An Analysis of Effect Sizes for Single-Subject Research:

A Statistical Comparison of Five Judgmental Aids

Lee L. Mason
The University of Texas at San Antonio

Effect sizes for single-subject research were examined to determine to what extent they measure similar aspects of the effects of the treatment. Seventy-five articles on the reduction of problem behavior in children with autism were recharted on standard celeration charts. Pearson product-moment correlations were then conducted between two previously unexamined effect sizes, celeration and celeration change, as well as three more common statistics, the mean baseline reduction, the percentage of non-overlapping data, and the percentage of zero data. Significant correlations were found for both celeration and celeration change, suggesting that these and other effect sizes measure somewhat similar aspects of the effect of the treatment. These findings and limitations are discussed within the broader context of evidence-based practices in education.

The Use of Judgmental Aids in Single-Subject Research

Recent legislation, such as the No Child Left Behind Act of 2001 and the Individuals with Disabilities Education Improvement Act of 2004, calls for the use of evidence-based practices to make curricular and instructional decision in the classroom. Underlying this legislation is the assumption that educators will select interventions that would provide the strongest benefit for their student population. Evidence-based practices are research-validated instructional techniques that have met rigorous standards for research design, methodological quality, and the magnitude of the effect. Randomized controlled trials and meta-analyses, which rely on statistical evaluation, typically identify evidence-based practices by examining effect sizes that measure the magnitude of the effect of an intervention (Cohen, 2001). On the other hand, single-subject research relies on the use of visual analysis in "reaching a judgment about the reliability or consistency of intervention effects by visually examining graphed data" (Kazdin, 1982, p. 232). As a result, comparisons across studies become somewhat more subjective. Furthermore, rather than determining effect sizes across groups of participants, single-subject designs compare the effect of an intervention with an alternative treatment or an adjoining phase.

Parker, Vannest, and Brown (2009) note that even the best visual analyses are commonly supported by simple statistical heuristics. According to Michael (1974), who preferred the plain English term "judgmental aids" rather than "statistics," these numbers are simply stimuli that more easily elicit responses from researchers and practitioners than raw data alone. For instance, oral reading fluency has been shown to be sensitive to instructional changes (Good & Kaminski, 2003; Shinn, 1989), and it is frequently used as a measure to evaluate the effects of reading interventions. However, sequential assessments with a single individual typically show some random variability or "bounce" in addition to the actual changes in reading skill. This variability in oral reading rate can reduce the measure's sensitivity to changes in reading skill, thereby hindering its effectiveness for monitoring progress in reading. In such cases, judgmental aids may be more helpful in describing the overall efficacy of the intervention.

Over the years, researchers have offered many suggestions for summarizing and synthesizing single-subject research in terms of trend, slope, and variability. Some of the many examples are the percentage of non-overlapping data (PND; Scruggs, Mastropieri, & Castro, 1987), the percentage of zero data (PZD; Scotti, Evans, Meyer, & Walker, 1991), the mean baseline reduction (MBLR; Kahng, Iwata, & Lewin, 2002), the C statistic (Nourbakhsh & Ottenbacher, 1994), the percentage of all non-overlapping data (PAND; Parker, Hagan-Burke & Vannest, 2007), Kruskal-Wallis W, and the improvement rate difference (IRD; Parker et al., 2009).

Campbell (2003; 2004) synthesized the literature for reducing problem behavior in persons with autism by quantifying 117 single-subject research articles and comparing the effect sizes for the PND, PZD, MBLR, and regression-based d metrics. Pearson's product-moment correlations between all four were found to be statistically significant, except for PZD and d. This finding suggests that each effect size provides a similar interpretation of the data, so that multiple measures (i.e., both PND and PZD) are unnecessary. Campbell (2004) calls for future research to continue comparing and contrasting additional effect sizes so as to better understand their use in summarizing single-subject research.

One measure of single-subject research, which has long been used to measure change in frequency over time, is celeration (Graf & Lindsley, 2002; McGreevy, 1983; White & Haring, 1980). A celeration line is a trend line, drawn through multiple behavioral frequencies on a standard celeration chart (SCC), which quantifies the amount of learning over a given period of time. A frequent criticism of visual analysis in single-subject research is the lack of formal decision rules for analyzing data (Nourbakhsh & Ottenbacher, 1994). However, with standard displays such as the SCC, multiple practitioners interpret the same data in a more consistent manner: They bring the viewer's reaction under control of the data, rather than the less pertinent features of the graph (e.g., scale; Johnson & Pennypacker, 1993).

Using the SCC, a specific value is computed for each celeration line, thereby providing a judgmental aid for comparing celerations. Celeration offers the rate of behavior over time as the measure of effect. Clearly, a reading intervention designed to increase words correct per minute with a celeration of x2.0 has a greater effect than a similar intervention with a celeration of x1.4. Even though celerations are frequently compared with one another to measure the effects of behavioral interventions, celeration has not yet been systematically compared with other types of single-subject effect sizes. The purpose of this study is therefore to examine the extent to which celeration and celeration change relate to PND, PZD, and MBLR. Specifically, this research sought to answer the following question: To what extent does celeration offer a unique effect size for single-subject research?

METHOD

Selection of Studies

Campbell (2003; 2004) identified the 117 articles used in this research. According to an a priori power analysis, this sample size was sufficient for computing a Pearson product-moment coefficient (r; Faul, Erdfelder, Lang, & Buchner, 2007) to examine the correlation between celeration and other measures of single-subject effect size. Individual data sets were selected, based on four criteria:

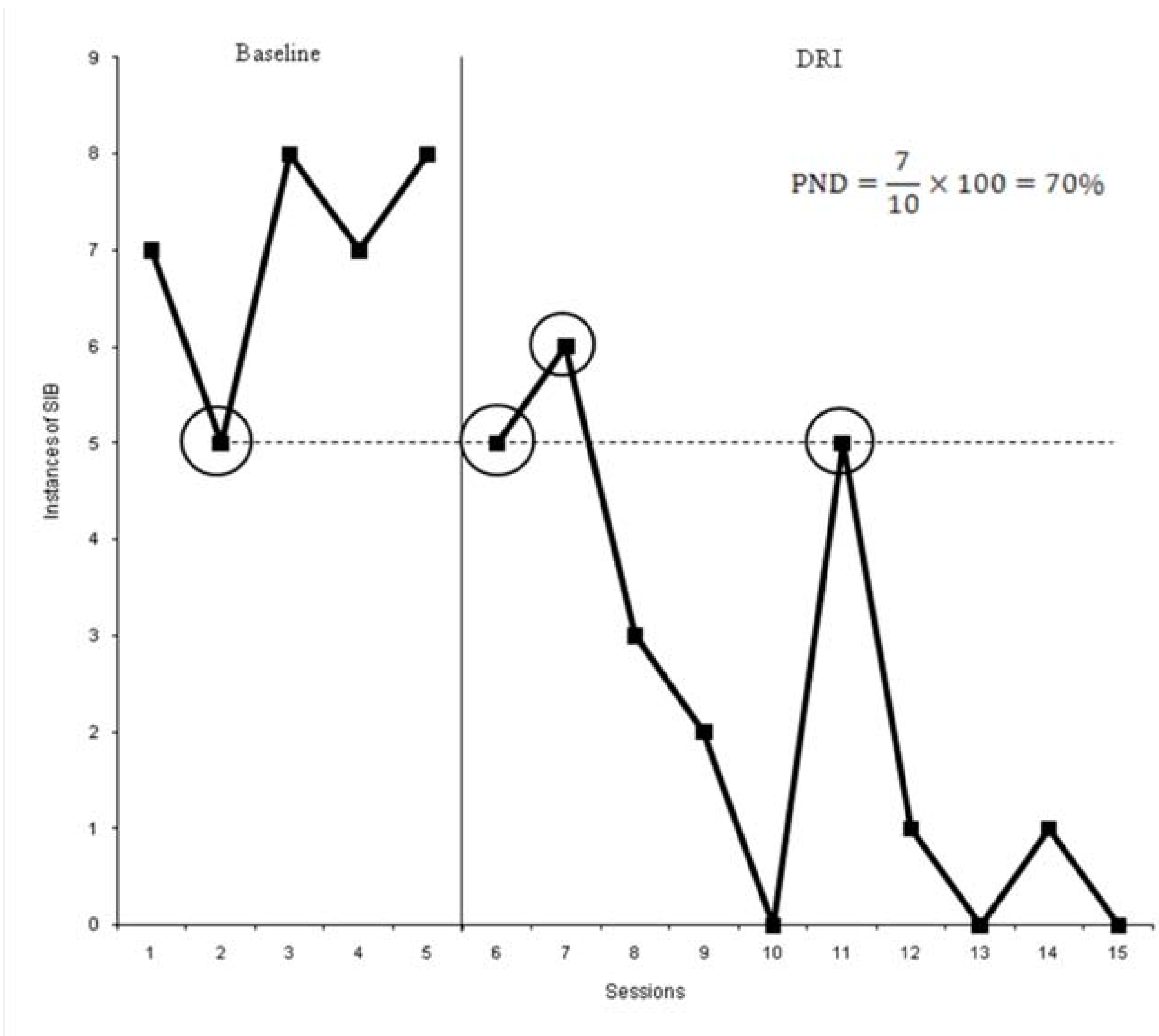
1. Only single-subject research was included to ensure that behavioral data for each participant were readily available.
2. Baseline and treatment phases in each single-subject design had to be presented as repeated measures.
3. Treatment targeted the reduction of problem behavior (e.g., self-injurious behavior, stereotypy, aggression, or property destruction).
4. At least one participant was diagnosed with autism.

If the article included multiple participants, only the behavers who fit these criteria were included in this analysis.

Single-Subject Effect Sizes

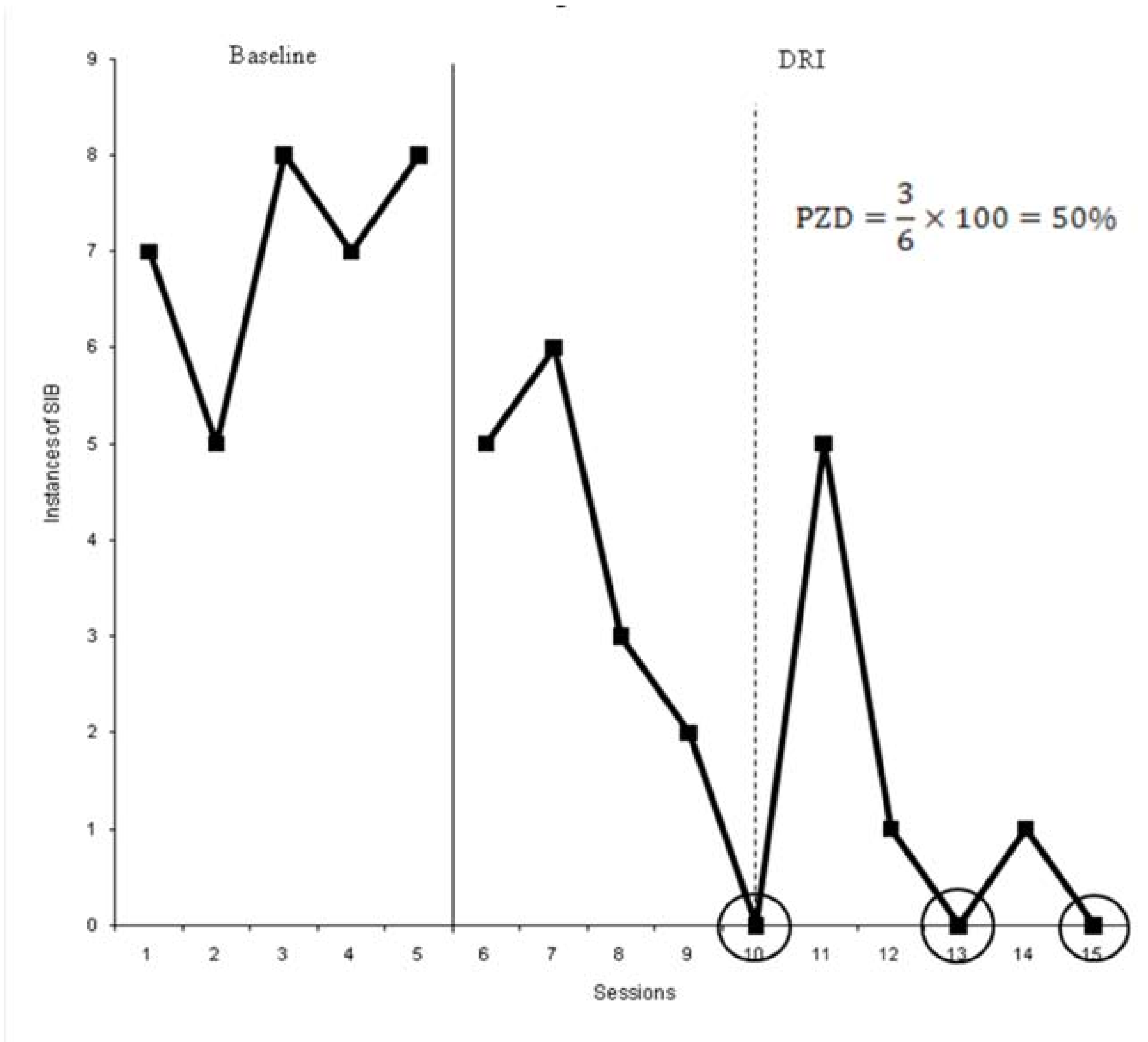
As noted, a variety of methods can be used to summarize single-subject data. Three of the more common methods found throughout single-subject literature are the percentage of non-overlapping data (PND), the percentage of zero data (PZD), and the mean baseline reduction (MBLR). The PND summarizes the effects of treatment by counting the number of data points in the intervention phase that do not overlap with the highest or lowest data points in the baseline phase, dividing by the total number of data points in the treatment phase, and multiplying by 100 (Scruggs et al, 1987; Scruggs, Mastropieri, Cook, & Escobar, 1986). Figure 1 shows hypothetical data on an intervention designed to reduce self-injurious behavior (SIB). The circled data point in the baseline phase represents the lowest level of SIB observed during baseline. A dashed line has been extended from this point into the intervention phase. The three data points circled in the intervention phase are those overlapping with the lowest data point in the baseline phase. The PND for this data set is 70%.

Figure 1. Hypothetical data demonstrating the calculation of percentage of non-overlapping data (PND).



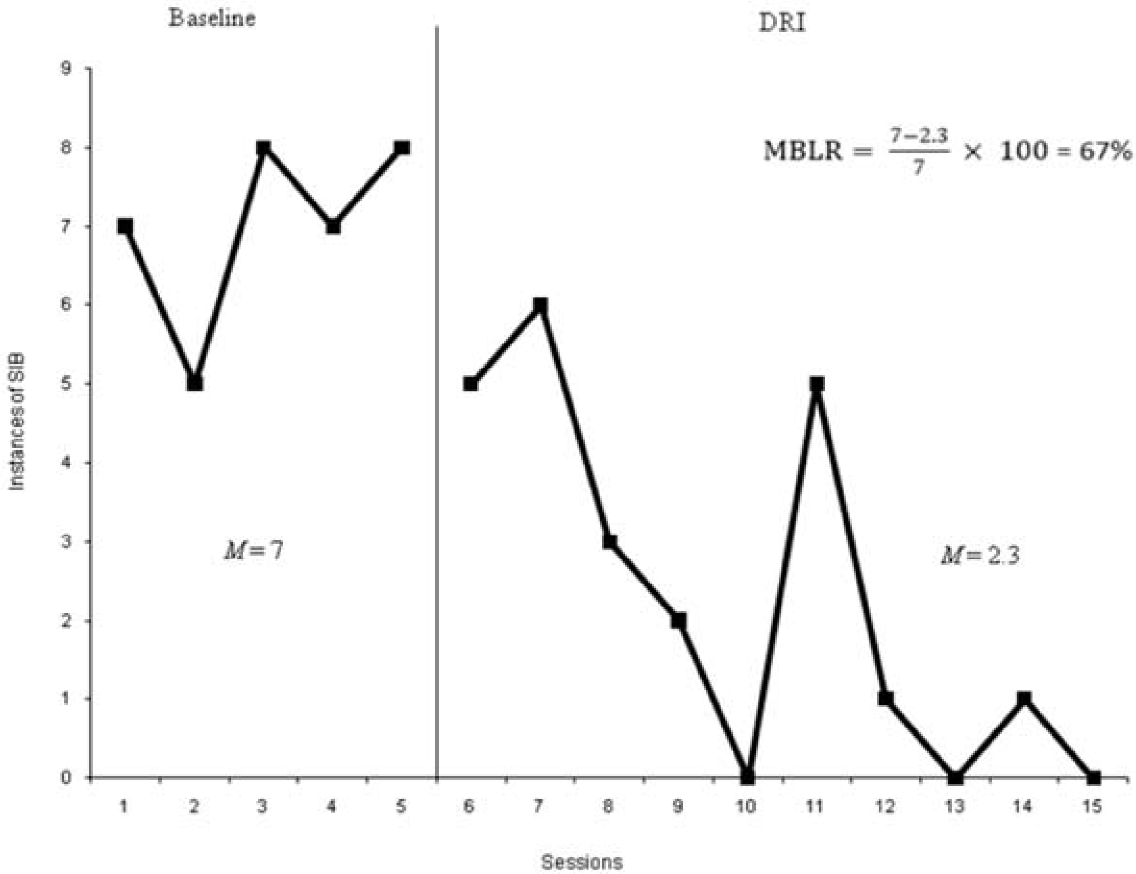
The PZD measures behavior reduction by locating the first data point in an intervention based on a count of zero; for the remainder of the phase, the percentage of data points remaining at zero is calculated (Scotti et al., 1991). Figure 2 presents the same hypothetical data. In this figure, the three data points that reach zero are circled, and a dashed line is drawn at the first zero data point. The PZD is calculated from this point forward and equals 50%.

Figure 2. Hypothetical data demonstrating the calculation of percentage of zero data (PZD).



The MBLR is found by subtracting the mean treatment value from the mean baseline value, next dividing by the mean baseline value, and then multiplying the result by 100 (Kahng et al., 2005). Figure 3 shows the hypothetical data set once again. The average count of the 5 observations in the baseline is 7, whereas the average of the 10 observations in treatment is 2.3. These are calculated to give a MBLR of 67%.

Figure 3. Hypothetical data demonstrating the calculation of mean baseline reduction (MBLR).



This analysis also examined the celeration line of the first treatment phase and the celeration change between the baseline and the intervention. To calculate the celeration lines and the MBLR, the graphically presented data were converted to raw numbers. Using a drafting divider, the distance between the horizontal axis and each data point was measured in millimeters and rounded to the nearest half-millimeter (Huitema, 1985). An approximate value was then produced by measuring this distance against the vertical axis of the same graph. This data-conversion procedure has been used with a high degree of reliability (Allison, Faith, & Franklin, 1995; Kahng et al., 2005; Skiba, Casey, & Center, 1985-86).

Recharting on the Standard Celeration Chart

To compare celeration with the above-listed effect sizes, the data in each article were recharted on the SCC. The only graphs considered were those with a behavior or product of a behavior on the vertical axes and a unit of time on the horizontal axes. Using the guidelines Porter (1985) provided, each of the 117 articles was screened and recharted. A summary of these procedures follows.

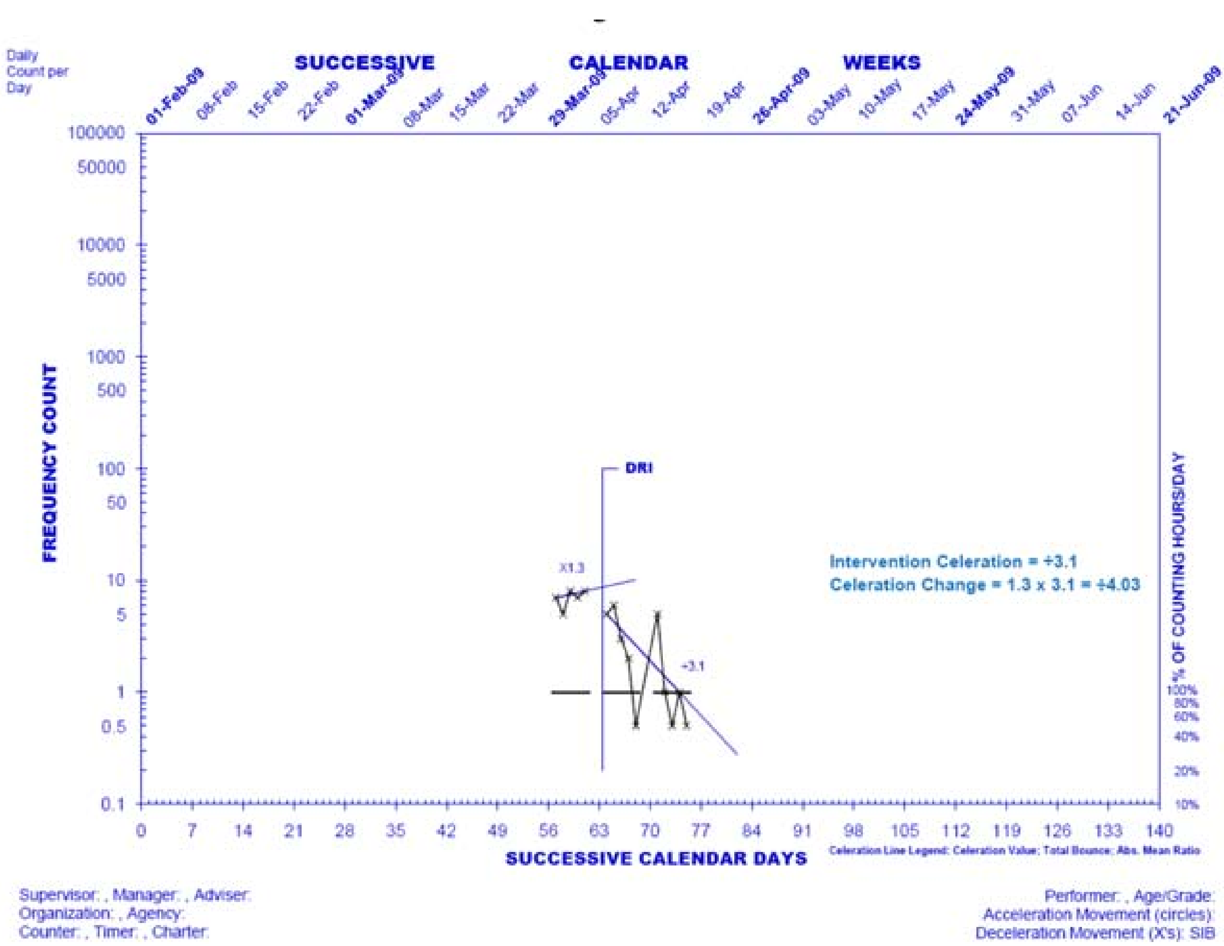
The Dpmin-11EC SCC was used to replot data from each article. This chart consists of calendar days along the horizontal axis, allowing for a comparison of studies that use various observation schedules (e.g., daily versus twice weekly). Additionally, the SCC measures frequency on the vertical axis, so that studies using different measures or interval lengths (e.g., number versus percent-interval) could be compared. Therefore, all the original details from the research are preserved on the SCC.

The frequencies were charted on the Microsoft Excel Standard Celeration Chart Template (Harder, 2008). A new chart was used for each data set from each study. In some cases, as with multi-element designs, the same baseline was used with multiple intervention phases – each replotted on its own chart. Record floors and ceilings were marked with dashes, and data points were placed between. Frequencies based on a count of zero were plotted ÷2 below the record floor (White & Neely, 2004).

Separate celeration lines were drawn for both the initial baseline and the first intervention phase. For the purposes of comparing effect sizes, both the celeration of the first intervention phase and the celeration change between the baseline and intervention phases were recorded for every chart. Celeration lines were automatically computed for each phase by the Excel Standard Chart Template, using the median slope method (White, 2005). The median slope is found by drawing lines passing through all possible pairs of data points, then selecting the line that falls in the middle of that array. If all the slopes in the distribution are arranged in numerical order and there is an odd number of scores, the median slope would be the score in the middle. With an even number of slopes, either the line representing the most conservative slope can be selected, or the two middle slopes can be averaged. White (2005) notes that the median slope is generally more useful in predicting future performance than other methods of calculating trend lines.

Celeration changes were determined by comparing the celeration of the baseline phase to the celeration of the intervention phase. Using the same hypothetical data as above, Figure 4 displays a celeration turn down from x1.3 to ÷3.1. This yields a celeration change of ÷4.03.

Figure 4. Hypothetical data demonstrating the calculation of celeration and celeration change.



Celeration lines were not calculated for any phase that had fewer than five daily frequencies. In cases where the intervention had fewer than five data points, the data set was excluded. If the baseline phase contained fewer than five data points but the intervention phase had at least five points, the intervention celeration was calculated, but the celeration change could not be determined.

Each article was closely examined to determine the frequency of observation. When this information was not provided, an assumption was made of once daily excluding weekends. When an article listed multiple sessions per day, only the initial daily data point was recharted. For example, if an article stated that two sessions were run each day, only the sequentially odd-numbered data points were replotted. Articles that listed a variety of sessions (i.e., between 3 and 5 sessions run daily) were excluded.

Additional information was required to rechart percent-interval data, including the total observation time and the interval length. Articles that did not include this information could not be recharted. Recharting percent-interval data requires converting each data point to an assumed frequency. However, three factors must be determined first: (a) the record floor, (b) the record ceiling, and (c) the total number of intervals observed in each session.

The minimum frequency that can be recorded during a session is called the record floor. In percent-interval graphs, this is the total observation time. For most articles, the observation time remained constant throughout the study. If observation time was given as a range (e.g., sessions ranging from 10 to 15 minutes), the shorter observation time was used as the record floor. When interrupted-interval recording procedures were used (e.g., a 5-second observe, a 5-second record cycle used for 10 minutes), only the actual observation time was used as the record floor.

The maximum frequency that can be recorded during a session is called the record ceiling. This is directly defined by the interval length used in each study. To find the record ceiling, divide 60 by the interval length (e.g., 60 divided by 6-second intervals yields a record ceiling of 10).

For converting a percentage of intervals to a frequency estimate, the total number of intervals observed in each session is needed. This can be found by multiplying the record floor by the record ceiling (e.g., a record floor of 10 multiplied by a record ceiling of 10 equals 100 intervals). A percentage of intervals can then be converted to the number of intervals by multiplying the percentage by the number of intervals observed (e.g., 75% of 100 intervals equals 75 intervals scored). Finally, dividing the number of intervals observed by the observation time yields a frequency estimate (Porter, 1985). This number can now be recharted on the SCC.

RESULTS

This study examined the extent to which celeration offers an independent effect size for single subject research. Of the original 117 articles Campbell (2003) identified, 75 fit the criteria for eligibility in this study. The data sets for two articles could not be located and were therefore not included in this analysis. The remaining articles examined 112 behavers, and a total of 176 behaviors that were recharted and included in this review. Interestingly, out of initial 117 articles, only two (Bierly & Billingsley, 1983; Sugai & White, 1986) originally plotted their data on standard celeration charts.

Correlation coefficients were computed among the five single-subject effect sizes by using the R statistical computing environment. The Bonferroni approach to control for Type 1 error was used across the 10 correlations, thereby requiring a p value of less than 0.005 to show statistical significance ($0.05 \div 10 = 0.005$; Green & Salkind, 2008). Table 1 shows that 4 out of the 10 correlations were statistically significant and were greater than or equal to 0.23. The largest correlation occurred between the celeration of the intervention phase and the celeration change of $r = 0.54$, $p < 0.001$. This is understandable since the intervention celeration is used to determine the celeration change.

A moderate correlation was found between the celeration of the intervention phase and the mean baseline reduction of $r = -0.33$, $p < 0.001$, and a small correlation was found between celeration change and MBLR, $r = -0.26$, $p = 0.002$. These negative coefficients can be explained by examining the manner in which each effect size was determined. For example, imagine the data set in which problem

Table 1
Relationship Between Single-Subject Effect Sizes

	Celeration	Celeration Change	MBLR	PND	PZD
Celeration	–				
Celeration Change	.54*	–			
MBLR	-.33*	-.26*	–		
PND	.05	-.12	.08	–	
PZD	.07	-.11	.23*	.06	–

Note: MBLR = Mean Baseline Reduction; PND = Percentage of Non-overlapping Data; PZD = Percentage of Zero Data.
*$p < .01$

behavior was high during baseline and immediately dropped to zero at the start of the intervention, where it remained. This would result in a high MBLR (e.g., 100%) and a low intervention celeration (e.g., x1.00). Conversely, a data set in which the baseline numbers were high, but gradually decreased over several intervention sessions, would result in a lower MBLR (e.g., 50%) and a greater celeration value (e.g., ÷4.00).

Another small correlation was found between MBLR and PZD, r = 0.23, p = 0.001. This is consistent with Campbell (2003, 2004), suggesting that these two effect sizes are measuring somewhat similar aspects of the effects of treatment. Conversely, no significant correlations were found between the intervention celeration or the celeration change and PND or PZD, indicating that these statistics measure different aspects of effectiveness.

DISCUSSION

Single-subject research has always relied on the graphical analysis of data to determine the effects of an intervention. This is primarily done by comparing level, trend, or variability across phases. Although several researchers have attempted to convert these effects into numbers that can be compared across studies, no single statistic appears to account for all methods of visual analysis. The data presented here suggest that celeration and celeration change are independent evaluations of single-subject research, which measure an effect that is entirely unrelated to PND and PZD. One reason for this may be because celeration measures slope, whereas the other statistics measure level or variability.

An interim step in determining effect size may be to select the appropriate statistic based on visual analysis. That is, multiple graphs demonstrating a change in level may then be compared using PND or PZD, whereas celeration or an improvement rate difference may be used to compare graphs showing a change in slope. What is important to note in the present study is that the mean baseline reduction did show some amount of correlation with both celeration and celeration change. Therefore, the effect sizes measuring level and slope are not mutually exclusive. To date, there has been no consensus on which effect sizes best represent raw data.

This research has other limitations that must be addressed. Most notably, recharting data does not result in a true frequency. Interval recording produces only an estimate of the actual frequency of behavior. Additionally, converting intervals to a percentage and back again results in some error (Porter, 1985). As a result, many of the charts included in this study were not precise.

Although the number of publications about individuals with autism continues to rise, there is an obvious dearth of data being presented on standard charts. Whether this is due to the multiply/divide scale on the SCC or the inability to manipulate axes is unclear. However, the ease with which it allows users to calculate a celeration line and compare data across charts makes a compelling argument for an increase in standard celeration charting. While the results of this study demonstrate that both celeration and celeration change are related to other single-subject effect sizes, future researchers are strongly urged to continue examining and comparing additional methods for synthesizing single-subject designs.

Salzberg, Strain, and Baer (1987), as well as Michael (1974), note that the idiosyncrasies and familiarity accompanying prolonged and intense interaction with time-series data do not occur in a one-number summary. The experimenter is forced to rely on theory and other people's research, and then attempt to draw conclusions about the relative merits of broad categories of intervention. Although Michael suggests that the use of these judgmental aids may produce a stimulus the teacher or behavior analyst can more easily react to, he cautions that these statistics may be worthwhile only when the time spent learning how to use such techniques and the effort in determining which one to use is relatively small compared with the simplifying effect achieved.

The term "effect size" has been used here to talk about comparing the effectiveness of interventions across single-subject research; however, other methods, such as metacharting, may also function to compare celerations. Lindsley, Calkin, and White (1993) emphasize the importance of analyzing chart collections, and Cooper, Kubina, and Malanga (1998) provide a variety of ways in which collections of standard celeration charts can be synchronized and displayed. Charting repeated measures not only helps users to stay connected with the data, but metacharting also allows them to make instructional or intervention decisions based on multiple sources of data (thereby also acting as a judgmental aid).

For celeration to truly function as a measure of the magnitude of effect for single-case interventions, future research should address the classification of large, medium, and small celeration effect sizes. Green and Salkind (2008) note that "as with all effect size indices, there is no good answer to the question 'What value indicates a strong relationship between two variables?'" (p. 259). Effect size is dictated by the discipline within which the research is conducted. For celeration charting, each SCC includes a celeration fan ranging from x16 to ÷16 that may act as a guideline for talking about the magnitude of a celeration (e.g., 1.4, 2.0, and 4.0 – irrespective of sign – can be interpreted as small, medium, and large effect sizes, respectively).

For years, educators and researchers have been using data, or practice-based evidence, to make instructional decisions in their classrooms and clinics. These measures help to demonstrate that adequate progress is being made towards a specified goal. Recent educational policy may have just begun mandating the use of evidence in the classroom, but the practice is hardly new. Many practitioners have argued that the prescription of evidence-based practices results in the loss of autonomy. However, the specific educational gains of each student are more important than the generalization of practices across settings and participants. Cook, Tankersley, and Landrum (2010) conclude that evidence-based practices "will not and should not ever take the place of professional judgment but can be used to inform and enhance the decision making of special education teachers" (p. 380). Ultimately, effect size and other statistics are simply additional judgmental aids to help practitioners make data-based decisions.

REFERENCES

Note: Asterisks denote articles that were included in the synthesis.

*Aiken, J. M., & Salzberg, C. L. (1984). The effects of a sensory extinction procedure on stereotypic sounds of two autistic children. *Journal of Autism and Developmental Disorders, 14*, 291-299.

*Allison, D. B., Basile, V. C., & MacDonald, R. B. (1991). Brief report: Comparative effects of antecedent exercise and lorazepam on the aggressive behavior of an autistic man. *Journal of Autism and Developmental Disorders, 21*, 89-94.

Allison, D.B., Faith, M.S., & Franklin, R.D. (1995). Antecedent exercise in the treatment of disruptive behavior: A meta-analytic review. *Clinical Psychology: Science and Practice, 2*, 279-303.

*Altmeyer, B. K., Williams, D. E., & Sams, V. (1985). Treatment of severe self-injurious and aggressive biting. *Journal of Behavior Therapy and Experimental Psychiatry, 16*, 169-172.

*Bebko, J. M., & Lennox, C. (1988). Teaching the control of diurnal bruxism to two children with autism using a simple cueing procedure. *Behavior Therapy, 19*, 249-255.

*Bierly, C., & Billingsley, F. F. (1983). An investigation of the educative effects of overcorrection on the behavior of an autistic child. *Behavioral Disorders, 9*, 11-21.

*Brawley, E. R., Harris, F. R., Allen, K. E., Fleming, R. S., Peterson, R. F. (1969). Behavior modification of an autistic child. *Behavioral Science, 14*, 87-97.

Campbell, J.M., (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification, 28*, 234-246.

Campbell, J.M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities, 24*, 120-138.

*Carr, E. G., & Newsom, C. (1985). Demand-related tantrums: Conceptualization and treatment. *Behavior Modification, 9*, 403-426.

*Casey, L. O. (1978). Development of communicative behavior in autistic children: A parent program using signs. *Journal of Autism and Childhood Schizophrenia, 8*, 45-94.

*Chapman, S., Fisher, W., Piazza, C. C., & Kurtz, P. (1993). Functional assessment and treatment of life-threatening drug ingestion in a dually diagnosed youth. *Journal of Applied Behavior Analysis, 26*, 255-256.

*Charlop-Christy, M. H., & Haymes, L. K. (1996). Using obsessions as reinforcers with and without mild reductive procedures to decrease inappropriate behaviors of children with autism. *Journal of Autism and Developmental Disorders, 26*, 527-546.

*Charlop-Christy, M. H., & Haymes, L. K. (1998). Using objects of obsession as token reinforces for children with autism. *Journal of Autism and Developmental Disorders, 28*, 189-198.

*Chock, P. N., & Glahn, T. J. (1983). Learning and self-stimulation in mute and echolalic autistic children. *Journal of Autism and Developmental Disorders, 13*, 365-381.

*Clarke, J.C. & Thomason, S. (1984). The use of an aversive smell to eliminate autistic self-stimulatory behavior. *Child & Family Behavior Therapy, 5*, 51-61.

*Clarke, S., Dunlap, G., Foster-Johnson, L., Childs, K.E., Wilson, D., White, R., & Vera, A. (1995). Improving the conduct of students with behavioral disorders by incorporating student interests into curricular activities. *Behavioral Disorders, 20*, 221-237.

Cohen, B.H. (2001). *Explaining psychological statistics* (2nd Ed). New York: John Wiley & Sons, Inc.

*Coleman, C.L., & Holmes, P.A. (1998). The use of noncontingent escape to reduce disruptive behaviors in children with speech delays. *Journal of Applied Behavior Analysis, 31*, 687-690.

Cook, B.G., Tankersley, M., & Landrum, T.J., (2010). Determining evidence-based practices in special education. *Exceptional Children, 75*, 365-383.

Cooper, J.O., Kubina, R., & Malanga, P. (1998). Six procedures for showing collections of standard celeration charts, *Journal of Precision Teaching and Celeration, 15*, 56-73.

*Day, H.M., Horner, R.H., & O'Neill, R.E. (1994). Multiple functions of problem behaviors: Assessment and intervention. *Journal of Applied Behavior Analysis, 27*, 279-289.

*Day, R.M., Rea, J.A., Schussler, N.G., Larsen, S.E., Johnson, W.L. (1988). A functionally based approach to the treatment of self-injurious behavior. *Behavior Modification, 12*, 565-589.

*Doke, L., Wolery, M., & Sumberg, C. (1983). Treating chronic aggression: Effects and side effects of response-contingent ammonia spirits. *Behavior Modification, 7*, 531-556.

*Doleys, D.M., Wells, K.C., Hobbs, S.A., Roberts, M.W., & Cartelli, L.M. (1976). The effects of social punishment on noncompliance: A comparison with timeout and positive practice. *Journal of Applied Behavior Analysis, 9*, 471-482.

*Ducharme, J.M., Lucas, H., & Pontes, E. (1994). Errorless embedding in the reduction of severe maladaptive behavior during interactive and learning tasks. *Behavior Therapy, 25*, 489-501.

*Eason, L.J., White, M.J., & Newsom, C. (1982). Generalized reduction of self-stimulatory behavior: An effect of teaching appropriate play to autistic children. *Analysis of Intervention in*

Developmental Disabilities, 2, 157-169.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

*Fischer, W.W., Ninness, H.A.C., Piazza, C.C., & Owen-DeSchryver, J.S. (1996). On the reinforcing effects of the content of verbal attention. *Journal of Applied Behavior Analysis, 29*, 235-238.

*Freeman, B.J., Moss, D., Somerset, T., & Ritvo, E.R. (1977). Thumbsucking in an autistic child overcome by overcorrection. *Journal of Behavior Therapy & Experimental Psychiatry, 8*, 211-212.

Good, R.H., & Kaminski, R.A. (2003). *Dynamic indicators of basic early literacy skills* (DIBELS; 6th Ed.). Longmont, CO: Sopris West.

*Gordon, R., Jandleman, J.S., Harris, S.L. (1987). The effects of contingent versus non-contingent running on the out-of-seat behavior of an autistic boy. *Child & Family Behavior Therapy, 8*, 37-44.

Graf, S., & Lindsley, O.R. *Standard Celeration Charting* 2002. Poland, OH: Graf Implements.

Green, S.B., & Salkind, N.J. (2008). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, NJ: Pearson Prentice Hall.

*Gunter, P., Brady, M. P., Shores, R. E., Fox, J. J., Owen, S., & Goldzweig, I. R. (1983). The reduction of aberrant vocalizations with auditory feedback and resulting collateral behavior change of two autistic boys. *Behavioral Disorders, 11*, 254-263.

*Handen, B.L., Apolito, P.M., & Seltzer, G.B., (1984). Use of differential reinforcement of low rates of behavior to decrease repetitive speech in an autistic adolescent. *Journal of Behavior Therapy & Experimental Psychiatry, 15*, 359-364.

Harder, S.R., White, O.R., and Born, S. (2008) Standard Celeration Chart: Daily per Minute, Version 7-3. Freeware charting template, archived at http://lists.psu.edu/cgi-bin/wa?A2=ind0805&L=SCLISTSERV&T=0&F=&S=&P=1580.

*Haring, T.G., & Kennedy, C.H. (1990). Contextual control of problem behavior in students with severe disabilities. *Journal of Applied Behavior Analysis, 23*, 235-243.

*Harris, S.L., & Wolchik, S.A. (1979). Suppression of self-stimulation: Three alternative strategies. *Journal of Applied Behavior Analysis, 12*, 185-198.

*Harris, S.L., Handleman, J.S., Fong, P.L. (1987). Imitation of self-stimulation: Impact on the autistic child's behavior and affect. *Child & Family Behavior Therapy, 9*, 1-21.

*Herbert, E.W., Pinkston, E.M., Hayden, M.L., Sajwaj, T.E., Pinkston, S., Cordua, G., & Jackson, C. (1973). Adverse effects of differential parental attention. *Journal of Applied Behavior Analysis, 6*, 15-30.

*Horner, R.H., & Day, H.M. (1991). The effects of response efficiency on functionally equivalent competing behaviors. *Journal of Applied Behavior Analysis, 24*, 719-732.

*Horner, R.H., Day, H.M., & Day, J.R. (1997). Using neutralizing routines to reduce problem behaviors. *Journal of Applied Behavior Analysis, 30*, 601-614.

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

Johnston, J.M., & Pennypacker, H.S. (1993). *Strategies and tactics of behavioral research* (2nd Ed). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kahng, S., Iwata, B., Lewin, A.B. (2002). Behavioral treatment of self-injury, 1964-2000. *American Journal on Mental Retardation, 107*, 212-221.

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

*Kennedy, C.H. (1994). Manipulating antecedent conditions to alter the stimulus control of problem behavior. *Journal of Applied Behavior Analysis, 27*, 161-170.

*Kern, L., Carberry, N., & Haidara, C. (1997). Analysis and intervention with two topographies of challenging behavior exhibited by a young woman with autism. *Research in Developmental Disabilities, 18*, 275-287.

*Krantz, P.J., MacDuff, M.T., & McClannahan, L.E. (1993). Programming participation in family activities for children with autism: Parents' use of photographic activity schedules. *Journal of Applied Behavior Analysis, 26*, 137-139.

*Lalli, J.S., Casey, S., & Kates, K. (1995). Reducing escape behavior and increasing task completion with functional communication training, extinction, and response chaining. *Journal of Applied Behavior Analysis, 28*, 261-268.

Lindsley, O.R., Calkin, A.B., & White, O.R. (1993). How to numerically and graphically summarize learning across classrooms, schools, and published precision teaching studies (metacharting). *Journal of Precision Teaching, 10*, 46-53.

*Luiselli, J.K., Medeiros, J., Jasinowski, C., Smith, A., & Cameron, M.J. (1994). Behavioral medicine treatment of ruminative vomiting and associated weight loss in an adolescent with autism. *Journal of Autism and Developmental Disorders, 24*, 619-629.

*Luiselli, J.K., Suskin, L., & McPhee, D.F. (1981). Continuous and intermittent application of overcorrection in a self-injurious autistic child: Alternating treatments design analysis. *Journal of Behavior Therapy & Experimental Psychiatry, 12*, 355-358.

*Maag, J. W., Wolchik, S. A., Rutherford, R. B., & Parks, T. P. (1986). Response covariation on self-stimulatory behaviors during sensory extinction procedures. *Journal of Autism and Developmental Disorders, 16*, 119-132.

*Marholin, D., & Townsend, N. M. (1978). An experimental analysis of side effects and response maintenance of a modified overcorrection procedure. *Behavior Therapy, 9*, 383-390.

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

*McKeegan, G. F., Estill, K., & Campbell, B. (1987). Elimination of rumination by controlled eating and differential reinforcement. *Journal of Behavior Therapy and Experimental Psychiatry, 18*, 143-148.

*McKeegan, G. F., Estill, K., & Campbell, B. M. (1984). Use of nonexclusionary timeout for the elimination of a stereotyped behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 15*, 261-264.

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

*Neufeld, A., & Fantuzzo, J. W. (1984). Contingent application of a protective device to treat the severe self-biting behavior of a disturbed autistic child. *Journal of Behavior Therapy and Experimental Psychiatry, 15*, 79-83.

Nourbakhsh, M.R., & Ottenbacher, K.J. (1994). The statistical analysis of single-subject data: A comparative examination. *Physical Therapy, 74*, 768-776.

*Oke, N. J., & Schreibman, L. (1990). Training social initiations to a high-functioning autistic child: Assessment of collateral behavior change and generalization in a case study. *Journal of Autism and Developmental Disorders, 20*, 479-497.

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

Parker, R.I., Vannest, K.J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children, 75*, 135-150.

*Piazza, C.C., Hanley, G.P., Bowman, L.G., Ruyter, J.M., Lindauer, S.E., & Saiontz, D.M. (1997). Functional analysis and treatment of elopement. *Journal of Applied Behavior Analysis, 30*, 653-672.

*Piazza, C.C., Moes, D.R., & Fisher, W.W. (1996). Differential reinforcement of alternative behavior and demand fading in the treatment of escape maintained destructive behavior. *Journal of Applied Behavior Analysis, 29*, 569-572.

*Piazza, C.C., Wayne, W.F., Hanley, G.P., LeBlanc, L.A., Worsdell, A.S., Lindauer, S.E., & Keeney, K.M. (1998). Treatment of pica through multiple analyses of its reinforcing functions. *Journal of*

Applied Behavior Analysis, 31, 165-189.

*Plummer, S., Baer, D.M., & LeBlanc, J.M. (1977). Functional considerations in the use of procedural timeout and an effective alternative. *Journal of Applied Behavior Analysis, 10*, 689-705.

Porter, K.L. (1985). Standard celeration chart summary of individual percentage-interval recording studies from the Journal of Applied Behavior Analysis 1968-1984. Unpublished doctoral dissertation, University of Kansas, Lawrence, KS.

*Ragland, E.U., Kerr, M.M., & Strain, P.S. (1978). Behavior of withdrawn autistic children. *Behavior Modification, 2*, 565-578.

*Reese, R.M., Sherman, J.A., & Sheldon, J.B. (1998). Reducing disruptive behavior of a group-home resident with autism and mental retardation. *Journal of Autism and Development Disorders, 28*, 159-165.

*Richman, D.M., Wacker, D.P., Asmus, J.M., & Casey, S.D. (1998). Functional analysis and extinction of different behavior problems exhibited by the same individual. *Journal of Applied Behavior Analysis, 31*, 475-478.

*Rincover, A. (1978). Sensory extinction: A procedure for eliminating self-stimulatory behavior in developmentally delayed children. *Journal of Abnormal Child Psychology, 6*, 299-310.

*Rincover, A., Cook, R., Peoples, A., & Packard, D. (1979). Sensory extinction and sensory reinforcement principles for programming multiple adaptive behavior change. *Journal of Applied Behavior Analysis, 12*, 221-233.

*Risley, T.R. (1968). The effects and side effects of punishing the autistic behaviors of a deviant child. *Journal of Applied Behavior Analysis, 1*, 21-34.

*Rolider, A., & Van Houten, R. (1985). Movement suppression time-out for undesirable behavior in psychotic and severely developmentally delayed children. *Journal of Applied Behavior Analysis, 18*, 275-288.

*Rosenthal-Malek, A., & Mitchell, S. (1997). The effects of exercise on the self-stimulatory behaviors and positive responding of adolescents with autism. *Journal of Autism and Developmental Disorders, 27*, 193-202.

*Russo, D.C., & Koegel, R.L. (1977). A method of integrating an autistic child into a normal public-school classroom. *Journal of Applied Behavior Analysis, 10*, 579-590.

Salzberg, C.L., Strain, P.S., Baer, D.M. (1987). Meta-analysis for single-subject research. *Remedial and Special Education, 8*, 43-48.

*Sasso, G.M., Melloy, K.J., Kavale, K.A. (1990). Generalization, maintenance, and behavioral covariation associated with social skills training through structured learning. *Behavioral Disorders, 16*, 9-22.

*Sasso, G.M., Teimers, T.M., Cooper, L.J., Wacker, D., Berg, W., Steege, M., Kelly, L., & Allaire, A. (1992). Use of descriptive and experimental analyses to identify the functional properties of aberrant behavior in school settings. *Journal of Applied Behavior Analysis, 25*, 809-821.

Scotti, J.R., Evans, I.M., Meyer, L.H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96*, 233-256.

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

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

Shinn, M. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.

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

*Smith, D.E.P. (1980). Is isolation room time-out a punisher? *Behavioral Disorders, 6*, 247-256.

*Smith, M.D. (1987). Treatment of pica in an adult disabled by autism by differential reinforcement of incompatible behavior. *Journal of Behavior*

Therapy and Experimental Psychiatry, 18, 285-288

*Solnick, J.V., Rincover, A., & Peterson, C.R. Some determinants of the reinforcing and punishing effects of timeout. *Journal of Applied Behavior Analysis, 10*, 415-424.

*Stahmer, A.C., & Schreibman, L. (1992). Teaching children with autism appropriate play in unsupervised environments using a self-management treatment package. *Journal of Applied Behavior Analysis, 25*, 447-459.

*Sugai, G., & White, W. J. (1986). Effects of using object self-stimulation as a reinforcer on the prevocational work rates of an autistic child. *Journal of Autism and Developmental Disorders, 16*, 459-471.

*Symons, F., & Davis, M. (1994). Instructional conditions and stereotyped behavior: The function of prompts. *Journal of Behavior Therapy & Experimental Psychiatry, 25*, 317-324.

*Tanner, B.A., & Zeiler, M. (1975). Punishment of self-injurious behavior using aromatic ammonia as the aversive stimulus. *Journal of Applied Behavior Analysis, 8*, 53-57.

*Tate, B.G., & Baroff, G.S. (1966). Aversive control of self-injurious behavior in a psychotic boy. *Behavior Research & Therapy, 4*, 281-287.

*Tomporowski, P.D. (1983). Training an autistic client: The effect of brief restraint on disruptive behavior. *Journal of Behavior Therapy & Experimental Psychiatry, 2*, 169-173.

*Tustin, R.D. (1995). The effects of advance notice of activity transitions on stereotypic behavior. *Journal of Applied Behavior Analysis, 28*, 91-92.

*Underwood, L.A., Figueroa, R.G., Thyer, B.A., & Nzeocha, A. (1989). Interruption and DRI in the treatment of self-injurious behavior among mentally retarded and autistic self-restrainers. *Behavior Modification, 13*, 471-481.

*Van Houten, R. Rolider, A. (1988). Recreating the scene: An effective way to provide delayed punishment for inappropriate motor behavior. *Journal of Applied Behavior Analysis, 21*, 187-192.

*Wells, K.C., Forehand, R., & Hickey, K. (1977). Effects of a verbal warning and overcorrection on stereotyped and appropriate behaviors. *Journal of Abnormal Child Psychology, 5*, 387-403.

White, O. (2005) Trend Lines. In G. Sugai & R. Horner (Eds.) *Encyclopedia of behavior modification and cognitive behavior therapy*, Vol. 3: Educational applications. Thousand Oaks, CA: Sage Publications, Inc.

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

White, O.R., & Neely, M. (2004). *The chart book: An overview of standard celeration chart conventions and practices*. Kansas City, KS: Behavior Research Company.

*Wong, S.E., Floyd, J., Innocent, A.J., & Woolsey, J.E. (1991). Applying a DRO schedule and compliance training to reduce aggressive and self-injurious behavior in an autistic man: A case report. *Journal of Behavior Therapy & Experimental Psychiatry, 22*, 299-304.

*Woods, T. S. (1981). Reducing severe aggressive and self-injurious behavior: A nonintrusive, home based approach. *Behavioral Disorders, 7*, 180-188.

*Wulbert, M., Barach, R., Perry, M., Straughan, J., Sulzbacher, S., Turner, K., & Wiltz, N. (1974). The generalization of newly acquired behaviors by parents and child across three different settings: A study of an autistic child. *Journal of Abnormal Child Psychology, 2*, 87-98.