, various SCDs (i.e., multiple-baseline and alternating treatment) can inform and hasten the development of scientific theory generalizable to larger populations (Horner et al., 2005).

The power of SCD research lies in detection and replication of experimental effects. Replications occur when introduction or removal of an independent variable repeatedly results in a change in the dependent variable (Kazdin, 2011). For example, if Participant A improved following intervention, then a replication occurs when Participant B shows comparable experimental effects. The more replications demonstrated, the more confidence engendered by the intervention for its application to a wide range of people.

SCD experimenters typically detect experimental effects and subsequent replication through the visual inspection of graphic displays for changes in level, trend, and variability within and between experimental phases (Kennedy, 2005). Level refers to the mean of data within the phase. Trend indicates the slope of data as increasing, decreasing, or remaining flat. The trend also has a magnitude usually expressed in qualitative terms: high, medium, and low (Kennedy). Variability occurs when data points vary from one another along a trend. Variability also finds its expression in qualitative terms with high, medium, and low.

To report experimental results in SCDs, the scientific literature has a rich tradition of relying on visual analysis of graphed data and a narrative description of the visual analysis, typically found in the results or discussion section of research articles. Graphed data in SCD visually depict a complex statistical process in participant behavior, such as changes in level, trend, and variability (Kennedy, 2005). Many authoritative texts provide evidence demonstrating visual analysis directly and efficiently communicates experimental findings and facilitates an open dialogue with the larger scientific community (Baer, 1977; Cooper et al., 2007; Gast, 2010; Kazdin, 2011; Michael, 1974; Parsonson & Baer, 1978, 1986; Sidman, 1960).

Graphic display of results and the accompanying narrative of the study allow experimenters to clearly communicate their interpretation of the data to the scientific community and parse out competing hypotheses (Cooper et al., 2007). The scrutiny of scientific peer review serves as one of the hallmarks of scientific discovery, resulting in a highly constrained and conservative accumulation of knowledge (Sagan, 1997). The public dialogue created by visual analysis serves as a strength of SCDs. Even with the fruitful track record of SCD, certain practices within the visual analysis framework may lead to imprecise and inefficient communication to the larger scientific community.

Graphic displays tend to vary across studies meaning the distance between rulings and other idiosyncrasies can influence interpretation of data (Kubina & Yurich, 2012). Research has also shown variance among how experimenters interpret visual data (DeProspero & Cohen, 1979; Franklin, Gorman, Beasley, & Allison, 1996; Gibson & Ottenbacher, 1988; Gottman & Glass, 1978; Hagopian, Fisher, Thompson, Owen-DeSchryver, Iwata & Wacker, 1997; Knapp, 1983; Ottenbacher, 1986; Wampold & Furlong, 1981). Additionally, limitations of visual analysis include a propensity to Type I errors, lack of universal decision rules (Campbell & Herzinger, 2010). Given variability with visual analysis among researchers, it is incumbent upon the scientific community to explore and implement superior visual displays and embrace metrics that provide the clearest and most information rich presentation of experimental findings.

The Standard Celeration Chart

One type of graphic display, the Standard Celeration Chart (SCC), allows chart readers to quickly and reliably interpret salient features of data such as trend, variability, and level because it features standard axes and rulings. Figure 1 displays a reproduction of the SCC. The creator of the SCC, Ogden Lindsley, discussed how nonstandardized graphs could negatively impact visual communication, "The teachers shared their progress on these behavior change projects by showing charts in class each week. It took 20 to 30 minutes to share one behavior project because most of this time was spent describing each teacher's unique charting and recording system" (Lindsley, 1990, p. 11). By using a standardized visual display, chart readers can instead focus on interesting characteristics of the data rather than decoding and analyzing chart construction and encounter possible artifacts created by idiosyncratic design features.

The SCC offers a number of advantages when describing behavior (for a full description see Graf & Lindsley, 2002; Kubina & Yurich, 2012; Lindsley, 2005; McGreevy, 1983; Pennypacker, Gutierrez, & Lindsley, 2003; West, Young, & Spooner, 1990; White & Haring, 1980). The horizontal axis displays universally accepted units of time: calendar days, weeks, months, or years depending on the chart. Maintaining the actual units of time present when the experimenters originally measured the behavior faithfully portrays a vivid and accurate depiction of a behavior changing over time. The proper graphical display of time tells experimenters whether observed changes occurred due to the independent variable or maturation, a historical record of growth over time. Using non-calendar time, such as sessions, obscures the view of how behavior unfolded in time and potentially increases error attributable to poorly constructed graphics.

In addition to maintaining authentic units of time along the horizontal axis, the vertical axis can accommodate a full set of human behavior frequencies, from 1 per day up to 1000 per minute. The daily SCC as shown in Figure 1 has six cycles starting at .001 and ending at 1000. The large range along the vertical axis shows proportional distance with a semilogarithmic or ratio scale.

The SCC creates proportional distance with the vertical axis scaling, decreasing distance

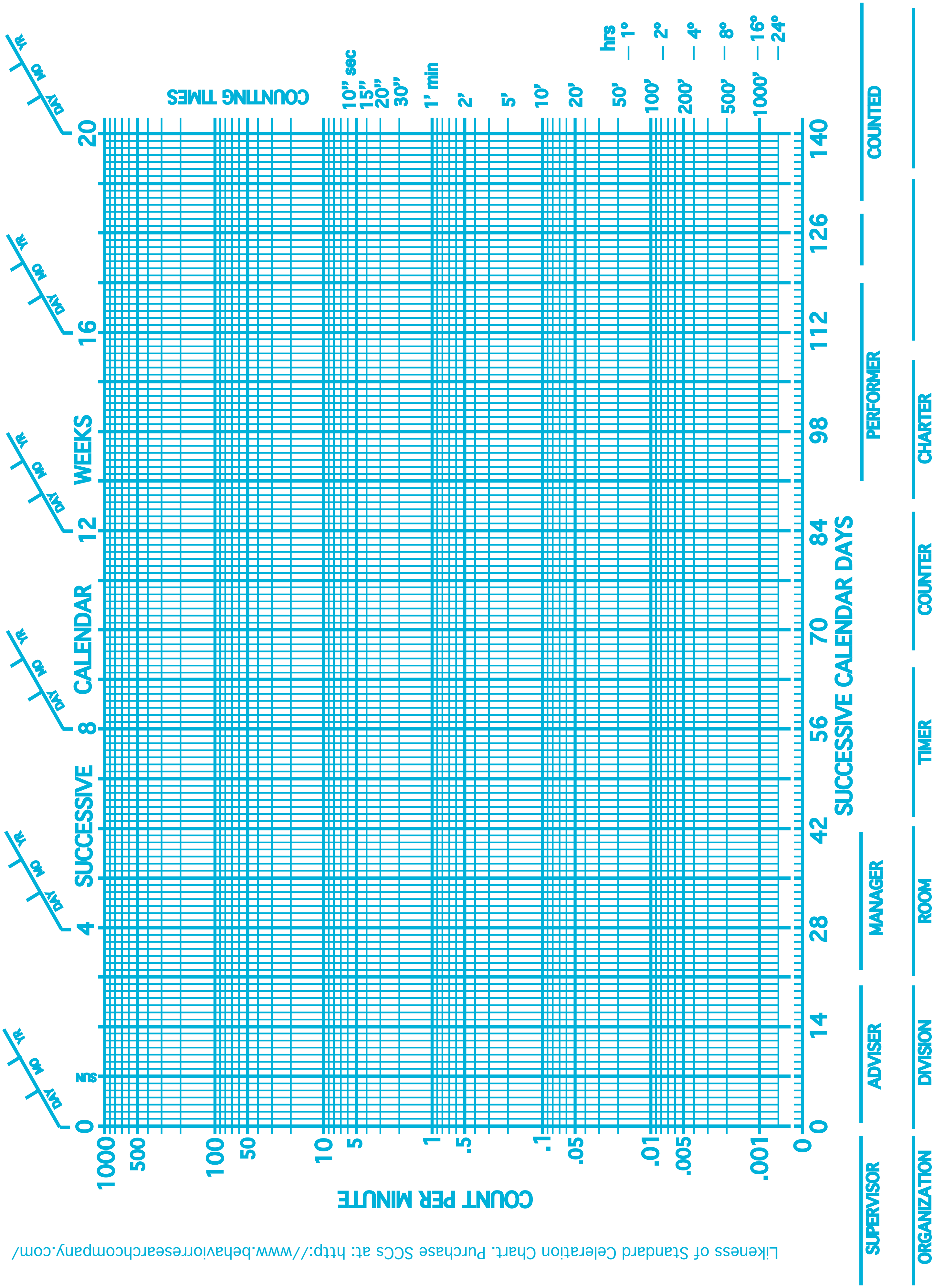

Figure 1. Likeness of daily Standard Celeration Chart.

between rulings compared to equal interval charts (Schmid, 1986). The distance between one and two has the same amount of space as going from two to four because both distances reflect a doubling of quantity (i.e., a factor of x2). Graphic displays with equal interval axes do not show proportional distance and allocate a third more distance when moving from two to four. The increased distance between rulings may inappropriately influence chart readers to conclude a more robust experimental effect. Given the smaller space between rulings, the SCC produces a conservative display of data and guards against experimenters attributing experimental effects to the distance between rulings rather than distance between data points due to intervention (Binder, 1996; Kubina & Yurich, 2012; White, 1987).

While the SCC has many other benefits, the present paper offers a rationale and a method for using the SCC to calculate a measure for changes in slope within phases, known as the celeration value, and comparing two celeration values across phases, called the celeration multiplier (Pennypacker, Gutierrez, Lindsley, 2003), also referred to as a celeration turn (Graf & Lindsley, 2002). Calculating the celeration multiplier on the SCC assists learners, parents, teachers, speech communication and language therapists, occupational therapists, school psychologists, school administrators, behavior analysts, and researchers who need to efficiently and effectively communicate and quantify findings to an audience of peers, professionals, interested chart readers, or the scientific community.

Quantifying Findings from SCDs

Beyond the celeration value and celeration multiplier, a larger discussion has enveloped statistics and the quantification of SCD experimental effects. An argument for statistics as desirable supplements for visual analysis rests on enhanced communication and uncovering of true experimental effects. In other words, descriptive narration for what the experimenter sees during visual analysis is supported or not supported by quantitative analysis. Statistics proposed in prior research include standard parametric tests (Gentile, Roden, & Klein, 1972; Shine & Bower, 1971) effect size (Corcoran, 1985; Gingerich, 1984; Gorman-Smith & Matson, 1985), regression analysis (Allison & Gorman, 1993; Busk & Serlin, 1992; Center, Skiba, & Casey, 1985-1986; Glass, Willson, & Gottman, 1975; White, Rusch, Kazdin, & Hartmann, 1989), randomization tests (Edgington, 1980, 1992), and application of meta-analytic techniques within individual studies (Busse, Kratchowill, & Elliot, 1995). Application of such statistics received criticism for a variety of reasons; (a) inappropriateness of SCD data to meet assumptions of independent measures, (b) autocorrelation of data (Busk & Marascuilo, 1988; Matyas & Greenwood, 1991; Sharpley & Alavosius, 1988), (c) complexity of some derived measures and lack of specific guidelines to select measures, (d) inadequate number of data points within studies (Huitema, 1985, 1986), and (e) potential for statistical inflation.

One statistic, the percentage of nonoverlapping data (PND), has garnered support among some researchers to quantify experimental effects and as a method to aggregate findings from multiple studies (Scruggs & Mastropieri, 2001; Scruggs, Mastropieri, & Casto, 1987). The PND statistic quantifies the amount of nonoverlapping data between baseline and experimental conditions (e.g., if 3 of 6 data points do not over lap, then the percentage of non overlap equals 50 percent), theoretically measuring change in level and variability across experimental phases. PND also avoids many of the critiques raised against prior statistics because it provides a measure of tangibility or convincingness rather than effect size (Scruggs, 1992; Scruggs, Mastropieri, & Castro, 1987). Scruggs and Mastropieri (1994) developed general interpretational guidelines; a PND greater than .70 indicates a robust effect, greater than .50 but below .70 demonstrates questionable effectiveness, and a PND below .50 means no effect. Despite its popularity, several limitations of the PND statistic have surfaced. Allison and Gorman (1994) found PND effected by the number of data points collected: PND decreases as the number of data points collected increases. Allison and Gorman's report also restated concerns regarding insensitivity to changes in slope and level and oversensitivity to atypical events such as outliers in baseline.

If research community continues to use PND to describe experimental effects, what additional measures might reduce some of its limitations? Parker, Hagan-Burke, and Vannest (2007) proposed the percentage of all non-overlapping data (PAND) as an alternative measure. The PAND statistic has several advantages, such as its use of all data points to calculate overlap, its easy application to multiple baseline designs, and its conversion to the widely used Pearson's Phi and Phi2. PAND, however, also has several shortcomings. PAND does not alleviate an insensitivity to detect change at the upper end of the scale and it does not account for slope in baseline. A recent study examined four overlap methods, including PND and PAND, by comparing the quantitative synthesis to visual analyst' judgments. The results showed all of the overlap methods had high levels of errors prompting the study authors to conclude statistics such as PND and PAND "should be abandoned" (Wolery, Busick, Reichow, & Barton, 2011).

An alternative or supplement to PND and PAND, the celeration value and celeration multiplier provides a more precise measure of slope within and between experimental phases. The celeration value and the celeration multiplier are quantitative measures describing different changes. On a daily Standard Celeration Chart, the celeration line quantifies the slope by stating how much the data set grew (i.e., multiplied) or decayed (i.e., divided) per week. The celeration value describes the direction and degree of change of a single celeration line with a quantity that has a multiplication or division sign. For example, a celeration line beginning at a data point of 15 and ending at a data point of 30 represents a celeration value of x2.0 or a weekly doubling. The celeration line also has a bracket with the amount of days distinguishing the time frame from one celeration value to the next (Kubina & Yurich, 2012). As an example, a celeration of x1.4 occurring over 23 days would appear as x1.4 [23 days].

The celeration value provides descriptive and quantitative information for a single phase of data. The celeration multiplier (steps to calculate appear later) provides the direction and degree of change between one celeration value and another. If a celeration in one phase comes to x2.1 [14 days} and then next phase has a celeration value of x1.35 [9 days], the celeration multiplier equals ÷1.55. The celeration multiplier communicates the comparison celeration (i.e., second celeration) has divided or turned down 1.55 times compared to the reference celeration (i.e., first celeration value).

A number of behavioral scientists (e.g., Kazdin, 1976; Graf & Lindsley, 2002; Kubina & Yurich, 2012; Lindsley, 2005; Pennypacker et al., 2003; White, 1974) have encouraged the use of the SCC to quantify changes of slope within and between experimental phases. A study by Mason (2010) confirming the importance of celeration values of specific phases and the subsequent celeration turn (or celeration multiplier) found "celeration and celeration change are independent evaluations of single-subject research, which measure an effect that is entirely unrelated to PND…" (Mason, 2010, p. 10). To explain the benefits of celeration values and celeration multipliers, White (2005) suggested recharting extant data on the SCC. The following sections explain the steps of determining a celeration line, celeration value, and celeration multiplier, then charts extant data onto SCCs to illustrate potential benefits.

Steps to Calculate Celeration Value and the Celeration Multiplier

Calculating celeration change measures require four steps. First, to chart or rechart data on a SCC the conventions for data include the use of dots to represent acceleration targets and X's for deceleration targets (Graf & Lindsley, 2002). The figures in the present manuscript come from a software drawing program. Most researchers and practitioners use paper Standard Celeration Charts available at the Behavior Research Company (BRCO) (i.e., http://www.behaviorresearchcompany.com/). However, other computer options have recently become available found at the BRCO website and elsewhere on the web, see http://precisionteaching.pbworks.com/

The second step for calculating celeration measures calls for determination of a celeration line. While various options exist (e.g., focus line method, freehand method), for research and producing consistency we recommend using the split-middle technique (Kazdin, 1976, 1982; White, 1974). To use the split middle, data must be divided in half by a vertical line (Figure 2). An equal number of data points must fall on each side. Next, each side is further divided by another vertical line. The data on the left and the right are halved. An equal number of data points must again fall on each side of the line. Then, the median data points for the first and second halves receive a horizontal line that intersects the small vertical line of each half. A celeration line is then used to connect the intersections. Finally, the celeration line requires adjustment until 50% of the data points fall on or above the line and 50% fall on or below the line. The slope of the trend does not change during adjustment.

When graphing data, preserving any days or time units between data points helps when deciding between split-middle and quarter-intersect methods for determining slope. The split-middle technique can provide a reasonable approximation of the data trend. The quarter-intersect method may provide a more reasonable approximation when the data set contains seven or fewer data points and when it contains large, unevenly dispersed gaps (White, 2005). The quarter-intersect technique resembles the split-middle except for two differences. The time units or calendar days along the horizontal axis determines placement of the vertical half lines regardless if an equal number of data points fall on each side. All time units or calendar days between the first and last data point count towards the calculation. Also, the quarter-intersect line is not adjusted to ensure 50% of the data points fall on or above the line and 50% fall on or below.

The third step for calculating the celeration value focuses on the celeration line. Because the SCC contains standard axes, the slope of the celeration line in relation to the axes determines the celeration value. Kazdin (1976, 1982), White (1974) and Pennypacker et al. (2003) recommend detailed procedures to determine celeration value. To start, a point on the horizontal axis where the celeration line rests, along with its position along the vertical axis, serves as an arbitrary starting point (day X). Figure 3 displays data recharted from Alison and Gorman (1993); on the first day of the phase (day X), the celeration line rests at 12 along the vertical axis. Next, the steps require selecting another point. On day 10 (x + 10), the celeration line rests on 2 along the vertical axis. Dividing the numerically larger value by the smaller value determines the celeration value. A multiplication sign (x) indicates an accelerating trend and a division sign (÷) indicates a decelerating trend. Another method for finding celeration values involves using a "Finder" (Kubina & Yurich, 2012; Pennypacker et al., 2003). Figure 3 displays a decelerating trend

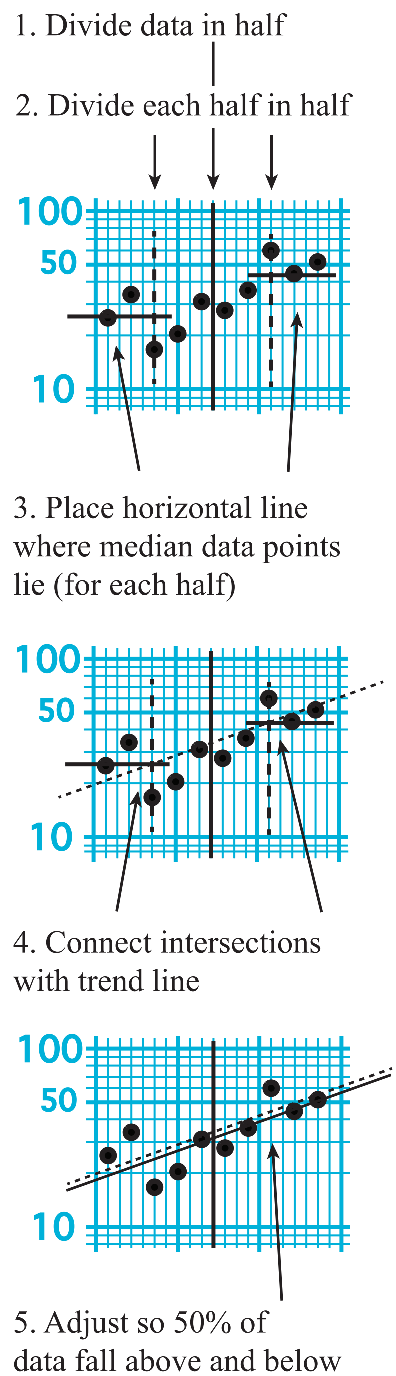

Figure 2. Five steps for drawing a celeration line with the split-middle technique.

of ÷6, or a ÷6 celeration line, during the first phase (i.e., celeration 1) and an accelerating trend of x6 during the second phase (i.e., celeration 2).

The fourth step for determining the degree of change between celeration values, known as the celeration multiplier, entails contrasting two celeration lines with measured celeration values. The reference celeration, of the first celeration, and the comparison celeration, the second celeration follow one of two rules depending on the values of the celeration. The rule for when two celeration values have the same trends or signs, both accelerating (i.e., x) or decelerating (i.e., ÷), is to take the larger celeration value divided by the smaller (Pennypacker et al., 2003). For instance, if the first phase, or reference celeration, had a celeration value of x3.0 and the second phase, or comparison celeration, of data has a celeration value of x5.5, x5.5 would be divided by x3.0 = x1.83 (affix sign of change contingent upon whether the second phase improves or worsens).

If celeration values have different trends or signs, one accelerating and the other decelerating, then the rule says to multiply the two values together and use the sign of change. Therefore, if the first celeration line (phase one) is dividing and the second celeration line (phase two) is multiplying the sign attached would be a multiply sign because the behavior grew or turned up in the second phase, (e.g., ÷1.8 and x2.4 would be ÷1.8 times x2.4 = x4.32). A multiplication sign (x) or a division sign (÷) always indicates an accelerating or decelerating change in slope relative to the preceding celeration (Graf & Lindsley, 2002; Kazdin, 1976, 1982; Pennypacker et al., 2003). Figure 3 shows a celeration multiplier, a x36 turn up. The first phase has a ÷6 celeration while the second phase has a x6 value, therefore 6 x 6 = 36. The celeration "turned up," therefore, the terms "turn up," "turn down," and "no turn" contextualize the change between the two celerations.

Recharted Data on SCC

Figures 3 to 6 display data previously used in the PND critique literature (Allison & Gorman, 1993; Scruggs, Mastropieri, & Castro, 1987). Each figure features a celeration line, celeration value, celeration multiplier, and PND statistic. The split-middle technique was used to determine the trend line. Alison and Gorman (1993) presented Figure 3 as an example of the insensitivity of PND to account for change in slope. The data in Figure 3 has a PND value of 0 because no data in the second phase overlap the data in the first phase. The x36 celeration turn, which indicates an exceptionally massive change in slope not detected by PND, further speaks to the limited power of PND.

Figure 4 shows data from Scruggs, Mastropieri, and Castro (1987). The data show the effect of a gradual change in slope, rather than a drastic change, on the PND statistic. The data has a PND value of .50, generally regarded as a questionable effect; however, the celeration multiplier of x3.0 turn up indicates an impactful change. Figure 5 presents more recharted data from Scruggs, Mastropieri, and Castro (1987). The figure notes the effect of an inappropriate baseline slope on the PND statistic. In this case of Figure 5, the PND statistic does not detect the accelerating slope in baseline. The data has a PND value of .83, yet, the celeration multiplier of a ÷1.7 turn down means the second phase produced a diminishing effect on the data in the second phase.

An ABAB design appears in Figure 6 (Scruggs, Mastropieri, & Castro, 1987). The data has an overall PND value of .45 indicating no noticeable effect, but the celeration multiplier between the first AB (i.e., celeration 1 and celeration 2) turns down very modestly by ÷1.1, turns down more significantly from celeration 2 to celeration 3, and significantly turns up, x6.1., in the last AB phase (i.e., celeration 3 to celeration 4). The celeration multipliers provide the clarity of precise numbers necessary for the quantitative evaluation of changes from phase to phase. Using celeration value for a single phase and the celeration multiplier for comparisons between phases inspires analytical confidence, thereby reducing the need to rely on a single statistic like PND.

Summary

The critical review of experimental findings by ones peers refines interpretation of results and leads to the accumulation of scientific knowledge. It is incumbent upon experimenters to add clarity to their findings by embracing techniques that precisely and efficiently display their data and quantity their results. The SCC along with celeration values and celeration multipliers enhance analysis of applied and experimental findings and communication to the scientific community. The present paper described some of the advantages of using the SCC, explained how to calculate two celeration measures, and contrasted celeration measures with another commonly used metric, PND (Scruggs & Mastropieri, 2001).

Users of SCCs benefit from a standard graphic display thereby facilitating the interpretation

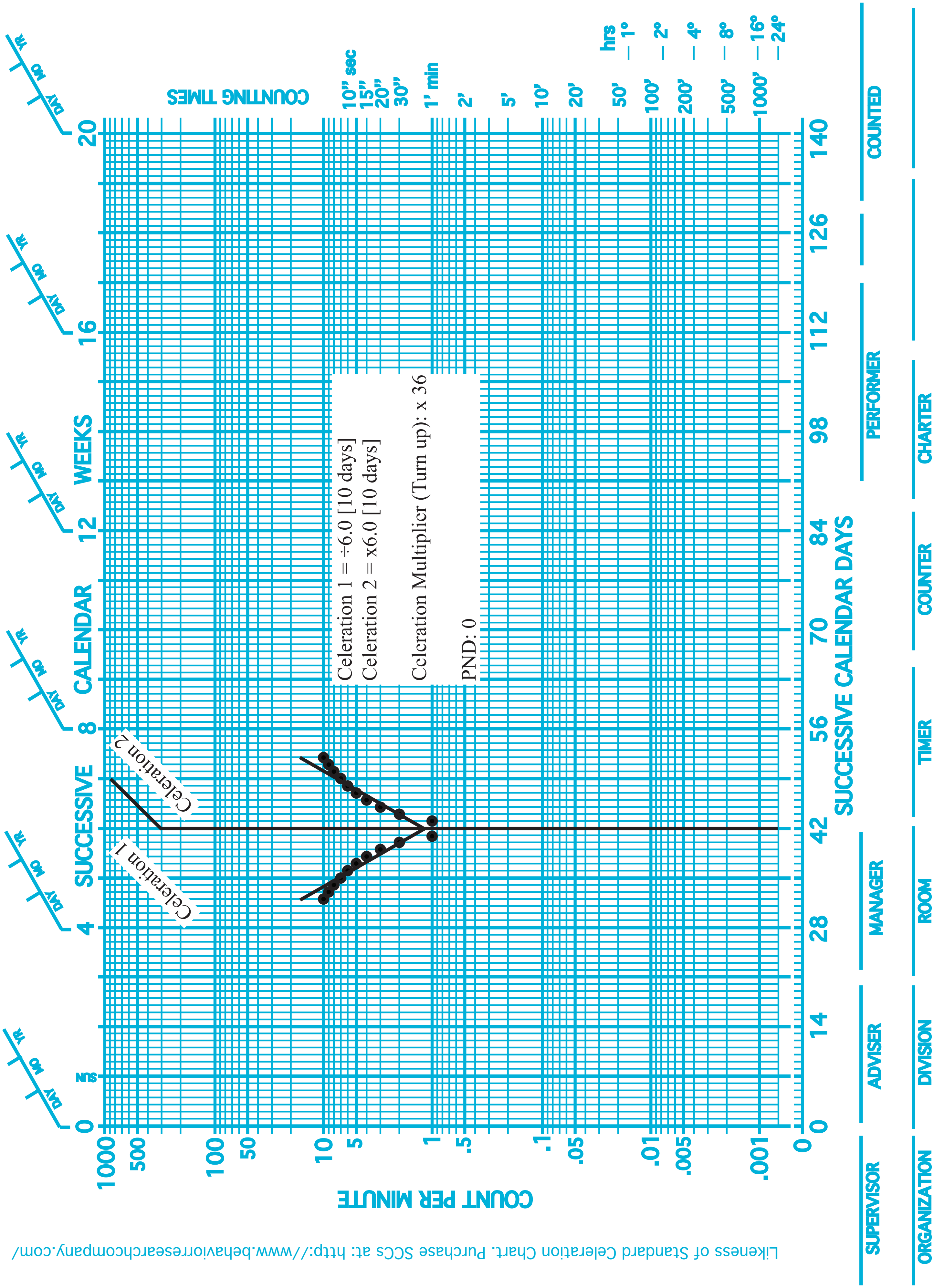

Figure 3. Contrary trends with a resulting PND of 0.

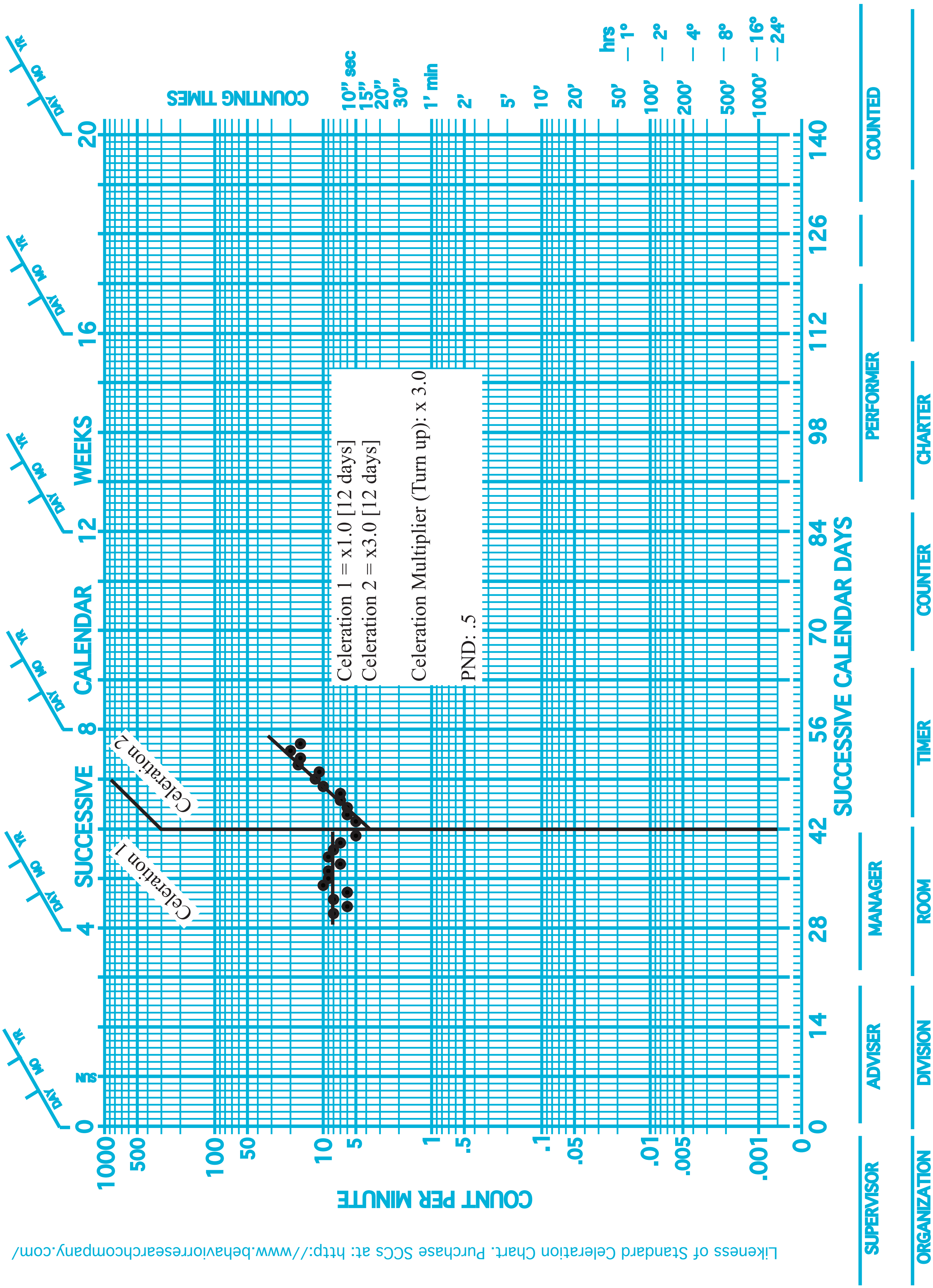

Figure 4. Recharted data from Scruggs, Mastropieri, and Castro (1987).

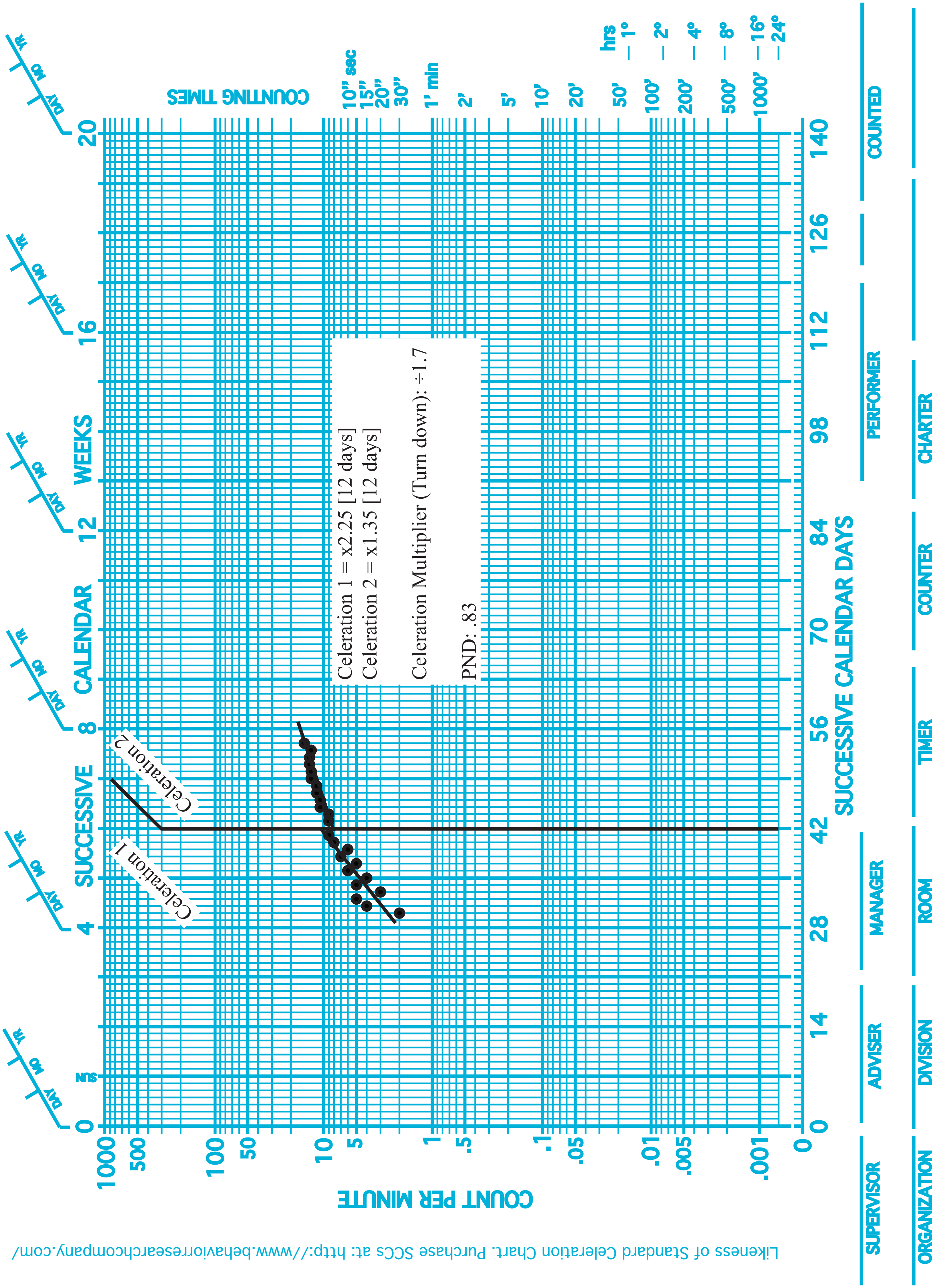

Figure 5. Additional recharted data from Scruggs, Mastropieri, and Castro (1987).

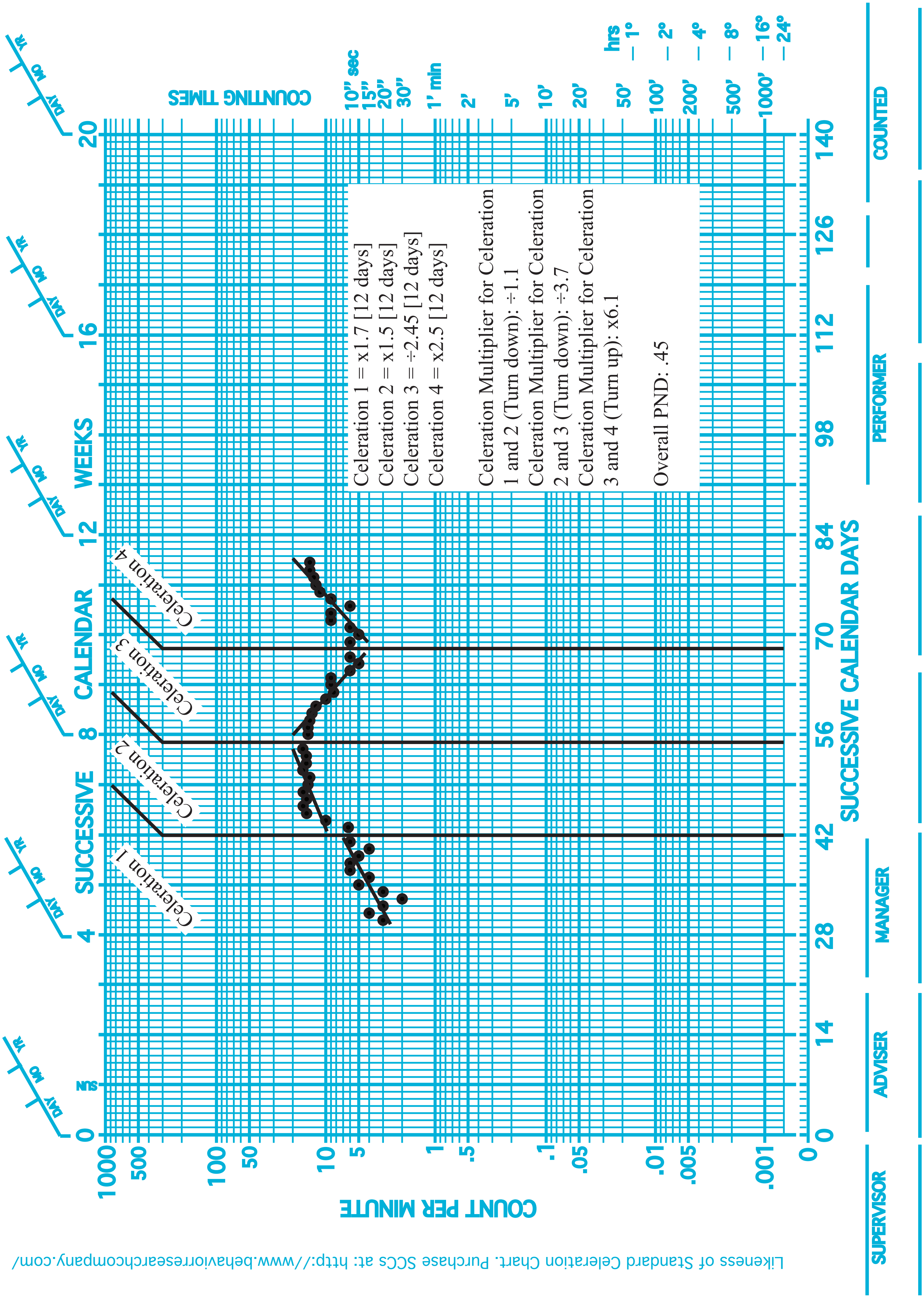

Figure 6. Recharted data from Scruggs, Mastropieri, and Castro (1987) showing an ABAB design.

of results and guarding against error attributable to graphics with limited analytical power. Commonly used nonstandard graphics, particularly poorly constructed equal interval charts, may misrepresent data and lead to inaccurate conclusions of experimental effects. The axes and multiply/divide nature of the SCC places emphasis on standard communication by researchers interested in truly significant changes in dependent variables.

In addition to graphic displays, quantifying the slope with the celeration value and comparing changes with the celeration multiplier further illuminate experimental findings. Researchers have embraced the PND statistic but examples from the present article, along with several other critiques (Allison & Gorman, 1994; Wolery et al., 2011) demonstrate its insensitivity to slope. Results from the present paper suggest researchers reporting PND should also report measures of slope or celeration to prevent inaccurate data analysis.

The celeration value offers something no slope of a line on an equal interval scaled chart can ever offer - a unit of measurement quantifying behavior change as count over time over time. Having celeration as a unit of behavior change places chart readers on the same level as other natural sciences that routinely quantify change with standard, absolute, universal measures (Pennypacker et al., 2003). When physicists wish to measure stress or pressure they use a unit called a pascal. A photometricist, a scientist who studies the measurement of light, evaluates illuminance with a lux. And many scientists who study electromagnetism, sound, and computing measure frequency with the hertz. With the advent of the SCC practitioners and researchers in education and psychology can measure change with celeration and change between phases with the celeration multiplier.

Not only does celeration elevate the science of education and psychology by providing a standard unit for change, all chart readers see a x2.0, ÷3.5, or a x1.25 in the same way. Celeration lines have standard angles of change with their associated numeric quantity expressing the precise magnitude of weekly change (weekly change only for a daily SCC: celeration for weekly charts show monthly change, celeration for monthly charts show sextuple monthly change, and yearly charts show quintuple yearly celeration). Therefore, when analyzing data within a phase SCC readers have visual and quantitative information communicating the effects of the presence or absence of the independent variable on the dependent variable.

The celeration multiplier plays a vital role when analyzing the change in celeration from one phase to another. Going from baseline to an intervention phase, for example, constitutes an instance where a quantity clearly communicates the magnitude of change. The celeration multiplier combines with turn information: celeration turns in one of three ways, up, down or no change (said turn up, turn down, no turn; Graf & Lindsley, 2002). A x1.2 celeration in baseline and a x2.4 celeration in an intervention phase comes to a x2.0 turn up. With the celeration multiplier researchers no longer need to search for statistic to quantify findings in single case design research.

Despite the effectiveness of the celeration value and celeration multiplier to serve as alternate statistics, a note of caution is warranted. The split-middle and quarter-intersect methods provide a trend or celeration of the data by serving as an approximation of the median slope of the data. Both methods contain susceptibility towards error, such as when graphs contain too few data points, heteroscedasticity of data, and the appropriateness of a single trend line to summarize complex data patterns more suited to multiple trend lines. The multiply/divide nature of the SCC reduces the chance for different types of error but does not eliminate it completely.

Also, researchers should place the values in context when interpreting celeration multiplier data because general interpretational guidelines have not been established. Future research should address significance concerns and also extend application of the celeration multiplier from AB designs to different design types, such as multiple baseline and multiple phase designs. The data presented in this paper, along with the PND critique literature, demonstrate a need to use efficient standard graphics and account for slope within single subject studies. The SCC and celeration measures provide a precise way to display data and quantify results. Quantification and the resulting command of the subject matter always leads to better science.

References

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research & Therapy, 31*, 621-631.

Allison, D. B., & Gorman, B. S. (1994). "Make things as simple as possible, but no simpler": A rejoinder to Scruggs and Mastropieri. *Behaviour Research and Therapy, 32*, 885-890.

Baer, D. M. (1977). Perhaps it would be better not to know everything. *Journal of Applied Behavior Analysis, 10*, 167-172.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163-197.

Busk, P. L., & Marascuilo, L. A. (1988). Autocorrelation in single-subject research: A counterargument to the myth of no autocorrelation. *Behavioral Assessment, 10*, 229-242.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratchowill & J. R. Levin (Eds.), *Single-case research design and analysis* (pp. 187-211). Hillsdale, NJ: Erlbaum.

Busse, R. T., Kratchowill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology, 33*, 269-285.

Campbell, J. M., & Herzinger, C. V. (2010). Statistics and single subject research methodology. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 417-453). New York: Routledge.

Center, B. A., Skiba, R. J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Special Education, 19*, 387-400.

Christensen, S. L., Ysseldyke, J. E., Thurlow, M. L. (1989). Critical instructional factors for students with mild handicaps: An integrative review. *Remedial and Special Education, 10*, 21-31.

Cooper, J.O., Heron, T.E., & Heward, W.L. (2007). *Applied behavior analysis* (2nd ed). Columbus: Pearson.

Corcoran, K. J. (1985). Aggregating the idiographic data of single-subject research. *Social Work Research and Abstracts*, 21, 9-12.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of variance model for the intrasubject replication design. *Journalof Applied Behavior Analysis, 12,* 563-570.

Edgington, E. S. (1980). *Randomization tests*. New York: Marcel Dekker.

Edgington, E. S. (1992). Nonparametric tests for single-case experiments. In T. R. Kratchowill & J. R. Levin (Eds.), *Single-case research design and analysis* (pp. 133-157). Hillsdale, NJ: Erlbaum.

Franklin, R. D., Gorman, B. S., Beasley, T. M., & Allison, D. B. (1996). Graphical display and visual analysis. In R. D. Franklin, D. B. Allison, & B. s. Gorman (Eds.), *Design and analysis of single case research* (pp. 119-158). Mahwah, NJ: Erlbaum.

Gast, D. L. (2010). *Single subject research methodology in behavioral sciences*. New York: Routledge.

Gentile, J. R., Roden, A. H., & Klein, R. D. (1972). An analysis-of-variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis, 5*, 193-198.

Gibson, G., & Ottenbacher, K. (1988). Characteristics influencing the visual analysis of single-subject data: An empirical analysis. *Journal of Applied Behavioral Science, 24*, 298-314.

Gingerich, W. J. (1984). Meta-analysis of applied time-series data. *Journal of Applied Behavioral Science, 20*, 71-79.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time series experiments*. Boulder: University of Colorado Press.

Gorman-Smith, D., & Matson, J. L. (1985). A review of treatment research for self-injurious and stereotyped responding. *Journal of Mental Deficiency Research, 29*, 295-308.

Graf, S. & Lindsley, O. (2002). *Standard celeration charting 2002*. Poland, OH: Graf Implements.

Gottman, J. M., & Glass, G. V. (1978). Analysis of time-series experiments. In T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change* (pp. 197-235). New York: Academic Press.

Hagopian, L. P., Fisher, W. W., Thompson, R. H., Owen-DeSchryver, J., Iwata, B. A., & Wacker, D. P. (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis, 30*, 313-326.

Horner, R. H., Carr, E. G., Halle, J., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179.

Huitema, B. E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment, 7*, 107-118.

Huitema, B. E. (1986). Autocorrelation in behavioral research: Wherefore art thou? In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 187-208). New York: Plenum Press.

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research*. New York: Routledge

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

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York: Oxford University Press.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.

Knapp, T. J. (1983). Behavior analysts' visual appraisal of behavior chance in graphic display. *Behavioral Assessment, 5*, 155-164.

Kubina, R. M., & Yurich, K. K. L. (2012). *The Precision Teaching Book*. Lemont, PA: Greatness Achieved Publishing Company.

Lindsley, O. R. (1990a). Precision teaching: By teachers for children. *Teaching Exceptional Children, 22*(3), 10-15.

Lindsley, O. R. (2005). Standard Celeration Chart System. In M. Hersen, G. Sugai, & R. Horner (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy. Volume III: Education applications* (pp. 1545-1548). Thousand Oaks,CA: Sage.

Mason, L. L. (2010). An Analysis of Effect Sizes for Single-Subject Research:A Statistical Comparison of Five Judgmental Aids. Journal of *Precision Teaching and Celeration Volume, 26*, 3-16.

Matyas, T. A. & Greenwood, K. M. (1991). Problems in the estimation of autocorrelation in brief time series and some implications for behavioral data. *Behavioral Assessment, 13*, 137-157.

McGreevy, P. (1983). *Teaching and learning in plain English* (2nd ed.). Kansas City, MO: Plain English Publications.

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

Ottenbacher, K. J. (1986). Reliability and accuracy of visually analyzing graphed data from single-subject designs. *American Journal of Occupational Therapy, 40*, 464-469.

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education, 40*, 194-204.

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 chang*e. New York: Academic Press.

Parsonson, B. S., & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 157-186). New York: Plenum.

Pennypacker, H. S., Gutierrez, A., & Lindsley, O.

R. (2003). *Handbook of the Standard Celeration Chart*. Concord, MA: Cambridge Center for Behavioral Studies.

Sagan, C. (1997). *The Demon-Haunted World: Science as a Candle in the Dark*. New York: Ballatine Books.

Schmid, C. F. (1986). Whatever happened to the semilogarithmic chart? *The American Statistician, 40*, 238-244.

Scruggs, T. E. (1992). Single subject methodology in the study of learning and behavioral disorders: Design, analysis, and synthesis. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 7, pp. 223-248). Greenwich, CT: JAI Press.

Scruggs, T. E., & Mastropieri, M. A. (1994). The utility of the PND statistic: A reply to Allison and Gorman. *Behaviour Research and Therapy, 32*, 879-883.

Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality, 9*, 227-244.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education, 8*(2), 24-33.

Sharpley, C. F., & Alavosius, M. P. (1988). Autocorrelation in behavioral data: An alternative perspective. *Behavioral Assessment, 10*, 243-251.

Shine, L. C., & Bower, S. M. (1971). A one-way analysis of variance for single-subject designs. *Educational and Psychological Measurement, 31*, 105-113.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Skinner, B. F., & Fester, C. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Wampold, B., & Furlong, M. (1981). The heuristics of visual inference. *Behavioral Assessment, 3*, 79-92.

West, R. P., Young, K. R., & Spooner, F. (1990). Precision teaching: An introduction. *Teaching Exceptional Children, 22*, 4-9.

White, O. R. (2005). *Notes on the median slope: A method for describing the relationships between two variables*. Unpublished manuscript.

White, O. R. (1987). Some comments concerning "the quantitative synthesis of single- subject research." *Remedial and Special Education, 8*, 34-39.

White, O. R. (1974). *The "split middle" a "quickie" method of trend estimation*. Seattle: University of Washington, Experimental Education Unit, Child Development and Mental Retardation Center.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus, OH: Merrill.

White, D. M., Rusch, F. R., Kazdin, A. E., & Hartmann, D. P. (1989). Application of meta-analysis in individual-subject research. *Behavioral Assessment, 11*, 281-296.

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (in press). Comparison of overlap methods for quantitatively synthesizing singlesubject data. *The Journal of Special Education*.