This is an Accepted Manuscript of an article published by Taylor & Francis Group in Educational Psychology in Practice on 18/11/2014, available online: http://www.tandfonline.com/ 0.1080/02667363.2014.975785.

Please use the following citation when referencing this work:

McGill, R. J., Baker, D., & Busse, R. T. (2015). Social Story™ interventions for decreasing challenging behaviours: A single-case meta-analysis 1995-2012. *Educational Psychology in Practice: Theory, Research, and Practice in Educational Psychology, 31,* 21-42. doi: 10.1080/02667363.2014.975785

Social Story™ Interventions for Decreasing Challenging Behaviors: A Single-Case Meta-Analysis 1995-2012

Ryan J. McGill

Texas Woman's University

Diana Baker

R. T. Busse

Chapman University

Abstract

A meta-analysis of the single-case research examining the efficacy of Social Story™ interventions for decreasing problem behaviors of children and adolescents with autism spectrum disorder was conducted by examining 27 outcome studies (n = 77) between 1995 and 2012 that yielded 64 intervention effects across three single-case outcome indicators. The overall mean visual analysis ratings and percentage of non-overlapping data scores indicated that the use of Social Story™ interventions resulted in small to negligible effects whereas the weighted effect size estimator ($\overline{ES}$ = .79) indicated moderate to large treatment effects. Moderator analysis indicated that intervention setting, intervention agent, length of treatment, and publication status were all associated with positive effects for behavioral outcomes, although the significance of these outcomes were not consistent across indicators. Implications for practitioners and clinicians, suggestions for future research, and limitations are discussed.

Keywords: meta-analysis, single-case design, autism spectrum disorder, social story

Social Story™ Interventions for Decreasing Challenging Behaviors: A Single-Case Meta-Analysis 1995-2012

As the referral and identification rates for autism spectrum disorder (ASD) among children and adolescents continues to rise there is a need to critically evaluate the efficacy of therapies that are intended to treat the social and behavioral symptoms that commonly present to professionals in clinical settings. Although there is no empirically validated "cure" for ASD, many symptoms of the disorder can be effectively managed through a combination of psychoeducational intervention programs. According to Sansosti, Powell-Smith, and Kincaid (2004), the primary focus of autism interventions is to reduce problem behaviors as well as improve pro-social behaviors that generalize across settings. Some authors (e.g., Francis, 2007; Gentry, 2003) have suggested utilizing interventions that directly teach social skills and perspective taking to children with ASD. Social Stories are a flexible intervention strategy that successfully incorporates both of these components. In a recent meta-analysis of school-based social skills interventions for children with ASD, Chenier and colleagues (2012) found Social Stories to be one of the more popular intervention techniques cited within the empirical literature.

Social Stories™

A Social Story™ (hereafter referred to as Social Story or Social Stories) is an intervention approach that combines visual structure and textual sequencing to teach social skills and reduce problem behavior. The approach has grown in popularity within the last two decades and is commonly utilized by professionals to facilitate the development of pro-social skills within school-based settings (Kuoch & Mirenda, 2003). According to Gray (2010), a social story can be utilized to assist individuals with ASD to interpret and navigate a challenging or

confusing social situation that often results in maladaptive behaviors. The theoretical rationale for Social Stories interventions is based upon the cognitive theory of ASD which posits that social difficulties are the result of a child's inability to understand the social behavior of others, a concept known as *theory of mind* (Baron-Cohen, 1995).

Social Stories are composed of various combinations of sentence forms. *Descriptive sentences* describe the salient features of a specific social situation that distinguish it from others; *directive sentences* specify desired behavioral responses from the individual within that social context; and *perspective sentences* describe the feelings of the individual and others within the target social situation. Sentences can be accompanied by pictures, which serve as an additional visual cue to the individual. Social Stories are often presented to an individual in advance of a target situation which has elicited problem behaviors in the past. Stories can be read to the individual by an adult or peer, read directly by the student, or recorded verbally and played for individuals who do not have the ability to read (Gray & Garand, 1993). Gray (2010) outlined several basic steps for developing Social Story interventions. First, clinicians must target a problematic social situation that is to be the focus of the intervention. Next, the contextual features (e.g., when, where, duration, antecedent, consequence) of the target situation must be defined. Information in the second step is primarily obtained through direct observations, data collection, and interviews with caregivers and teachers. Finally, an individualized Social Story is developed to emphasize a specific skill and the behavioral steps necessary to successfully negotiate the target situation.

Review of Previous Research Synthesis

The flexibility and ease of Social Stories makes for an attractive intervention option for clinicians seeking to improve social outcomes for children and adolescents with ASD. Though

Social Story interventions are often viewed by stakeholders as feasible and effective (Smith, 2001), the empirical research findings for them have been mixed. In a 2005 review of the efficacy of ASD interventions Olley indicated that the Social Story research base was largely anecdotal, mainly consisting of descriptive evaluations of intervention effectiveness. Although intuitively appealing, anecdotal evidence is insufficient for establishing the efficacy of applied intervention technologies (Kratochwill & Shernoff, 2004).

The Social Story research base has greatly expanded over the last decade and now includes several research syntheses. A descriptive review of the effectiveness of Social Story interventions for children with ASD was conducted by Sansosti and colleagues in 2004. Using rudimentary search parameters the authors were able to locate 10 studies that met inclusion guidelines. Though evidence of positive intervention effects was found in several of the studies, the authors determined that the evidence-base for the effectiveness of Social Stories was limited. Additionally, the use of Social Stories as a standalone behavioral intervention for children and adolescents with ASD was not advised. Similar results have been obtained in other research reviews utilizing similar analytic methods (e.g., Ali & Frederickson, 2006; Nichols, Hupp, Jewell, & Ziegler, 2005; Rust & Smith, 2006).

Reynhout and Carter (2006) later provided the first quantitative review to date, a comprehensive meta-analysis that included a total of 16 published and unpublished studies. The synthesis differed from previous reviews by including studies comprised of individuals diagnosed with additional developmental disabilities other than ASD, and included studies that utilized between-group experimental designs. In addition to a descriptive analysis, percentage of non-overlapping data (PND) was calculated for studies that utilized single-case design, and standardized means difference effect sizes were calculated for the group design studies. A mean

effect size of .99 was found for the group designs, indicating that Social Stories were largely effective at increasing target social skills at the group level. Conversely, a mean PND of 43% was obtained from a synthesis of 12 single-case studies indicating that the efficacy of Social Stories was less robust at the individual case level. Although the authors speculated that overlap between the baseline and treatment phases may have attenuated the resulting PND statistics, subsequent removal of problematic data points did not have an appreciable impact on the recalculated PND value. Due to these inconsistent results, the authors deferred clinical recommendations in lieu of advocating for more rigorous scientific approaches for evaluating the Social Story research base.

In an effort to remediate some of the deficiencies of previous reviews, an updated meta-analysis was completed by Kokina and Kern in 2010. The study improved on previous reviews through utilizing a more systematic search process and inclusion criteria, and evaluating the potential effects of various intervention components (e.g., use of functional analysis procedures, assessment of social validity). Additionally, PND was utilized as a common estimate of effects across the studies which allowed for a more consistent evaluation. The authors evaluated a total of 18 studies that were published between 2002 and 2009. Across the studies, a total median PND score of 62% was obtained. According to the guidelines for interpreting the PND statistic that have been proposed by Scruggs and Mastropieri (1998), outcomes indicated questionable intervention effectiveness.

More recently Test, Richter, Knight, and Spooner (2011) conducted a review of single-case Social Story research. This synthesis differed significantly from previous reviews by evaluating each of the included studies according to recently published technical guidelines for single-case research (Horner, Carr, Halle, McGee, Odom, & Wolery, 2005). Although 28 studies

were reviewed, a meta-analysis (using the PND statistic) was completed on only 18 of the included studies. Separate PNDs were calculated for intervention, maintenance, and generalization phases. Athough the authors made no attempt to synthesize the data, they provided combined PND statistics for 8 of the 18 cases evaluated. From those cases, a mean PND of 50% was found, indicating negligible intervention effects.

Previous research on the efficacy of Social Stories has yielded several common themes. Most authors have argued that despite the promise of Social Stories, the poor results that have been obtained from empirical examinations of its efficacy prevent it from being considered an evidence-based intervention. Despite more positive portrayals in descriptive reviews, quantitative evidence for the effectiveness of Social Stories has been limited. Nevertheless, more recent quantitative reviews indicate that the effectiveness of various Social Story interventions is variable and contingent upon a number of individual design factors. That is, the intervention may be highly effective within specific clinical settings but not in others. Additional investigation is needed to determine the conditions that are responsible for this variability.

Limitations of Previous Studies

Although previous research syntheses have highlighted important findings in regard to the effectiveness of Social Stories within clinical settings, a more specific examination of the Social Stories literature is needed. Although originally developed to synthesize between-group research designs, meta-analytic procedures for combining the results of single-case studies have been established in the empirical literature for over 20 years (Du Paul, Eckert, & Vilardo, 2012; Kratochwill & Levin, 1992). Of the studies that were reviewed above, only one can be considered a conventional meta-analysis (Kokina & Kern, 2010).

Despite the development of numerous single-case effect sizes, researchers have only assessed single-case Social Story effects using the PND statistic. This evaluative approach with single-case data is problematic due to the fact that different results have been obtained in studies analyzing the same dataset with multiple single-case indicators. In a simulation study conducted by Parker, Vannest, and Davis (2011), significant discrepancies were observed between several quantitative indicators, which indicate that the choice of a particular effect size estimate by a practitioner may have a substantial impact on the outcome of a treatment evaluation. Furthermore, many studies fail to provide adequate data for independently calculating a PND statistic. This is common in fields that traditionally eschew quantitative data analysis such as applied behavior analysis where most single-case interventions are evaluated using subjective evaluative methods (e.g., visual inspection). The incorporation of additional single-case indicators is necessary for the quantitative synthesis of studies that have been excluded from previous reviews. To date, a meta-analysis of the single-case Social Story literature using these procedures has yet to be completed.

Previous quantitative reviews have also been confounded by the inclusion of individuals with different primary clinical diagnoses despite the fact that Social Stories are utilized predominately to treat social skills deficits in children with ASD (Gray, 2010). The last comprehensive review of the effectiveness of Social Stories that included only individuals diagnosed with ASD as their primary clinical diagnosis was completed by Rust and Smith in 2006. Many relevant intervention studies have been published since that time.

Most important, previous reviews have not comprehensively investigated (e.g., assessing for potential moderator variables) the effects of Social Stories on decreasing problem behaviors. According to Cooper, Heron, and Heward (2007), the goals of behavioral intervention are to

simultaneously decrease maladaptive behaviors as well as elicit appropriate replacement behaviors. Although there are some data to indicate that Social Stories may be effective at promoting social skills in children and adolescents with ASD, a more thorough investigation is needed to determine whether it is an effective treatment for reducing problem behaviors.

Finally, the single-case research design quality utilized in Social Story research has been a persistent concern raised in previous studies. Despite these concerns, only one review to date (Test et al., 2011) has attempted to evaluate the quality of studies utilizing conventional single-case design quality indicators (e.g., Horner et al., 2005). Although consensus guidelines have subsequently been adopted for evaluating the quality of single-case research by a panel of national experts (Kratochwill et al., 2012), these standards have yet to be utilized to evaluate the quality of more recently published studies.

Purpose of Current Meta-Analysis

The purpose of the present meta-analysis is to provide a quantitative review of Social Story interventions for students diagnosed with ASD from research conducted between 1995 and 2012. The primary objective is to report single-case effect sizes for outcomes across several indices. A secondary objective is to assess the potential effects of moderating variables on intervention outcomes. As a result of previous research, it was hypothesized that small to moderate effects in reducing problem behaviors would be demonstrated. Additionally, it was hypothesized that discrepancies in treatment outcome would not be invariant across the single-case indicators.

Method

Search Procedures and Study Selection

Multiple techniques were utilized to locate potential studies for inclusion in the meta-analysis. First, the ERIC, PsychARTICLES, and PsychInfo databases were searched for published articles on November 9th and 10th of 2012. The search for the articles published in the English language was not limited by any specified date range. The search was conducted using the following terms, with the results listed in parentheses: social stories (254); intervention (192); behavior (154); autism (133). Articles that did not utilize decreasing problem behaviors as one of their outcome criteria were eliminated, leaving a total of 23 articles. Second, an ancestral search on the reference lists of identified studies was conducted to identify additional studies for potential inclusion, yielding 4 additional studies. Third, a serial search was conducted on the table of contents for the following peer-reviewed journals: *Autism*, *Autism: The International Journal of Research and* Practice, *Focus on Autism and Other Developmental Disabilities*, *Intervention in School and Clinic*, Journal *of Applied Behavior Analysis*, *Journal of Autism and Developmental Disorders*, and *Journal of Positive Behavioral Interventions*. Finally, a subsequent search was conducted on the ProQuest Dissertations and Theses database on November 12, 2012 using variations of the same keyword search terms described above. As a result of that screening 11 dissertations and theses were identified for potential inclusion based upon title screening. This resulted in a total of 38 studies, spanning a 17 year period (1995-2012) that were available for review.

Criteria for Inclusion

At a general level, studies had to examine the effects of Social Story interventions on decreasing target behaviors for children and adolescents with autism in home and school settings

to be included in the meta-analysis. Additionally, seven specific selection criteria were utilized to further screen potential studies for inclusion:

1. Participants were described as school-aged children in preschool through grade 12 and diagnosed with autism. Participants who were described as having comorbid disorders were included although the primary diagnosis had to be clearly identified as ASD.
2. The study utilized a validated Social Story methodology as a primary component in the intervention package. Studies that utilized multiple-component intervention packages were included and coded accordingly for potential moderator analysis.
3. The intervention program monitored specific problem behavior(s) that were targeted for decrease. Studies that simultaneously sought to increase pro-social behaviors were included however those effects were not incorporated into the meta-analysis.
4. The study utilized a single-case research design methodology with a minimum of one baseline and one treatment phase and a minimum of three data points or progress monitoring probes using the same outcome measure.
5. The study reported a single-case effect size or provided case data that could be used to calculate an effect size statistic (e.g., graphs for visual inspection).
6. Studies were readily classified into one of four target behavior categories that were defined as follows:
    a. *Physically aggressive behaviors*: An intervention that focused primarily on reducing behavior(s) threatening or causing physical harm toward others (e.g., kicking, biting, hitting).
    b. *Verbal behaviors*: An intervention that focused on reducing verbalizations and/or vocal sounds (e.g., yelling, screaming, crying).

c. *Stereotypic behaviors*: Ritualized, restrictive, and repetitive behaviors (e.g., rocking, hand flapping, lining up objects).

d. *Multiple behaviors*: Studies that simultaneously targeted multiple problem behaviors from the categories described above.

7. The study had to provide sufficient descriptive information to evaluate research design quality using the following standards adopted from the *What Works Clearing House* guide for evaluating single-case research (Kratochwill, et al., 2010):

a. Independent and dependent variables are operationally defined.

b. Treatment integrity is assessed.

c. Independent variable must be systematically manipulated, with the researcher determining when and how the independent variables change.

d. Each outcome variable must be measured systematically over time by more than one assessor, and the study needs to collect inter-assessor agreement in each phase and on at least 20% of the data points in each condition (i.e., baseline, intervention) and the inter-assessor agreement must meet minimum thresholds (e.g., 80%).

e. The study must include at least three attempts to demonstrate an intervention effect, at three points in time, or with three different phase repetitions.

f. Phases must have a minimum of three data points to qualify as an attempt to measure an effect.

g. Study must provide an estimate of intervention effect.

Each study identified from preliminary screening was reviewed by the first and second authors to determine if it met inclusion criteria. Of the reviewable studies found through the

primary search, 27 out of the 38 studies (71%) met the inclusion criteria and were included in the meta-analysis. Studies were excluded for one or more of the following reasons: (a) studies were literature reviews or other qualitative research synthesis (18%); (b) studies did not include participant samples that included individuals with a primary diagnosis of ASD (18%); (c) studies did not utilize single-case design as a research methodology (45%); and (d) studies did not contain case data where Social Story interventions were utilized to decrease target behaviors (18%). Of the 27 studies included in the meta-analysis, 77 cases were evaluated.

Coding and Reliability

The first and second authors independently coded inclusion criteria, study information, and effect size calculations. Both coders were doctoral students in school psychology and trained in coding procedures. Reliability and procedural fidelity was assessed by the lead author who had prior single-case effect size and meta-analytic research experience. Treatment integrity (Peterson, Homer, & Wonderlich, 1982) was coded as being assessed if studies reported integrity data for 25% or more of intervention sessions. Social validity (Wolf, 1978) was coded as assessed when studies explicitly sought to obtain information regarding the acceptability and/or perceived importance of the intervention by stakeholders. Reliability assessment was coded as assessed if the study reported a systematic attempt to assess the accuracy of dependent variable data collection for 25% or more of intervention sessions. Functional assessment was coded as assessed if the study described procedures for determining relevant descriptive-level data (e.g., antecedent, consequences, setting events, duration, latency, frequency) about the target behavior as part of intervention plan development.

For primary screening criteria and study/sample characteristic coding, a total of 7 studies were randomly selected to examine procedural reliability, which represented 25% of the total

studies (n = 27) reviewed. Coders agreed on 97% of the 16 study and sample categories in the reliability subsample. For the quantitative effect size computations, raters agreed on 96% of the evaluations with a total of 7 disagreements. All of the discrepancies were remediated individually through verbal conference prior to data analysis.

Outcome Indicators and Data Analyses

To compare and contrast our findings with conclusions from previous Social Story research syntheses, three single-case indicators were used to evaluate individual studies. The use of multiple indicators to evaluate single-case outcomes is consistent with best practice guidelines within the technical literature (Beretvas & Chung, 2008). Additionally, treatment evaluations based upon a triangulation of metrics safeguard against the previously noted threats to statistical conclusion validity that are introduced by overreliance on a single indicator or method bias. These evaluative techniques have been utilized to assess the effects of social skills interventions (Chenier et al., 2012) and are being utilized here to assess the single-case Social Story literature. Indicators were selected on the basis of their widespread use within the single-case research community as well as their versatility for use by practitioners and clinicians. Individual indicators and their computation are described in more detail below.

Visual Analysis Ratings. Visual inspection is a prominent assessment method utilized in applied behavior analysis and consists of subjective inspection of visually presented data to answer the question of whether intervention data indicate that an intervention was effective. For this review, Visual Analysis Rating (VAR; Busse & Kennedy; 2011) was adapted to evaluate Social Story treatment outcomes. Two researchers, who were trained in single-case visual inspection, reviewed studies that provided graphed data and rated study outcomes according to the following guidelines: *significant decrease in target behaviors from baseline* (+2); *moderate*

decrease in target behaviors from baseline (+1*); little to no decrease in target behaviors from baseline* (0); *moderate increase in target behaviors from baseline* (-1); *significant increases in target behaviors from baseline* (-2). It is important to note that the evaluators were trained to take into consideration important elements of visual inspection such as trend, mean shifts, variability in the data, and immediacy of effects. The evaluators assigned an overall VAR using the following guidelines: For reversal designs, separate VARs were assigned to each baseline-treatment interval and then averaged to calculate the overall VAR for the study and/or case; for multiple participant studies, the VARs for each case were averaged to calculate the overall VAR for the study. We averaged the overall study VARs from each researcher to create a grand VAR for each study. It should be noted that VAR is similar to the visual inspection techniques that were described by Chenier and colleagues (2012) in their evaluation of single-case social skills interventions for children with autism. The strengths of VAR are that it is a relatively simple method, and it may be useful with limited data points. The limitations are that it may be less robust at discriminating small effects and its reliability and validity have not been established within the technical literature.

PND. PND (Scruggs, Mastropieri, & Casto, 1987) is one of the most popular quantitative indicators found in single-case intervention studies (Bellini, Peters, Benner, & Hopf, 2007). Whereas there are several variations of overlap indices, the most parsimonious method is to divide the total number of treatment data points that do not overlap with the lowest baseline data point by the total number of treatment points, and then multiplying the product by 100. The resulting percentage is interpreted as an estimate of the magnitude of change in the dependent variable that is associated with the independent variable. PNDs greater than or equal to 80 are indicative of a strong effect, 60 to 79 is a moderate effect, and PNDs below 60 indicate

negligible effect. We calculated overall PND statistics for each individual case that met inclusion criteria providing appropriate data. For reversal designs, PNDs were calculated for each baseline-treatment interval and then averaged for each case. Case PNDs were averaged to provide an overall PND estimate for studies using multiple participant designs. The strengths of PND are that it is relatively simple to compute, and it is unaffected by nonlinearity and heterogeneity. The limitations are that it is potentially oversensitive to atypical baseline data, it is adversely affected by trends, and that it ignores all baseline data except for a single data point.

Mean Difference Effect Size. The "no-assumptions" effect size described by Busk and Serlin (1992) was used to evaluate mean phase differences for studies that provided appropriate data for its calculation. We calculated effect sizes by subtracting the baseline mean from the treatment mean and dividing by the standard deviation of the baseline data, with no assumptions made about the homogeneity of the phase variances, producing an un-weighted effect size similar to those calculated for between-subject design studies (Hedges & Olkin, 1985). Because the studies being evaluated were focused on decreasing problem behaviors, we converted negative effect size values to positive values. Effect sizes were calculated for each individual case that met inclusion criteria and provided appropriate data. For reversal designs, effect sizes were calculated for each baseline-treatment interval and then averaged for each case. For each un-weighted effect size estimate we calculated 95% confidence intervals using procedures described for standardized means difference effect sizes found in Grisson and Kim (2005), with baseline data points substituted for number of control group participants and treatment data points substituted for number of treatment group participants. The strengths of the no assumptions method are that it provides a means for quantifying single-case outcomes, and the effect size can be interpreted much like a $\underline{z}$ score. The limitations are that the effect size

estimates do not account for trend or autocorrelation, and the phase means may be affected by outliers. Cohen (1988) recommended that practitioners interpret effect sizes of .80 as "large," .50 as "moderate," and .20 and below as "small." Although these guidelines were developed for between-group analysis, they are utilized as conventional guidelines for interpreting single-case effect size estimates in the current study.

The effect size estimates for each of the included studies were inversely weighted according to modified procedures for parametric effect sizes outlined in Hedges and Olkin (1985). Weighted effect sizes were calculated by substituting the number of baseline data points for the number of participants in the control group and the number of treatment data points for the number of participants in the treatment group. Tests of homogeneity/heterogeneity (e.g., Q and I^2) were conducted to determine if a fixed-effects model could be assumed. In a fixed-effects model, differences between effect sizes from one study to another are attributed to variation that would be expected from random sampling error. The Q test follows a chi-square distribution and is computed by summing the squared deviations of each study effect size estimate from the overall effect size estimate. If a Q test is significant it indicates that the weighted estimator should not be interpreted as a single effect parameter. Conversely, a non-significant Q test indicates it is appropriate to interpret the observed value as an estimate of the common population effect size, despite between-study differences that may be observed. The I^2 statistic describes the percentage of total variation across studies that is due to heterogeneity rather than chance and is often presented as a supplemental descriptive statistic for Q. Higgins and Thompson (2002) suggested the following guidelines for interpretation: $I^2 = 25\%$ (small heterogeneity), $I^2 = 50\%$ (medium heterogeneity), and $I^2 = 75\%$ (large heterogeneity). Negative I^2 values are automatically set to zero.

Fail-Safe *N* Analyses

Originally developed as a measure to estimate the number of unpublished studies in the "file-drawer" that would be needed to bring the mean effect size down to a pre-determined criterion level, the results of fail-safe analyses is commonly interpreted as a quantitative estimate of *publication bias*. According to Rothstein, Sutton, and Borenstein (2005), publication bias occurs whenever published studies that are included in a systematic review are unrepresentative of the population of completed studies. Fail-safe *N* analyses using the formula for standardized mean difference effect sizes described by Orwin (1983) were used to estimate the relative stability of the single-case effect size findings.

Results

Characteristics of the Studies

Characteristics of the studies included in the meta-analysis are provided in Table 1. Most of the studies that were included were published from 2006 to 2012 (52%) and were published in peer-refereed journals (70%). The majority of Social Story interventions were provided in written format (62%), utilized advanced forms of single-case design (e.g., reversal designs, multiple baseline designs), and lasted from 0 to 3 weeks in duration. In contrast to the findings from previous reviews (e.g., Sansosti, Powell-Smith, & Kincaid, 2004), only three studies from the current meta-analysis utilized simple AB designs. The majority of studies reported procedures for obtaining reliability data (85%) and conducted treatment integrity (52%) and social validity (59%) assessments. Follow-up assessment was conducted in 63% of the studies, and less than a quarter of the studies included in the meta-analysis (11%) reported the use of functional assessment procedures as a part of intervention development for individual

participants. Over three-quarters of the studies (78%) reported using Gray's criteria in developing individual Social Stories.

Characteristics of Participants and Setting

Characteristics of the sample and settings can be found in Table 2. A total of 77 participants were included in the study from a broader sample of 108 individuals from the 27 included studies. The majority of the studies (59%) were comprised of 2 to 5 participants. Most of the studies (63%) were comprised of all male samples. Individual participants ranged in age from 2 to 15 years old. Almost three-quarters of the studies (74%) took place in a school-based setting and over half (59%) were targeted toward elementary aged children (i.e., 6-12 years of age). The agent responsible for delivering the intervention to participants was relatively well distributed across the studies with almost over half of the studies (58%) reporting the intervention agent as either the child(ren)'s teacher or researcher.

Single-Case Design Technical Standards

Table 3 summarizes the quality standards for single-case design research according to Kratochwill et al. (2010) for the 27 studies that were evaluated for the current meta-analysis. Given the fact that previous literature reviews have called into question the quality of Social Story research design, it was of interest to determine whether the quality of research has improved over time. The Kratochwill et al. (2010) technical standards were adopted by the United States Department of Education Institute of Educational Sciences (IES) for evaluating the quality of single-case research that was submitted to the *What Works Clearing House*. The standards represent consensus guidelines that were adopted by an expert panel of single-case researchers that was convened by the IES in 2010.

For the purposes of this review, seven single-case quality indicators were extracted from the Kratochwill et al. (2010) document for evaluating the studies included in the meta-analysis. As a result of that assessment it was found that an average of 5.44 quality indicators (78%) was met per study. Although the independent variable was systematically manipulated by the researcher and inter-reliability data were provided in 85% of the studies, only 67% of the studies operationally defined relevant study variables. To assess design quality over time a comparison was made between studies published from 2005 and beyond and those published prior to 2005. The selection of 2004 as a reference marker was not arbitrary as it was the year in which the first review was published (Sansosti et al., 2004) that called into question the technical adequacy of Social Story research. Studies published from 1995 to 2004 (n = 11) met an average of 5 quality indicators, with only one of the studies (9% of the studies) meeting all seven design standards. In contrast, studies published from 2005 to 2012 met an average of 5.75 (n = 16) quality indicators, with eight studies (50% of the studies) meeting all seven of the design standards. Overall, the 15% increase in average indicators met as well as the nominal increase in number of studies meeting all of the standards across the periods provide some evidence of systematic improvements in single-case research quality over time.

VAR

The aggregate VAR ratings obtained from the 26 studies that provided adequate graphic displays is provided in Table 4. Across the studies, a grand VAR of .68 (range 0 to +2) was obtained. Using the rating guidelines suggested by Busse and Kennedy (2011), the grand VAR suggests minimal decreases in target behaviors from baseline across the studies. Independent ratings for each of the studies were all within one qualitative level of each other, providing some evidence of relative stability in ratings across evaluators. Baseline data variability was noted in

many of the studies which obfuscated potential intervention effects. Nevertheless, these findings are consistent with previous Social Story literature reviews. Interestingly, the restricted positive range in aggregate VAR values raises a question regarding potential publication bias, given that no studies were located in which problem behaviors increased after the provision of Social Stories.

PND

The PND values obtained from the 26 studies that provided adequate data for evaluating the presence of overlap can be found in Table 4. The mean PND was 51% (range 0%-100%). Using the interpretive guidelines suggested by Scruggs and Mastropieri (1998), this mean PND value reflects negligible intervention effects. The wide range of PND scores indicates potential contamination due to floor and ceiling effects (Scruggs, Mastropieri, & Casto, 1987). In reviewing the PND scores in Table 4, seven studies (26%) were flagged as potential outliers (as defined by PND values within 5% or less of the extreme ends of the score range). When confounded by such disparate data, it is possible that the mean PND value may not provide an accurate measure of treatment effectiveness.

Mean Difference Effect Size

There were 12 studies that provided sufficient data for calculating single-case effect sizes. Un-weighted effect sizes and confidence intervals for individual studies can be found in Table 4. The grand un-weighted mean for the studies was 1.67, which corresponds to an average decrease in problem behavior of over one and a half standard deviations. However, un-weighted aggregates treat all component studies equally, which is a significant limitation due to the fact that some studies provided as few as 20 total data points whereas others were composed of as many as 288. To take into account this sample variability and provide a more unbiased estimate

of treatment effects, a fixed-effects model was utilized to weight each individual study effect size prior to obtaining a weighted grand effect size estimator. The weighted estimator for the studies was significantly less than the un-weighted value ($\overline{ES}$ = .79, 95% CI [.67, .91]); using Cohen's (1988) guidelines for interpretation, the weighted estimator provides evidence of moderate to large treatment effects. Means difference effect sizes were significantly heterogeneous (Q_{TOTAL} [11] = 203.76, $p < .001$; $I^2 = 95\%$); thus additional moderator analyses were conducted.

Moderator Analysis

The purpose of the moderator analysis was to identify study characteristics or grouping variables which may have been responsible for significant outcome variability in the meta-analysis sample. In the current study, significant variability was observed across several outcome indicators. Due to the fact that advanced meta-analytic procedures have yet to be developed for non-parametric single-case indicators such as VAR and PND, moderator analysis for those particular indicators was limited to calculating within-class mean values for each of the grouping variables. We used fixed-effects procedures for analysis of heterogeneous distributions described by Lipsey and Wilson (2001) to analyze the mean difference effect sizes. Summary statistics for the effects of intervention setting, intervention agent, target behaviors, use of functional assessment procedures, length of intervention, and publication type are provided in Table 5. Although of interest, assessment of additional moderator variables was not possible due to the fact that there were not enough studies to adequately represent the categories for data analysis.

A statistically significant effect was not found for the use of functional assessment procedures ($Q_{BETWEEN}$ [1] = 1.96, $p = .16$). A statistically significant effect for intervention setting was obtained ($Q_{BETWEEN}$ [1] = 48.11, $p < .001$), with school-based interventions ($k = 20$, $\overline{ES}$ = 1.46) producing larger treatment effects than home-based settings ($k = 5$, $\overline{ES}$ = .53). No

effect sizes were available for studies that provided interventions in multiple settings; this category was not included in the moderator analysis. Aggregate VAR and PND values were not significant for any of the setting categories. A significant effect was also found for intervention agent ($Q_{BETWEEN}$ [2] = 91.25, $p < .001$), with Social Story treatments delivered by researchers (k = 8, $\overline{ES}$ = 1.99) providing larger effects than those delivered by teachers (k = 8, $\overline{ES}$ = .82) and parents (k = 3, $\overline{ES}$ = .44). Single-case effect sizes were only able to be calculated for one study with an intervention delivered by a classroom aide (ES = .60) and one study with an intervention that included multiple agents (ES = 1.44); neither was included in the moderator analysis. Nevertheless, teacher and classroom aide interventions were significant across multiple indicators. There was a statistically significant difference across target behaviors ($Q_{BETWEEN}$ [1] = 76.72, $p < .001$), with interventions targeting verbal behaviors (k = 5, $\overline{ES}$ = 2.03) providing larger effects than those targeting multiple behaviors (k = 15, $\overline{ES}$ = .99). An effect size was only able to be calculated for one study that targeted physical behaviors (ES = .20) and no effect sizes were available for studies that targeted stereotypic behaviors; these categories were not included in the moderator analysis. The effects of interventions that targeted verbal behaviors were significant across multiple outcome indicators. A statistically significant effect was also found for intervention length ($Q_{BETWEEN}$ [1] = 5.66, p = .02), with interventions of three weeks or less producing larger treatment effects (k = 18, $\overline{ES}$ = .86) than those that were longer (k = 9, $\overline{ES}$ = .51). However, it should be noted that these effects were not significant across any of the other indicators. Finally, a statistically significant effect for publication status ($Q_{BETWEEN}$ [1] = 26.48, $p < .001$) was obtained with published studies (k = 19, $\overline{ES}$ = 1.12) reporting larger treatment outcomes than those obtained from dissertations (k = 8, $\overline{ES}$ = .50). Again, VAR and PND results were not significant for this grouping variable.

Fail-Safe *N* Analyses

Fail-safe *N* analysis indicated that 35 additional studies with an average effect size of 0 would have to be located to reduce the weighted estimator to .20. The results of this analysis indicate that a relatively large number of previously unidentified studies would have to be located to have a substantial impact on the current meta-analysis results.

Discussion

The results of this meta-analysis indicate that Social Story interventions for children and adolescents diagnosed with ASD provide small to large effects for decreasing problem behaviors, with results varying across intervention setting, intervention agent, target behavior, length of treatment, and publication type. Specifically, large effects were found for interventions delivered within school-based settings ($\overline{ES}$ = 1.46), interventions delivered by researchers ($\overline{ES}$ = 1.99), and interventions that targeted verbal behaviors ($\overline{ES}$ = 2.03). However, the significance of these effects was not consistent across single-case outcome indicators. Across multiple indicators, moderate to large effects were found for interventions delivered by teachers, interventions delivered by classroom aides, and interventions targeting verbal behaviors.

Not surprisingly, non-parametric indicators (e.g., VAR, PND) were much more conservative than the effect sizes in evaluating intervention outcomes. VAR ratings were only moderately significant for one of the moderator categories (interventions delivered by classroom aide) and PND values were only moderately significant for three intervention categories (interventions delivered by teachers, interventions delivered by classroom aides, and interventions targeting verbal behaviors). Conversely, weighted effect size estimators met conventional guidelines for moderate to large effects for 67% of the moderator categories.

According to Burns and Wagner (2008), it is not uncommon for single-case effect sizes to be inflated to the point of calling into question the use of conventional interpretive guidelines (e.g., Cohen, 1988). However, the weighted estimator that was obtained in this meta-analysis was significantly lower than those that have been obtained in other single-case intervention reviews (e.g, Burns & Wagner, 2008; DuPaul, DuPaul, Eckert, & Vilardo, 2012) and fell within the bounds of conventional between-subjects interpretive guidelines. Nevertheless, it should be noted that treatment outcomes were not consistent across outcome indicators. Whereas VAR and PND indicated small to negligible treatment effects overall, means difference effect sizes indicated moderate to large effects. Unfortunately there is not a mechanism for reconciling variability across multiple single-case indicators. Given the congruence between the VAR and PND results with those that have been obtained in previous meta-analytic reviews, we believe that the positive single-case effect size results should be interpreted with caution. These results are consistent with previous single-case evaluation research (e.g., Brossart, Parker, Olson, & Mahadevan, 2006) and illustrate the critical importance of indicator selection when evaluating single-case intervention designs.

Implications for Practice

Given the small to negligible effects on behavioral functioning that was observed across settings on several of the indicators, there is insufficient evidence for justifying the use of Social Story treatments as a primary intervention for decreasing problem behaviors for children with autism. Although we stipulate that Social Stories were developed primarily as a positive behavior intervention technology, it is important to keep in mind that the goals for practitioners when selecting and implementing interventions in school and other clinical settings is to simultaneously promote adaptive replacement behaviors while reducing maladaptive behaviors

(Cooper, Heron, & Heward, 2007). Given the current findings, we encourage practitioners to utilize other therapies with more evidentiary support (e.g., applied behavior analysis, social skills training) when addressing challenging autistic behaviors. Despite the popularity of Social Stories with many clinical professionals (e.g., special education teachers, speech and language pathologists), there is insufficient evidence to support its use as an evidence-based intervention.

Implications for Research

Although the Social Story intervention literature has grown over the last two decades there are still important gaps that need to be addressed. First, given the fact that intervention programs for individuals with severe problem behaviors often involve multiple components (Kern, Benson, & Clemens, 2010), more research is needed to examine the efficacy of Social Stories as part of a comprehensive treatment package for addressing the social and behavioral needs of individuals with autism. Second, although our survey of single-case quality indicators indicates that the technical quality of single-case Social Story intervention research has improved significantly over the last five years, many studies failed to incorporate functional assessment procedures (89%) and follow-up assessment (63%). Although a statistically significant effect was not found for the use of functional assessment in the moderator analysis, significant differences in effects across indicators were noted between interventions that utilized functional assessment procedures ($\overline{ES}$ = 1.71, VAR = .92) and those that did not ($\overline{ES}$ = .73, VAR = .65). Because only 3 single-case studies could be located that utilized functional assessment procedures, it may behoove future investigators to more critically examine the potential impact of such procedures on the efficacy of Social Stories. Finally, although the current meta-analysis as well as the review by Test and colleagues (2011) evaluated the technical quality of the single-case studies that were reviewed, there has yet to be an examination of the potential role that

study quality may have in relation to intervention outcomes. Given the persistent advice within the Social Story literature for more rigorous research designs, an examination of the role that study design may have on outcomes is needed.

Limitations

Although these data have direct implications for practice and future research, some limitations should be considered. First, this meta-analysis was limited to studies that utilized single-case design. Although most Social Story intervention studies utilize single-case methods to evaluate intervention efficacy, traditional between-subjects Social Story research has been published. Although the results of this meta-analysis were consistent with previous reviews of the literature which have synthesized between-subjects and single-case studies, the results of this review are limited to only providing an update to the literature with respect to the latter. Second, to utilize fixed-effects meta-analytic procedures to evaluate the single-case effect sizes it was necessary to violate some of the assumptions of parametric statistical analysis. Whereas we acknowledge that the utilization of parametric statistical techniques with single-case data is controversial, the impact of these practices is yet to be clearly understood (Huitema, 2011). As was previously discussed, the problem with the "no assumptions" approach has more to do with how to interpret the results than it has to do with the use of the method (Burns & Wagner, 2008). It is not uncommon for investigators to find effect size values that far exceed conventional thresholds for large effects when working with single-case data. Although the effect size results in the current study were more optimistic than more conservative estimators, the use of weighting procedures helped to temper the effects of some outlier estimates. Nevertheless, we believe that such cautions are worth noting and are buoyed by the recent development of a

variety of powerful non-parametric indicators from which practitioners and investigators can now select if they remain opposed to the use of parametric indicators with single-case data.

Conclusions

This meta-analysis examined 27 single-case intervention studies that yielded 64 individual outcomes across three single-case effect size indicators. The results indicated that Social Story interventions were associated with small to large decreases in target problem behaviors of children and adolescents with ASD. Consistent with previous studies, some evidence was found for larger effects when taking into account individual study variables. Nevertheless, the current meta-analysis indicates that the effectiveness of Social Story interventions is limited and that they should be used cautiously as a therapy for treating problem behaviors in children and adolescents with ASD.

References

References marked with an asterisk indicate studies included in the meta-analyses.

*Aboulafia, Y. (2012). *Social stories: efficacy and generalization for young children with autism spectrum disorders* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3493570).

*Adams, L., Gouvousis, A., VanLue, M., & Waldron, C. (2004). Social story intervention: improving communication skills in a child with an autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities, 19,* 87-94. doi: 10.1177/10883576040190020301

*Agosta, E., Graetz, J. E., Mastropieri, M. A., & Scruggs, T. E. (2004). Teacher-researcher partnerships to improve social behavior through Social Stories. *Intervention in School and Clinic, 39,* 276-287. doi: 10.1177/10534512040390050401

Ali, S. & Frederickson, N. (2006). Investigating the evidence base of social stories. *Educational Psychology in Practice, 22,* 355-377. doi: 10.1080/02667360600999500

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: Massachusetts Institute of Technology Press.

*Beh-Pajooh, A., Ahmadi, A., Shokoohi-Yekta, M., & Asgary, A. (2011). The effect of social stories on reduction of challenging behaviours in autistic children. *Procedia: Social and Behavioral Sciences, 15,* 351-355. doi: 10.1016/j.sbspro.2011.03.100

Bellini, S., Peters, J., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders. *Journal of Remedial and Special Education, 28,* 153-162. doi: 10.1177/07419325070280030401

Beretvas, S. N., & Chung, H. (2008). A review of single-subject design meta-analyses:

Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention, 2,* 129-141. doi: 10.1080/17489530802446302

Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30,* 531-563. doi: 10.1177/0145445503261167

*Brownell, M. D. (2002). Musically adapted social stories to modify behaviors in students with autism: Four case studies. *Journal of Music Therapy, 39*, 117-144. doi: 10.1093/jmt/39.2.117

Burns, M. K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37,* 126-136. Retrieved from http://www.nasponline.org

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill and J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187-212). Hillsdale, NJ: Erlbaum.

Busse, R. T., & Kennedy, K. (2011, April). *Methods for assessing response to intervention*. Workshop presented at the annual convention of the California School Psychologists Association, Costa Mesa, CA.

*Chan, J. M., & O' Reilly, M. F. (2008). A Social Stories™ intervention package for students with autism in inclusive classroom settings. *Journal of Applied Behavior Analysis, 41,* 405-409. doi: 10.1901/jaba.2008.41-405

Chenier, J., Hunter, K., Fischer, A. J., Patty, E., O'Leary, K., York, H., Robichaux, N., Liu, J., Dressler, T., Hartman, K., & Gresham, F. (2012, February). *Meta-analysis of social skills interventions using single subject design.* Paper presented at the meeting of the National

Association of School Psychologists, Philadelphia, PA.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

*Crozier, S., & Tincani, M. J. (2005). Using a modified social story to decrease disruptive behavior of a child with autism. *Focus on Autism and Other Developmental Disabilities, 20*, 150-157. doi: 10.1177/10883576050200030301

*Demiri, V. (2004). *Teaching social skills to children with autism using social stories: An empirical study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3132918)

DuPaul, G. J., Eckert, T. L., & Vilardo, B. (2012). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis 1996-2010. *School Psychology Review, 41,* 387-412. Retrieved from http://www.nasponline.org

Francis, K. (2007). Autism interventions: A critical update. *Developmental Medicine and Child Neurology, 47,* 493-499. doi: 10.1111/j.1469-8749.2005.tb01178.x

Gentry, J. A. (2003). Social skills interventions for students with autism. *Communique (31) 6,* 18-19. Retrieved from http:www.nasponline.org

*Gilles, A. A. (2008). *Treatment of sleep disturbances in children with autistic disorder: Utilization of behavioral intervention, social story, and picture activity schedule* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3346421)

*Graetz, J. E., Mastropieri, M. A., & Scruggs, T. E. (2009). Decreasing inappropriate behavior

for adolescents with autism spectrum disorders using modified social stories. *Education and Training in Developmental Disabilities, 44,* 91-104. Retrieved from http://daddcec.org

Gray, C. (2010). *The new Social Story™ book*. Arlington, TX: Future Horizons.

Gray, C. A., & Garand, J. D. (1993). Social stories: Improving responses of students with autism with accurate social information. *Focus on Autistic Behavior, 8*, 1-10. doi: 10.1177/108835769300800101

Grisson, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. New York: Psychology Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in meta-analysis. *Statistics in Medicine, 21,* 1539-1538. doi: 10.1002/sim.1186

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

Huitema, B. E. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies* (2nd ed.). Hoboken, NJ: John Wiley.

Kern, L., Benson, J. L., & Clemens, N. H. (2010). Strategies for working with severe challenging and violent behavior. In G. Gimple-Peacock, R. A. Ervin, E. J. Daly III, & K. W. Merrell (Eds.), *Practical handbook of school psychology: Effective practices for the 21st century* (pp. 459-474). New York: Guilford Press.

Kokina, A., & Kern, L. (2010). Social Story™ interventions for students with autism spectrum disorders: A meta-analysis. *Journal of Autism and Developmental Disorders, 40,* 812-826. doi: 10.1007/s10803-009-0931-0

Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Erlbaum.

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

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2012). Single-case intervention research design standards. *Remedial and Special Education, 34,* 26-38. doi: 10.1177/0741932512452794

Kratochwill, T. R., & Shernoff, E. S. (2004). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Review, 33,* 34-48. Retrieved from http://www.nasponline.org

*Kuttler, S., Myles, B. S., & Carlson, J. K. (1998). The use of social stories to reduce precursors to tantrum behavior in a student with autism. *Focus on Autism and Other Developmental Disabilities, 13*, 176-182. doi:10.1177/108835769801300306

*Kuoch, H., & Mirenda, P. (2003). Social story interventions for young children with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities, 18,* 219-227. doi: 10.1177/10883576030180040301

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Lorimer, P. A., Simpson, R. L., Myles, B. S., & Ganz, J. B. (2002). The use of social stories as a preventative behavioral intervention in a home setting with a child with autism. *Journal*

of Positive Behavior Interventions, 4, 53-60. doi: 10.1177/109830070200400109

*Mancil, G. R., Haydon, T., & Whitby, P. (2009). Differentiated effects of paper and computer-assisted Social Stories™ on inappropriate behavior in children with autism. *Focus on Autism and other Developmental Disabilities, 24,* 205-215. doi: 10.1177/1088357609347324

*Marr, D., Mika, H., Miraglia, J., Roerig, M., & Sinnotta, R. (2007). The effect of sensory stories on targeted behaviors in preschool children with autism. *Physical & Occupational Therapy in Pediatrics, 27(1),* 63-79. doi: 10.1080/J006v27n01_05

Nichols, S., Hupp, S. D. A., Jewell, J. D., & Zeigler, C. S. (2005). Review of Social Story interventions for children diagnosed with autism spectrum disorders. *Journal of Evidence-Based Practices for Schools, 6,* 90-120.

*Norris, C., & Dattilo, J. (1999). Evaluating effects of a social story intervention on a young girl with autism. *Focus on Autism and Other Developmental Disabilities, 14*, 180-186. doi: 10.1177/108835769901400307

Olley, J. G. (2005). Curriculum and classroom structure. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Vol. 2. Assessment, interventions, and policy* (3rd ed., pp. 863-881). Hoboken, NJ: John Wiley.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics, 8,* 157-159. doi: 10.3102/10769986008002157

*Ozdemir, S. (2008). The effectiveness of social stories on decreasing disruptive behaviors of children with autism: Three case studies. *Journal of Autism and Developmental Disorders, 38*, 1689-1696. doi: 10.1007/s10803-008-0551-0

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review

of nine nonoverlap techniques. *Behavior Modification, 35,* 303-322. doi: 10.1177/0145445511399147

*Pasiali, V. (2004). The use of prescriptive therapeutic songs in a home-based environment to promote social skills acquisition by children with autism: Three case studies. *Music Therapy Perspectives, 22,* 11-20. Retrieved from http://search.proquest.com.libpro xy.chapman.edu/docview/199502818?accountid=10051

*Perry, E. C. (2011). *A comparison of social stories with and without a reinforcement component* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3455478)

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15,* 477-492. doi: 10.1901/jaba.1982.15-477

*Quilty, K. M. (2007). Teaching paraprofessionals how to write and implement Social Stories for children with autism spectrum disorders. *Remedial and Special Education, 28,* 182-189. doi: 10.1177/07419325070280030701

Reynhout , G., & Carter, M. (2006). Social Stories™ for children with disabilities. *Journal of Autism and Developmental Disorders, 36,* 445-469. doi: 10.1007/s10803-006-0086-1

*Reynhout, G. & Carter, M. (2007). Social Story™ efficacy with a child with autism spectrum disorder and moderate intellectual disability. *Focus on Autism and other Developmental Disorders, 22,* 173-182. doi: 10.1177/10883576070220030401

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.). *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 1-7). New York: John Wiley.

Rust, J., & Smith, A. (2006). How should the effectiveness of Social Stories to modify the behaviour of children on the autistic spectrum be tested? *Autism: The International Journal of Research and Practice, 10,* 125-138. doi: 10.1177/1362361306062019

Sansosti, F. J., Powell-Smith, K. A., & Kincaid, D. (2004). A research synthesis of social story interventions for children with autism spectrum disorders. *Focus on Autism and other Developmental Disorders, 19,* 194-204. doi: 10.1177/10883576040190040101

*Scapinello, S. S. (2009). *Effectiveness of Social Stories™ for children with autism spectrum disorders* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 304918659)

*Scattone, D., Wilczynski, S. M., Edwards, R. P., & Rabian, B. (2002). Decreasing disruptive behaviors of children with autism using social stories. *Journal of Autism and Developmental Disorders, 32*, 535-543. doi: 10.1023/A:1021250813367

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22,* 221-242. doi: 10.1177/01454455980223001

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and special education, 8,* 24-33. doi: 10.1177/074193258700800206

Smith, C. (2001). Using social stories to enhance behaviour in children with autism spectrum difficulties. *Educational Psychology in Practice, 17,* 337-345. doi: 10.1080/02667360120096688

*Swaggert, B., Gagnon, E., Bock, S. J., Earles, T. L. Quinn, C., Myles, B., & Simpson, R. L. (1995). Using social stories to teach social and behavioral skills to children with autism. *Focus on Autistic Behavior, 10,* 1-16. doi: 10.1177/108835769501000101

Test, D. W., Richter, S., Knight, V., & Spooner, F. (2011). A comprehensive review and meta-analysis of the social stories literature. *Focus on Autism and other Developmental Disabilities, 26,* 49-62. doi: 10.1177/1088357609351573

*Watts, K. S. (2008). *The effectiveness of a social story intervention in decreasing disruptive behavior in Autistic children* (Doctoral dissertation, The Ohio State University). Retrieved from https://etd.ohiolink.edu

*Wheeler, K. L. (2005). *The power of social stories: A strategy for students with autism spectrum disorder* (Master's thesis). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 1429649)

Wolf, M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203-214. doi: 10.1901/jaba.1978.11-203

*Wright, L. A. (2007). *Utilizing social stories to reduce problem behavior and increase pro-social behavior in young children with autism* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3349076)

Table 1

Frequency and Percentage of Study Characteristics (n = 27)

Characteristics	Frequency	Percentage
Year of Publication		
1995-2000	3	11
2001-2005	10	37
2006-2012	14	52
Study Source		
Dissertation or thesis	8	30
Published study	19	70
Type of Social Story Intervention[a]		
Written	16	62
Pictorial	1	4
Audio/Musical	2	8
Multiple components	7	26
Study Design		
AB	3	11
ABAB	5	19
MB	11	40
Other	8	30
Intervention Length in Weeks		
0-3	18	67
4-6	8	30
>6	1	3
Follow-Up Assessment		
Yes	10	37
No	17	63
Reliability Assessment		
Yes	23	85
No	4	15
Treatment Integrity Assessment		
Yes	14	52
No	13	48
Social Validity Assessment		
Yes	16	59
No	11	41
Functional Assessment		
Yes	3	11
No	24	89
Gray's Criteria[b]		
Yes	21	78
No	6	22

Note. MB = Multiple baseline design.
[a]Percentage based upon 26 studies which reported Social Story intervention type.
[b]Gray (2010) has provided criteria for completing social stories. Evaluation was based upon whether researchers made specific reference to the criteria in the text during story development.

Table 2

Frequency and Percentage of Sample Participants and Setting Characteristics (n =77)

Characteristics	Frequency	Percentage
Number of Participants[a]		
1	7	26
2-5	16	59
>5	4	15
Gender		
All Male	17	63
All Female	9	33
Both Males and Females	1	4
Intervention Setting		
Home	5	19
School-based	20	74
Multiple	2	7
School-Level		
Pre-School	5	18
Elementary	16	59
Secondary	1	4
Combined	5	19
Intervention Agent		
Parent	3	11
Teacher	8	29
Classroom Aide	3	11
Researcher	8	29
Multiple	5	19

Note. Values may not sum to 100 due to rounding. [a]Some cases were excluded from individual studies in accordance with established inclusion procedures.

Table 3

Studies Meeting What Work's Clearing House Single-Case Design Quality Standards

	Quality Standards for Single-Case Design[1]							
Study	A	B	C	D	E	F	G	Met
Swaggert et al. (1995)	✓					✓	✓	3
Kutler, Myles, & Carlson (1998)	✓		✓	✓	✓	✓		5
Norris & Dattilo (1999)		✓	✓	✓		✓	✓	5
Brownell (2002)	✓		✓	✓	✓		✓	5
Lorimer, Simpson, Myles & Ganz (2002)			✓	✓	✓	✓		4
Scattone et al. (2002)	✓	✓	✓	✓	✓	✓	✓	7
Kuoch & Mirenda (2003)	✓	✓	✓	✓	✓	✓		6
Adams et al. (2004)			✓	✓	✓	✓	✓	5
Agosta et al. (2004)	✓					✓	✓	3
Demiri (2004)		✓	✓	✓	✓	✓	✓	6
Pasiali (2004)	✓		✓	✓	✓	✓	✓	6
Crozier & Tincani (2005)	✓	✓	✓	✓	✓	✓	✓	7
Wheeler (2005)						✓	✓	2
Marr et al. (2007)			✓	✓		✓	✓	4
Quilty (2007)	✓	✓	✓	✓	✓	✓	✓	7
Reynhout & Carter (2007)	✓			✓		✓	✓	4
Wright (2007)	✓	✓	✓	✓	✓	✓	✓	7
Chan & O' Reilly (2008)	✓	✓	✓	✓	✓	✓		6
Gilles (2008)			✓		✓	✓	✓	4
Ozdemir (2008)	✓	✓	✓	✓	✓	✓	✓	7
Watts (2008)	✓	✓	✓	✓	✓	✓	✓	7
Graetz, Mastropieri, & Scruggs (2009)	✓	✓	✓	✓	✓	✓	✓	7
Mancil, Haydon, & Whitby (2009)	✓	✓	✓	✓	✓	✓		6
Scapinello (2009)			✓	✓	✓	✓	✓	5
Beh-Pajooh et al. (2011)			✓	✓	✓	✓	✓	5
Perry (2011)	✓	✓	✓	✓	✓	✓	✓	7
Aboulafia (2012)	✓	✓	✓	✓	✓	✓	✓	7
Studies Meeting Standard (*n* = 27)	**18**	**14**	**23**	**23**	**21**	**26**	**22**	

Note. [1]Descriptions for standards can be found in methods section of the current review and were adopted from Kratochwill et al. (2010).

Table 4

Summary of Single-Case Social Story Study Outcomes for Decreasing Challenging Behaviors

Study	*N*[a]	Behavior	VAR	PND	ES	95% ES CI
Swaggert et al. (1995)	3	Physical	0	0%		
Kutler, Myles, & Carlson (1998)	1	Physical	1.75	95%		
Norris & Dattilo (1999)	1	Stereotypic	0	8%		
Brownell (2002)	4	Verbal	1.33	88%	4.33	2.72-5.94
Lorimer, Simpson, Myles & Ganz (2002)	1	Verbal	.75	43%	1.44	0.92-1.96
Scattone, Wilczynski, Edwards & Rabian (2002)	3	Multiple	1.00	83%		
Kuoch & Mirenda (2003)	3	Multiple	.75	44%		
Adams, Gouvousis, VanLue, & Waldron (2004)	1	Multiple	0	5%	0.34	0.07-0.61
Agosta, Graetz, Mastropieri, & Scruggs (2004)	1	Verbal	.50	80%	1.62	0.93-2.31
Demiri (2004)	4	Multiple	0	5%		
Pasiali (2004)	3	Multiple	.25	50%	0.93	0.46-1.40
Crozier & Tincani (2005)	1	Verbal	2.00	100%	2.45	1.31-3.59
Wheeler (2005)	2	Multiple	1.00	100%		
Marr et al. (2007)	5	Physical	0	37%		
Quilty (2007)	3	Multiple	0	5%	0.60	0.09-1.11
Reynhout & Carter (2007)	1	Stereotypic	0	40%		
Wright (2007)	4	Multiple	0	17%	0.57	0.53-0.61
Chan & O' Reilly (2008)	2	Multiple	1.50	88%	3.81	2.67-4.95
Gilles (2008)[b]	11	Physical			0.20	0.00-0.40
Ozdemir (2008)	3	Multiple	2.00	100%		
Watts (2008)	6	Verbal	0.17	10%		
Graetz, Mastropieri, & Scruggs (2009)	3	Multiple	1.00	75%		
Mancil, Haydon, & Whitby (2009)	3	Physical	1.00	64%		
Scapinello (2009)	15	Multiple	0.50	42%	1.89	1.42-2.36
Beh-Pajooh, et al. (2011)	3	Multiple	0.33	50%	1.83	1.32-2.34
Perry (2011)	4	Multiple	1.50	70%		
Aboulafia (2012)	7	Multiple	0.42	25%		

Note. VAR = visual analysis ratings; PND = percentage of non-overlapping data; ES = un-weighted single-case effect size.
[a]Some cases were excluded from individual studies in accordance with established inclusion procedures.
[b]Graphic displays did not have adequate features (e.g., phase lines) for estimating VAR and PND. Baseline and treatment descriptive statistics were provided.

Table 5

Summary Statistics for Effects of Moderator Variables for Intervention Outcomes

Category	*k*	VAR	PND	ES	95% ES CI *LL*	95% ES CI *UL*
Intervention Setting						
Home	5	.38	35%	.53*	.39	.67
School-Based	20	.75	56%	1.46**	1.24	1.68
Multiple	2	.59	35%			
Intervention Agent						
Parent	3	.25	24%	.44	.29	.59
Teacher	8	.84	65%*	.82**	.48	1.16
Classroom Aide	3	1.00*	68%*			
Researcher	8	.69	50%	1.99**	1.70	2.28
Multiple	5	.38	30%			
Behavior						
Physical	5	.69	49%			
Verbal	5	.95	64%*	2.03**	1.62	2.44
Stereotypic	2	0	24%			
Multiple	15	.68	51%	.99**	.83	1.15
Treatment Weeks						
0-3	18	.64	48%	.86**	.73	.99
>3	9	.75	56%	.51*	.25	.77
Functional Assessment						
Yes	3	.92	49%	1.71**	1.15	2.27
No	24	.65	51%	.73*	.61	.85
Publication Type						
Published	19	.75	56%	1.12**	.95	1.29
Dissertation	8	.51	44%	.50*	.34	.66

Note. *k* = number of studies. VAR = mean visual analysis rating; PND = mean percentage of non-overlapping data; ES = weighted effect size estimator. *LL* = lower limit, *UL* = upper limit, MB = multiple baseline design.
*moderate positive effect, **large positive effect