Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Am J Intellect Dev Disabil. Author manuscript; available in PMC 2017 February 17.

Published in final edited form as:

Am J Intellect Dev Disabil. 2016 May ; 121(3): 169–193. doi:10.1352/1944-7558-121.3.169.

Comparing Single Case Design Overlap-Based Effect Size Metrics From Studies Examining Speech Generating Device Interventions

Mo Chen,
University of Minnesota, Minneapolis, MN

Jolene K. Hyppa-Martin,
University of Minnesota, Minneapolis, MN and University of Minnesota Duluth, Duluth, MN

Joe E. Reichle, and
University of Minnesota, Minneapolis, MN

Frank J. Symons
University of Minnesota, Minneapolis, MN

Abstract

Meaningfully synthesizing single case experimental data from intervention studies comprised of individuals with low incidence conditions and generating effect size estimates remains challenging. Seven effect size metrics were compared for single case design (SCD) data focused on teaching speech generating device use to individuals with intellectual and developmental disabilities (IDD) with moderate to profound levels of impairment. The effect size metrics included percent of data points exceeding the median (PEM), percent of nonoverlapping data (PND), improvement rate difference (IRD), percent of all nonoverlapping data (PAND), Phi, nonoverlap of all pairs (NAP), and Tau_{novlap}. Results showed that among the seven effect size metrics, PAND, Phi, IRD, and PND were more effective in quantifying intervention effects for the data sample (N = 285 phase or condition contrasts). Results are discussed with respect to issues concerning extracting and calculating effect sizes, visual analysis, and SCD intervention research in IDD.

Keywords

effect size metric; speech generating devices; single case design; intellectual and developmental disabilities; evidence-based practices

Evidence-based practice (EBP) has become a standard for an increasing number of educational and human services disciplines (American Speech-Language-Hearing Association [ASHA], 2014; Jenson, Clark, Kircher, & Kristjansson, 2007; Meline & Wang, 2004; Odom, 2009). Meta-analyses contribute to EBP by providing a high level of scrutiny when evaluating an evidence base (Orlikoff, Schiavetti, & Metz, 2015). However, for some

Correspondence concerning this article should be addressed to Mo Chen, University of Minnesota, Department of Educational Psychology, 56 East River Road, Minneapolis, MN, 55455, United States (chen1641@umn.edu).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

domains of applied research, meta-analyses can be more challenging to implement because of the lack of agreed upon effect size metrics (Lipsey & Wilson, 2001; Shadish, Hedges, & Pustejovsky, 2014).

One area where this is particularly evident involves intervention research with low-incidence populations including individuals with intellectual and developmental disabilities (IDD) with moderate to profound levels of impairment. In the specific area of speech generating device (SGD) interventions for individuals with IDD with moderate to profound levels of impairment, for example, the population is extremely heterogeneous. This has resulted in a preponderance of single case design (SCD) studies for which there is not a standard effect size metric. In this article, we used the results from SCD SGD intervention studies as a vehicle to further evaluate different non-parametric effect size metrics from SCD studies. Our rationale was, in part, pragmatic. That is, the literature specific to SGD and severe disabilities is a manageable literature. Additionally, SGDs represent an area of a relatively rapid growth with tablet-based systems emerging and proliferating (Kuster, 2012; McNaughton & Light, 2013). In the following sections we provide a brief overview of the literature addressing SGD and SCD. Subsequently, we provide a rationale for the SCD effect size comparisons that were conducted.

The Need for Effect Size Clarity for Applied Intervention Research With Low-Incidence Populations Using SGDs

Within the past decade, SGD interventions have become more accessible to individuals with IDD with moderate to profound levels of impairment. In part, this is the result of increased attention to the use of SGDs with “beginning” communicators, including older children and adults with IDD with significant levels of impairment who are still developing basic communication skills (Johnston, Reichle, Feeley, & Jones, 2012; Reichle, Beukelman, & Light, 2002). Additionally, the use of SGDs with individuals with IDD experiencing more significant levels of impairment has increased as a result of advances in technology (Fernandez, 2011; Hershberger, 2011; McNaughton & Light, 2013). Only 5 years ago, high-tech, touch screen SGDs were often obtained from specialized SGD manufacturers, after a complete evaluation from a licensed speech and language pathologist, followed by a funding request and approval (Beukelman & Mirenda, 2013). Today, consumers can immediately obtain an SGD by purchasing a portable, touch-screen computer from a retailer and downloading one of the numerous low-cost or free augmentative and alternative communication (AAC) software applications that are readily available (Kuster, 2012). Consequently, interventionists are reporting increased rates of care partners pursuing SGD interventions for individuals with IDD with significant levels of impairment (Gosnell, Costello, & Shane, 2011).

With increasing use, there is a need for EBP guidelines addressing outcomes specific to SGDs and IDD. Systematic reviews related to SGD interventions often involve the analysis of studies in which SCDs were used. Although many parametric or non-parametric effect size metrics have been proposed for SCD studies (see Table 1), there has not been wide-spread agreement on which effect size metric(s) best serve the evaluation of SCD research

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Gast & Ledford, 2014; Wolery, Busick, Reichow, & Barton, 2010). As a result, the majority of the extant systematic reviews for SCD studies involving SGDs have not reported effect size metrics (e.g., Gevarter et al., 2013; Kagohara et al., 2013; Lancioni, O'Reilly, et al., 2007; Mirenda, 2003; Rispoli, Franco, van der Meer, Lang, & Camargo, 2010; Schlosser & Blischak, 2001; Schlosser & Sigafoos, 2006; Snell, Chen, & Hoover, 2006; Snell et al., 2010; van der Meer & Rispoli, 2010), or have reported a single overlap-based effect size metric (e.g., Percentage of Nonoverlapping Data [PND] in Branson & Demchak, 2009, in Millar, Light, & Schlosser, 2006, in Schlosser & Wendt, 2008, and in Stephenson & Limbrick, 2013; and Improvement Rate Difference [IRD] in Ganz et al., 2012).

Existing research comparing different effect size metrics for SCDs (see Table 2 for a brief overview) has provided important evidence for the utility of several newer effect size metrics (e.g., IRD, PAND [percent of all nonoverlapping data], Phi, NAP [nonoverlap of all pairs], Tau_{novlap} [Kendall's tau nonoverlap]) by analyzing their correlations with some earlier metrics such as PEM [percent of data points exceeding the median], PND, and Pearson R^2; their ability to differentiate intervention effects of different magnitudes; and their agreement with visual analysis judgments. Among these studies, some used phase contrasts exclusively from withdrawal or multiple baseline designs (e.g., Ma, 2006; Parker & Hagan-Burke, 2007; Parker, Hagan-Burke, & Vannest, 2007), although most used phase contrasts for which no specific research design information was provided (e.g., Campbell, 2004; Parker & Vannest, 2009; Parker, Vannest, & Brown, 2007; Parker, Vannest, & Davis, 2011; Parker, Vannest, Davis, & Sauber, 2011; Wolery et al., 2010). Schlosser, Sigafoos, and Koul (2009) reported that important SGD-related research questions often include comparing the relative effect of treatments such as interventions involving SGDs, manual signs, and nonelectronic picture exchange systems (e.g., van der Meer et al., 2013). Comparison designs (e.g., alternating treatment designs) are critical for these types of SGD-related research questions, but data from studies using comparisons designs, in which different intervention conditions are compared, have been largely overlooked in the existing effect size metric comparison studies.

Given the current state of evidence addressing SCD effect size metrics and prior SCD effect size metric comparisons, our goal was to better understand the performance of the various effect size metrics across different SCD design types used in SGD research. To accomplish this, we identified and compared seven effect size metrics that were previously reported on by Parker, Vannest, and Davis (2011): PEM, PND, IRD, PAND, Phi, NAP, and Tau_{novlap}. In the current investigation we chose to focus on non-parametric effect size metrics because they are more often used in practice by analyzing the overlap of data points across phases as compared to parametric effect size metrics (Maggin et al., 2011; Shadish et al., 2014). Additionally, the calculation of non-parametric effect size metrics is more aligned with the advocacy for a "bottom-up" approach to analysis of SCD data in comparison to a "top-down" approach (Parker & Vannest, 2012). Specifically, the bottom-up approach refers to combining data from individual phase contrasts to form one or more effect size metrics that represent the entire design; whereas the top-down approach refers to using data from the entire design to form an omnibus effect size metric by use of statistical models, such as hierarchical modeling, randomization, or complex multiseries regression (Parker & Vannest, 2012). Therefore, although top-down models seem promising, the bottom-up approach is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

advocated for use given its easier accessibility for interventionists and behavior analysts, higher consistency with the logic of visual analysis, and more intuitive and meaningful results (Parker & Vannest, 2012).

The specific analysis objectives for this investigation included an analysis of (a) the intercorrelation among the seven effect size metrics, (b) the discriminability of seven effect size metrics with respect to the magnitude of intervention effects, and (c) the agreement between the seven effect size metrics and visual analyses. These three areas have been frequently explored in previous effect size metric comparison research for other applied study domains and are relevant to the judgment of relative utility of effect size metrics for SCD studies (e.g., Parker & Vannest, 2009; Parker, Vannest, & Davis, 2011; Wolery et al., 2010; also see Table 2).

Method

Search Procedures

Electronic searches were conducted to identify articles addressing SGD interventions with the targeted population. The following search procedures were executed in January 2014 and yielded a total sample of 220 studies to which inclusion and exclusion criteria were applied.

First, an SGD was defined as an electronic, aided, augmentative and/or alternative communication device that provides auditory stimuli via digitized and/or synthesized speech output (adapted from Schlosser et al., 2009). Six electronic databases were searched including: Academic Search Premier, Education Resources Information Center (ERIC; Access via CSA), PsychINFO, Medline (Ovid), Education Full Text, and LLBA (Linguistics & Language Behavior Abstracts). The search was limited to English-language peer-reviewed studies published between 1985 and 2013. Historically, SGDs have been referred to by a variety of terms, including *voice output communication aid* (Olive et al., 2007), *VOCA* (e.g., Sigafoos, Drasgow, et al., 2004), and *voice output device* (e.g., Dicarlo & Banajee, 2000). In other sources, no specific term was used to describe the device itself. Instead, the SGD was referred to as *assistive technology* (e.g., Kagohara, 2010), or by the method it was accessed by the communicator such as *microswitch* (e.g., Lancioni et al., 2006). To ensure that a sufficiently broad search was conducted, six separate multiword search terms were entered into each database including speech generating device, voice output communication aid, voice output device, VOCA and *assistive technology, communication device* and *assistive technology*, and *microswitch* and *communication*.

Next, an archival search strategy was used. Hand or electronic searches of the tables of contents of six journals published between 1985 and 2013 were conducted using the same search terms as the electronic database search listed previously. The journals were *Augmentative and Alternative Communication* (AAC); *Journal of Applied Behavior Analysis* (JABA); *Journal of Developmental and Physical Disabilities* (JDPD); *Journal of Speech, Language, and Hearing Research* (JSLHR); *Language, Speech, and Hearing Services in Schools* (LSHSS); and *American Journal of Speech-Language Pathology* (AJSLP). These journals were selected because the first three yielded the most returns from the initial database search and the second three represented the American Speech Language

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hearing Association journals that historically have published a breadth of communication and SGD intervention literature.

Finally, an ancestral search was conducted through the reference lists of 11 recent SGD-related literature reviews (Branson & Demchak, 2009; Ganz et al., 2012; Gevarter et al., 2013; Kagohara et al., 2013; Lancioni, O'Reilly, et al, 2007; Millar et al., 2006; Rispoli et al., 2010; Schlosser & Sigafoos, 2006; Snell et al., 2006; Stephenson & Limbrick, 2013; van der Meer & Rispoli, 2010).

Inclusion/Exclusion Criteria Applied to Articles Identified

To be included in the analysis, an article was required to (a) be published between 1985 and 2013, (b) employ an experimental single case design (e.g., withdrawal design, multiple baseline or multiple probe design, alternating treatment design, or combined design), (c) utilize at least one SGD throughout an intervention, (d) serve at least one learner with a developmental disability/delay who had moderate to profound levels of intellectual impairment (i.e., an IQ of 50 or below based on standardized tests; DSM-5 [American Psychiatric Association, 2013], or an author[s]' description of a participant as experiencing a moderate, severe, or profound levels of intellectual impairment. In multiple-participant studies, only those participants meeting this criterion were included in the data analysis. Seventy-six percent of the included participants had their cognitive status described narratively with IQ from standardized testing used for the remainder), (e) report on one or more relevant dependent variables of the participant's communication comprehension or production skills, (f) graphically display the participant's acquisition-related data with a minimum of one communication skill involving an SGD, (g) permit an effect size calculation (e.g., sufficient resolution in relevant graph[s], or raw data to permit graphing); (h) be peer reviewed, and (i) be published in English.

Exclusion criteria were (a) quasi-experimental (AB design) studies (e.g., Russell & Beard, 1992), phase designs without a return to baseline following an intervention condition (e.g., AB [e.g., Reichle & Ward, 1985; Russell & Beard, 1992], ABC [e.g., Sigafoos, O'Reilly, Seely-York, et al., 2004], ABCD [e.g., one participant in Logan et al., 2001); (b) group design studies (no group design studies were found in the current sample of 220 articles); (c) articles providing only maintenance or generalization data with no acquisition data (e.g., Fragale, O'Reilly, Aguilar, & Pierce, 2012; Sigafoos, Didden, & O'Reilly., 2003; van der Meer et al., 2011); (d) phase design studies that did not provide baseline data for the targeted communication skill (e.g., Radstaake et al., 2013; Wacker et al., 1990); (e) articles that did not report at least one communication skill involving SGD use as the dependent measure (e.g., Ferris & Fabrizio, 2008); (f) data displayed with insufficient resolution for data extraction (e.g., Schlosser, Belfiore, Nigam, Blischak, & Hetzroni, 1995); (g) SGDs used for noncommunicative purposes (e.g., leisure purposes; Lancioni et al., 2007); (h) preference-only studies assessing learners' preference for the SGD and other different AAC strategies (e.g., the study 2 in the article by Sigafoos et al., 2009); and (i) individuals who experienced disability onset after age 21 (based on the definition for developmental disability from the U.S. Department of Health and Human Services Developmental Disabilities Assistance and Bill of Rights Act (2000; e.g., Lancioni, O'Reilly, Singh, Buonocunto, et al., 2009).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Preparation for the Data Sample

Based on the inclusion and exclusion criteria, 285 phase or condition contrasts from 45 studies across 21 different journals were identified. These consisted of 181 AB phase contrasts from seven studies with withdrawal design (i.e., ABAB, ABABA, or ABA design), 14 studies with multiple-baseline designs (i.e., across participants [$n = 9$], across teaching tasks or materials [$n = 2$], across communication partners [$n = 1$], across participants and settings [$n = 1$], and across participants and teachers [$n = 1$]), and 12 studies with multiple-probe design (i.e., across responses [$n = 6$], across participants [$n = 2$], across devices [$n = 1$], across participants and teaching materials [$n = 1$], across participants and time periods [$n = 1$], and across time periods and settings [$n = 1$]). The remaining 39 AB phase contrasts and 65 condition contrasts were derived from six studies with alternating treatment design and six studies with combined designs (i.e., the combination of multiple-probe and alternating treatment designs [$n = 3$], the combination of multiple-baseline and alternating treatment designs [$n = 2$], and the combination of withdrawal and alternating treatment designs [$n = 1$]). The median number of data points per phase or condition contrast was 15, and the interquartile range (IQR) was 8 to 22. The median number of data points in Phase A (i.e., the baseline condition) was 5 (IQR: 3-9), and in Phase B (i.e., the intervention condition) was 8 (IQR: 4-16). The 285 phase or condition contrasts were derived from 72 graphs, among which 57 graphs were from withdrawal designs, multiple-baseline or multiple-probe designs, and 15 graphs were from alternating treatment designs or combined designs.

To prepare the data sample, the graphs in the identified 45 studies were saved as individual .jpeg images. These were uploaded into PlotDigitizer® (Huwaldt, 2010) data extraction software and digitally converted into numerical data. Then, the numerical data points were extracted from the graphs and downloaded into a spreadsheet that displayed the data from each graph. Although the PlotDigitizer® software identifies values up to the 5th decimal place, values of each dependent measure were rounded to the decimal place that was reported in the original study. For instance, if a target dependent measure of the frequency of requests was reported using a whole number in a study, the data extracted through PlotDigitizer® were also rounded to the whole number. Additionally, if the original data were graphed in a cumulative manner (e.g., Sigafoos & Drasgow, 2001, sessions two through 28), the cumulative data were extracted through PlotDigitizer® and then the data were further transformed into noncumulative data by subtracting the first session data from the second session, subtracting the second session data from the third session, and so on, until only the data for each respective session were obtained.

For graphs using withdrawal, multiple-baseline, or multiple-probe designs, each of the adjacent AB data series was extracted and treated separately (i.e., each intervention phase was compared with the immediately preceding baseline). If a combination of withdrawal and multiple-baseline or multiple-probe design was used, each adjacent AB series involved was extracted. For studies using phase designs such as ABA, ABAC, or ABACA, each AB phase contrast was extracted based on the same principle described earlier. That is, the data for each intervention phase and its immediately preceding baseline phase were extracted. When a return to baseline was involved, the resulting adjacent BA data contrast was also extracted, with the designation of baseline and intervention phases remaining unchanged. For graphs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

using alternating treatment designs, the baseline-and-condition comparison and/or between-condition comparison were extracted. Baseline-and-condition comparison referred to the comparison of a dependent variable involving SGD use between the adjacent baseline and intervention conditions. Between-condition comparisons referred to the comparisons between two different concurrently implemented intervention conditions, in which at least one of the two intervention conditions involved SGD use.

Some examples of a between-condition comparison included an intervention condition with an SGD and an intervention condition without an SGD (e.g., Sigafoos & Drasgow, 2001), the comparison between an intervention condition with an SGD and an intervention condition with another AAC strategy such as manual signs or picture symbols (e.g., Soto, Belfiore, Schlosser, & Haynes, 1993; van der Meer, Kagohara, et al., 2012), and the comparison between two intervention conditions in which some aspect of the SGD differed (e.g., a condition with an SGD producing long utterances versus a condition with an SGD producing short utterances [Sigafoos et al., 2011]). If three or more intervention conditions were alternated, each pairwise comparison was extracted, with at least one intervention condition among each pair involved SGD use. Additionally, if one intervention condition involving SGD use was superior to other condition(s) and then a final phase was implemented involving continued use of the superior condition alone, data from the final phase were not extracted for analysis. For data extraction and later effect size metric computations of between-condition comparisons, the condition involving the use of an SGD served as the intervention condition whereas the condition involving no use of an SGD was considered as the baseline condition. If both conditions involved the use of an SGD, the condition that was reported to be the more complex or sophisticated use of the SGD was considered as the intervention condition whereas the other condition was considered as the baseline condition. For example, the use of an SGD with voice output was considered as the intervention condition, whereas the use of an SGD without voice output was considered as the baseline condition (e.g., Sigafoos et al., 2011).

For data sets using combined designs, if a combination of alternating treatment and withdrawal, multiple-baseline, or multiple-probe designs were implemented, data were extracted based on the rules established for alternating treatment designs (e.g., Kennedy & Haring, 1993). Data from maintenance and/or generalization conditions were not analyzed in the current investigation.

Computation of Effect Sizes

The seven effect size metrics computed included PEM, PND, PAND, Phi, IRD, NAP, and Tau_{novlap} (Parker, Vannest, & Davis, 2011). All computations were conducted in a programmed Excel spreadsheet (Microsoft Company, 2007) based on the computation procedures for each index (available from the first author upon request). IRD, NAP, and Tau_{novlap} can also be computed by an online calculator at http://www.singlecaseresearch.org/calculators (Vannest, Parker, & Gonen, 2011). After the required raw data were entered onto the Excel spread sheet, the Excel spread sheet computed and exported the results automatically.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Procedures for Visual Analysis

One judge independently rated all 45 studies comprised of 285 phase/condition contrasts, and a second judge independently rated 30 studies comprised of 219 phase/condition contrasts. Both raters had completed doctoral-level specialized coursework on SCD research methods and had collaborated on several SCD studies. Procedures for visual analysis implemented by Petersen-Brown, Karich, and Symons (2012) were replicated in the current investigation. Specifically, four indicators were adopted to make visual analysis judgments and if changes in at least two out of four indicators were detected in a phase or condition contrast, an intervention effect was coded for the specific contrast. The indicators were: immediacy, variability, trend, and level.

For phase contrasts, we utilized the same operational definitions for the four indicators as those described in Petersen-Brown et al. (2012). *Immediacy* was defined as whether there was a difference between last three data points in Phase A versus the first three data points in Phase B. *Variability* was defined as whether there was a difference in the data fluctuation about the mean in Phase A versus the mean in Phase B. *Trend* was defined as whether there was a difference between the slope of the data in Phase A versus that in Phase B. *Level* was defined as whether there was a difference between the mean of Phase A and the mean of Phase B.

The study by Petersen-Brown et al. (2012) did not involve condition contrasts. Consequently, for condition contrasts we applied the same logic evident in the phase contrast definitions and developed the corresponding operational definitions. *Immediacy* was defined as whether there was a difference between the first three data points of two conditions being compared. *Variability* was defined as whether there was a difference in data fluctuation about the mean of Condition A versus the mean of Condition B. *Trend* was defined as whether there was a difference between the slope of the data in Condition A versus that in Condition B. *Level* was defined as whether there was a difference between the mean of Condition A and the mean of Condition B.

In addition to conducting visual analysis for each of the 285 phase or condition contrasts, we adopted the same criteria as described in Petersen-Brown et al. (2012) and made a holistic judgment of the intervention effect of each of the 45 studies included. That is, for studies with four or more contrasts involved, the whole study was judged to demonstrate a large intervention effect if 75% or more of the contrasts resulted in an intervention effect after application of the criteria described above. A small effect was recorded if at least 50% but no more than 75% of the contrasts showed an intervention effect. No effect was recorded if less than 50% of the contrasts showed an intervention effect. For studies with fewer than four contrasts, the entire study was judged to demonstrate a large intervention effect if all contrasts resulted in an intervention effect as described above. A small effect was recorded if at least 50% but not all of the contrasts showed an intervention effect. No effect was recorded if less than 50% of the contrasts resulted in an intervention effect. For studies that involved condition contrasts, the holistic judgment of the study was based only on the condition contrasts, regardless of whether there were phase contrasts extracted from the study. The core research question in comparison designs is to compare the effects of different intervention conditions (Barlow & Hayes, 1979). Thus, conducting a visual

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

analysis for the comparison of different intervention conditions and comparing that analysis with the results of effect size metrics for the condition contrasts was conceptually aligned with our research questions.

Statistical Data Reduction

The 45 studies were separated into two data subsets. The first dataset included the phase contrasts that were derived from the 23 studies that did not include any condition contrasts (i.e., the studies with withdrawal, multiple baseline, and multiple probe designs). The second dataset included the phase and condition contrasts that were derived from the 12 studies that included condition contrasts (i.e., the studies with alternating treatment design or combined designs that were previously described). The original dataset was separated into these two subsets because the results tended to be different between these two general types of SCDs. Next, a correlation analysis using Pearson's *r* was conducted to examine the intercorrelation among the metrics for the two datasets. Subsequently, uniform probability distributions for the seven metrics were plotted to test the ability of each effect size metric in discriminating the intervention effects of different magnitudes for the two datasets (Parker & Vannest, 2009).

Consistent with the procedures used in Petersen-Brown et al. (2012), two steps were implemented to compare the agreement between effect size metrics and visual analysis judgments for all 45 studies. First, receiver operating characteristic (ROC) analysis was conducted to find the cutoff score for each effect size metric to differentiate intervention effects based on the visual analysis results of the 45 studies. To measure the accuracy of a ROC analysis, area under the curve (AUC) was reported, which should be above .80 to be acceptable (Muller et al., 2005). The sensitivity and specificity values were also reported. Sensitivity represents the degree to which a metric correctly detects the true effect, while specificity represents the degree to which a metric correctly rules out no effect (Petersen-Brown et al., 2012). Second, after the corresponding cutoff score was identified, kappa coefficients were calculated to gauge the convergence between effect size metrics and visual analysis results.

Interobserver Agreement and Fidelity

To ensure fidelity in the application of established procedures throughout this study, interobserver agreement (IOA) was computed for search procedures, the application of inclusion and exclusion criteria, the computation of effect size metrics, and visual analyses. Unless otherwise specified, IOA was computed using the formula of agreements being divided by agreements plus disagreements x100. Throughout the agreement comparison process, any disagreements were discussed and a final consensus was reached.

Search procedures—Once the initial systematic search was completed, a second independent observer (a doctoral-level graduate student) conducted an independent electronic search using the same keywords, which yielded 99.7% IOA. Subsequently, the independent observer also reviewed the six journals used in the archival search, resulting in 97.0% IOA. Finally, the independent observer implemented an ancestral search of two SGD-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

related literature reviews randomly selected from the 11 literature reviews scrutinized by the first author, which produced 100% IOA.

Inclusion and exclusion criteria—Among the initial pool of 220 articles yielded by electronic, journal, and ancestral searches, a total sample of 79 (35.9%) articles was randomly selected for IOA. Two independent observers reviewed each of the 79 articles to determine whether each article should be included based on the inclusion and exclusion criteria. This yielded an IOA of 98.7%. Subsequently, the two observers reviewed each of these 79 articles to determine which individual participants within each study met inclusion criteria for the present study. This yielded an IOA of 100%, using the same item-by-item agreement formula.

Interrater agreement of effect size metric computation—After the initial calculation of effect size metrics for the final 45 studies, a research assistant independently re-extracted numerical data from graphs through PlotDigitizer® and recomputed the effect size metrics for 23% of the total 285 phase or condition contrasts. A difference larger than . 01 was set as the criterion for disagreement (the criterion recommended by Parker, Hagan-Burke, & Vannest, 2007). Two raters reached agreement on 64 phase or condition contrasts with an agreement coefficient of 98.2%.

Interrater agreement of visual analysis—Two thirds of the studies (30 out of 45) were randomly selected to compute IOA on visual analysis. The phase or condition contrasts that comprised these studies represented 77% (219 out of 285) of the total phase or condition contrasts. The two doctoral student visual analysts agreed on their rating results for 212 out of the 219 independently analyzed phase or condition contrasts, resulting in an IOA of 97.0%, and on 30 out of 30 holistic judgments of each study's intervention effect, resulting in an IOA of 100%.

Results

Intercorrelations Among Effect Size Metrics

Pearson's *r* intercorrelations (see Table 3) among effect size metrics for the 181 phase contrasts from dataset of studies that only involved phase contrasts (i.e., did not compare interventions) were moderate to strong (range: .7 to .9). For the 104 phase or condition contrasts from studies that involved intervention comparisons, intercorrelations were somewhat lower (range: .4 to .9.) reflecting less consistency among the seven effect size metrics for this group of contrasts. In both datasets (i.e., studies with and without intervention comparisons) PEM, NAP, and Tau_{novlap} were consistently highly correlated (r = .9). Similarly, PAND, Phi, and IRD were consistently highly correlated (r = .9). The high correlation between NAP and Tau_{novlap} is understandable given that they can be mutually transformed and the computations for both are based on examining the pairwise comparisons of data points between Conditions A and B (Parker, Vannest, & Davis, 2011). The high correlation between PAND and Phi is also understandable given that Phi can be computed based on PAND (Parker, Hagan-Burke, & Vannest, 2007). PAND and Phi were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

correlated similarly with the other five effect size metrics. NAP and Tau_{novlap} were also correlated similarly with the other five effect size metrics.

The correlation between PEM and four other metrics (i.e., PND, IRD, PAND, and Phi) in the dataset that only involved phase contrasts ranged from .7 to .8. In the dataset that included intervention comparisons, it ranged from .4 to .5. Similarly, in the dataset that only involved phase contrasts, the correlation of PND with PAND and Phi was higher than in the dataset that included intervention comparisons. However, the correlation between PND and IRD was similarly high across both datasets. Also, in the dataset that involved condition contrasts, the correlations between the two metrics of PAND and Phi and the two metrics of NAP and Tau_{novlap} were lower. In general, across both datasets PAND, Phi, IRD, PND were more correlated whereas NAP, Tau_{novlap}, and PEM were more correlated.

Discriminability for Intervention Effects

Parker and Vannest (2009) observed that the utility of an effect size metric depends, at least partially, on its capability of effectively discriminating the intervention effects of different magnitudes among a data sample (i.e., discriminability). They plotted the uniform probability distribution proposed by Cleveland (1985) for each effect size metric to indicate its discriminability for intervention effects. The criteria for a metric having high discriminability included the appearance of 45 degree diagonal lines; no floor or ceiling effects; and no gaps, clumping, or flat segments (Chambers, Cleveland, Kleiner, & Tukey, 1983). The uniform probability distribution plots (see Figure 1 and Figure 2) for the seven effect size metrics for the contrasts from both datasets (i.e., studies with and without intervention comparisons) were examined using the same criteria proposed by Chambers et al. (1983).

As Figure 1 shows, for the contrasts from investigations that only involved phase contrasts, none of the seven probability distributions completely met the criteria for high discriminability. There were clear ceiling effects across seven effect size metrics around their 60^{th} percentile, suggesting that these metrics may not discriminate well among 40% of the most successful interventions. No obvious floor effects were found for the seven effect size metrics that were the focus of this investigation. In contrast, the IRD and Phi distributions were relatively superior to the others because they were slightly closer to the diagonal line (see Figure 1), especially between zero and the 60^{th} percentile.

As Figure 2 shows, the contrasts from the studies that involved intervention comparisons exhibited clear ceiling effects across the metrics around their 80^{th} percentile, suggesting that these metrics may not discriminate well among 20% of the most successful interventions. No obvious floor effects were found. Overall, IRD and Phi distributions were superior to the others because they were much closer to the diagonal line (see Figure 2). Notably, the superiority of IRD and Phi in terms of discriminability seemed better reflected for the dataset that included both phase and condition contrasts.

Comparison of Effect Size Metrics With Visual Analysis Outcomes

Agreement between effect size metrics and visual analysis has been considered a critical criterion for determining the relative utility of the effect size metrics for SCD studies (Gast

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

& Ledford, 2014; Wolery et al., 2010). In terms of the visual analysis for the holistic judgment of each of the 45 studies' intervention effect, only five studies were rated as having a small effect, 40 studies were rated as having a large effect, and no studies were rate as having no effect. Table 4 displays the cutoff scores for each effect size metric in differentiating small and large intervention effects, the corresponding AUC, as well as sensitivity and specificity values. Overall, the cutoff scores differed across these seven effect size metrics, ranging from .47 to .77. Their AUC results were all above the reasonable level of .80. In particular, each of the seven metrics had almost perfect specificity in terms of ruling out a small effect.

Table 5 displays the agreement between the visual analysis and the effect size metrics, as well as the corresponding kappa coefficients for all 45 studies. The kappa coefficients were all above .40 (range: .40 - .67). Values between .41 and .60 are usually considered moderate (Landis & Koch, 1977, Petersen-Brown et al., 2012). Based on this criterion, six of the seven effect size metrics (i.e., all except PEM) reasonably differentiated whether a study's intervention effect was large or small.

Overall, the seven effect size metrics tended to have a low level of false positives (i.e., when the visual analysis judged it as small effect, but effect size metric indicated a large effect based on its cutoff score). There were no false positives for any of the seven effect size metrics. However, all of the seven effect size metrics tended to have high levels of false negatives (i.e., when the visual analysis judged it as large effect, but effect size metric indicated a small effect based on its cutoff score). The percentage of false negatives for the seven effect size metrics ranged from 10% to 25%.

Discussion

This study used the outcomes from SCD studies in which SGD interventions were implemented with persons with IDD with moderate to profound levels of impairment to compare seven effect size metrics (i.e., PEM, PND, IRD, PAND, Phi, NAP, and Tau_{novlap}). Overall we examined 285 SCD phase or condition contrasts from a sample of 45 studies published between 1985 and 2013. Findings relevant to the seven effect size metrics are first discussed, followed by a discussion of what we believe are the major implications for meta-analysis and SCD related to evidence-based practice and individuals with IDD. Finally, limitations and future research directions are summarized.

Relative Utility of the Seven Effect Size Metrics

Potential pattern consistency among effect size metrics—For the data sample analyzed in this study, the intercorrelations among the seven effect size metrics suggested that two subgroups of these measures may yield similar outcomes. The first group included PAND, Phi, IRD, and PND, and the second included NAP, Tau_{novlap}, and PEM. Consequently, for meta-analyses of SCD intervention studies, once the choice between these two groups has been made; the specific effect size metric chosen may be a somewhat less important methodological decision. For instance, some existing meta-analyses in the field (e.g., Branson & Demchak, 2009; Millar, Light, & Schlosser, 2006; Schlosser & Wendt, 2008; Stephenson & Limbrick, 2013) used PND as the effect size metric whereas others

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(e.g., Ganz et al., 2012) used IRD as the effect size metric. The relatively high correlation between PND and IRD in the current study provided some evidence that the results from these meta-analyses may be comparable in terms of quantifying effect size.

In this study, NAP was less correlated with PAND (r = .7 for the dataset that only involved phase contrasts and .4 for the dataset that included condition contrasts). This result was not consistent with results reported by Parker and Vannest (2009) in which NAP and PAND were highly correlated (r = .9). Given that the correlation between NAP and PAND differed across the two datasets (i.e., studies with and without intervention comparisons) in the current study, it is reasonable to suspect that the correlations among effect size metrics may vary as a function of the data samples selected for effect size analysis and this may explain the differences between the results reported by Parker and Vannest (2009) and the current study. Notably, Parker and Vannest (2009) used data obtained through a convenience sample of published studies, whereas the data for the current study was obtained through a systematic search of a specific content area that was known to include a diversity of SCDs that compare clinically relevant interventions (i.e., they require alternating treatment designs). Parker and Vannest (2009) did not include alternating treatment designs in their convenience sample, whereas the data sample in the current study included both SCDs that only had phase contrasts, as well as alternating treatment designs with condition contrasts. Therefore, the degree to which specific, highly correlated effect size metrics are interchangeable may at least partly depend on how the dataset is selected.

Discrimination of intervention effect magnitude—Results from the probability distributions (Figure 1 & Figure 2) should be considered with caution. Visual analysis results for each study suggest the possibility of publication bias. None of the 45 studies examined for this investigation were judged as having no intervention effect, only five were judged as having a small effect, and 40 were judged as having a large intervention effect. Due to potential publication bias toward studies yielding positive intervention effects (Ickowicz, 2014) and the general characteristics of SCD research (Wolery, Dunlap, & Ledford, 2011), it is unlikely that the results of the seven effect size metrics were normally distributed. Overall, the analysis of the uniform probability plot for the seven effect size metrics revealed that none of the seven effect size metrics fully discriminated intervention effects ranging from a very small to very large magnitude. In particular, at around 60^{th} to 80^{th} percentile, the maximum had been reached. However, when the dataset was comprised of both phase and condition contrasts, the relative superiority among these effect size metrics in discriminating intervention effects became more clear. This finding suggests that the discriminability of these metrics may be also a function of the data sample and design selected. When only phase contrasts were included, due to potential publication bias, there seemed to be little room for the effect size metrics to reflect their discrimination ranges. When condition contrasts were included, there seemed to be more room for these effect size metrics to show their capacity in discriminating intervention effects of different magnitude. One potential reason might be that performance data from condition contrasts comparing two different interventions may provide more variable data sets than those generated by phase contrasts in which no intervention is compared to an intervention, the latter likely contribute to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

publication bias in favor of studies featuring phase contrasts of high magnitude with less variability.

The probability distributions (see Figures 1 and 2) revealed less ceiling effect when the plotted dataset involved studies including both phase and condition contrasts. While there was still some evidence of ceiling effect, this difference suggested that these effect size metrics may be better able to discriminate intervention effects of small to moderate magnitude. One explanation for this finding is related to the limits of non-parametric overlap-based effect size metrics in differentiating the magnitude of intervention effects when two phase or condition contrasts have the same data overlap patterns (Gast & Ledford, 2014; Wolery et al., 2010). For example, consider an intervention study in which two participants obtained the same baseline data of 0, 1, 1, 0, 0. The intervention data for one participant were 6, 7, 5, 7, 6, and the intervention data for the second participant were 60, 70, 50, 70, 60. Clearly, the magnitude of the change between baseline and intervention was greater for the second participant. Because overlap is the major consideration when computing a non-parametric effect size metric, the same effect size would be obtained for both the first and second participants. Consequently, non-parametric, overlap-based effect size metrics may more accurately be referred to as effect size *estimators* (Shadish, Rindskopf, & Hedges, 2008) and this limitation of overlap-based effect size metrics may help explain the results revealed by the probability distributions in the current study.

In spite of the moderate discriminability of the seven effect size metrics, comparatively, IRD and Phi seemed to better meet the criteria for discriminating the magnitude of intervention effects among the current data sample. Their uniform probability distributions seemed to be closer to the diagonal line, especially through the 60^{th} or 80^{th} percentile (see Figure 1 and Figure 2; these results were similar to Parker and Hagan-Burke, 2007). That is, although IRD and Phi were also limited in differentiating among the large intervention effects, they seemed to be the most effective in differentiating the intervention effects of small to moderate magnitude. Thus, IRD and Phi may be more promising metrics in terms of their ability to differentiate intervention effect magnitude.

Agreement with visual analysis judgments—Parker and Hagan-Burke (2007) suggested that given the holistic nature of visual analysis as well as its historical significance for SCD research, statistical analysis would not be a substitute for visual analysis but could augment it. From this perspective, the utility of effect size metrics for SCD research is partly dependent upon whether the inferences about intervention effects from the two analysis approaches are in agreement at a high level. In this study, the cutoff scores for each effect size metric to differentiate large versus small intervention effects in accordance with visual analysis results ranged from .47 for Tau_{novlap} to .77 for PAND. Notably, our cutoff score for NAP was .73 which was lower than the .96 cutoff score reported by Petersen-Brown et al. (2012) who used the same computation procedure. This may suggest that cutoff scores might also be influenced by the data sample selected. That said, when cutoff scores were identified using this procedure, most of the effect size metrics (i.e., all of them except PEM) corresponded well with the visual analysis findings. In particular, IRD corresponded best with visual analysis findings, having the highest kappa coefficient of .67, followed by PAND and Phi which both had kappa coefficients of .56.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Effect size metric comparison summary—As described earlier, given the current data sample, there may be two subgroups among the seven effect size metrics. One group consisted of PAND, Phi, IRD, and PND. The other group consisted of NAP, Tau_{novlap}, and PEM. IRD and Phi may be slightly superior over the other five effect size metrics in terms of differentiating the magnitude of intervention effects. IRD was most consistent with visual analysis judgment. Therefore, it seems likely that the effect size metrics from the group including PAND, Phi, IRD, and PND may better represent the current data sample, especially IRD and Phi, than effect size metrics from the other group.

Implications for Meta-Analysis in the Field of SGD or Related Communication Interventions in Low-Incidence Populations for Which SCD Will Be Used

Findings from the current study have several implications for conducting meta-analyses in the field of SGD interventions involving participants with an IDD. First, prior to conducting meta-analyses to inform EBP in a particular content area, the rationale for selecting any one particular effect size metric should be established because different metrics may lead to different conclusions. Second, although the general effect size metric comparison research may provide us with some confidence in utilizing certain effect size metric(s) for meta-analyses, it is likely that the results from the extant comparison research studies may not be readily generalizable. In particular, a majority of the comparison research was based on convenience samples of published studies, which is not the case for meta-analyses that require systematic searches. Third, there may be “subsets” of effect metrics within which the effect sizes may be comparable and interchangeable, as shown in the current data sample. Knowing that some SCD effect sizes may be comparable or interchangeable, could create some confidence in lining up and interpreting the results from meta-analyses for SGD interventions involving SCD studies.

Finally, the current study suggests that it is feasible to utilize different effect size metrics to more fully represent data from comparison design studies (e.g., alternating treatment designs). To do this, one critical procedure is to clearly define the *baseline* and *intervention* conditions for the two or more intervention conditions that are being alternated or compared within the study. In the current study, we defined the condition involving no use or less complex or sophisticated use of SGDs as the baseline, while the condition involving the use or more complex or sophisticated use of SGDs as the intervention condition. In this manner, we were able to apply the same analysis procedures that have been commonly applied to phase contrasts. Applying this type of procedure could support the generation of informative and meaningful meta-analyses by more easily allowing the inclusion of between-condition contrasts generated by comparison design studies.

Limitations and Directions for Future Studies

We did not randomly sample from the universe of SCD intervention studies using SGDs in interventions for individuals with more diverse developmental disabilities. Consequently, the generality of our findings with respect to effect size metrics applied to SGD interventions is limited to the sample of studies we located that included learners with IDD with moderate to profound levels of impairment. It may be helpful to explore whether similar findings would be found for other curricular areas. Second, our findings are specific to the metrics we

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

evaluated. Future work should systematically address the relative performance of other effect size metrics, including parametric effect size metrics such as ANOVA-based procedures like Cohen's *d* and Hedges's *g*. Doing so would expand the evidence addressing the potential equivalence or relative superiority of different effect size metrics. Determining the most representative metric, which could then be more consistently applied in meta-analyses, would better permit aggregation of existing datasets.

Third, we aimed to examine the relative utility of seven non-parametric effect size metrics in representing a data sample systematically searched from peer-reviewed studies involving SGD interventions for individuals with IDD. However, we did not appraise the degree of experimental control in each of the included studies, nor the EBP status of the SGD field. To investigate whether the practice of SGD interventions for individuals with IDD is evidence-based, some systematic methods for evaluating EBP associated with SCD studies are available, such as the procedural and coding manual for review of evidence-based interventions outlined by the Task Force on Evidence-Based Interventions in School Psychology of American Psychological Association (2002), and the What Works Clearinghouse (WWC) standards for SCD studies (Kratochwill et al., 2010). The WWC standards, for example, suggest that for a literature base to be determined as scientifically validated, it needs to be gauged against standards in terms of design, visual analysis, quantitative measurement, and replication (Kratochwill et al., 2010; Maggin, Briesch, & Chafouleas, 2013). Although we acknowledge these procedures for standardizing the EBP evaluation, the present study did not seek to validate the evidence status for the SGD field and instead only examined effect size. For instance, effect size metrics were computed for data from ABA designs, which does not meet the WWC standard of three intra-individual replications. In part, this was because including effect size metrics pertaining only to studies that meet predetermined standards and demonstrate strong experimental control may bring about other limitations (Manolov, Sierra, Solanas, & Botella, 2014; Nickerson, 2000). Because our focus was on the comparison of effect size metrics, we hope our results may inform one aspect of the EBP evaluation for SCD studies, that is, quantifying effects.

Fourth, although the studies we identified covered diverse experimental SCDs (e.g., withdrawal designs, multiple-baseline or multiple-probe designs, and alternating treatment designs), it remains premature for us to suggest exactly which effect size metric(s) may or may not be appropriate for certain designs, although our findings suggest that IRD and Phi may better serve studies involving both phase and condition contrast (e.g., studies using comparison designs). Notably, some other designs, such as the changing criterion design, were not included in the current study. Thus, future research could identify several groups of studies representing each type of single case research designs and compare the relative utility of different effect size metrics for the different designs.

Finally, we conducted visual analysis using the approach proposed by Petersen-Brown et al. (2012). Other approaches to visual analysis, possibly more rigorous ones, are available, such as the one delineated in the study by Maggin et al. (2013) based on the WWC standards. One potential future research direction could be to examine whether the agreement between effect size metric(s) and visual analysis may vary as a function of how visual analysis is conducted.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In general, with increasing requirements in accountability highlighted by the Individuals with Disabilities Education Act (IDEA, 2004), and a continued recognition for the importance of evidence-based practice in the SGD content area associated with data utilized in this investigation (e.g., ASHA, 2014), improving our understanding of effect size metrics for SCD studies should continue to be addressed because of the premium placed on meta-analyses in providing evidence of *what works*, as well as to contribute to current methods of judging intervention effect sizes among SCD communication intervention studies for samples and populations of individuals living with significant IDD.

Acknowledgments

This research was supported, in part, by grant #2-T73MC12835-03-00 from the Maternal & Child Health Bureau (MCHB) of the U.S. Department of Health and Human Services awarded to the University of Minnesota. The authors would like to acknowledge Dr. John Hoch for his consultation in conducting ROC analysis.

References

References marked with an asterisk indicate studies from which the phase/condition contrasts were obtained.

Allison DB, Gorman BS. Calculating effect sizes for meta-analysis: The case of the single case. Behavior Research and Therapy. 1993; 31:621–631. DOI: 10.1016/0005-7967(93)90115-B

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. American Psychiatric Publishing; Arlington, VA: 2013.

American Psychological Association. Task Force on the Evidence-Based Interventions in School Psychology. Procedural and coding manual for review of evidence-based interventions. 2002. Retrieved from http://www.indiana.edu/~ebi/EBI-Manual.pdf

American Speech-Language-Hearing Association. Evidence-based practice. 2014. Retrieved from: http://www.asha.org/members/ebp/

*. Banda DR, Copple KS, Koul RK, Sancibrian SL, Bogschutz RJ. Video modeling interventions to teach spontaneous requesting using AAC devices to individuals with autism: A preliminary investigation. Disability and Rehabilitation. 2010; 32(16):1364–1372. DOI: 10.3109/09638280903551525 [PubMed: 20465397]

Barlow DH, Hayes SC. Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behavior Analysis. 1979; 12(2):199–210. DOI: 10.1901/jaba.1979.12-199 [PubMed: 489478]

*. Bellon-Harn ML, Harn WE. Scaffolding strategies during repeated story-book reading: An extension using a voice output communication aid. Focus on Autism and Other Developmental Disabilities. 2008; 23(2):112–124. DOI: 10.1177/1088357608316606

Beretvas SN, Chung H. A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. Evidence-Based Communication Assessment and Intervention. 2008; 2(3):129–141. DOI: 10.1080/17489530802446302

Beukelman, DR., Mirenda, P. Augmentative and alternative communication: Supporting children and adults with complex communication needs. 4th ed. Paul Brookes; Baltimore, MD: 2013.

Branson D, Demchak M. The use of augmentative and alternative communication methods with infants and toddlers with disabilities: A research review. Augmentative and Alternative Communication. 2009; 25(4):274–286. DOI: 10.3109/07434610903384529 [PubMed: 19883287]

Busk, PL., Serlin, RC. Meta-analysis for single-case research. In: Kratochwill, TR., Levin, JR., editors. Single-case research designs and analysis: New directions for psychology and education. Lawrence Erlbaum; Hillsdale, NJ: 1992. p. 159-185.

*. Cannella-Malone HI, DeBar RM, Sigafoos J. An examination of preference for augmentative and alternative communication devices with two boys with significant intellectual disabilities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Augmentative and Alternative Communication. 2009; 25(4):262–273. DOI: 10.3109/07434610903384511 [PubMed: 19883289]

Campbell JM. Statistical comparison of four effect sizes for single-subject designs. Behavior Modification. 2004; 28:234–246. DOI: 10.1177/0145445503259264 [PubMed: 14997950]

Chambers, J., Cleveland, W., Kleiner, B., Tukey, P. Graphical methods for data analysis. Wadsworth; Emeryville, CA: 1983.

*. Choi H, O'Reilly M, Sigafoos J, Lancioni G. Teaching requesting and rejecting sequences to four children with developmental disabilities using augmentative and alternative communication. Research in Developmental Disabilities. 2010; 31:560–567. DOI: 10.1016/j.ridd.2009.12.006 [PubMed: 20079604]

*. Chung Y-C, Carter EW. Promoting peer interactions in inclusive classrooms for students who use speech-generating devices. Research & Practice for Persons with Severe Disabilities. 2013; 38(2):94–109. DOI: 10.2511/027494813807714492

Cleveland, W. Elements of graphing data. Wadsworth; Emeryville, CA: 1985.

*. Dattilo J, Camarata S. Facilitating conversation through self-initiated augmentative communication treatment. Journal of Applied Behavior Analysis. 1991; 24(2):369–378. DOI: 10.1901/jaba.1991.24-369 [PubMed: 1890052]

*. Davis CA, Reichle J, Southard K, Johnston S. Teaching children with severe disabilities to utilize nonobligatory conversational opportunities: An application of high-probability requests. Journal of the Association for Persons with Severe Handicaps. 1998; 23(1):57–68. DOI: 10.2511/rpsd.23.1.57

Dicarlo CF, Banajee M. Using voice output devices to increase initiations of young children with disabilities. Journal of Early Intervention. 2000; 23(3):191–199. DOI: 10.1177/1053815100023003080 1

*. Durand VM. Functional communication training using assistive devices: Effects on challenging behavior and affect. Augmentative and Alternative Communication. 1993; 9:168–176. DOI: 10.1080/07434619312331276571

*. Durand VM. Functional communication training using assistive devices: Recruiting natural communities of reinforcement. Journal of Applied Behavior Analysis. 1999; 32(3):247–267. DOI: 10.1901/jaba.1999.32-247 [PubMed: 10513023]

*. Dyches TT. Effects of switch training on the communication of children with autism and severe disabilities. Focus on Autism and Other Developmental Disabilities. 1998; 13(3):151–162. DOI: 10.1177/108835769801300303

*. Falcomata TS, Ringdahl JE, Christensen TJ, Boelter EW. An evaluation of prompt schedules and mand preference during functional communication training. The Behavior Analyst Today. 2010; 11(1):77–84. DOI: 10.1037/h0100690

Fernandez B. iTherapy: The revolution of mobile devices within the field of speech therapy. SIG. 2011; 16:35–40. Perspectives on School-Based Issues, 12(2). DOI: 10.1044/sbi12.2.35

Ferris K, Fabrizio MA. Comparison of error correction procedures involving a speech-generating device to teach a child with autism new tacts. The Journal of Speech-Language Pathology and Applied Behavior Analysis. 2008; 3:185–198. DOI: 10.1037/h0100246

*. Flores M, Musgrove K, Renner S, Hinton V, Strozier S, Franklin S, Hil D. A comparison of communication using the Apple iPad and a picture-based system. Augmentative and Alternative Communication. 2012; 28(2):74–84. DOI: 10.3109/07434618.2011.644579 [PubMed: 22263895]

Fragale CL, O'Reilly MF, Aguilar J, Pierce N. The influence of motivating operations on generalization probes of specific mands by children with autism. Journal of Applied Behavior Analysis. 2012; 45(3):565–577. DOI: 10.1901/jaba.2012.45-565 [PubMed: 23060669]

Ganz JB, Earles-Vollrath TL, Heath AK, Parker RI, Rispoli MJ, Duran JB. A meta-analysis of single case research studies on aided augmentative and alternative communication systems with individuals with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2012; 42:60–74. DOI: 10.1007/s10803-011-1212-2 [PubMed: 21380612]

Gast, DL., Ledford, JR. Single case research methodology: Applications in special education and behavioral sciences. 2nd ed. Routledge; New York, NY: 2014.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gevarter C, O'Reilly MF, Rojeski L, Sammarco N, Lang R, Lancioni GE, Sigafoos J. Comparing communication systems for individuals with developmental disabilities: A review of single-case research studies. Research in Developmental Disabilities. 2013; 34:4415–4432. DOI: 10.1016/j.ridd.2013.09.017 [PubMed: 24377101]

Gosnell J, Costello J, Shane H. Using a clinical approach to answer "What communication Apps should we use?". Perspectives on Augmentative and Alternative Communication. 2011; 20(3):87–96. DOI: 10.1044/aac20.3.87

*. Hanley GP, Iwata BA, Thompson RH. Reinforcement schedule thinning following treatment with functional communication training. Journal of Applied Behavior Analysis. 2001; 34(1):17–38. DOI: 10.1901/jaba.2001.34-17 [PubMed: 11317985]

Hershberger D. Mobile technology and AAC apps from an AAC developer's perspective. Perspectives on Augmentative and Alternative Communication. 2011; 20(1):28–33. DOI: 10.1044/aac20.1.28

Herzinger CV, Campbell JM. Comparing functional assessment methodologies: A quantitative synthesis. Journal of Autism and Developmental Disorders. 2007; 37:1430–1445. DOI: 10.1007/s10803-006-0219-6 [PubMed: 17004118]

Hintze, J. *NCSS and PASS* [Computer software]. Number Cruncher Statistical Systems; Kaysville, UT: 2004.

*. Hunt P, Sota G, Maler J, Muller E, Goetz L. Collaborative teaming to support students with augmentative and alternative communication needs in general education classrooms. Augmentative and Alternative Communication. 2002; 18:20–35. DOI: 10.1080/aac.18.1.20.35

Huwaldt, JA. PlotDigitizer. 2010. http://plotdigitizer.sourceforge.net/

Ickowicz A. Paucity of negative clinical trials reports and publication bias. Journal of the Canadian Academy of Child and Adolescent Psychiatry. 2014; 23(1):7. [PubMed: 24516471]

Individuals With Disabilities Education Improvement Act. 2004. Pub. L. No. 108–446

Jenson WR, Clark E, Kircher JC, Kristjansson SD. Statistical reform: evidence based practice, meta-analyses, and single subject designs. Psychology in the Schools. 2007; 44:483–494. DOI: 10.1002/pits.20240

*. Johnston SS, McDonnell AP, Nelson C, Magnavito A. Teaching functional communication skills using augmentative and alternative communication in inclusive settings. Journal of Early Intervention. 2003; 25(4):263–280. DOI: 10.1177/105381510302500403

Johnston, SS., Reichle, J., Feeley, K., Jones, E. AAC strategies for individuals with moderate and severe disabilities. Paul Brookes; Baltimore, MD: 2012.

Kagohara DM. Three students with developmental disabilities learn to operate an iPod to access age-appropriate entertainment videos. Journal of Behavioral Education. 2010; 20(1):33–43. DOI: 10.1007/s10864-010-9115-4

Kagohara DM, van der Meer L, Ramdoss S, O'Reilly MF, Lancioni GE, Davis TN, Sigafoos J. Using iPods and iPads in teaching programs for individuals with developmental disabilities: A systematic review. Research in Developmental Disabilities. 2013; 34:147–156. DOI: 10.1016/j.ridd.2012.07.027 [PubMed: 22940168]

*. Kennedy CH, Haring TG. Teaching choice making during social interactions to students with profound multiple disabilities. Journal of Applied Behavior Analysis. 1993; 26:63–76. DOI: 10.1901/jaba.1993.26-63 [PubMed: 8473259]

Kratochwill, TR., Hitchcock, J., Horner, RH., Levin, JR., Odom, SL., Rindskopf, DM., Shadish, WR. Single-case designs technical documentation. 2010. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kuster, JM. Internet: In search of the perfect speech-language App?. The ASHA Leader. 2012. http://www.asha.org/Publications/leader/2012/120403/Internet–In-Search-of-the-Perfect-Speech-Language-App/

*. Lancioni GE, O'Reilly MF, Oliva D, Coppa MM. Using multiple micro-switches to promote different responses in children with multiple disabilities. Research in Developmental Disabilities. 2001; 22:309–318. DOI: 10.1016/S0891-4222(01)00074-9 [PubMed: 11523954]

Lancioni GE, Singh NN, O'Reilly MF, Sigafoos J, Didden R, Oliva D, Lamartire ML. Effects of microswitch-based programs on indices of happiness of students with multiple disabilities: A new

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

research evaluation. American Journal on Mental Retardation. 2007; 112(3):167–176. DOI: 10.1352/0895-8017(2007)112[167:EOMPOI]2.0.CO;2 [PubMed: 17542654]

*. Lancioni GE, Singh NN, O'Reilly MF, Sigafoos J, Oliva D, Baccani S. Teaching 'yes' and 'no' responses to children with multiple disabilities through a program including microswitches linked to a vocal output device. Perceptual and Motor Skills. 2006; 102:51–61. DOI: 10.2466/pms.102.1.51-61 [PubMed: 16671596]

Lancioni GE, O'Reilly MF, Cuvo AJ, Singh NN, Sigafoos J, Didden R. PECS and VOCAs to enable students with developmental disabilities to make requests: An overview of the literature. Research in Developmental Disabilities. 2007; 28:468–488. DOI: 10.1016/j.ridd.2006.06.003 [PubMed: 16887326]

Lancioni GE, O'Reilly MF, Singh NN, Buonocunto F, Sacco V, Colonna F, Bosco A. Technology-based intervention options for post-coma persons with minimally conscious state and pervasive motor disabilities. Developmental Neurorehabilitation. 2009; 12(1):24–31. DOI: 10.1080/17518420902776995 [PubMed: 19283531]

*. Lancioni GE, O'Reilly MF, Singh NN, Sigafoos J, Oliva D, Severini L. Enabling two persons with multiple disabilities to access environmental stimuli and ask for social contact through microswitches and a VOCA. Research in Developmental Disabilities. 2008; 29:21–28. DOI: 10.1016/j.ridd.2006.10.001 [PubMed: 17174529]

*. Lancioni GE, O'Reilly MF, Singh NN, Sigafoos J, Oliva D, Smaldone A, Chiapparino C. Persons with multiple disabilities access stimulation and contact the caregiver via microswitch and VOCA technology. Life Span and Disability. 2009; 2:119–128.

*. Lancioni GE, O'Reilly MF, Singh NN, Sigafoos J, Didden R, Oliva D, Groeneweg J. Persons with multiple disabilities accessing stimulation and requesting social contact via microswitch and VOCA devices: New research evaluation and social validation. Research in Developmental Disabilities. 2009; 30:1084–1094. DOI: 10.1016/j.ridd.2009.03.004 [PubMed: 19361954]

*. Lancioni GE, Singh NN, O'Reilly MF, Sigafoos J, Didden R, Smaldone A, La Martire ML. Helping a man with multiple disabilities to use single vs repeated performance of simple motor schemes as different responses. Perceptual and Motor Skills. 2010; 110(1):105–113. DOI: 10.2466/pms.110.1.105-113 [PubMed: 20391876]

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. DOI: 10.2307/2529310 [PubMed: 843571]

*. Light JC, Binger C, Agate TL, Ramsay KN. Teaching partner-focused questions to individuals who use augmentative and alternative communication to enhance their communicative competence. Journal of Speech, Language, and Hearing Research. 1999; 42:241–255. DOI: 10.1044/jslhr.4201.241

Lipsey, MW., Wilson, DB. Practical Meta-analysis. SAGE; Thousand Oaks, CA: 2001.

*. Logan KR, Jacobs HA, Gast DL, Smith PD, Daniel J, Rawls J. Preferences and reinforcers for students with profound multiple disabilities: Can we identify them? Journal of Developmental and Physical Disabilities. 2001; 13(2):97–122. DOI: 10.1023/A:1016624923479

Ma H. An alternative method for quantitative synthesis of single subject researches: Percentage of data points exceeding the median. Behavior Modification. 2006; 30:598–617. DOI: 10.1177/0145445504272974 [PubMed: 16894232]

Maggin DM, Briesch AM, Chafouleas SM. An application of the What Works Clearinghouse Standards for evaluating single-subject research: Synthesis of the self-management literature base. Remedial and Special Education. 2013; 34:44–58. DOI: 10.1177/0741932511435176

Maggin DM, Swaminathan H, Rogers HJ, O'Keefe BV, Sugai G, Horner RH. A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. Journal of School Psychology. 2011; 49:301–321. DOI: 10.1016/j.jsp.2011.03.004 [PubMed: 21640246]

Manolov R, Sierra V, Solanas A, Botella J. Assessing functional relations in single-case designs: Quantitative proposals in the context of the evidence-based movement. Behavior Modification. 2014; 38(6):878–913. DOI: 10.1177/0145445514545679 [PubMed: 25092718]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

*. McMillan JM. Teachers make it happen: From professional development to integration of augmentative and alternative communication technologies in the classroom. Australasian Journal of Special Education. 2008; 32(2):199–211. DOI: 10.1080/10300110802047467

McNaughton D, Light J. The iPad and mobile technology revolution: Benefits and challenges for individuals who require Augmentative and Alternative Communication. Perspectives on Augmentative and Alternative Communication. 2013; 29(2):107–116. DOI: 10.3109/07434618.2013.784930

Meline T, Wang B. Effect-size reporting practices in AJSLP and other ASHA journals 1999-2003. American Journal of Speech-Language Pathology. 2004; 13:202–207. DOI: 10.1044/1058-0360(2004/021) [PubMed: 15339229]

Microsoft Company. Microsoft Office Excel 2007. Microsoft Press; 2007.

Mirenda P. Toward functional augmentative and alternative communication for students with autism: Manual signs, graphic symbols, and voice output communication aids. Language, Speech, and Hearing Services in Schools. 2003; 34:203–216. DOI: 10.1044/0161-1461(2003/017)

Muller MP, Tomlinson G, Marrie TJ, Tang P, McGreer A, Low DE, Gold WL. Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? Clinical Infectious Diseases. 2005; 40(8):1079–1086. DOI: 10.1086/428577 [PubMed: 15791504]

Millar DC, Light JC, Schlosser RW. The impact of augmentative and alternative communication intervention on the speech production of individuals with developmental disabilities: A research review. Journal of Speech, Language, and Hearing Research. 2006; 49:248–264. DOI: 10.1044/1092-4388(2006/021)

Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. Psychological Methods. 2000; 5:241–301. DOI: 10.1037//1082-989X.5.2.241 [PubMed: 10937333]

*. Northup J, Wacker DP, Berg WK, Kelly L, Sasso G, DeRaad A. The treatment of severe behavior problems in school settings using a technical assistance model. Journal of Applied Behavior Analysis. 1994; 27:33–47. DOI: 10.1901/jaba.1994.27-33 [PubMed: 8188562]

Odom SL. The tie that binds: Evidence-based practice, implementation science, and outcomes for children. Topics in Early Childhood Special Education. 2009; 29(1):53–61. DOI: 10.1177/0271121408329171

*. O'Keefe BM, Dattilo J. Teaching the response-recode form to adults with mental retardation using AAC systems. Augmentative and Alternative Communication. 1992; 8:224–233. DOI: 10.1080/07434619212331276213

Olive ML, de la Cruz B, Davis TN, Chan JM, Lang RB, O'Reilly MF, Dickson SM. The effects of enhanced milieu teaching and a voice output communication aid on the requesting of three children with autism. Journal of Autism and Developmental Disorders. 2007; 37:1505–1513. DOI: 10.1007/s10803-006-0243-6 [PubMed: 17066309]

Orlikoff, RF., Schiavetti, NE., Metz, DE. Evaluating research in communication disorders. 7th ed. Pearson Education, Inc; Boston, MA: 2015.

Parker RI, Hagan-Burke S. Median-based overlap analysis for single case data. Behavior Modification. 2007; 31(6):919–936. DOI: 10.1177/0145445507303452 [PubMed: 17932244]

Parker RI, Hagan-Burke S, Vannest K. Percentage of all non-overlapping data (PAND): An alternative to PND. The Journal of Special Education. 2007; 40:194–204. DOI: 10.1177/00224669070400040101

Parker RI, Vannest KJ. An improved effect size for single-case research: Nonoverlap of all pairs. Behavior Therapy. 2009; 40:357–367. DOI: 10.1016/j.beth.2008.10.006 [PubMed: 19892081]

Parker RI, Vannest KJ. Bottom-up analysis of single-case research design. Journal of Behavioral Education. 2012; 21(3):254–265. DOI: 10.1007/s10864-012-9153-1

Parker RI, Vannest KJ, Brown L. The improvement rate difference for single-case research. Exceptional Children. 2009; 75:135–150.

Parker RI, Vannest KJ, Davis JL. Effect size in single-case research: A review of nine nonoverlap techniques. Behavior Modification. 2011; 35(4):303–322. DOI: 10.1177/0145445511399147 [PubMed: 21411481]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Parker RI, Vannest KJ, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-*U*. Behavior Therapy. 2011; 42:284–299. DOI: 10.1016/j.beth.2010.08.006 [PubMed: 21496513]

Petersen-Brown S, Karich AC, Symons FJ. Examining estimates of effect using non-overlap of all pairs in multiple baseline studies of academic intervention. Journal of Behavioral Education. 2012; 21:203–216. DOI: 10.1007/s10864-012-9154-0

Radstaake M, Didden R, Lang R, O'Reilly M, Sigafoos J, Lancioni GE, Curfs LMG. Functional analysis and functional communication training in the classroom for three children with Angelman Syndrome. Journal of Developmental and Physical Disabilities. 2013; 25:49–63. DOI: 10.1007/s10882-012-9302-4

Reichle, J., Beukelman, DR., Light, JC. Exemplary practices for beginning communicators: Implications for AAC. Paul H. Brookes; Baltimore, MD: 2002.

*. Reichle J, Dettling EE, Drager KDR, Leiter A. Comparison of correct responses and response latency for fixed and dynamic displays: Performance of a learner with severe developmental disabilities. Augmentative and Alternative Communication. 2000; 16:154–163. DOI: 10.1080/07434610012331279014

Reichle J, Ward M. Teaching discriminative use of an encoding electronic communication device and signing exact English to a moderately handicapped child. Language, Speech, and Hearing Services in Schools. 1985; 16:58–63. DOI: 10.1044/0161-1461.1601.58

Rispoli MJ, Franco JH, van der Meer L, Lang R, Camargo SPH. The use of speech generating devices in communication interventions for individuals with developmental disabilities: A review of the literature. Developmental Neurorehabilitation. 2010; 13(4):276–293. DOI: 10.3109/17518421003636794 [PubMed: 20629594]

Russell T, Beard L. Computer assisted speech as an alternative communication system for the severely mentally handicapped. Baylor Educator. 1992; 17:43–50.

*. Schepis MM, Reid DH, Behrmann MM, Sutton KA. Increasing communicative interactions of young children with autism using a voice output communication aid and naturalistic teaching. Journal of Applied Behavior Analysis. 1998; 31:561–578. DOI: 10.1901/jaba.1998.31-561 [PubMed: 9891394]

*. Schepis MM, Reid DH. Effects of a voice output communication aid on interactions between support personnel and an individual with multiple disabilities. Journal of Applied Behavior Analysis. 1995; 28(1):73–77. DOI: 10.1901/jaba.1995.28-73 [PubMed: 7706152]

*. Schepis MM, Reid DH, Behrman MM. Acquisition and functional use of voice output communication by persons with profound multiple disabilities. Behavior Modification. 1996; 20(4):451–468. DOI: 10.1177/01454455960204005 [PubMed: 8875815]

Schlosser RW, Belfiore PJ, Nigam R, Blischak D, Hetzroni O. The effects of speech output technology in the learning of graphic symbols. Journal of Applied Behavior Analysis. 1995; 28(4):537–549. DOI: 10.1901/jaba.1995.28-537 [PubMed: 14743828]

Schlosser RW, Sigafoos J. Augmentative and alternative communication interventions for persons with developmental disabilities: narrative review of comparative single-subject experimental studies. Research in Developmental Disabilities. 2006; 27:1–29. DOI: 10.1016/j.ridd.2004.04.004 [PubMed: 16360073]

Schlosser, RW., Sigafoos, J., Koul, RK. Speech output and speech-generating devices in autism spectrum disorders. In: Mirenda, P., Iacono, T., editors. Autism spectrum disorders and AAC. Paul Brookes; Baltimore, MD: 2009. p. 141-169.

Schlosser RW, Wendt O. Effects of augmentative and alternative communication intervention on speech production in children with autism: A systematic review. American Journal of Speech-Language Pathology. 2008; 17:212–230. DOI: 10.1044/1058-0360(2008/021) [PubMed: 18663107]

Shadish WR, Hedges LV, Pustejovsky JE. Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. Journal of School Psychology. 2014; 52(2):123–147. DOI: 10.1016/j.jsp.2013.11.005 [PubMed: 24606972]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Shadish WR, Rindskopf DM, Hedges LV. The state of the science in the meta-analysis of single-case experimental designs. Evidence-Based Communication Assessment and Intervention. 2008; 2(3): 188–196. DOI: 10.1080/17489530802581603

*. Sigafoos J, Roberts-Pennell D. Wrong-item format: A promising intervention for teaching socially appropriate forms of rejecting to children with developmental disabilities? Augmentative and Alternative Communication. 1999; 15:135–140. DOI: 10.1080/07434619912331278635

Sigafoos J, Didden R, O'Reilly M. Effects of speech output on maintenance of requesting and frequency of vocalizations in three children with developmental disabilities. Augmentative and Alternative Communication. 2003; 19(1):37–47. DOI: 10.1080/0743461032000056487

*. Sigafoos J, Drasgow E. Conditional use of aided and unaided AAC: A review and clinical case demonstration. Focus on Autism and Other Developmental Disabilities. 2001; 16(3):152–161. DOI: 10.1177/108835760101600303

Sigafoos J, Drasgow E, Halle JW, O'Reilly M, Seely-York S, Edrisinha C, Andrews A. Teaching VOCA use as a communicative repair strategy. Journal of Autism and Developmental Disorders. 2004; 34(4):411–422. DOI: 10.1023/B:JADD.0000037417.04356.9c [PubMed: 15449516]

*. Sigafoos J, O'Reilly M, Seely-York S, Edrisinha C. Teaching students with developmental disabilities to locate their AAC device. Research in Developmental Disabilities. 2004; 25:371–383. DOI: 10.1016/j.ridd.2003.07.002 [PubMed: 15193671]

*. Sigafoos J, Ganz J, O'Reilly M, Lancioni G. Evidence-based practice in the classroom: Evaluating a procedure for reducing perseverative requesting in an adolescent with autism and severe intellectual disability. Australasian Journal of Special Education. 2008; 32(1):55–65. DOI: 10.1080/10300110701839931

Sigafoos J, Green VA, Payne D, Son S-H, O'Reilly M, Lancioni GE. A comparison of picture exchange and speech-generating devices: Acquisition, preference, and effects on social interaction. Augmentative and Alternative Communication. 2009; 25:99–109. DOI: 10.1080/07434610902739959 [PubMed: 19444681]

Sigafoos J, O'Reilly MF, Seely-York S, Weru J, Son SH, Green VA, Lancioni GE. Transferring AAC intervention to the home. Disability and Rehabilitation. 2004; 26(21–22):1330–1334. DOI: 10.1080/09638280412331280361 [PubMed: 15513733]

*. Sigafoos J, Wermink H, Didden R, Green VA, Schlosser RW, O'Reilly MF, Lancioni GE. Effects of varying lengths of synthetic speech output on augmented requesting and natural speech production in an adolescent with Klinefelter syndrome. Augmentative and Alternative Communication. 2011; 27(3):163–171. DOI: 10.3109/07434618.2011.610355 [PubMed: 22008029]

Snell ME, Brady N, McLean L, Ogletree BT, Siegel E, Sylvester L, Sevcik R. Twenty years of communication intervention research with individuals who have severe intellectual and developmental disabilities. American Journal of Intellectual and Developmental Disabilities. 2010; 115(5):364–380. DOI: 10.1352/1944-7558-115-5.364

Snell ME, Chen L-Y, Hoover K. Teaching augmentative and alternative communication to students with severe disabilities: A review of intervention research 1997-2003. Research & Practice for Persons with Severe Disabilities. 2006; 31(3):203–214. DOI: 10.1177/154079690603100301

*. Soto G, Belfiore PJ, Schlosser RW, Haynes C. Teaching specific requests: A comparative analysis on skill acquisition and preference using two augmentative and alternative communication aids. Education and Training in Mental Retardation. 1993; 28:169–178.

*. Steege MW, Wacker DP, Cigrand KC, Berg WK, Novak CG, Reimers TM, DeRaad A. Use of negative reinforcement in the treatment of self-injurious behavior. Journal of Applied Behavior Analysis. 1990; 23(4):459–467. DOI: 10.1901/jaba.1990.23-459 [PubMed: 2150070]

Stephenson J, Limbrick L. A review of the use of touch-screen mobile devices by people with developmental disabilities. Journal of Autism and Developmental Disorders. 2015; 45:3777–3791. DOI: 10.1007/s10803-013-1878-8 [PubMed: 23888356]

U.S. Department of Health and Human Services. Developmental Disabilities Assistance and Bill of Rights Act. 2000. Retrieved from http://wwhelp.wwrc.net/wwwebhelp/developmental_disabilities_assistance_and_bill_of_rights_act_dd_act_.htm

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

*. van Acker R, Grant SH. An effective computer-based requesting system for persons with Rett syndrome. Communication Disorders Quarterly. 1995; 16(2):31–38. DOI: 10.1177/152574019501600205

van den Noortgate W, Onghena P. Hierarchical linear models for the quantitative integration of effect sizes in single-case research. Behavior Research Methods, Instruments & Computers. 2003; 35:1–10. DOI: 10.3758/BF03195492

van den Noortgate W, Onghena P. A multilevel meta-analysis of single-subject experimental design studies. Evidence-Based Communication Assessment and Intervention. 2008; 2(3):142–151. DOI: 10.1080/17489530802505362

*. van der Meer L, Didden R, Sutherland D, O'Reilly MF, Lancioni GE, Sigafoos J. Comparing three augmentative and alternative communication modes for children with developmental disabilities. Journal of Developmental and Physical Disabilities. 2012; 24:451–468. DOI: 10.1007/s10882-012-9283-3

van der Meer L, Kagohara D, Achmadi D, Green VA, Herrington C, Sigafoos J. Teaching functional use of an iPod-based speech-generating device to individuals with developmental disabilities. Journal of Special Education Technology. 2011; 26(3):1–11. DOI: 10.1177/016264341102600301

*. van der Meer L, Kagohara D, Achmadi D, O'Reilly MF, Lancioni GE, Sutherland D, Sigafoos J. Speech-generating devices versus manual signing for children with developmental disabilities. Research in Developmental Disabilities. 2012; 33:1658–1669. DOI: 10.1016/j.ridd.2012.04.004 [PubMed: 22554812]

van der Meer LAJ, Rispoli M. Communication interventions involving speech-generating devices for children with autism: A review of the literature. Developmental Neurorehabilitation. 2010; 13(4): 294–306. DOI: 10.3109/17518421003671494 [PubMed: 20629595]

*. van der Meer L, Sutherland D, O'Reilly MF, Lancioni GE, Sigafoos J. A further comparison of manual signing, picture exchange, and speech-generating devices as communication modes for children with autism spectrum disorders. Research in Autism Spectrum Disorders. 2012; 6:1247–1257. DOI: 10.1016/j.rasd.2012.04.005

*. van der Meer L, Kagohara D, Roche L, Sutherland D, Balandin S, Green VA, Sigafoos J. Teaching multi-step requesting and social communication to two children with autism spectrum disorders with three AAC options. Augmentative and Alternative Communication. 2013; 29(3):222–234. DOI: 10.3109/07434618.2013.815801 [PubMed: 23879660]

Vannest, KJ., Parker, RI., Gonen, O. Single case research: Web based calculators for SCR analysis. Version 1.0. Texas A&M University; College Station, TX: 2011. [Web-based application]Retrieved from http://www.singlecaseresearch.org

*. Wacker DP, Wiggins B, Fowler M, Berg WK. Training students with profound or multiple handicaps to make requests via microswitches. Journal of Applied Behavior Analysis. 1988; 21(4):331–343. DOI: 10.1901/jaba.1988.21-331 [PubMed: 2976066]

*. Wacker DP, Harding JW, Berg WK, Lee JF, Schieltz KM, Padilla YC, Shahan TA. An evaluation of persistence of treatment effects during long-term treatment of destructive behavior. Journal of the Experimental Analysis of Behavior. 2011; 96(2):261–282. DOI: 10.1901/jeab.2011.96-261 [PubMed: 21909168]

Wacker DP, Steege MW, Northup J, Sasso G, Berg W, Reimers T, Donn L. A component analysis of functional communication training across three topographies of severe behavior problems. Journal of Applied Behavior Analysis. 1990; 23:417–429. DOI: 10.1901/jaba.1990.23-417 [PubMed: 2150069]

Wolery M, Busick M, Reichow B, Barton EE. Comparison of overlap methods for quantitatively synthesizing single-subject data. The Journal of Special Education. 2010; 44(1):18–28. DOI: 10.1177/0022466908328009

Wolery M, Dunlap G, Ledford JR. Single-case experimental methods: Suggestions for reporting. Journal of Early Intervention. 2011; 33:103–109. DOI: 10.1177/1053815111418235

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

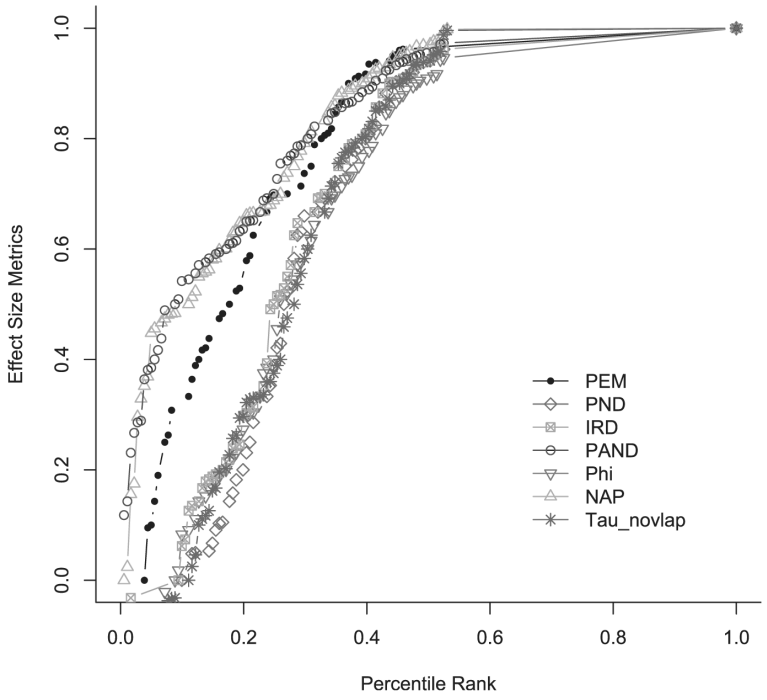

Figure 1.

Uniform probability plots for seven effect size metrics on 181 phase contrasts from studies that did not involve intervention comparisons. PEM = percent of data points exceeding the median; PND = percent of nonoverlapping data; IRD = improvement rate difference; PAND = percent of all nonoverlapping data; NAP = nonoverlap of all pairs; Tau_{novlap} = Kendall's tau nonoverlap.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

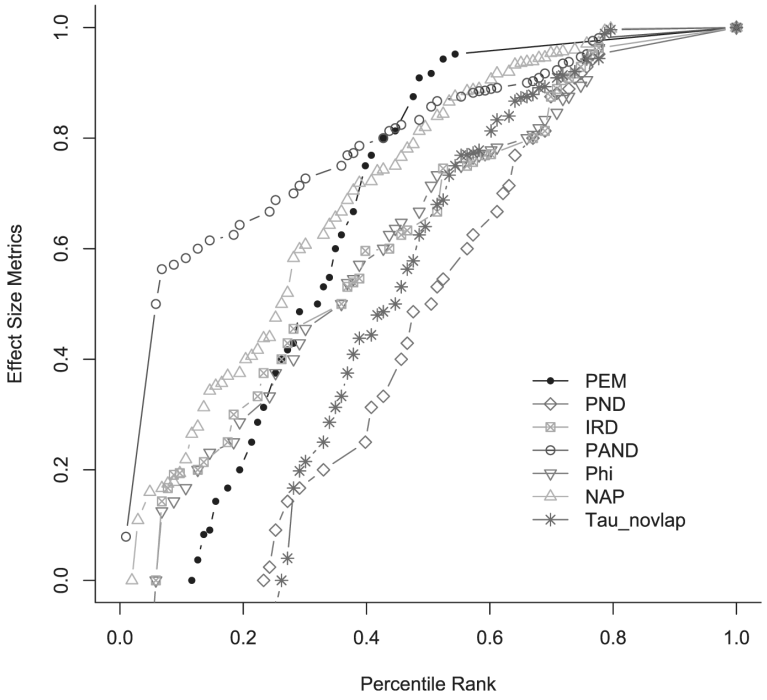

Figure 2.

Uniform probability plots for seven effect size metrics on 104 phase/condition contrasts from studies that involved intervention comparisons. PEM = percent of data points exceeding the median; PND = percent of nonoverlapping data; IRD = improvement rate difference; PAND = percent of all nonoverlapping data; NAP = nonoverlap of all pairs; Tau_{novlap} = Kendall's tau nonoverlap.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Major Effect Size Metrics for Single Case Design (SCD) Studies

Category	Examples	Brief Description
Parametric	Standardized Mean Difference (SMD)	The mean difference between Phase B and Phase A, divided by the standard deviation of Phase A (Busk & Serlin, 1992).
	Pearson R or R^2; Regression-Based Standardized Mean Difference (d_{REG})	The use of a linear-regression technique to remove trend from repeated observations by calculating predicted values based on data in Phase A only. The resulting adjusted R^2 value is converted to d_{REG} via a standard formula (Allison & Gorman, 1993).
	Cohen's d, Hedges's g	A measure of standardized difference between the mean of Phase A and that of Phase B (Beretvas & Chung, 2008).
	Multilevel modeling	The application of Hierarchical Linear Models (HLM) for synthesizing SCD data (van den Noortgate & Onghena, 2003; 2008).
Non-Parametric	Mean Baseline Difference (MBD) or Mean Baseline Reduction (MBLR)	The difference between the mean of Phase A and the mean of Phase B, divided by the mean of Phase A and multiplying by 100 (Herzinger & Campbell, 2007).
	Percent of Data Exceeding the Median of Baseline (PEM)	The percentage of data points in Phase B exceeding the median of data points in Phase A (Ma, 2006).
	Percentage of Data Exceeding a Median Trend (PEM-T)	The percentage of data points exceeding the trend line (Wolery et al., 2010).
	Kruskal-Wallis W	A rank-based measure of agreement between Phase A and Phase B (Hintze, 2004).
	Percentage of Nonoverlapping Data (PND)	The percentage of data points in Phase B exceeding the single highest data point in Phase A (Scruggs, Mastropieri, & Casto, 1987).
	Percentage of Zero Data (PZD)	The percentage of data points in Phase B that remain at zero since the first data point of zero including the first zero (Scotti, Evans, Meyer, & Walker, 1991).
	Percentage of All Nonoverlapping Data (PAND)	The percentage of data points that overlap between Phase A and Phase B out of the total number of data points across Phase A and Phase B, subtracted from 100%. Overlapping data points refer to minimum number that would have to be transferred across phases for complete data separation (Parker, Hagan-Burke, & Vannest, 2007).
	Phi	A metric derived from PAND, can be easily calculated as PAND × 2 – 1 (Parker, Hagan-Burke, & Vannest, 2007).
	Improvement Rate Difference (IRD)	The difference of the improvement rate (IR) in Phase B and that of Phase A. The IR for each phase is defined as the number of "improved data points" divided by the total data points in that phase (Parker & Hagan-Burke, 2007).
	Pairwise Data Overlap (PDO) and Pairwise Data Overlap Squared (PDO^2); Nonoverlap of All Pairs (NAP); Tau_{novlap}	All these metrics examine the overlap between Phase A and Phase B by looking at the pairwise comparisons of data points in Phase A and Phase B. However each metric has its own specific calculation formula (see Parker & Vannest, 2009; Parker, Vannest, & Davis, 2011; Wolery et al., 2010).
	Tau-U Family	The family consists of four specific metrics: Tau_{novlap}; A vs. B + $Trend_B$ (a metric examining overlap between Phase A and Phase B, as well as Phase B trend); A vs. B – $Trend_A$ (a metric examining overlap between Phase A and Phase B, with Phase A trend controlled); A vs. B + $Trend_B$ – $Trend_A$ (a metric examining overlap between Phase A and Phase B, as well as Phase B trend, with Phase A trend controlled; Parker, Vannest, Davis, & Sauber, 2011).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Am J Intellect Dev Disabil. Author manuscript; available in PMC 2017 February 17.

Table 2

Effect Size Metric Comparison Research for Single Case Design (SCD) Studies

Author(s)	Effect Size Metrics Compared	# of Datasets	Source of Datasets	Designs Involved	Major Analyses
Campbell (2004)	PND, PZD, MBLR, and regression-based *d* (d_{REG})	117	A systematic search of studies on the reduction of problem behavior	Not specified	1. Statistical distribution (mean, standard deviation, minimum value, maximum value) 2. Intercorrelation analysis
Ma (2006)	PEM, PND	202	A systematic search of studies on self-control	Reversal design; or multiple-baseline design	1. Statistical distribution (mean and standard deviation) 2. Correlation with original authors' judgment of intervention effects
Parker & Hagan-Burke (2007)	PEM, PND, Pearson *R*, Kruskal-Wallis *W*, and IRD	165	A convenience sample of 35 articles	Multiple phase design; or multiple baseline designs	1. Statistical distribution (mean and standard deviation) 2. Intercorrelation analysis 3. Discriminability analysis through analyzing a uniform probability plot (Cleveland, 1985) 4. Agreement with visual analysis
Parker, Hagan-Burke, & Vannest (2007)	PAND, Phi, PND, and R^2	75	A convenience sample of 49 studies	Multiple baseline design	1. Statistical distribution (percentile ranks) 2. Intercorrelation analysis 3. Statistical power analysis
Parker, Vannest & Brown (2009)	IRD, PND, R^2, and Kruskal-Wallis *W*	166	A random sample of published studies	Not specified	1. Intercorrelation analysis 2. Discriminability analysis
Parker & Vannest (2009)	NAP, PND, PEM, PAND, and R^2	200	A random sample of 44 articles	AB pre-experimental designs	1. Statistical distribution (Percentile ranks) 2. Discriminability analysis 3. Intercorrelation analysis 4. Agreement with visual analysis
Wolery, Busick, Reichow, & Barton (2010)	PND, PDO^2, PEM, and PEM-T	160	A random sample of articles from the *Journal of Applied Behavior Analysis*	Not specified	1. Agreement with visual analysis
Parker, Vannest, Davis, & Sauber (2011)	Tau-*U* family	382	A convenience sample of published articles	Not specified	1. Statistical distribution (percentile ranks) 2. Discriminability analysis
Parker, Vannest, & Davis (2011)	PEM, PND, IRD, PAND, Phi, NAP, and Tau_{novlap}	200	A convenience sample of more than 60	Not specified	1. Statistical distribution (percentile ranks)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author(s)	Effect Size Metrics Compared	# of Datasets	Source of Datasets	Designs Involved	Major Analyses
			articles		

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Intercorrelations Among Seven Effect Size Metrics for 181 Phase Contrasts From Studies That Did Not Involve Intervention Comparisons (Left Lower/Not Bold) and 104 Phase/Condition Contrasts From Studies That Involved Intervention Comparisons (Right Upper/Bold), Respectively

Metric	PEM	PND	IRD	PAND	Phi	NAP	Tau_{novlap}
PEM	1	**.5**	**.5**	**.4**	**.4**	**.9**	**.9**
PND	.7	1	**.8**	**.7**	**.7**	**.6**	**.6**
IRD	.8	.9	1	**.9**	**.9**	**.5**	**.5**
PAND	.7	.9	.9	1	**.9**	**.4**	**.4**
Phi	.7	.9	.9	.9	1	**.4**	**.4**
NAP	.9	.7	.8	.7	.7	1	**.9**
Tau_{novlap}	.9	.7	.8	.7	.7	.9	1

Note. PEM = percent of data points exceeding the median; PND = percent of nonoverlapping data; IRD = improvement rate difference; PAND = percent of all nonoverlapping data; NAP = nonoverlap of all pairs; Tau_{novlap} = Kendall's tau nonoverlap.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Cutoff Scores for Effect Size Metrics in Differentiating Large and Small Effects Based on Visual Analysis, and Corresponding AUC, Sensitivity, and Specificity Values for All 45 Studies

Metric	Cutoff Score	AUC	Sensitivity	Specificity
PEM	0.72	0.87	0.76	0.98
PND	0.51	0.96	0.88	1.00
IRD	0.49	0.95	0.88	1.00
PAND	0.77	0.90	0.80	0.99
Phi	0.54	0.90	0.80	0.99
NAP	0.73	0.90	0.80	0.99
Tau_{novlap}	0.47	0.90	0.80	0.99

Note. PEM = percent of data points exceeding the median; PND = percent of nonoverlapping data; IRD = improvement rate difference; PAND = percent of all nonoverlapping data; NAP = nonoverlap of all pairs; Tau_{novlap} = Kendall's tau nonoverlap.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Agreement Between Visual Analysis and Effect Size Metrics, and Corresponding Kappa Coefficients for All 45 Studies

Metric	Magnitude of Effect	Visual Analysts' Judgments		Kappa Coefficient
		Small	Large	
PEM	Small	5	10	0.40
	Large	0	30	
PND	Small	5	8	0.47
	Large	0	32	
IRD	Small	5	4	0.67
	Large	0	36	
PAND	Small	5	6	0.56
	Large	0	34	
Phi	Small	5	6	0.56
	Large	0	34	
NAP	Small	5	7	0.51
	Large	0	33	
Tau_{novlap}	Small	5	8	0.47
	Large	0	32	

Note. PEM = percent of data points exceeding the median; PND = percent of nonoverlapping data; IRD = improvement rate difference; PAND = percent of all nonoverlapping data; NAP = nonoverlap of all pairs; Tau_{novlap} = Kendall's tau nonoverlap.