

# Examining Estimates of Effect Using Non-Overlap of All Pairs in Multiple Baseline Studies of Academic Intervention

Shawna Petersen-Brown · Abbey C. Karich · Frank J. Symons

Published online: 3 May 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** The inclusion of single-case design (SCD) studies in meta-analytic research is an important consideration in identifying effective evidence-based practices. Various SCD effect sizes have been previously suggested; non-overlap of all pairs (NAP) is a recently introduced effect size. Preliminary field tests investigating the adequacy of NAP are promising, but no analyses have been conducted using only multiple baseline designs (MBDs). This preliminary study investigated typical values of NAP in MBDs, investigated agreement with visual analysis, and suggested cut scores for interpreting a NAP effect size. Typical values of NAP in MBDs were larger compared to a previous meta-analysis of studies using AB, MBD, or ABAB withdrawal designs, and agreement of suggested cut scores and visual analysis was moderate.

**Keywords** Effect size · Single subject · Multiple baseline design · Meta-analysis

## Introduction

Systematically identifying effective interventions for students is of paramount importance to educators. The evidence-based practice movement has gained momentum in the past decade with developments in psychology, education, and prevention science. Federal law defines scientifically based research as “rigorous, systematic and objective procedures to obtain valid knowledge” (No Child Left

---

S. Petersen-Brown (✉) · A. C. Karich · F. J. Symons  
University of Minnesota, Minneapolis, MN, USA  
e-mail: pete6786@umn.edu

A. C. Karich  
e-mail: kari0060@umn.edu

F. J. Symons  
e-mail: symon007@umn.edu

Behind Act of 2001, p. 126), which includes research that is “evaluated using experimental or quasi-experimental designs” (p. 541). Recent Race to the Top initiatives confirm the importance of using evidence to adopt educational practices and policies in the outlined priority goals of the Department of Education’s performance plan (Analytical Perspectives 2010). The adoption and implementation of evidence-based practices must be a responsibility of not only practitioners but also researchers in evaluating the feasibility and effectiveness of interventions that are integrated into practice settings (Kratochwill and Shernoff 2003).

Identifying effective interventions is based increasingly on the outcomes of meta-analyses (Cooper et al. 2009), and practitioners should examine meta-analytic studies to inform policy decisions (Kavale and Forness 2000). One of the problems with relying on meta-analyses is that they often exclude data generated from small N and single-case design (SCD) studies evaluating interventions. SCDs can provide a strong basis for establishing causal inference, and these designs are widely used in applied and clinical disciplines in psychology and education (Kratochwill et al. 2010). The traditional approach to analysis of SCD data involves visual comparison within and across conditions of a study (Parsonson and Baer 1978); however, when aggregating SCD data in meta-analysis, statistical approaches to calculating effect sizes are needed.

As of yet, there is no agreed upon effect size (ES) metric from SCDs for use in quantitative syntheses of intervention studies. Researchers have proposed a variety of ES metrics for synthesizing single-case research including percentage of non-overlapping data (PND; Scruggs et al. 1987), percentage of all non-overlapping data (PAND; Parker et al. 2007), and percentage exceeding median (PEM; Ma 2006). Each of these ES metrics utilizes non-overlapping data between phases as an indicator of performance differences. Although potentially useful, weaknesses of existing non-overlap indices cited by Parker and Vannest (2009) include “(a) lack of a known underlying distribution (PND); (b) weak relationship with other established effect sizes (PEM); (c) low ability to discriminate among published studies (PEM, PND); (d) low statistical power for small N studies (PND, PAND, PEM); and (e) open to human error in hand calculations from graphs (PND, PAND, PEM)” (p. 357).

In an attempt to address these weaknesses (with the possible exception of human error), Parker and Vannest (2009) proposed a fourth index of non-overlapping data, non-overlap of all pairs (NAP). NAP calculates a percentage of non-overlapping data by investigating the extent to which each data point in phase A (baseline) overlaps with each data point in phase B (intervention). Initial research on NAP is promising; however, little is known about interpreting NAP values in relation to detecting and evaluating treatment effects, specifically within multiple baseline designs (MBD). Preliminary evaluations of NAP utilized AB contrasts between phases of published articles using a variety of designs, including AB, ABAB, and multiple baseline designs (MBDs) (Parker and Vannest 2009). However, a recent review of research found that SCDs are increasing in prevalence overall, with a growing emphasis on MBDs (Hammond and Gast 2010). As discussed above, the evidence-based movement urges the identification of effective educational interventions in the field. A majority of academic intervention research in education is conducted with MBDs, perhaps because most academic skills are nonreversible

(Bramlett et al. 2010). An ES that can accurately allow educators to interpret the results of an aggregation of MBD studies is needed to identify academic interventions as evidence-based on a larger scale. When behaviors are nonreversible, it is less likely that a behavior will overlap with baseline data because data are not expected to return to baseline levels, therefore potentially inflating non-overlap indices within MBDs as compared to AB case study and ABAB withdrawal designs examined by Parker and Vannest (2009).

MBDs are considered more complex than withdrawal or reversal designs because a number of responses are identified and measured over time by varying the length of baseline observations across two or more behaviors, settings, or participants (Baer et al. 1968). Because observations in MBDs extend across two or more baselines, the design typically yields more data points (e.g., 60–80 total; Parker et al. 2007) also contributing to the complexity of the design over withdrawal or reversal designs. Moreover, specific parameters of effect detection (sensitivity, specificity) are rarely investigated in SCD ES evaluation studies and have never been investigated specific to MBDs to the author's knowledge. In this context, sensitivity evaluates how accurate an approach is at detecting a true effect while specificity refers to the ability of an approach to rule out the absence of an effect (McNeily and Hanley 1982).

The purpose of the current study was to investigate the use of NAP with MBDs due to the increasing use of MBDs to evaluate the effectiveness of academic interventions, the need for ESs to allow for quantification of intervention effects, and the potential for SCD non-overlap ESs to differ in MBDs. The following research questions guided the analyses: (a) What are typical NAP values within MBDs?, (b) To what extent is NAP sensitive and specific in detecting intervention effects in SCD studies using MBDs?, (c) What constitutes a large and small NAP effect size in MBDs?, and (d) To what extent does a framework for interpreting NAP ES estimates within MBDs agree with interpretations based on visual analysis?

## Method

### Data Collection

The PsycINFO, ERIC, and Academic Search Premier databases were searched for articles on March 26, 2011, using the terms *multiple baseline AND reading intervention*, *multiple baseline AND math intervention*, *multiple baseline AND writing intervention*, *multiple probe AND reading intervention*, *multiple probe AND math intervention*, and *multiple probe AND writing intervention*. The following criteria were used to select articles to include in the current meta-analysis:

1. The study was published in a peer-reviewed journal between 2000 and 2011; this date range was selected because a previous study aggregated academic intervention research prior to 2000 (Swanson et al. 1999);
2. The study investigated an intervention to enhance reading, writing, or math performance;

3. The study used a multiple baseline or multiple probe design with three or more participants, settings, behaviors, or sets of materials;
4. The study provided sufficient data to compute NAP;
5. The study's participants were school-age (3–21 years); and
6. The study was written in English.

The search terms identified 127 articles, 85 of which met the above criteria and were included in the current study. Forty-two studies were excluded, most for the following reasons: an insufficient number of baselines, outcomes focused on social-emotional behaviors rather than academic behaviors, and duplicate publications. In cases in which multiple outcomes or multiple comparison sets were used (e.g., multiple subjects in a multiple baseline across materials study), each outcome or comparison set was used. This resulted in 176 comparison sets.

### Coding of Articles

Studies that met inclusion criteria were systematically reviewed and coded using a coding form created in a spreadsheet program. The authors coded study design characteristics and a decision about the effect based on visual analysis.

#### *Study Design Characteristics*

Design characteristics included the type of multiple baseline employed and the number of participants, settings, or behaviors included. The type of multiple baseline employed was coded as follows; across participants, across settings, across behaviors, a combination of the three aforementioned conditions, across materials, and a category deemed as “other.”

#### *Visual Analysis*

Visual analysis was also conducted to decide whether an effect existed. As suggested by the single-case technical documentation of What Works Clearinghouse, two of the authors evaluated changes in level, trend, variability, and immediacy of change between the baseline and first treatment phase (Kratochwill et al. 2010). Change in level was determined by looking at the last data point in baseline and the first data point of the intervention phase. Change in variability refers to the fluctuation of the data around the mean from baseline to intervention phases. The effects of the intervention were analyzed by immediacy of change, or the rapidity of the effect after the onset and/or withdrawal of the intervention. Immediacy of the effect is the extent to which the level, trend, and variability of the last three data points in the phases are different from the first three data points in the intervention phase. If changes in at least two of these four elements were detected between phases, the authors coded an intervention effect. Agreement was established by rating two studies collaboratively, and then the authors coded the remaining studies individually. Authors coded the articles separately and then

convened to make adjustments in coding to minimize potential interobserver error throughout the coding process.

The data for each study were judged to demonstrate a large intervention effect if 75% or more of the baselines within the MBD showed an intervention effect as described above. (If a study had fewer than four baselines, each needed to demonstrate an effect to be coded as a large effect.) Multiple baseline data in which between 50% and 75% of baselines demonstrated an effect as described above were judged as a small effect.

### Effect Size Calculation

NAP (Parker and Vannest 2009) was used to estimate ES for each set of participants. As mentioned above, NAP calculates a percentage of non-overlapping data by investigating the extent to which all possible pairs of data points across phases overlap. Each data point in the baseline phase A was compared to each data point in the intervention phase B to determine whether overlap occurred between phases. Each pair of data points that overlapped completely was assigned a value of '1', and each pair of data points that tied was assigned a value of '0.5'. Adding the overlap sums, subtracting from the total possible number of comparison pairs (i.e., the number of data points in phase A multiplied by the number of data points in phase B), and dividing by the total possible number of comparison pairs derives NAP for that baseline to intervention phase change. For example, a study with 5 baseline data points and 11 intervention data points has 55 total possible pairs. If one data point from the baseline overlaps with two data points and ties with one data point in the intervention phase, there are 2.5 total overlaps. Subtracting 2.5 from 55 is 52.5, and 52.5 divided by 55 is equal to 0.95; 0.95 is the NAP value for this example. Readers are directed to Parker and Vannest (2009) for further information on how to calculate NAP as well as other SCD ESs (PND, PEM, and PAND). NAP was then averaged across each baseline of each outcome within each study.

### Interobserver Agreement

Interobserver agreement was calculated for 25% of included studies and outcomes. Percentage agreement was calculated as agreements divided by agreements plus disagreements multiplied by 100% for coding variables and visual analysis decisions. Percentage agreement between the two raters was 98.9% for coding study variables and 92.2% for decisions from visual analyses.

### Analyses

Kappa coefficients and receiver operating characteristic (ROC) analysis were used to answer the primary research questions for this study. The kappa coefficient (Cohen 1960) is a measure of observer agreement taking chance into account. Kappa adjusts for chance in the calculation of observer agreement by subtracting chance from observed agreement. ROC analysis was used to assess the extent to which a measure (in this case, NAP) finds the same dichotomous outcome as a

“gold standard” measure (in this case, visual analysis). ROC has become widely used in psychology to assess the accuracy of diagnostic tests and procedures (Swets et al. 2000).

The sensitivity and specificity of an assessment are important concepts in ROC analysis. Accuracy in a ROC analysis is measured by area under the curve (AUC); the curve is the ROC curve. An AUC value above 0.80 indicates a reasonable measure, based on possible values between 0.50 and 1.0 and in accordance with criteria used in past investigations (e.g., Muller et al. 2005). The accuracy is determined by how well the measure (NAP) detects false positives and false negatives by the ROC curve, also known as the “sensitivity and specificity” curve. Sensitivity refers to NAP’s ability to detect true effects, while specificity refers to NAP’s ability to rule out when there is no effect. There is a trade-off between sensitivity and specificity; as sensitivity increases, specificity decreases, and vice versa. The measure’s accuracy is determined by the relationship between sensitivity and specificity in a ROC analysis.

## Results

The first research question addressed typical NAP values within MBDs. The overall mean of NAP estimates across 176 comparisons was 0.92 (SD = 0.10, range = 0.51–1.00), and the median was 0.96. NAP estimates found in the current sample were generally quite high, and the distribution was negatively skewed (skew = -1.69). ESs varied somewhat according to study characteristics (see Table 1), but were still close to 1.00. Median NAP ESs were lower for reading interventions (median=0.94) when compared to math and writing interventions (medians of 0.99 and 0.99, respectively).

**Table 1** Median NAP values by study characteristics

	N	Median value
Multiple baseline type	176	
Across participants	92	0.96
Across settings	1	0.95
Across behaviors	52	0.97
Other (e.g., across materials)	31	0.95
Number of multiple baselines	176	
Three	81	0.97
Four	49	0.95
Five	7	0.89
Six	18	0.97
Seven	9	1.00
Eight or more	12	0.94
Academic area	176	
Reading	127	0.94
Writing	30	0.99
Math	19	0.99

**Table 2** Classification of effect sizes based on NAP effect size and visual analysis

Visual analysis	NAP effect size			Total
	No effect	Small	Large	
No effect	36	3	7	46
Small	15	10	7	32
Large	12	15	71	98
Total	63	28	85	176

Medians did not differ considerably by the type of multiple baseline design or the number of baselines included in the design with one exception. MBDs employing five baselines yielded a lower NAP of 0.89.

The second research question addressed the sensitivity and specificity of NAP in detecting intervention effects in SCDs using MBDs. The AUC was 0.86 for the large effect of 75% of the baselines demonstrating an effect from visual analyses, and 0.82 for the small effect criterion of 50%. AUC values range from 1.00 (a perfect measure) to 0.50 (a measure that correctly classifies individual cases at a chance level; Zweig and Campbell 1993). The NAP criteria resulted in AUC values of above 0.80, which suggests a reasonable measure of effect.

The third research question addressed the identification of NAP values that correspond to large and small effects. Based on this favorable AUC estimate of the visual analysis cut scores, the coordinates of the curve were examined to determine what NAP score would also indicate an effect. A visual analysis cut score of at least 75% of the baselines showing an effect resulted in a NAP score of 0.96, which yielded a specificity of 0.81 and sensitivity of 0.73. A visual analysis cut score of 50% or more of the baselines showing an effect resulted in a NAP cut score of 0.93, which resulted in specificity of 0.83 and sensitivity of 0.78.

The last research question addressed the agreement between NAP ESs and interpretations based on visual analysis. Based on the results from the ROC analysis, a large effect cutoff of 0.96 and a small effect cutoff of 0.93 were investigated further using a contingency table and kappa coefficient. The contingency table is shown in Table 2. It suggests that NAP effect size cutoffs agreed to a greater extent with visual analysis when identifying studies with a large effect or no effect, but that agreement was much lower when identifying studies with a small effect. The kappa coefficient corresponding to this analysis was 0.45; values between 0.41 and 0.60 are generally considered moderate (although these criteria are arbitrary; Landis and Koch 1977).

## Discussion

The current analysis supports the suggestion that non-overlap ES metrics are larger when MBDs are used. The median NAP from the current data was 0.96, whereas the median NAP in the initial field test was somewhat lower at 0.84 (Parker and Vannest 2009). Estimates of ES cutoffs also varied between studies. Large and

small ES cutoffs suggested in this study were 0.96 and 0.93, respectively. In Parker and Vannest's (2009) introduction of NAP, they found that NAP ESs of 0-0.65, 0.66-0.92, and 0.93-1.0 corresponded to small, moderate, and large effects based on expert visual judgment. Based on the initial findings reported here, it would be reasonable to hypothesize that the higher NAP ESs may be due, at least in part, to the non-reversible nature of academic behaviors investigated through MBD. However, alternate hypotheses should also be considered, including the possibility that random error led to higher NAP estimates in the current investigation or that the interventions being tested had already been shown effective in previous research. Additional research is needed to directly test these hypotheses.

NAP ES calculations appeared to have acceptable sensitivity and specificity in the current study. NAP ES designations agreed with visual analysis decisions over 80% of the time among the multiple baseline studies in the sample. While these results are promising, the kappa coefficient calculated in this analysis is moderate, and it may be especially difficult for the NAP effect size to identify a small effect as it was defined in this study. Additionally, ceiling effects are a potential concern since NAP values in this set of multiple baseline studies were generally quite high. As a result of the high NAP values, the difference between a small and large effect as defined in this study was quite small.

Past research has suggested values that may imply a meaningful effect for a variety of single-case ES metrics (Brossart et al. 2006; Parker et al. 2005; Parker and Hagan-Burke 2007; Parker and Brossart 2003). Parker and Vannest's (2009) paper introducing NAP also suggested typical values and potential ES cutoffs based on Cohen's guidelines for *d*. However, Parker and Vannest (2009) showed that Cohen's guidelines may not be relevant for single-case design. The current study may lend additional support to the notion of incompatibility of Cohen's guidelines with single-case designs because NAP values of 0.93 and 0.96 corresponded to Cohen's *d* values of 1.63 and 2.08, which are well above the small and large cutoffs Cohen suggested. The studies used in Parker and Vannest (2009) included AB comparisons from a variety of designs. The current study investigated published data from MBDs only and found that there may be important distinctions between typical ES values for MBD studies and the broader universe of SCD designs that include MBDs, AB, ABA, and ABAB designs.

There are three key limitations worth noting. Only those studies that evaluated interventions in reading, math, and writing, were published between 2000 and 2011, and were available in full-text online were included. Therefore, the results are specific to the studies sampled. The effect size (NAP) calculation and the visual analysis were performed by the same two individuals (first two authors). Therefore, the possibility exists that there was bias influencing the visual inspection (this possibility is made less plausible by the fact that study datasets were visually inspected independently and a high IOA was obtained). An additional limitation common to all non-overlap effect sizes (NAP, PND, PAND) is lack of sensitivity to the magnitude of the discrepancy between baseline and intervention phases, only noting whether or not there is a discrepancy.

Future work should investigate the use of the NAP ES as well as its sensitivity and specificity with a variety of published studies, including studies published prior



to 2000, studies of non-academic interventions, and studies that used phase change designs. Additionally, most studies investigating ES indices for single-case research use fabricated data and/or a convenience sample. Further research should attempt to include a larger and more representative sample to investigate ESs. The data from the current study suggest potentially useful interpretive criteria for NAP. However, more research is needed on NAP and other single-subject effect sizes in order for single-subject research to be effectively included in meta-analytic research.

## References

### \* Denotes studies included in the meta-analysis

- \*Alber-Morgan, S. R., Matheson Ramp, E., Anderson, L. L., & Martin, C. M. (2007). Effects of repeated readings, error correction, and performance feedback on the fluency and comprehension of middle school students with behavior problems. *The Journal of Special Education, 41*, 17–30.
- Analytical Perspectives (2010) *Budget of the United States Government*. Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/budget/fy2012/assets/spec.pdf>.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–97.
- \*Bergeron, J. P., Lederberg, A. R., & Connor, C. M. D. (2009). Building the alphabetic principle in young children who are deaf or hard of hearing. *Volta, 109*, 87–119.
- \*Beveridge, B. R., Weber, K. P., Derby, K. M., & McLaughlin, T. (2005). The effects of a math racetrack with two elementary students with learning disabilities. *International Journal of Special Education, 20*, 58–65.
- \*Binger, C., Maguire-Marshall, M., & Kent-Walsh, J. (2011). Using aided AAC models, recasts, and contrastive targets to teach grammatical morphemes to children who use AAC. *Journal of Speech, Language, and Hearing Research, 54*, 160–176.
- \*Blankenship, T. L., Ayres, K. M., & Langone, J. (2005). Effects of computer-based cognitive mapping on reading comprehension for students with emotional behavior disorders. *Journal of Special Education Technology, 20*, 15–23.
- \*Bliss, S., Skinner, C., & Adams, R. (2006). Enhancing an English language learning fifth-grade student's sight-word reading with a time-delay taped-words intervention. *School Psychology Review, 35*, 663–670.
- \*Bonfiglio, C. M., Daly, E. J., Persampieri, M., & Andersen, M. (2006). An experimental analysis of the effects of reading interventions in a small group reading instruction context. *Journal of Behavioral Education, 15*, 92–108.
- \*Bouck, E. C., Bassette, L., Taber-Doughty, T., Flanagan, S. M., & Szwed, K. (2009). Pentop computers as tools for teaching multiplication to students with mild intellectual disabilities. *Education and Training in Developmental Disabilities, 44*, 367–380.
- \*Boulineau, T., Fore, C., Hagan-Burke, S., & Burke, M. D. (2004). Use of story-mapping to increase the story-grammar text comprehension of elementary students with learning disabilities. *Learning Disability Quarterly, 27*, 105–121.
- Bramlett, R., Cates, G. L., Savina, E., & Lauinger, B. (2010). Assessing effectiveness and efficiency of academic interventions in school psychology journals: 1995–2005. *Psychology in the Schools, 47*, 114–125.
- Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*, 531–563.
- \*Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children, 28*, 237–249.

- \*Burns, M. K., Ganuza, Z. M., & London, R. M. (2009). Brief experimental analysis of written letter formation: Single-case demonstration. *Journal of Behavioral Education, 18*, 20–34.
- \*Carlson, B., McLaughlin, T., Derby, K. M., & Blecher, J. (2009). Teaching preschool children with autism and developmental delays to write. *Electronic Journal of Research in Educational Psychology, 7*, 225–238.
- \*Cieslar, W., McLaughlin, T., & Derby, K. M. (2008). Effects of the copy, cover, and compare procedure on the math and spelling performance of a high school student with behavioral disorder: A case report. *Preventing School Failure, 52*(4), 45–52.
- \*Clarfield, J., & Stoner, G. (2005). The effects of computerized reading instruction on the academic performance of students identified with ADHD. *School Psychology Review, 34*, 246–255.
- \*Cohen, E. J., & Brady, M. P. (2011). Acquisition and generalization of word decoding in students with reading disabilities by integrating vowel pattern analysis and Children’s literature. *Education and Treatment of Children, 34*, 81–113.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- \*Crabtree, T., Alber-Morgan, S. R., & Konrad, M. (2010). The effects of self-monitoring of story elements on the reading comprehension of high school seniors with learning disabilities. *Education and Treatment of Children, 33*, 187–203.
- \*Daly, E. J., Garbacz, S. A., Olson, S. C., Persampieri, M., & Ni, H. (2006). Improving oral reading fluency by influencing students’ choice of instructional procedures: An experimental analysis with two students with behavioral disorders. *Behavioral Interventions, 21*(1), 13–30.
- \*Daly, E. J., Johnson, S., & LeClair, C. (2009). An experimental analysis of phoneme blending and segmenting skills. *Journal of Behavioral Education, 18*, 5–19.
- \*De La Paz, S. (2001). Teaching writing to students with attention deficit disorders and specific language impairment. *The Journal of Educational Research, 95*, 37–47.
- \*Dimling, L. M. (2010). Conceptually based vocabulary intervention: Second graders’ development of vocabulary words. *American Annals of the Deaf, 155*, 425–448.
- \*Dowrick, P. W., Kim-Rupnow, W. S., & Power, T. J. (2006). Video feedforward for reading. *The Journal of Special Education, 39*, 194–207.
- \*Dufrene, B. A., Reisener, C. D., Olmi, D. J., Zoder-Martell, K., McNutt, M. R., & Horn, D. R. (2010). Peer tutoring for reading fluency as a feasible and effective alternative in response to intervention systems. *Journal of Behavioral Education, 19*, 1–18.
- \*Duhon, G. J., House, S. E., Poncy, B. C., Hastings, K. W., & McClurg, S. C. (2010). An examination of two techniques for promoting response generalization of early literacy skills. *Journal of Behavioral Education, 19*, 62–75.
- \*Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.
- \*Fevre, D. M., Moore, D. W., & Wilkinson, I. A. G. (2003). Tape assisted reciprocal teaching: Cognitive bootstrapping for poor decoders. *British Journal of Educational Psychology, 73*, 37–58.
- \*Fritschmann, S., Deshler, D. D., & Schumaker, B. (2007). The effects of instruction in an inference strategy on the reading comprehension skills of adolescents with disabilities. *Learning Disability Quarterly, 30*, 245–262.
- \*Fung, I. Y. Y., Wilkinson, I. A. G., & Moore, D. W. (2003). L1-assisted reciprocal teaching to improve ESL students’ comprehension of English expository text. *Learning and Instruction, 13*, 1–31.
- \*Ganz, J. B., & Flores, M. M. (2009). The effectiveness of direct instruction for teaching language to children with autism spectrum disorders: Identifying materials. *Journal of Autism and Developmental Disorders, 39*, 75–83.
- \*Glover, P., McLaughlin, T., Derby, K. M., & Gower, J. (2010). Using a direct instruction flashcard system with two students with learning disabilities. *Electronic Journal of Research in Educational Psychology, 8*, 457–472.
- \*Gyovai, L. K., Cartledge, G., Kourea, L., Yurick, A., & Gibson, L. (2009). Early reading intervention: Responding to the learning needs of young at-risk English language learners. *Learning Disability Quarterly, 32*, 143–162.
- Hammond, D., & Gast, D. L. (2010). Descriptive analysis of single subject research designs: 1983–2007. *Education and Training in Autism and Developmental Disabilities, 45*, 187–202.

- \*Hanser, G. A., & Erickson, K. A. (2007). Integrated word identification and communication instruction for students with complex communication needs. *Focus on Autism and Other Developmental Disabilities*, 22, 268–278.
- \*Harris, P. J., Oakes, W. P., Lane, K. L., & Rutherford, R. (2009). Improving the early literacy skills of students at risk for internalizing or externalizing behaviors with limited reading skills. *Behavioral Disorders*, 34, 72–90.
- \*Hindin, A., & Paratore, J. R. (2007). Supporting young children's literacy learning through home-school partnerships: The effectiveness of a home repeated-reading intervention. *Journal of Literacy Research*, 39, 307–333.
- \*Hines, S. J. (2009). The effectiveness of a color coded, onset rime decoding intervention with first Grade students at serious risk for reading disabilities. *Learning Disabilities Research & Practice*, 24, 21–32.
- \*Hitchcock, C. H., Prater, M. A., & Dowrick, P. W. (2004). Reading comprehension and fluency: Examining the effects of tutoring and video self-modeling on first-grade students with reading difficulties. *Learning Disability Quarterly*, 27, 89–103.
- \*Jacobson, L. T., & Reid, R. (2010). Improving the persuasive essay writing of high school students with ADHD. *Exceptional Children*, 76, 157–174.
- \*Jitendra, A. K., Edwards, L. L., Starosta, K., Sacks, G., Jacobson, L. A., & Choutka, C. M. (2004). Early reading instruction for children with reading difficulties. *Journal of Learning Disabilities*, 37, 421–439.
- \*Jones, K. M., Wickstrom, K. F., Noltemeyer, A. L., Brown, S. M., Schuka, J. R., & Therrien, W. J. (2009). An experimental analysis of reading fluency. *Journal of Behavioral Education*, 18, 35–55.
- \*Joseph, L. M., & Hunter, A. D. (2001). Differential application of a cue card strategy for solving fraction problems: Exploring instructional utility of the cognitive assessment system. *Child Study Journal*, 31, 123–136.
- \*Joseph, L. M., & Orlins, A. (2005). Multiple uses of a word study technique. *Reading Improvement*, 42, 73–79.
- Kavale, K. A., & Forness, S. R. (2000). Policy decisions in special education: The role of meta-analysis. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 281–326). Mahwah, NJ: Erlbaum.
- \*Kourea, L., Cartledge, G., & Musti-Rao, S. (2007). Improving the reading skills of urban elementary students through total class peer tutoring. *Remedial and Special Education*, 28, 95–107.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case design technical documentation*. Retrieved from What Works Clearinghouse website: [http://ies.ed.gov/ncee/wwc/pdf/wwc\\_scd.pdf](http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf).
- Kratochwill, T. R., & Shernoff, E. S. (2003). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Quarterly*, 18, 389–408.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- \*Lane, K. L., Graham, S., Harris, K. R., Little, M. A., Sandmel, K., & Brindle, M. (2010). Story writing: The effects of self-regulated strategy development for second-grade students with writing and behavioral difficulties. *Journal of Special Education*, 44, 107–128.
- \*Lane, K. L., Harris, K. R., Graham, S., Weisenbach, J. L., Brindle, M., & Morphy, P. (2008). The effects of self-regulated strategy development on the writing performance of second-grade students with behavioral and writing difficulties. *The Journal of Special Education*, 41, 234–253.
- \*Lane, K. L., Little, M. A., Redding-Rhodes, J., Phillips, A., & Welsh, M. T. (2007). Outcomes of a teacher-led reading intervention for elementary students at risk for behavioral disorders. *Exceptional Children*, 74, 47–70.
- \*Lawson, T. R., & Greer, R. D. (2006). Teaching the function of writing to middle school students with academic delays. *Journal of Early and Intensive Behavior Intervention*, 3, 151–170.
- \*Lee, Y., & Vail, C. O. (2005). Computer-based reading instruction for young children with disabilities. *Journal of Special Education Technology*, 20, 5–18.
- \*Li, D. (2007). Story mapping and its effects on the writing fluency and word diversity of students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 5(1), 77–93.
- \*Lovelace, S., & Stewart, S. R. (2007). Increasing print awareness in preschoolers with language impairment using non-evocative print referencing. *Language, Speech, and Hearing Services in Schools*, 38, 16–30.

- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches. *Behavior Modification*, *30*, 598–617.
- \*Marvin, K. L., Rapp, J. T., Stenske, M. T., Rojas, N. R., Swanson, G. J., & Bartlett, S. M. (2010). Response repetition as an error-correction procedure for sight-word reading: A replication and extension. *Behavioral Interventions*, *25*, 109–127.
- \*Mason, L. H., Kubina, R. M, Jr, Valasa, L. L., & Cramer, A. M. (2010). Evaluating effective writing instruction for adolescent students in an emotional and behavior support setting. *Behavioral Disorders*, *35*, 140–156.
- \*Mayfield, K. H., & Vollmer, T. R. (2007). Teaching math skills to at-risk students using home-based peer tutoring. *Journal of Applied Behavior Analysis*, *40*, 223–237.
- \*McCallum, R. S., Fearrington, J. Y., & Skinner, C. H. (2011). Increasing math assignment completion using solution-focused brief counseling. *Education and Treatment of Children*, *34*, 61–80.
- \*McCurdy, M., Daly, E., Gortmaker, V., Bonfiglio, C., & Persampieri, M. (2007). Use of brief instructional trials to identify small group reading strategies: A two experiment study. *Journal of Behavioral Education*, *16*, 7–26.
- \*McCurdy, M., Skinner, C., Watson, S., & Shriver, M. (2008). Examining the effects of a comprehensive writing program on the writing performance of middle school students with learning disabilities in written expression. *School Psychology Quarterly*, *23*, 571–586.
- McNeily, B. J., & Hanley, J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36.
- \*Mesmer, E. M., Duhon, G. J., Hogan, K., Newry, B., Hommema, S., Fletcher, C., et al. (2010). Generalization of sight word accuracy using a common stimulus procedure: A preliminary investigation. *Journal of Behavioral Education*, *19*, 47–61.
- \*Miao, Y., Darch, C., & Rabren, K. (2002). Use of precorrection strategies to enhance reading performance of students with learning and behavior problems. *Journal of Instructional Psychology*, *29*, 162–174.
- \*Mims, P. J., Browder, D. M., Baker, J. N., Lee, A., & Spooner, F. (2009). Increasing comprehension of students with significant intellectual disabilities and visual impairments during shared stories. *Education and Training in Developmental Disabilities*, *44*, 409–420.
- \*Morgan, L., & Goldstein, H. (2004). Teaching mothers of low socioeconomic status to use decontextualized language during storybook reading. *Journal of Early Intervention*, *26*, 235–252.
- Muller, M. P., Tomlinson, G., Marrie, T. J., Tang, P., McGreer, A., Low, D. E., et al. (2005). Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? *Clinical Infectious Diseases*, *40*, 1079–1086.
- \*Musti-Rao, S., & Cartledge, G. (2007). Effects of a supplemental early reading intervention with at-risk urban learners. *Topics in Early Childhood Special Education*, *27*, 70–85.
- \*Musti-Rao, S., Hawkins, R. O., & Barkley, E. A. (2009). Effects of repeated readings on the oral reading fluency of urban fourth-grade students: Implications for practice. *Preventing School Failure*, *54*(1), 12–23.
- No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2001).
- \*Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology*, *39*, 335–355.
- \*Odo, M., Barnett, D. W., Hawkins, R. O., & Musti-Rao, S. (2010). Reciprocal peer tutoring and repeated reading: Increasing practicality using student groups. *Psychology in the Schools*, *47*, 842–858.
- \*Ortiz Lienemann, T., Graham, S., Leader-Janssen, B., & Reid, R. (2006). Improving the writing performance of struggling writers in second grade. *The Journal of Special Education*, *40*, 66–78.
- Parker, R. I., Brossart, D. F., Callicott, K. J., Long, J. R., Garcia de Alba, R., Baugh, F. G., et al. (2005). Effect sizes in single case research: How large is large? *School Psychology Review*, *34*, 124–136.
- Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case data: A comparison of several methods. *Behavior Therapy*, *34*, 189–211.
- Parker, R. I., & Hagan-Burke, S. (2007). Useful effect size interpretations for single case research. *Behavior Therapy*, *38*, 95–105.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND). *The Journal of Special Education*, *40*, 194–204.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357–367.

- Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 101–165). Orlando, FL: Academic Press, Inc.
- \*Persampieri, M., Gortmaker, V., Daly, E. J., I. I. I., Sheridan, S. M., & McCurdy, M. (2006). Promoting parent use of empirically supported reading interventions: Two experimental investigations of child outcomes. *Behavioral Interventions*, 21, 31–57.
- \*Petursdottir, A. L., McMaster, K., McComas, J. J., Bradfield, T., Braganza, V., Koch-McDonald, J., et al. (2009). Brief experimental analysis of early reading interventions. *Journal of School Psychology*, 47, 215–243.
- \*Poncy, B. C., & Skinner, C. H. (2011). Enhancing first-grade students' addition-fact fluency using classwide cover, copy, and compare, a sprint, and group rewards. *Journal of Applied School Psychology*, 27, 1–20.
- \*Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2010). An investigation of detect, practice, and repair to remedy math fact deficits in a group of third grade students. *Psychology in the Schools*, 47, 342–353.
- \*Pullen, P. C., Lane, H. B., Lloyd, J. W., Nowak, R., & Ryals, J. (2005). Effects of explicit instruction on decoding of struggling first grade students: A data-based case study. *Education & Treatment of Children*, 28, 63–76.
- \*Resetar, J. L., Noell, G. H., & Pellegrin, A. L. (2006). Teaching parents to use research-supported systematic strategies to tutor their children in reading. *School Psychology Quarterly*, 21, 241–261.
- \*Ridge, A. D., & Skinner, C. H. (2011). Using the TELLS prereading procedure to enhance comprehension levels and rates in secondary students. *Psychology in the Schools*, 48, 46–58.
- \*Saddler, B. (2006). Increasing story-writing ability through self-regulated strategy development: Effects on young writers with learning disabilities. *Learning Disability Quarterly*, 29, 291–305.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research. *Remedial and Special Education*, 8, 24.
- \*Simon, R., & Hanrahan, J. (2004). An evaluation of the touch math method for teaching addition to students with learning disabilities in mathematics. *European Journal of Special Needs Education*, 19, 191–209.
- \*Spooner, F., Rivera, C. J., Browder, D. M., Baker, J. N., & Salas, S. (2009). Teaching emergent literacy skills using cultural contextual story-based lessons. *Research and Practice for Persons with Severe Disabilities*, 34, 102–112.
- \*Stagliano, C., & Boon, R. T. (2009). The effects of a story-mapping procedure to improve the comprehension skills of expository text passages for elementary students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 7(2), 35–58.
- \*Strong, A. C., Wehby, J. H., Falk, K. B., & Lane, K. L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students with emotional and behavioral disorders. *School Psychology Review*, 33, 561–582.
- Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: The Guilford Press.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1–26.
- \*Test, D. W., & Ellis, M. F. (2005). The effects of LAP fractions on addition and subtraction of fractions with students with mild disabilities. *Education and Treatment of Children*, 28, 11–25.
- \*Valleley, R. J., & Shriver, M. D. (2003). An examination of the effects of repeated readings with secondary students. *Journal of Behavioral Education*, 12, 55–76.
- Van \*Norman, R. K., & Wood, C. L. (2008). Effects of prerecorded sight words on the accuracy of tutor feedback. *Remedial and Special Education*, 29, 96–107.
- \*Viel Ruma, K., Houchins, D. E., Jolivet, K., Fredrick, L. D., & Gama, R. (2010). Direct instruction in written expression: The effects on English speakers and English language learners with disabilities. *Learning Disabilities Research & Practice*, 25, 97–108.
- \*Walker, B., Shippen, M. E., Alberto, P., Houchins, D. E., & Cihak, D. F. (2005). Using the expressive writing program to improve the writing skills of high school students with learning disabilities. *Learning Disabilities Research & Practice*, 20, 175–183.
- \*Whalon, K., & Hanline, M. F. (2008). Effects of a reciprocal questioning intervention on the question generation and responding of children with autism spectrum disorder. *Education and Training in Developmental Disabilities*, 43, 367–387.

- \*Windingstad, S., Skinner, C. H., Rowland, E., Cardin, E., & Fearington, J. Y. (2009). Extending research on a math fluency building intervention: Applying taped problems in a second-grade classroom. *Journal of Applied School Psychology, 25*, 364–381.
- \*Xin, Y. P., Wiles, B., & Lin, Y. Y. (2008). Teaching conceptual model-based word problem story grammar to enhance mathematics problem solving. *The Journal of Special Education, 42*, 163–178.
- \*Ziolkowski, R. A., & Goldstein, H. (2008). Effects of an embedded phonological awareness intervention during repeated book reading on preschool children with language delays. *Journal of Early Intervention, 31*, 67–90.
- Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*, 561–577.