Single Subject Research: A Synthesis of Analytic Methods

Fahad Alresheed, Brittany L. Hott, and Carmen Bano
Texas A&M University-Commerce

Historically, the synthesis of single subject design has employed visual inspection to yield significance of results. However, current research is supporting different techniques that will facilitate the interpretation of these intervention outcomes. These methods can provide more reliable data than employing visual inspection in isolation. This article compares the different techniques, compares the benefits of utilizing these techniques in addition to visual inspection, the limitations of each technique being reviewed, and evidence for combining traditional statistical measures with visual inspection.

Keywords: Single Subject Research, Visual Inspection, Nonoverlap Methods, Quantitative Research Synthesis

Single subject research is an experimental design that strives to record relationships between independent and dependent variables (Gast, 2010; Kennedy, 2005). Each participant serves as both the control and the experimental condition by using either reversal or multiple-baseline designs (Horner, Carr, Halle, McGee, Odom, & Wolery, 2005). It is a popular design used in a wide variety of settings; however, it is primarily applied in educational research. Traditionally, visual inspection has been used to interpret and understand results (Horner, et al., 2005; Park, Marascuilo, & Gaylord-Ross, 1990; Parker, Hagan-Burke, & Vannest, 2007). Visual inspection involves interpreting performance based on visual interfaces, such as line graphs, by noting level and slope changes, and differences between baseline and intervention data (Scruggs, Mastropieri, & Regan, 2006). Given the high subjectivity of visual inspection, using other statistical measures would add more objectivity when interpreting results (Park et al., 1990; Scruggs et al., 1987).

A variety of methods for interpreting the results of single subject research have emerged in the last two decades. Nonoverlap methods can provide more objective analysis of data within single subject research (Alresheed, & Shipman, 2012; Mastropieri & Scruggs, 2013).

Visual Inspection

Visual inspection has been the most common method of analysis for evaluating the effect of an intervention in single-subject designs (e.g., Alberto & Troutman, 2009; Bengali & Ottenbacher, 1998; Mastropieri & Scruggs, 1986). This has been true even though visual inspection reliability has often been questioned (Gast, 2010; Kazdin, 2011; Park et al., 1990). Further, Park at al. (1990) found that inter rater reliability for visual inspection is weak and recommend using statistical procedures in addition to visual inspection.

Given, decisions made by looking at a graph can be subjective and inconsistent; this has caused a problem within single-subject research for a long time. However, visual analyses should not be abandoned. As Franklin, Gorman, Beasely, and Allison (1996) point out, visual analyses of single-subject data are necessary tools. Interpretation of treatment effect requires both, visual analysis and statistical analysis (Robey, Schultz, Crawford, & Sinner, 1999). Most researchers using single case designs still base their inferences on visual analysis, but several quantitative methods have been proposed. Each has flaws, but some methods are likely to be more useful than others (Kratochwill et al., 2010). The fact that limitations can be identified in any quantitative research method for synthesizing single-subject research should not be taken as evidence that subjective, qualitative reviews themselves are without flaws (Scruggs & Mastropieri, 2013). While researchers are still developing methods (Campbell & Herzinger, 2010), a lack of consensus regarding the most appropriate metric remains (Maggin, O'Keeffe, & Johnson, 2011). Two types of strategies have been proposed for assessing the magnitude of effect in single-subject research: regression and non-regression approaches. Non-regression models include simpler nonoverlap methods (Parker, Vannest, & Davis, 2011).

The number of nonoverlap methods for SCR has increased considerably over the past decade, and these methods can be easily confused. Some of these methods are very similar and some are closely related to other well-known statistical summaries. According to Campbell (2013), the introduction of PND (a nonoverlap method) was a "first wave" of SCD meta-analytic methods that was followed by a "second wave" of improved methods such as IRD and PAND. The "third wave" most likely will take the shape of more sophisticated linear modeling techniques, such as hierarchical linear modeling (HLM). All nonoverlap methods share the benefit of blending well with visual analysis of graphed data. In addition, all nonoverlap techniques are easy to use. They all can be calculated with a pencil from a data plot. Some appear more complex than others but after initial practice prove to be user friendly for consumers in schools and clinics (Parker et al., 2011). The purpose of the following research questions were addressed:

1. How do we synthesize single-subject research?
2. What are the benefits of each method?
3. What are the limitations of each method?
4. Are there benefits of combining statistical methods with visual inspection?

Method

Two researchers independently reviewed Sage Journal Online, Science Direct, SpringerLink, Taylor & Francis, Eric, Education Journals-ProQuest, Academic Search Complete, Education Research Complete, PsychINFO, and Wiley Online Library databases. Search terms included were: (a) single subject research, (b) evidence-based practices, (c) meta-analysis, (d) quantitative synthesis, (e) synthesis of single-subject research, (f) PND, (g) PAND, (h) PEM, (i) PEM-T, (j) MBLR, (k) PZD, (l) IRD, (m) PDO, (n) NAP, and (o) PNCD. The articles had to be published in a peer-reviewed journal. After the search, ten methods to synthesize single-subject research were found. The authors employed fictional data to create 14 graphs to illustrate the process of calculation of each method.

Results
Percentage of Non-overlapping Data (PND)

Scruggs, Mastropieri, and Casto (1987) used the percentage of non-overlapping data (PND) to calculate the effect sizes of single-subject research.

PND is the oldest method created for synthesizing single-subject studies. Although research has shown that PND has some empirical limitations, this estimator is one of the most commonly reported effect size for the quantitative synthesis of single-subject research (Maggin, Briesch, & Chafouleas, 2013) It is also the best known overlap method. According to Scruggs et al. (1987), a very important criterion to decide whether a treatment is effective is the percentage of overlapping data between treatment and baseline. If performance during an intervention phase does not overlap with performance during the baseline phase the treatment is considered effective. PND can easily be calculated in the great majority of cases, and provides a good measure of treatment effectiveness (Kazdin, 1978). PND is calculated by counting the number of treatment data points that exceeds the highest baseline data point and dividing this number by the total number of data points in the treatment phase. In mathematical language, the formula is the following:

$$PND = \frac{\text{Number of intervention data exceeding the highest baseline data point}}{\text{Total number of data points in the intervention phase}} \times 100$$

Figure 1. Formula for PND Calculation

PND scores range from 0 to 100%. A PND of less than 50 reflects unreliable treatment. A PND of 50-70% reflects questionable effectiveness. A PND of 70-90% reflects a fairly effective treatment. Finally, a PND of more than 90% reflects a highly effective treatment (Wendt, 2007). See figure 2 that shows an example of PND calculation.
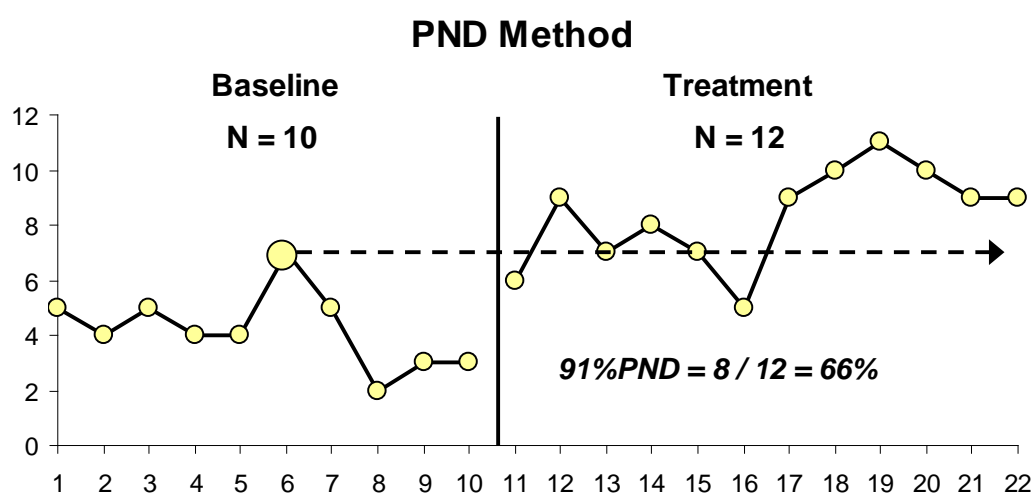

Figure 2. Example of PND Calculation

PND, however, has some weaknesses. According to Allison and Gorman (1993), PND might not show effects in at least three situations. Figures 2, 3, and 4 from show examples of these three situations (Allison & Gorman, 1993).

1. When outliers are present in the baseline phase. Outliers can distort the magnitude of effect estimates provided by PND (Campbell, 2013; Manolov & Solanas, 2009). See figure 3 that indicates a situation in which the treatment has a clear positive effect but, because of the presence of an outlier in the baseline phase, the PND returns a value of zero, indicating no effect.
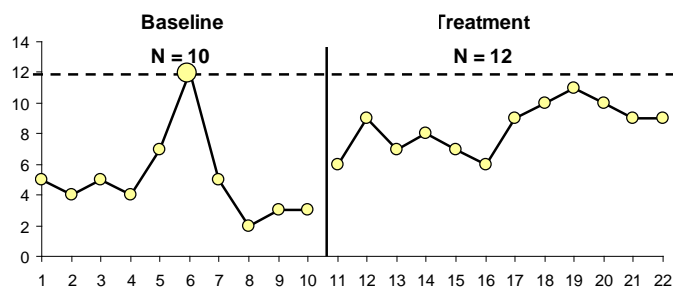

Figure 3. Presence of an Outlier in PND Calculation

2. When the treatment has a detrimental effect. See figure 4 that indicates a situation in which the treatment clearly has a detrimental effect (note that the goal of treatment was to increase the behavior) but, because of the presence of an outlier in the treatment phase, the PND returns a value of 10 indicating a small positive effect.
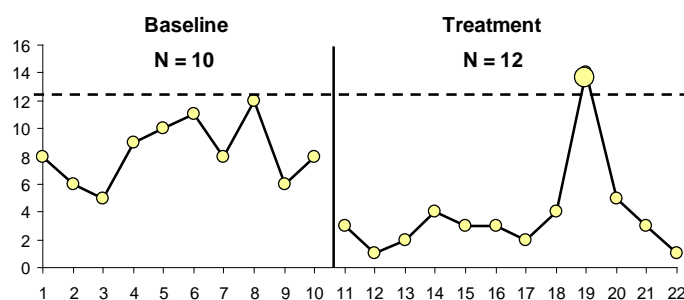

Figure 4. Detrimental Effect Not Shown in PND Calculation

3. When trend is present in the data. See figure 5 that indicates a situation in which the treatment has no effect but merely allows a pre-existing trend to continue. However, the PND value is 100 indicating the maximum possible effect.
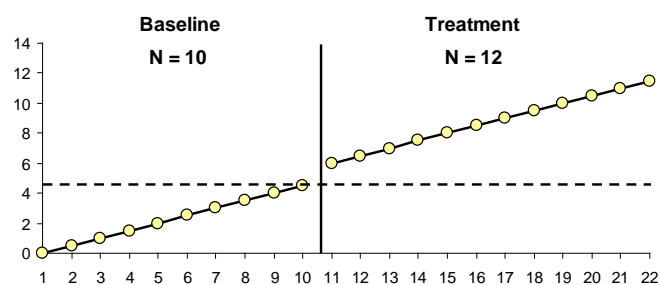
Figure 5. Pre-Existing Trend in PND Calculation

In addition, PND shows changes in level but ignores changes in slope. Figure 6 shows a situation in which the treatment clearly reverses a downward trend but the PND shows a value of zero indicating no effect (Allison & Gorman, 1993).
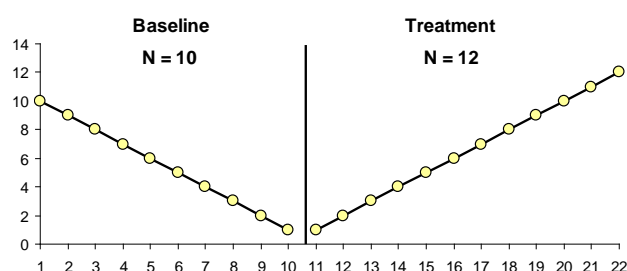

Figure 6. Downward Trend Reversal in PND Calculation

Four limitations of PND have been identified by Parker, Hagan-Burke, and Vannest (2007). First, PND is neither an effect size nor related to an accepted effect size, so it needs its own interpretation guidelines. Second, PND has unknown reliability, as it lacks a known sampling distribution. The third weakness is that PND ignores all baseline data except for one data point, which because of its extremity, is likely the most unreliable. The fourth limitation is that PND lacks sensitivity or discrimination ability, as it nears 100%, for very successful interventions.

Percentage of all Non-overlapping Data (PAND)

Parker, Hagan-Burke, and Vannest (2007) recently offered a variation on PND, the Percentage of All Non-Overlapping Data (PAND). With this approach, the total number of data points that do not overlap between baseline and intervention phases is identified.

Additionally, Park et al. (2007) indicate overlapping data points are the minimum number of points that would have to be transferred across phases for complete data separation. PAND is calculated by identifying the number of overlapping data points and dividing this number by the total number of points. The final number is subtracted from 100. In

mathematical language, the formula is the following:

$$\text{PAND} = 100 - \frac{\text{Number of overlapping data points}}{\text{Total number of data points}}$$

Figure 7. Formula for PAND Calculation

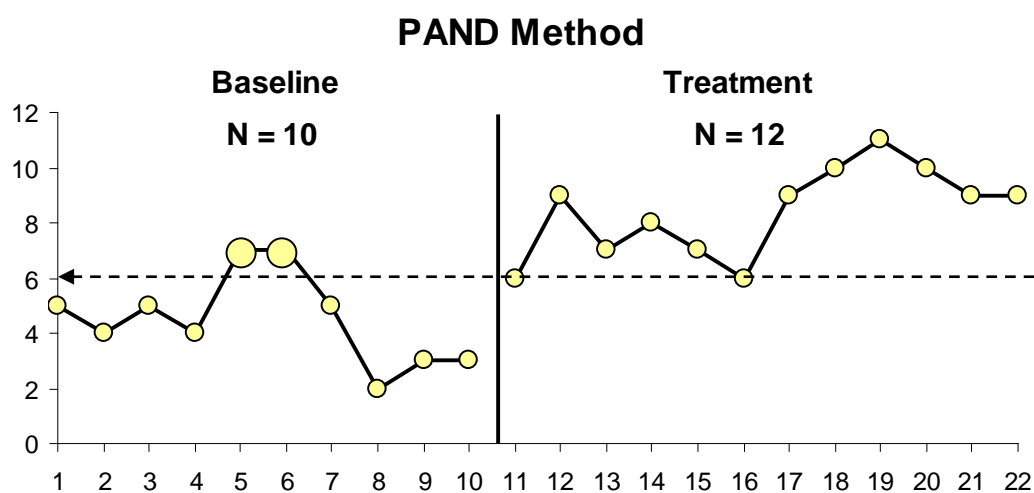See figure 8 that shows an example of PAND calculation.

PAND Method

Figure 8. Example of PAND Calculation

Like PND, PAND reflects non-overlap data between baseline and treatment phases but they are different in important ways. PAND uses all data from both phases, avoiding the PND focus on one unreliable data point (Parker et al., 2007). PAND can be translated to Pearson's Phi and Phi2, to determine effect size (Wendt, 2009). Also, Phi2 can be transformed to Cohen's d, a recognized effect size in another metric (Parker et al., 2007). The data requirements for PAND are minimal, mainly, a minimum of 20 data points. Two limitations cited for PND are not solved by PAND. The first is insensitivity at the upper end of the scale. When there is no data overlap between phases, both PND and PAND give a 100% score, regardless of the distance between

the two data clusters. PAND's second limitation is that it doesn't control positive baseline trend. Like PND, it does not try to adjust for prior rate of improvement. Before identifying a causal link between intervention and behavior, one must consider any positive baseline trend (Parker, Cryer, & Byrns, 2006). A large effect size alone does not imply that change was due to the intervention.

Percentage of Data Exceeding the Median (PEM)

A third non-parametric statistical method is percentage of data exceeding the median (PEM). The null hypothesis of the PEM approach is that if the treatment has no effect, the data points in the treatment phase will concentrate around the middle line. The data points have 50% of chance

of being above and 50% chance of being below the median of the previous baseline phase (Ma, 2006). PEM identifies the percentage of data points exceeding the median of the baseline phase (Ma, 2006). PEM is calculated by first locating the median point or point between the two median points in the baseline data. The median is the middle in the distribution (Cohen & Swerdlik, 2009). Then, a line will be drawn from the median into the treatment phase. Finally, the number of

data points above the median line will be counted. If the behavior is expected to increase, the points above the line will be counted. If the behavior is expected to decrease, the points below the line will be counted (Wendt, 2009). The total number will be divided by the total of data points in the treatment phase. See figure 9 that shows an example of PEM calculation.
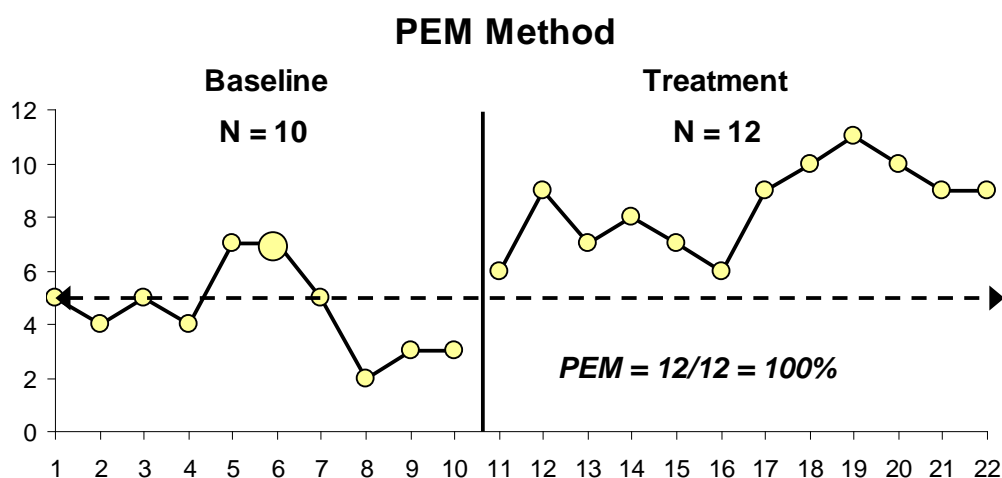

Figure 9. Example of PEM Calculation

PEM scores range from 0 to 1. A PEM of less than .7 reflects treatment that is not effective. A PEM of .7 to .9 reflects moderate effectiveness. A PEM of .9 to 1 reflects a highly effective treatment (Wendt, 2007). A critical advantage of PEM over PND is the fact that PEM reflects an effect size in the presence of floor or ceiling baseline data points. PEM has also some important limitations: It is insensitive to magnitude of data points above the median and it does not consider variability and trend.

Percentage of Data Exceeding the Median Trend (PEM-T)

Percentage of data exceeding a median trend (PEM-T) is another

nonparametric method. PEM-T is the only overlap method considering the trend in the data of the baseline. PEM-T is calculated by using first the split-middle technique (White & Haring, 1980), a common approach to determine trend in SSD data. After drawing a trend line in the baseline phase using split-middle technique, this line is extended to the treatment phase. The data points in the treatment phase above the trend line are counted and the percentage of non-overlap is calculated. See figure 10 that shows an example of PEM-T calculation.
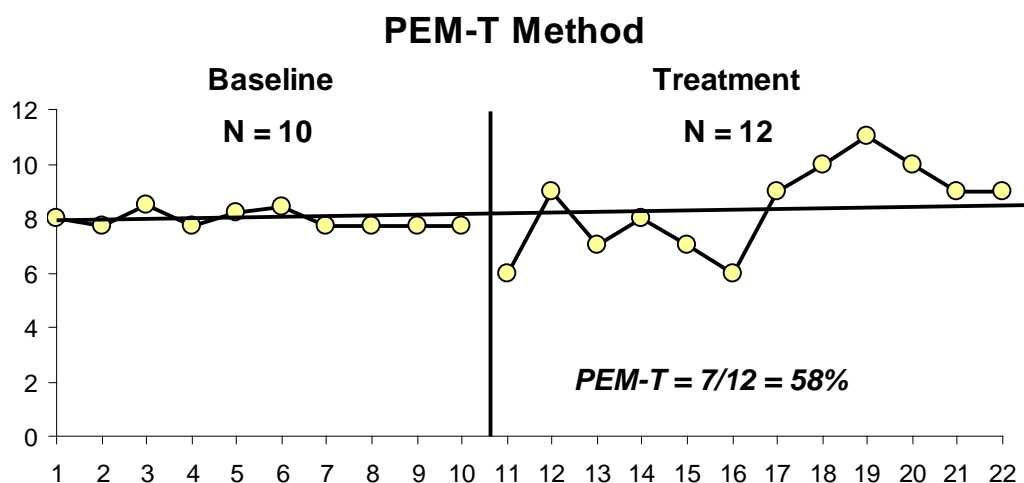
PEM-T Method

Figure 10. Example of PEM-T calculation

Mean Baseline Difference (MBLR)

Mean baseline reduction (MBLR) or mean baseline difference (MBD), also called percentage reduction, is another frequently used non-parametric statistical method. It measures the average reduction or increase of behavior from baseline (O'Brien & Repp, 1990). When the goal of the treatment is to reduce behavior, MBLR is calculated by subtracting the mean of the treatment points from the mean of the baseline points, then dividing by the mean of the baseline points and multiplying by 100 (Campbell, 2003).

$$MBLR = \frac{\text{Mean of baseline} - \text{Mean of treatment}}{\text{Mean of baseline}} \times 100$$

Figure 11. Formula for MBLR Calculation (Behavior Reduction)

When the goal of the treatment is to improve behavior, MBLR is calculated by subtracting the mean of the baseline points from the mean of the treatment points, then dividing by the mean of the baseline points and multiplying by 100 (Campbell, 2003).

$$MBLR = \frac{\text{Mean of treatment} - \text{Mean of baseline}}{\text{Mean of baseline}} \times 100$$

Figure 12. Formula for MBLR Calculation (Behavior Improvement)

A variation of this approach is using the same formula including only the last three points in the treatment and the last three points in the baseline. See figure 13 that shows an example of MBLR calculation.
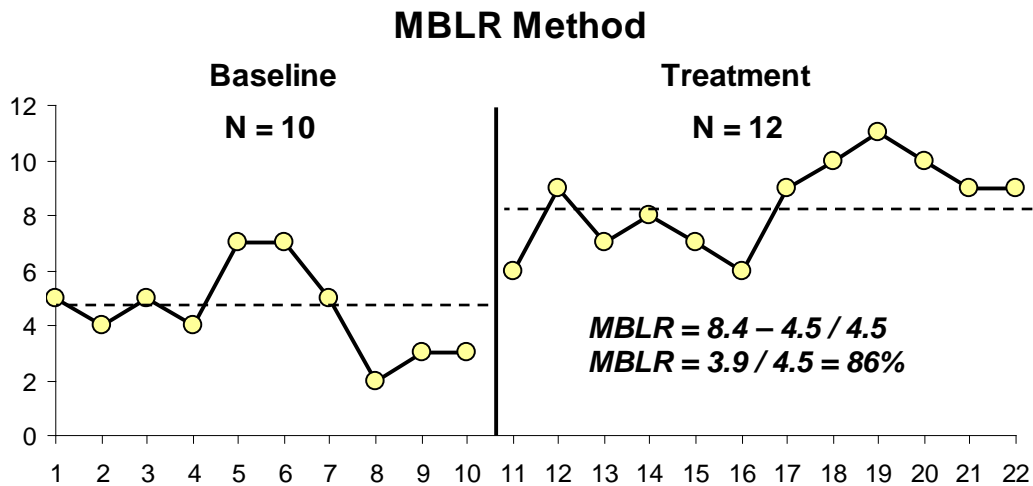
MBLR Method

Baseline
N = 10

Treatment
N = 12

MBLR = 8.4 − 4.5 / 4.5
MBLR = 3.9 / 4.5 = 86%

Figure 13. Example of MBLR calculation

Percentage of Zero Data (PZD)

Percentage of zero data (PZD) is the degree to which behavior is eliminated in treatment (Harvey, Boer, Meyer, & Evans, 2009) so it fits only certain scales and goals (Parker, Vannest, & Davis, 2011). It is easy to calculate but can be distorted if treatment is terminated immediately after zero data point occurs (Harvey et al., 2009). Because it calculates the degree to which intervention completely suppresses targeted behavior (Scotti, Evans, Meyer, & Walker, 1991), future quantitative reviews that examine treatments designed to eliminate problem behaviors in the context of a single-subject design should include PND and PZD metrics. PND and PZD scores have been found to be non-redundant indicators of treatment outcome (Campbell, 2003). However, when reduction of symptoms (vs. suppression of symptoms) is the focus of intervention, PZD will not constitute an appropriate measure to summarize treatment effects (Campbell, 2004). PZD is calculated by locating the first data point in the treatment phase that reaches zero and calculating the per cent of data points recorded in the treatment phase, including the first zero, that remain at zero (Scotti, et al., 1991). The PZD score is considered a more stringent indicator of treatment efficacy as it requires target behaviors to reach and stay at zero levels throughout treatment (Campbell, 2004).

$$PZD = \frac{\text{Number of 0 treatment data points after first 0}}{\text{Number of treatment data points after first 0}} \times 100$$

Figure 14. Formula for PZD Calculation

Just like PND, PZD statistics are overly sensitive to outliers and trend (Allison & Gorman, 1993). See figure 15 that shows an example of PZD calculation.
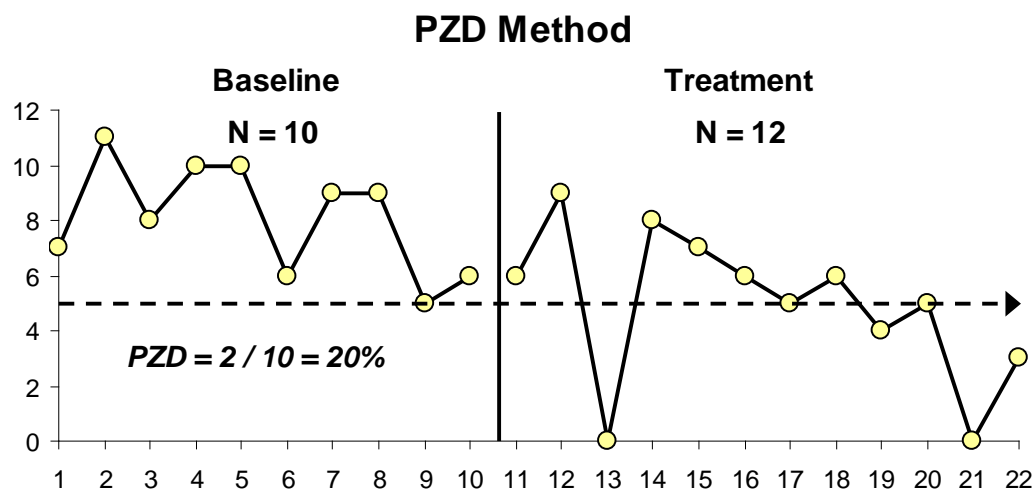
PZD Method

Figure 15. Example of PZD calculation

PZD scores range from 0 to 100%, with higher scores indicating more effective treatments (Shogren, Fagella-Luby, Bae, & Wehmeyer, 2004). A PZD score of less than 18% reflects an ineffective treatment. A score of 18% to 54% reflects a treatment of questionable effectiveness. A score between 55% and 80% reflects fair effectiveness. Finally, a score higher than 80% reflects a highly effective treatment (Scotti et al., 1991).

Improvement Rate Difference (IRD)

Improvement rate difference (IRD) is a measure -closely related to PAND - that expresses the difference in successful performance between baseline and intervention phases. IRD has a solid record of use in evidence-based medicine, under the name of risk reduction or risk difference (Parker, Vannest, & Brown, 2009).

According to Parker et al. (2009), IRD's advantages include (a) accessible interpretation as the difference in improvement rates between baseline and treatment phases; (b) simple hand-calculation, easily explained to most educators; (c) compatibility with PND from visual analysis; (d) known sampling distribution, so confidence intervals are available; (e) proven track record (as risk difference) in hundreds of evidence-based medical research studies; (f) few data distribution assumptions; and (g) application to complex single-case research designs and multiple data series.

To calculate IRD, two improvement rates (IRs) must be calculated first (Cochrane Collaboration, 2006; Sackett, Richardson, Rosenberg, & Haynes, 1997). The IR for each phase is defined as the number of "improved data points" divided by the total data points in that phase:

$$IR = \frac{\# \text{ improved data points}}{\# \text{ total data points}}$$

Figure 16. Formula for IR Calculation

An improved data point in baseline is defined as one that ties or exceeds any data point in the treatment phase. An improved data point in the treatment phase is defined as any which exceeds all data points in the baseline phase. "Exceeds" refers to higher levels of behaviors we wish to increase (e.g., homework completion) and to lower levels of behaviors we wish to decrease (e.g., tantrums). Improved data points are identified visually (Parker et al., 2009).

IRD is then calculated as the difference of the improvement rate of the treatment phase minus the improvement rate of the baseline phase: IRD = IRT - IRB

See figure 17 that shows an example of IRD calculation.

IRD Method

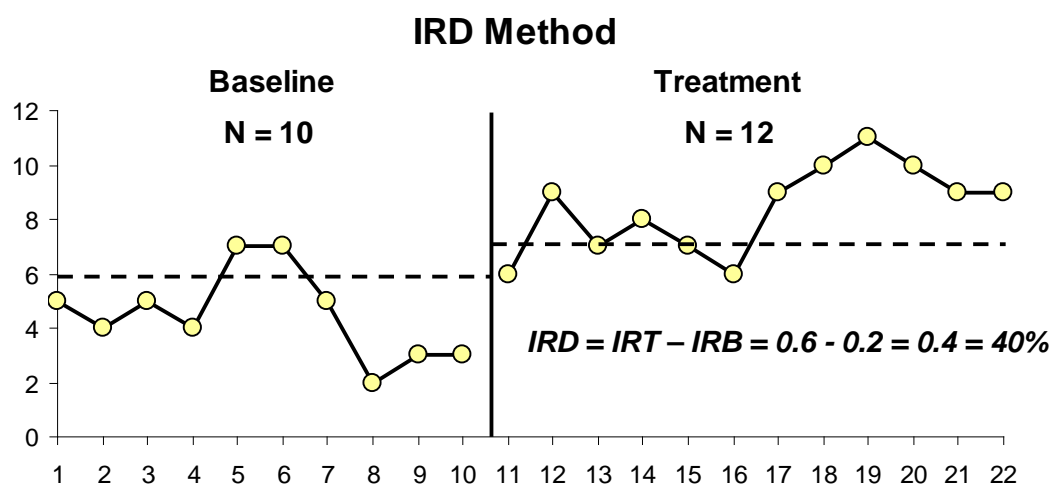$$IRD = IRT - IRB = 0.6 - 0.2 = 0.4 = 40\%$$

Figure 17. Example of IRD calculation

IRD scores range from 0 to 100% or 1.00. A negative IRD score is possible, indicating deterioration below baseline levels (Parker et al., 2009). From comparing visual ratings with IRD, Parker et al. (2009) estimated tentative benchmarks. Very small and questionable effects scored about .50 and below. Moderate-size effects had IRD scores of around .50 to .70. Effects rated as large and very large generally received IRD scores of .70 or .75 and higher.

When a baseline trend is prominent, a calculated effect size cannot fairly represent treatment effectiveness. In those cases, parametric (Allison & Gorman, 1993) or nonparametric (White & Haring, 1980) techniques can control the baseline trend. After applying a trend-compensating formula, IRD can safely be used without modification (Parker et al., 2009).

Pairwise Data Overlap (PDO)

Pairwise Data Overlap (PDO) calculates the overlap of all possible paired data comparisons between baseline and intervention phases (Parker & Vannest, 2007). To calculate PDO, the following steps must be followed (Wolery, Busick, Reichow, & Barton, 2008):

Compare baseline data point with all intervention data points

For each data point in baseline, count the number of nonoverlapping (nol) data points in treatment phase.

1. Sum all counts for all data points in Step 2.
2. Count the number of data points in the baseline.
3. Count the number of data points in the treatment phase.

4. Multiply the two counts (Steps 4 and 5) to determine the total number of
5. pairwise comparisons.

6. Divide the sum from Step 3 by the product from Step 4

$$PDO = \frac{9+9+9+9+9+4+7+9+9+9}{10 \ X \ 9} = 92\%$$

Figure 18. Formula for PDO Calculation

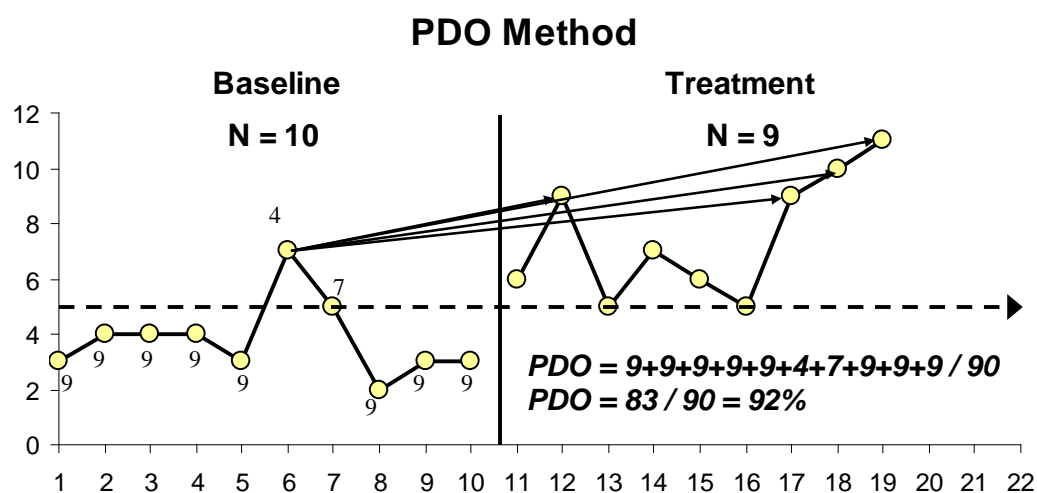See figure 19 that shows an example of PDO calculation.

Figure 19. Example of PDO calculation

PDO has advantages and limitations. PDO produces more reliable results with single subject designs than other non-parametric indices, and relates closely to established effect sizes (e.g., Pearson r, Kruskal-Wallis W). However, it takes slightly longer to calculate since it requires that individual data point results be written and added, making calculation laborious for long and crowded data series (Wendt, 2009).

Non-overlap of All Pairs (NAP)

NAP was developed mainly to improve upon existing SCR overlap-based methods: PND, PAND, and PEM.

Nonoverlap of all pairs (NAP) is interpreted as the percentage of all pairwise comparisons across baseline and treatment phase, which show improvement across phases or, more simply, the percentage of data, which improve across phases (Parker, Vannest, & Davis, 2011). A simpler wording is the per cent of non-overlapping data between baseline and treatment phases. The concept of score overlap is identical to that used by visual analysts of SCR graphs and is the same as is calculated in the other overlap indices, PAND, PEM and PND (Parker & Vannest, 2009). NAP is a "complete" nonoverlap

index as it individually compares all baseline and treatment phase data points. (Parker, Vannest, & Davis, 2011).

Although easily calculated by computer software, hand calculation requires some practice. The NAP hand-calculation method compares each baseline phase data point with each treatment phase data point. Parker and Vannest (2009) give a clear explanation of the hand-calculation process. First, it is necessary to calculate the number of Pos following these steps:

1. First, the total of paired comparisons (Pairs) across phases is calculated as number of points in baseline x number of points in treatment phase = 6 x 7 = 42.
2. Next, the "overlap zone" between phases is identified and within that zone only pairs that show decline (Neg) and ties (Ties) are counted.
3. These (Neg, Ties) are subtracted from number of Pairs to get the number of Pos.

Once the number of Pos is obtained, NAP is calculated using the following formula:

$$NAP = \frac{Pos + (.5 \ x \ no. \ of \ Ties)}{No. \ of \ Pairs}$$

Figure 20. Formula for NAP Calculation

See figure 21 that shows an example of NAP calculation.

NAP Method

Pairs = 10 x 12 = 120
Pos = 120 - 3 = 117
NAP = 117 + .5 x 1 / 120 = 117.5 / 120 = .98 = 98%
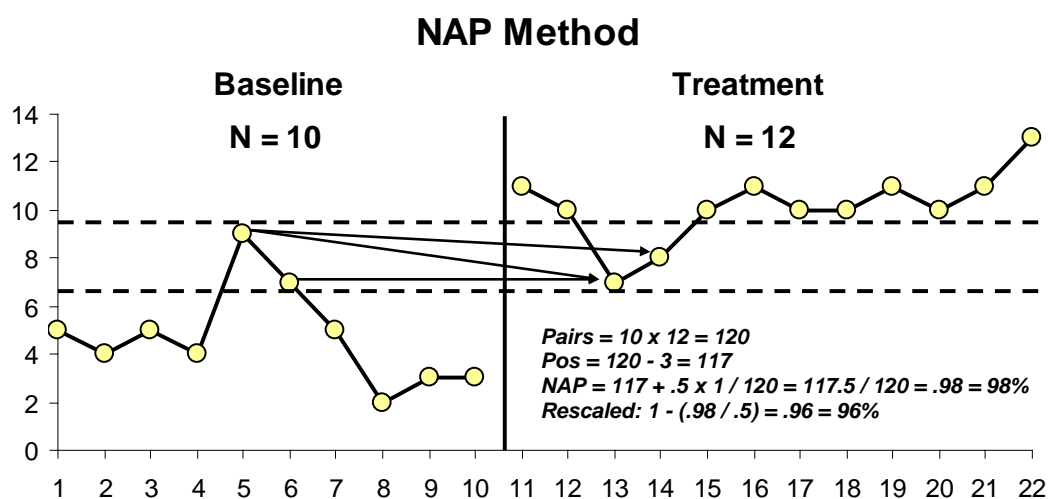Rescaled: 1 - (.98 / .5) = .96 = 96%

Figure 21. Example of NAP calculation

NAP is scaled from 50% to 100%, where 50% is a chance-level result. To rescale NAP to a 0% to 100% scale, this formula must be used: NAP 0-100 = 1 – (NAP50-100 / .5) (Parker, Vannest, & Davis, 2011). NAP also is obtained directly as the AUC per cent from a ROC analysis. NAP also can be calculated from intermediate output of the Wilcoxon Rank-Sum Test, usually located in statistical packages within "Two Sample t-Test" or "Non-Parametric Test" modules. Parker and Vannest (2009) offer a tentative NAP range based on visual judgements of 200 data sets: weak effects: 0–.65; medium effects: .66–.92; large or strong effects: .93–1.0. Transforming NAP to a zero chance level gives these corresponding ranges: weak effects: 0–.31; medium effects: .32–.84; large or strong effects: .85–1.0.

NAP should offer five comparative advantages: First, NAP should discriminate better among results from a large group of published studies. Earlier research indicated less than optimal discriminability by the other three nonoverlap indices (PND, PAND, PEM). A second advantage is less human error in calculations than the other three hand-calculated indices. On uncrowded graphs, PND, PAND and PEM are calculated with few errors, but not so on longer, more compacted graphs. A third advantage sought from NAP was stronger validation by R^2, the leading effect size in publication. Since NAP entails more data comparisons than other nonoverlap indices, it should relate more closely to R^2, which makes fullest use of data. The fourth anticipated advantage of NAP was stronger validation by visual judgments. The reason for that expectation was that visual analysis relies on multiple and complex judgments about the data, which should be difficult to capture with simpler indices such as PEM and PND. NAP is not a test on means or medians, but rather

on location of the entire score distribution, and is not limited to a particular hypothesized distribution shape (Parker & Vannest, 2009).

Percentage of Non-overlapping Corrected Data (PNCD)

The last non-parametric statistical method, PNCD, is a modification of PND. The data-correction procedure focuses on removing the baseline trend from data prior to estimating the change produced in the behavior as a result of intervention. Unstable baselines have been regarded as undesirable, but they can be common in applied settings in which the introduction of the treatment is subjected to factors that cannot always be controlled by the practitioners (Manolov & Solanas, 2009). Although a professional might be reluctant to initiate the intervention when there is trend in data, treatment administration may be imposed by institutional time schedules, a client's availability, and so on. In such a case, some kind of statistical control is advisable (Kazdin, 1978).

As mentioned previously in this chapter, PND is the most frequently applied overlap methods for quantifying treatment effectiveness in single-case studies and also in meta-analyses. Despite its attractiveness to psychologists, PND is not trouble-free (Allison & Gorman, 1994). Therefore, the main objective of the developers of PNCD was to propose a modification of the PND procedure that was intended to overcome some of PND's limitations (Manolov & Solanas, 2009). A data-correction procedure is to be implemented prior to applying the PND in order to eliminate from the data a possible pre-existing trend that was not related to the introduction of the intervention. Since the proposal is basically a modification of PND—adding an initial data-correction step—the procedure is called "percentage of nonoverlapping corrected data" (Manolov & Solanas, 2009). Manolov and

Solanas (2009) describe the following procedure to calculate PNCD:

1. Difference the baseline data points and obtain the differenced series. This means subtracting the previous data point from each data point. In the example, the differenced series has the following data points:

   0 (4 – 4), 1 (5 – 4), -2 (3 – 5), 4 (7 – 3)

2. Compute the mean of the previous series. The average of 0, 1, -2, and 4 is 0.75.

3. Compute the trend-correction factor for each data point: the mean of the differenced series, multiplied by Tt (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). In the example, the correction factor values are:

.75 x 1, .75 x 2, . . . , .75 x 10.

4. Perform the data correction subtracting the corresponding correction factor from each original data point. After the correction phase, A consists of:

   (.75 x 1), 2.5 (.75 x 2), 2.75 (.75 x 3), 0 (.75 x 4), and 3.25 (.75 x 5). After the correction phase, B consists of:

   2.5 (.75 x 6), 2.75 (.75 x 7), 3 (.75 x 8), .25 (.75 x 9), and 1.5 (.75 x 10).

5. Apply PND: None of the phase B data points is greater than the phase A highest value (3.25) and, therefore, PNCD = 0%.
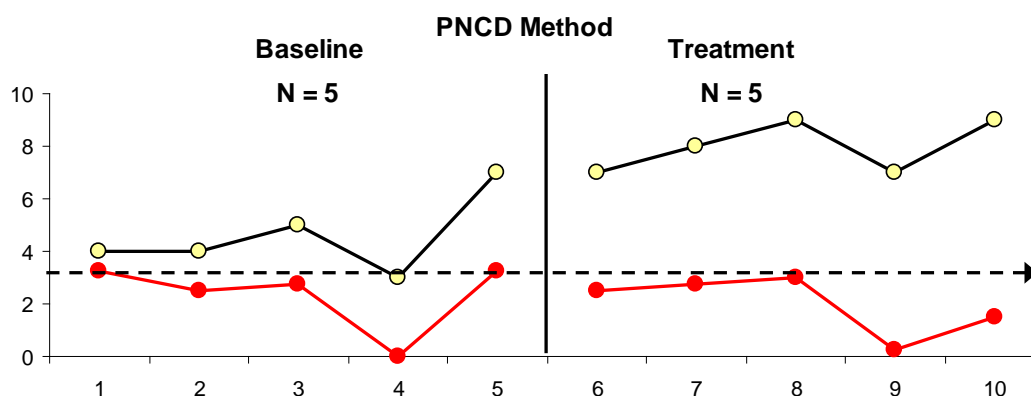
See figure 22 that shows the example of PNCD calculation.

Figure 22. Example of PNCD calculation

The first four steps are expected to remove any trend from the data and thus to avoid inflation in the percentages obtained by means of PND. Trend is not estimated from the whole data series, since a change in level between the phases may be confounded for trend, and such a correction may remove the intervention effect.

Discussion

The previously reviewed ten non-regression measures have shown to be concrete and objective. These methods are able to accurately provide a measure of effect that is more objective than visual inspection alone. Using statistical analysis to interpret results offers more concrete and objective interpretation; however, each method has its unique limitations that need to be taken into consideration before applying it to a specific study. A study conducted by Wolery, Busick, Reichow, & Barton, (2009) showed weak results when solely using an overlap method. One method may be more appropriate for one

study than another, depending on data gathered and desired outcomes. However, there is consensus regarding the most appropriate method for estimating the magnitude of treatment effect for single-subject research. Thus, the reporting of multiple metrics is highly recommended to determine whether consistent results were observed (Maggin, Briesch, & Chafouleas, 2013). While all of these non-regression measures have their limitations, evidence has also been provided to support combining statistical analysis with visual inspection. A study conducted by Park, Marascuilo, and Gaylord-Ross (1990) showed that inter rater reliability for only visual inspection was low but, when combining visual inspection and statistical analysis, inter-rater reliability strengthened. Future research in combining these two analyses together is warranted.

References

Alberto P. A, Troutman A. C. (2009). *Applied behavior analysis for teachers.* Upper Saddle River, NJ: Prentice Hall.

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavioral Research Theory. 31*(6), 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 & Therapy, 32*, 885-890.

Alresheed, F. M., & Shipman, K. S. (2012, November). *Alternative Methods for Synthesizing Single Subject Research.* Poster presented at Texas A&M University, Galveston, TX.

Bengali, M. K., & Ottenbacher, K. J. (1998). The effect of autocor- relation on the results of visually analyzing data from single- subject designs. *American Journal of Occupational Therapy, 52,* 650-655.

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.

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

Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education, 34,* 20–25.

Campbell, J.M. & Herzinger, C.V., (2010). Statistics and single subject research methodology. D. L. Gast (Ed.) *Single subject research design in behavioral sciences.* Lawrence Erlbaum Publishers.

Cochrane Collaboration. (2006). *Cochrane handbook for systematic reviews of interventions*. Retrieved from http://www.cochrane.org/index_author s_ researchers.htm

Cohen, R. J., & Swerdlik, M. E. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York: McGraw-Hill.

Franklin, R. D., Gorman, B. S., Beasely, T. M., & Allison, D. B. (1996). *Graphical display and visual analysis. Design and Analysis of Single-Case Research.* Mahwah, NJ: Lawrence Erlbaum.

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

Harvey, S. T., Boer, D., Meyer, L. H., & Evans, I. M. (2009). Updating a meta-analysis of intervention research with challenging behavior: Treatment validity and standards of practice. *Journal of Intellectual & Developmental Disabilities, 34*, 67–80.

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

Kazdin, A. E. (1978). Methodological and interpretive problems of single-case experimental designs. *Journal of Consulting and Clinical Psychology, 46*, 629-642.

Kazdin, A. E. (2011). *Single-case research design.* New York, NY: Oxford University Press.

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

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*, 598-617.

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self- management literature-base. *Remedial and Special Education, 34,* 44–58.

Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985-2009. *Exceptionality: A Special Education Journal, 19*(2), 109-135.

Manolov, R., & Solanas, A. (2009). Percentage of nonoverlapping corrected data. *Behavior Research Methods, 41,* 1262-1271.

Mastropieri, M. A., & Scruggs, T. E. (1986). Early intervention for socially withdrawn children. *The Journal of Special Education, 19,* 429-441.

O'Brien, S., & Repp, A. C. (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe or profound retardation. *Journal of the Association for Persons With Severe Handicaps, 15,* 148–159.

Park, H., Marascuilo, L., & Gaylord-Ross, R. (1990). Visual inspection and statistical analysis of single-case designs. *Journal of Experimental Education, 58,* 311–320.

Parker, R. I., Cryer, J. & Byrns, G. (2006). Controlling trend in single case research. *School Psychology Quarterly, 21*(3), 418-440.

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

Parker, R. I., & Vannest, S., (2007). *Pairwise data overlap for single case research.* Unpublished manuscript.

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). *Behavior Therapy, 40*(4), 357–367.

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

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single case research: A review of nine non-overlap techniques. *Behavior Modification, 35*, 302–322.

Robey, R. R., Schultz, M. C., Crawford, A. B., & Sinner, C. A. (1999). Review: Single-subject clinical outcome research: designs, data, effect sizes, and analyses. *Aphasiology*, *13*(6), 445-473.

Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine: How to practice and teach EBM*. London: Churchill Livingstone.

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. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34,* 9–19.

Scruggs, T. E., Mastropieri, M. A., & Regan, K. S. (2006). Statistical analysis for single subject research designs. *Applications of Research Methodology: Advances in Learning and Behavioral Disabilities , 19,* 33-53.

Shogren, K. A., Fagella-Luby, M. N., Bae, S. J., & Wehmeyer, M. I. (2004). The effect of choice making as an intervention for problem behavior: A meta-analysis. *Journal of Positive Behavior Interventions, 4*, 228–237.

Wendt, O. (2007, November). *Meta-analysis of single-subject experimental designs*. Paper presented at the Panel "Roles of Single-Subject Experimental Designs in EBP Times" at the Annual Convention of the American Speech-Language-Hearing Association (ASHA), Boston, MA.

Wendt, O. (2009, May). Calculating effect sizes for single-subject experimental designs. Paper presented at the Ninth Annual International Campbell Collaboration Colloquium, Oslo, Norway.

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

Wolery, M., Busick, M., Reichow, R., & Barton, E. E. (2008). *Quantitative synthesis of single subject research.* Paper presented at the Conference on Research Innovations in Early Intervention, San Diego, CA.