Education and Training in Autism and Developmental Disabilities, 2015, 50(2), 127–141

What is the Proof? A Methodological Review of Studies That Have Utilized Social Stories

Justin B. Leaf, Misty L. Oppenheim-Leaf, Ronald B. Leaf, Mitchell Taubman, John McEachin, Tracee Parker, Andrea B. Waks, and Toby Mountjoy
Autism Partnership Foundation

Abstract: Social stories are a commonly empirically evaluated and implemented procedure to increase pro-social behaviors and decrease aberrant behaviors for individuals diagnosed with an autism spectrum disorder. Despite their widespread use there have been questions raised to the soundness of the research methodology and the results which have been demonstrated within these research studies. This paper is a methodological review of 41 studies that evaluated social stories for individuals diagnosed with autism. We classified each study as one that utilized either a case study design, a reversal design, or a multiple baseline design. After classification we evaluated each study across multiple methodological dimensions and used this analysis to determine if a study showed either a clear demonstration, partial demonstration, or if there was no clear demonstration that the social story was responsible for behavior change. Results of this analysis indicated that the majority of studies either showed only a partial demonstration or no clear demonstration that the social story procedure was responsible for the behavior change. Based upon this analysis recommendations for clinicians and future researchers are discussed.

In 1993, Gray and Garand introduced social stories as a strategy that could be used to increase positive behaviors and reduce aberrant behaviors for individuals diagnosed with autism spectrum disorder (ASD). Social stories are a systematic form of intervention in which a teacher writes a brief text that describes a targeted behavior to be displayed by the student, when the student should display the desired behavior, why the student should display the targeted behavior, and how displaying the behavior will affect others in his or her environment.

In describing social stories, Gray and Garand specified several important guidelines. First, social stories should only be implemented with children and adolescents who fall in the "trainable mentally impaired range or higher who possess basic language skills" (Gray & Garand, 1993, p. 103). Second, each social story should be individualized. Third, social stories should only address one behavior at a time. Fourth, Gray and Garand suggested that, when reading a social story, the teacher and the participant should sit side by side. Fifth, social stories should use only positive language. Finally, they suggested specific types of sentences that should be used within a social story (Gray & Garand, 1993; Gray, 1994, 2004). Each social story should consist of at least four sentence types (i.e., descriptive, perspective, affirmative, and directive). Gray (2000, 2002, 2003) later added additional sentence types, which included: cooperative sentences, control sentences, and partial sentences. Gray (1995) later recommended that for every directive sentence there should be a total of two to four of the other sentence types.

Since the original article (i.e., Gray & Garand, 1993), there have been numerous articles published that have evaluated social stories on teaching social behavior (e.g., Adams, Gouvousis, VanLue, & Waldron, 2004; Barry & Burlew, 2004; Crozier & Tincani, 2007), teaching behaviors that are not inherently social (e.g., sitting) (e.g., Bledsoe, Myles, & Simpson, 2003; Hagiwara & Myles, 1999), and on decreasing aberrant behavior(s) (e,g., Agosta, Graetz, Mastropieria, & Scruggs, 2004 Chan & O'Reily, 2008; Crozier & Tincani, 2005;

Correspondence concerning this article should be addressed to Justin B. Leaf, Autism Partnership Foundation, 200 Marina Drive, Seal Beach, CA 90740. Email: Jblautpar@aol.com

Lorimer, Simpson, Myles, & Ganz, 2002; Kuttler, Myles, & Carlson, 1998). Additionally, there have been several curriculum books written on how to write and effectively implement social stories (Gray, 2000, 2002, 2003). Finally, several reviews have been published examining the effectiveness and efficiency of social stories (e.g., Kokina, & Kern, 2010; Sanosti, Powell-Smith, & Kincaid, 2004).

One of the first reviews on social stories was written by Sansosti et al, (2004), which evaluated studies that implemented social stories using single subject methodology. Sansosti and colleagues determined that the effects of social stories were limited in the articles they reviewed, due to possible confounding variables, weak methodological control, and weak treatment effects. Ali and Frederickson (2006) also acknowledged the limitations of the social story research, but stated that "the approach has promise and warrants further research" (Ali & Frederickson, 2006, p. 372).

In a more recent review, Kokina and Kern (2010) evaluated 18 different studies of social stories on a variety of dimensions (e.g., age, diagnosis, number of sessions, story format, use of comprehension questions). Similar to previous reviews, the authors concluded that research on social stories has questionable results in terms of effectiveness, and that the social story research has methodological flaws, which further limits the results.

Despite the lack of empirical evidence demonstrating efficacy, social stories are implemented with high frequency by teachers, parents and clinicians to individuals diagnosed with ASD (Reynhout & Carter, 2009). Reynhout and Carter (2009) surveyed 45 teachers regarding the perceived efficacy of social stories. The authors used a questionnaire with a 5-point Likert scale to examine various questions about social stories. The results from this survey revealed that 93.4% of the teachers surveyed considered social stories to be an effective intervention (48.9% agreed and 48.9% strongly agreed). Furthermore, the National Standards Report Project (The National Autism Center, 2009) for evidence based practices stated that story based intervention packages is an *established* procedure. Thus, it appears, that despite methodological flaws, social stories remains a widely used and endorsed procedure.

When evaluating research, it is imperative not only to look at the effects of the intervention but to evaluate if the researchers implemented the methodology appropriately. For example, if the researchers implemented the intervention condition at the wrong time (e.g., the data in baseline trending the desired way of intervention), it would not be clear if the change in behavior was due to the intervention or if the change in behavior was going to occur naturally. The purpose of this paper is to evaluate the correct use of single subject designs in research that utilized social stories for individuals diagnosed with autism. To do this, we evaluated research published in peer reviewed journals from the years 1993 (i.e., the year of the first publication on social stories) to 2012. This paper will specifically look at case study designs, reversal designs, and multiple baseline designs and evaluate the experimental rigor that was utilized. We will also utilize visual inspection to determine whether a clear behavior change was shown across the different studies. Based on the soundness of the research methodology and the extent of behavior change that was demonstrated, we will determine the documented effectiveness of the social story procedure for changing the participant(s) behavior(s). Finally, we will make recommendations to clinicians and researchers.

Search Procedure and Inclusion Criterion

Two researchers independently conducted multiple searches to identify studies that implemented social stories with individuals diagnosed with ASD. First, we conducted electronic searches on ERIC and PsycINFO using several combinations of keywords. Second, we conducted a visual search of the following journals: *Journal of Applied Behavior Analysis, Journal of Autism and Developmental Disabilities, Research in Autism Spectrum Disorders, Focus on Autism and Other Developmental Disabilities, and Autism.* We reviewed these journals as they commonly publish interventions for individuals diagnosed with autism. Third, we examined any references that were cited directly in the articles that were retrieved.

To be included in this review each research study had to meet four inclusion criteria. First, the article had to include at least one partici-

pant diagnosed with ASD. Second, the study had to utilize single subject methodology. Third, any comparative study was not included. Finally, all research studies had to be published or made available online in a peer reviewed journal from January 1993 to December 2012. After the two independent reviews, the two researchers identified a combined 41 studies that met this criterion. Next, the two researchers independently assigned each of the 41 studies to case-study design, reversal design, or multiple baseline design. The two researchers demonstrated 100% agreement regarding the categorical assignment of each research article.

Designs

Case Study Design

Nine studies were classified as a case study design. Case study designs are the most basic of all single subject designs. The most commonly implemented case study design is the AB case study design, which reports baseline data prior to intervention. There also exists a B (intervention) only case study design, but it provides no way of judging the impact of an intervention as no information is provided regarding baseline performance. The AB case study design begins with an initial baseline period (A) when the researcher measures the behavior of the participant prior to the intervention. Once the participant demonstrates consistent responding or responding trending in the opposite direction to that of the desired treatment effect, the researcher implements the treatment condition (B). Intervention is considered to be "successful" if the learner's behavior rapidly and significantly changes during the intervention condition.

Although the case study design is considered part of the single subject methodology, it is a weak design (Bailey & Burch, 2002). This is because, without replication either within or across participants, the case study design does not rule out many confounding variables (e.g., history, incidental occurrences, maturation, etc.) and it is difficult to determine if the intervention was truly responsible for the behavior change or if the behavior change was due to some other variable. Researchers can maximize the believability of this design in several ways. First, they can conduct a longer baseline period. Second, the baseline data must be stable or trending in the opposite direction of the desired treatment effect prior to intervention. Third, there should be little to no overlap between the baseline data and intervention data. Finally, the participant's behavior change should occur rapidly once the intervention condition begins (Bailey & Burch, 2002).

Reversal Designs

Thirteen studies were classified as utilizing a reversal design (e.g., ABA, ABAB design, ABABAC design). The reversal design is the most robust of all of the single subject designs. Essentially, the design demonstrates that the targeted behavior can be "turned on" or "turned off" at the researcher's will. The design typically starts with an initial baseline period (A) where the researcher measures the behavior of the participant prior to the intervention. Once the participant demonstrates consistent responding or responding trending in the opposite direction to that of the desired treatment effect, the researcher implements the intervention condition (B). Following implementation of the intervention, once the participant responds at a stable rate or displays the behavior at a rate trending in the desired direction, the researcher switches back to the baseline condition. The researcher continues to reverse the conditions to demonstrate functional control.

In order to show functional control the researcher must adhere to certain guidelines. First, in all baseline conditions the behavior must be stable or trending in the opposite direction of the desired treatment effect before switching conditions. Second, in all intervention conditions the behavior needs to be stable, trending in the desired direction, and at higher levels than during the baseline condition, prior to switching back to the baseline condition. Third, researchers should only implement this design with behaviors that can easily be reversed and for behaviors that will not cause injury to the participant or to others.

Multiple Baseline Designs

Nineteen of the studies were classified as utilizing a multiple baseline design or a variation

of the multiple baseline design (e.g., multiple probe design). The multiple baseline design is well suited when the target behavior is nonreversible or when it would be unethical to reverse the targeted behavior. A multiple baseline design is used when a researcher wants to measure an intervention across two or more participants, two or more behaviors for the same participant, or two or more settings (e.g., home and school). When the researcher implements this design they start by collecting baseline measures for each of the participants, behaviors, or settings. The researcher then implements intervention in a stepwise fashion across the multiple participants, behaviors, or settings. The researcher does not implement intervention for a new participant, behavior, or setting until there has been a significant change of behavior for the previous participant, for the previous behavior, or in the previous setting. Additionally, the researcher should not implement each intervention until the baseline behavior is stable or trending in the correct direction. This way, each baseline demonstrates that the intervention is responsible for the behavior change; more baselines (e.g., more participants) allows for the opportunity to replicate the treatment effects.

Measures across Research Designs

Measures for Case Study Design

We evaluated the nine studies that utilized the case study design across six main variables. First, we identified how many sessions were conducted in the baseline condition for each skill targeted in each of the studies. Second, we assessed whether the baseline data was stable or trending in the correct direction prior to switching to the intervention condition. The definition of stability or trending in the correct direction consisted of four components: (a) the participant displaying the same rate of behavior for the final two baseline sessions prior to switching to the intervention condition; (b) the participant displaying a rate of behavior that is trending in the opposite direction of the desired treatment effect for the final two baseline sessions prior to switching to the intervention condition; (c) the participant displaying the same rate of behavior or behavior at a rate trending in the opposite direction of the desired treatment effect for three out of the last four baseline sessions prior to switching to the intervention condition, without the last data point trending in the wrong direction prior to the switch; and (d) if two or more dependent variables were being measured, all measures had to meet the criterion stated above unless the researcher specifically stated that one of the dependent variables was the main dependent variable.

Third, we evaluated whether the treatment effect occurred immediately. This was defined as the behavior changing in the desired direction for two of the first three intervention sessions, or the third data point being higher (for increasing behaviors) or lower (for decreasing behaviors) than all data points in the baseline condition. Fourth, we compared the number of intervention data points that overlapped with the baseline data or that was trending in the wrong direction of the desired treatment effect. Fifth, we evaluated if the researchers used objective or subjective data. Finally, we evaluated whether the social story procedure was combined with any other procedure (e.g., video modeling, prompting, scripts) at any point within the study.

Measures for Reversal Design

We evaluated the 13 studies that utilized reversal design across five main variables. First, we calculated the percentage of all baseline conditions where the participant's behavior(s) were stable or trending in the opposite direction of the desired treatment effect. The same definition was utilized as for the case-study design. We calculated the number of baseline conditions where stability or trending in the correct direction was shown and divided by the total number of baseline conditions where an intervention condition directly followed the baseline condition.

Second, we calculated the percentage of all intervention conditions where the participant's behavior was stable or trending in the desired direction prior to reintroduction of a baseline condition. This definition consisted of three components: (a) the last two data points, prior to switching back to the baseline condition, was stable or trending in the direction of the desired treatment condition and higher than 85% of all of the baseline data

points or (b) three of the last four data points, prior to switching back to the baseline condition, was higher than 85% of all of the baseline data points and the last data point was not trending in the opposite direction of the desired treatment effect; and (c) if two or more dependent variables were being measure, all measures had to meet the criterion stated above unless the researcher specifically stated that one of the dependent variables was the main dependent variable. We calculated the number of intervention conditions where stability or trending in the correct direction was shown and divided by the total number of intervention conditions where a baseline condition directly followed the intervention.

Third, we utilized visual inspection to determine the percentage of intervention conditions where there was a clear behavior change from the baseline conditions. This definition consisted of three components: (a) for positive behaviors, 75% of all intervention data points demonstrated an increase from all baseline data points; or (b) for aberrant behaviors, 75% of all intervention data points demonstrated a decrease from baseline behaviors; or (c) a clear level change in the behavior could be observed through visual inspection. We calculated the number of intervention conditions where a change in behavior was demonstrated divided by the total number of intervention conditions. Fourth, we evaluated if the researchers used objective or subjective data. Finally, we evaluated whether the social story procedure was combined with any other procedure at any point within the study.

Measures for Multiple Baseline Designs

We evaluated the 19 studies that utilized multiple baseline designs across five main variables. First, we calculated the percentage of all baseline conditions where the participant's behavior(s) were stable or trending in the desired direction (same definition as stated above). We calculated the number of baseline conditions where stability or behavior trending in the correct direction was shown and divided by the total number of baseline conditions.

Second, we determined the percentage of all intervention conditions where the researcher correctly staggered intervention. The definition of correctly staggering intervention was that the researchers intervened upon a new skill, participant, or in a new setting only when the data from the previously intervened upon skill, participant, or setting was trending in the correct direction or stable and higher than 80% of all baseline points, without the previous two data points trending in the incorrect direction. We calculated the number of interventions that were correctly staggered divided by the total number of intervention conditions that were staggered.

Third, we evaluated if there was a clear change in behavior during the intervention condition utilizing the same definition that was utilized for reversal design studies. We calculated the number of intervention conditions where a change in behavior was demonstrated divided by the total number of intervention conditions. Fourth, we evaluated if the researchers used objective or subjective data. Finally, we evaluated whether the social story procedure was combined with any other procedure.

Level of Demonstration (Analysis of Measures)

We used the measures described above to create three levels of demonstration for studies in each research design category. This analysis was used to identify whether or not the researchers implemented the single subject methodology with enough evidence to convince the reader that the social story procedure was responsible for the behavior change. The three levels of demonstration were: 1) convincing demonstration; 2) partial demonstration; 3) no convincing evidence that the social story procedure was responsible for changing the targeted behavior. Across all studies, if the researcher(s) did not report a measure (e.g., not report baseline data) then that measure was considered as not occurring. In addition, a study had to meet all criteria of a demonstration level (see Table 1, 2, and 3) in order to be characterized as that level of demonstration. If a study met criteria from different demonstration levels, that study was categorized as the lowest level of demonstration for which it met all criteria. All criteria were considered necessary in order to demonstrate that the social story procedure was responsible for behavior change; thus, not meet-

TABLE 1

Measures and Demonstration Levels for Case Study Designs

Level of Demonstration	*Type of Data*	*Length of Baseline*	*Baseline Trending*	*Effect Immediate*	*Overlapping Data*	*Combined with other procedures*
Convincing Evidence	Objective	3 or more sessions of baseline	Stable or trending in correct direction	Behavior change demonstrated within 3 sessions	20-0% overlapping data between baseline and intervention	Not combined with other procedures
Partial Evidence	Objective	1 or 2 sessions of baseline	Stable or trending in correct direction	Behavior change demonstrated within 3 sessions	40-21% overlapping data between baseline and intervention	Combined with other procedures
No Convincing Evidence	Subjective	0 sessions of baseline or baseline not reported	No stability or not trending in the correct direction	Behavior change occurring after 3 sessions	100 to 41% overlapping data between baseline and intervention	Combined with other procedures

ing any single criterion was considered a flaw of the study's demonstration that the social story procedure was responsible for the behavior change. Table 1 displays the different scoring criterion and levels of demonstration for studies where a case study design was utilized. Table 2 displays the different scoring criterion and levels of demonstration for studies where a reversal design was utilized. Table 3 displays the different scoring criterion and levels of demonstration for studies where a multiple baseline design was utilized.

Results

Table 4 provides the overall results on the level of demonstration for studies across the three designs evaluated. Table 5, 6, and 7 provide information about how each of the studies scored across the measures assessed and the level of demonstration each research study was classified as. Within each table the rationale for why a study was classified at a given level is highlighted within a cell (category). If any study does not have any highlights in any cell it means that study was determined to have a convincing level of demonstration.

Levels of Demonstration

Convincing Demonstration. Out of the 41 studies reviewed, only 3 (7.3%) achieved a convincing level of demonstration that the social story was responsible for the participant changing his or her behavior; all of these studies utilized a multiple baseline design. Thus, 3 out of the 19 studies (15.7%) that utilized a multiple baseline design were classified as a convincing level of demonstration.

Delano and colleagues (2006) were the first researchers that we identified to demonstrate a convincing demonstration that a social story procedure was effective in changing the behavior for participants diagnosed with autism. In this study, the authors used a multiple probe design across three participants to show that social stories could be used to teach appropriate social engagement while decreasing inappropriate social engagement or no social engagement.

In 2008, Chan and O'Reily published the second study that showed a clear demonstra-

TABLE 2

Measures and Demonstration Levels for Reversal Designs

Level of Demonstration	*Type of Data*	*Baseline conditions stable or trending correctly prior to intervention*	*Intervention conditions stable and higher than baseline or trending the correct way prior to reversal*	*Clear behavior change during intervention conditions*	*Combined with other procedures*
Convincing Evidence	Objective	100% of all baseline conditions	100% of all intervention conditions	100 to 80% of all intervention conditions	Not combined with other procedures
Partial Evidence	Objective	99 to 50% of all baseline conditions	99 to 50% of all intervention conditions	79 to 50% of all intervention conditions	Combined with other procedures
No Convincing Evidence	Subjective	49 to 0% of all baseline conditions	49 to 0% of all intervention conditions	49 to 0% of all intervention conditions	Combined with other procedures

tion that social stories were effective in changing behaviors. The social story did include role-plays after the story was read to the students. The researchers utilized a multiple probe design across behaviors (e.g., raising hand, social interaction, and vocalizations) to change two students' behaviors.

In 2011, Richter and Test published the final study that showed a clear demonstration that social stories were effective in changing participants' behaviors. In this study the authors utilized a multimedia social story to teach adults how to answer questions on a 16 question quiz. The researchers utilized a multiple probe design across participants and results showed a clear increase in the participants' ability to answer questions correctly; however, there was no measure of participants' behavior other than answering questions on the quiz.

Partial Demonstration. Out of the 41 studies evaluated, 17 studies (41.4%) displayed a partial level of demonstration that the social story was responsible for the participant changing his or her behavior. None of the studies that utilized a case study design were able to show partial demonstration. Seven of the 13 studies (53.8%) that utilized a reversal design were able to show partial demonstration and 10 out of the 19 studies (52.6%) that utilized a multiple baseline design were able to show partial demonstration.

There were seven studies that used a reversal design that were classified as partial demonstration. Four of these studies implemented multiple procedures (i.e., music, prompts, and reminders) along with the social story procedure. Four of the studies, at least once, introduced the intervention when the participant's baseline behavior was not stable or was trending in the wrong direction. Three of these studies, at least once, reversed back to the baseline condition when the intervention data was not stable or was trending in the

TABLE 3

Measures and Demonstration Levels for Multiple Baseline Designs

Level of Demonstration	*Type of Data*	*Baseline conditions stable or trending in the correct direction prior to intervention*	*Appropriate staggering of interventions*	*Clear behavior change in all intervention conditions*	*Combined with other procedures*
Convincing Evidence	Objective	100% of all baseline conditions	100 to 75% correctly staggering intervention	100 to 80% of all intervention conditions	Not combined with other procedures
Partial Evidence	Objective	99 to 66% of all baseline conditions	74 to 50% correctly staggering intervention	79 to 50% of all intervention conditions	Combined with other procedures
No Convincing Evidence	Subjective	65 to 0% of all baseline conditions	49 to 0% correctly staggering intervention	49 to 0% of all intervention conditions	Combined with other procedures

TABLE 4

Results: Levels of Demonstration

Design	Number of Studies	Level of Convincing		
		No Convincing Evidence	*Partial Evidence*	*Convincing Evidence*
Case Studies	9	9 (100%)	0	0
Reversals	13	6 (46.2%)	7 (53.8%)	0
Multiple Baselines	19	6 (31.6%)	10 (52.6%)	3 (15.8%)
Total	41	21 (51.2%)	17 (41.5%)	3 (7.3%)

wrong direction. Finally, three of these studies did not show a clear change in the participants' behaviors.

There were 10 studies that used a multiple baseline design that were classified as partial demonstration. Four studies implemented multiple procedures (i.e., video modeling, prompts, video feedback) along with the social story procedure; however, Thiemann and colleagues (2001) was the only study placed in the partial demonstration category for this reason alone. Two studies using a multiple baseline design, at least once, introduced the intervention at an inappropriate time; three did not stagger interventions correctly; and five did not demonstrate improvement in participant behavior.

No Clear Demonstration. Out of the 41 studies evaluated, 21 (51.2%) studies showed no convincing evidence that social stories were responsible for behavior change. Nine of the nine studies that used a case study design were unable to show a clear demonstration and failed to meet multiple criteria to be considered as partial demonstration. Four of the studies utilized additional procedures other than social stories.

There were six studies that used a reversal design that were classified as no convincing evidence that the social story was responsible for the behavior change. Three of these studies were placed in this level of demonstration due to their lacking of behavioral stability or behavior trending in the correct direction during baseline. Three studies were placed in this level of demonstration due to no behavior change demonstrated and one study was placed in this level due to changing from intervention to baseline conditions without behavior stability or trending in the correct direction.

There were six studies that used a multiple baseline design that were classified as no convincing evidence that the social story was responsible for the behavior change; five of the six studies were unable to show a clear change in the behavior. Five of the six studies had multiple reasons why they were unable to show convincing evidence, including: intervention not being staggered correctly, not showing a clear behavior change, and baseline conditions not trending correctly.

Discussion and Recommendations

The purpose of this paper was to evaluate researchers' execution of single subject methodologies for studies that implemented social stories for children and adolescents diagnosed with ASD. In total, 41 studies were reviewed. Results of this evaluation indicated that, due to the poor implementation of the various research designs, the vast majority of research studies (92.7%) were unable to offer a convincing demonstration of the effectiveness of social story procedures. Fifty one percent of studies were unable to show any clear demonstration that the social story procedure was responsible for changing participants' behaviors and 41% of studies were only able to show partial demonstration that the social story procedure was responsible for changing participants' behaviors. Only 7.3% of studies, through the proper implementation of single subject methodology, were able to show a convincing demonstration that the social story procedure was responsible for changing participant behavior. Thus, the results of this pa-

TABLE 5

Results for Studies Utilizing Case Study Designs

Authors and Year	*Number of Participants*	*Was Social Stories Combined with Other Procedures*	*Behaviors Targeted*	*Measurement System*	*Length of Baseline Condition*	*Was the Baseline Stable or Trending the Correct Direction*	*Was the Effect immediate (within 3 sessions)*	*Percentage of sessions where the intervention condition overlapped with the baseline condition or behavior was in the wrong direction*	*Level of Demonstration*
Bernad-Ripoll 2007	1	Video Modeling	Labeling of emotions, explanation of action responses	Event Recording	10 Sessions	No	No	100%	No Convincing Evidence
						Yes	Yes	0%	
Hutchins & Prelock, 2006	2	Comic Book Strips	Being nice towards sibling, stopping insisting or pestering others to continue to play	Subjective Rating Scale	4 Sessions	Yes	Yes	18.4%	No Convincing Evidence
					9 Sessions	Yes	No	100%	
Moore 2004	1	No	Sleeping in bed	Anecdotal	None Reported	Not Reported	Anecdotal Evidence Only	Not Reported	No Convincing Evidence
Norris & Dattilo1999	1	No	Social interaction (appropriate, inappropriate, None)	Estimated Frequency	5 Sessions	Yes	No	80.5%	No Convincing Evidence
O'Connor 2009	1	No	Swimming	Anecdotal	Not Reported	Not Reported	Anecdotal Evidence Only	Not Reported	No Convincing Evidence
Okada, Ohtake, & Yangihara 2010	1	No	Elbow Position Morning	10 Second Whole Interval	4 Sessions	Yes	No	69%	No Convincing Evidence
			Head Position Morning					78%	
			Elbow Position Lunch					100%	
			Head Position Lunch					24%	
Reynhout & Carter, 2007	1	No	Tapping hands	10 S Partial Interval	7 Sessions	Yes	No	58%	No Convincing Evidence
Rogers & Myles, 2001	1	Comic Strip Conversation	Number redirections, minutes tardy	Anecdotal	0 Sessions	N/A	Anecdotal Evidence Only	Not Reported	No Convincing Evidence
Swaggart et al., 1995	3 (Only 1) Using Single Subject Design	Prompts	Greeting, aggression	Percentage of Opportunity	9 Sessions	No	Yes	11.1%	No Convincing Evidence
		Response Cost System		Frequency	51 Sessions	No	No	Too Difficult to Determine with Published Graph	

TABLE 6

Results of Studies Utilizing Reversal Design

Author and Year	*Number of Participants*	*Design*	*Was Social Stories Combined with Other Procedures*	*Behaviors Targeted*	*Measurement System*	*Percentage of Baseline Conditions that were stable or trending correctly prior to condition change*	*Percentage of Intervention Conditions that were stable or trending correctly prior to condition change*	*Percentage of intervention conditions where there was a clear change in behavior?*	*Level of Demonstration*
Adams et al., 2004	1	ABAB	No	Crying, falling to floor, hitting, screaming	Frequency	37.5%	0%	0%	No Convincing Evidence
Agosta et al., 2004	1	ABA	No	Screaming, Sitting	15 S Partial time Interval	0%	100%	100%	No Convincing Evidence
Bledsoe et al., 2003	1	ABAB	No	Spilling food, wiping his mouth	Frequency	0%	100%	50%	No Convincing Evidence
Brownell 2002	4	ABAC	Music	Repetitive statements, instructional repetitions, loud voice	Frequency	62.5%	87.5%	100%	Partial Evidence
Crozier et al., 2005	1	ABAC	Modified Procedure and Prompts	Talking out	Frequency	100%	100%	100%	Partial Evidence
Crozier & Tincani, 2007	3	ABAB, ABAB, ABCACB	Prompts	Sitting, talking to peers, play	Frequency	66.7%	100%	77.7%	Partial Evidence
Hung, 2011	1	ABAB	None	Shouting	Frequency	50%	100%	100%	Partial Evidence
Ivey, Heflin, & Alberto, 2004	3	ABAB	Prompts	Novel behaviors, Participation Skills	Frequency	67%	66.6%	0%	No Convincing Evidence
Kuoch & Mirenda, 2003	3	ABA, ABA, ABACA	Reminders	Problem behaviors, cheating	Rate per minute	100%	67%	100%	Partial Evidence
Kuttler et al., 1998	1	ABAB	No	Precursor to tantrums	Frequency	25%	100%	100%	No Convincing Evidence
Lorimer et al., 2002	1	ABAB	No	Inappropriate verbalizations, tantrums	Frequency	75%	100%	75%	Partial Evidence
Mancil, Haydon, Whitby, 2009	3	ABABCBC	No	Screaming	Frequency	80%	53.3%	46.7%	No Convincing Evidence
Reichow et al., 2009	1	ABAB	No	Initiations to adults and peers	Frequency	100%	50%	100%	Partial Evidence

TABLE 7

Results for studies utilizing a Multiple Baseline Design

Author and Year	*Number of Participants*	*Design*	*Was Social Stories Combined with Other Procedures*	*Behaviors Targeted*	*Measurement System*	*Percentage of Baseline Conditions that were stable or trending correctly prior to condition change*	*Percentage of Intervention Conditions that were staggered correctly?*	*Percentage of Intervention Conditions were there was a clear behavior change?*	*Level of Demonstration*
Barry & Burlew, 2004	2	Multiple Baseline across Participants with a Reversal	Prompting	Appropriate Play	Duration	100%	100%	100%	Partial Evidence
Chan & O'Reilly, 2008	2	Multiple Probe across Behaviors	No	Raising hand, Social interaction, vocalizations	Frequency	100%	100%	100%	Convincing Evidence
Delano & Burlew, 2006	3	Multiple Probe Across Participants	No	Appropriate, inappropriate and No social engagement	Duration and Frequency	100%	100%	100%	Convincing Evidence
Dodd, Hupp, Jewell, & Krohn, 2008	2	Multiple Baseline Design Across Participants	No	Social Skills	Frequency	100%	100%	66%	Partial Evidence
Graetz, Mastropieri, & Scruggs, 2009	5 (2 Dropped)	Multiple Baseline Across Participants	No	Time on floor, high pitched voice, hands in mouth	Interval System	66.6%	100%	100%	Partial Evidence
Hagiwara & Myles, 1999	3	Multiple Baseline Across Settings	Prompting	Washing Hands	Duration, percentage of steps	100%	0%	0%	No Convincing Evidence
Hanley-Hocdorfer, Bray, Kehle, Elinoff, 2010	4	Multiple Baseline Across Participants	No	Verbal Imitations, Contingent Responses	Frequency	75%	0%	0%	No Convincing Evidence
Klett & Turan, 2012	3	Multiple Baseline Across Participants	No	Menstrual Care	Percentage of Steps Correct	100%	50%	33%	No Convincing Evidence
Ozdemir 2008	3	Multiple Baseline Across Participants	No	Disruptive Behaviors	15 S Partial Interval	66%	100%	100%	Partial Evidence
Quilty 2007	3	Multiple Baseline Across Participants	No	Vocalization, maladaptive behaviors, silly behaviors	Frequency	0%	0%	0%	No Convincing Evidence
Richter & Test, 2011	3	Multiple Probe Across Participants	No	Correct Responses on a Quiz	Responses	100%	100%	100%	Convincing Evidence
Samuels & Stansfield, 2011	4	Multiple Baseline Across Skills	No	Greetings, Inappropriate humor language, Attempts to touch, Hands in underwear, Videotaping	Frequency	62.5%	0%	25%	No Convincing Evidence
Sansosti & Powell-Smith, 2006	3	Multiple Baseline Across Participants	No	Social Engagement	15 S Partial Interval	100%	100%	66%	Partial Evidence
Sansosti & Powell-Smith 2008	3	Multiple Baseline Across Participants	Video Modeling and Prompting	Joining in, Maintaining conversation	15 S Partial Interval	100%	50%	100%	Partial Evidence

TABLE 7—(Continued)

Author and Year	*Number of Participants*	*Design*	*Was Social Stories Combined with Other Procedures*	*Behaviors Targeted*	*Measurement System*	*Percentage of Baseline Conditions that were stable or trending correctly prior to condition change*	*Percentage of Intervention Conditions that were staggered correctly?*	*Percentage of Intervention Conditions were there was a clear behavior change?*	*Level of Demonstration*
Scattone, Wilczynski, Edwards, & Rabian, 2002	3	Multiple Baseline Across Participants	No	Disruptive behaviors	10 S Partial Interval	100%	100%	66%	Partial Evidence
Scattone, Tingstrom & Wilczynski, 2006	3	Multiple Baseline Across Participants	No	Appropriate social interaction	10 S Partial Time Interval	100%	50%	66%	Partial Evidence
Scattone, 2008	1	Multiple Baseline Across Behaviors	Video Modeling Smiling Initiations	Eye Contact	10 S Partial Time Interval	100%	50%	66%	Partial Evidence
Schneider & Goldstein, 2010	3	Multiple Baseline Across Behaviors	Visual Schedule	On Task Behavior	10-15 Sec Momentary Time Sample	0%	50%	66%	No Convincing Evidence
Thiemann & Goldstein, 2001	5	Multiple Baseline Across Behaviors	Video Feedback Written Cues	Contingent responses, securing attention, comments	Frequency	100%	100%	100%	Partial Evidence

per show that, in regards to studies that use single subject methodology, there is little empirical evidence to support that social stories are an effective procedure for children and adolescents diagnosed with ASD.

These results are both similar and different from previous reviews on social stories (e.g., Kokina & Kern, 2010). The majority of reviews have indicated that there are methodological limitations to the research evaluating the social story procedure, which limits the conclusions we can make as to the effectiveness of the procedure. Most reviews, however, also go on to state that social stories may be a promising procedure. There appears to be little data to support the claim that social stories are a promising procedure. It is not enough for a research study to show a positive change in behavior; researchers must be able to demonstrate that the positive change in behavior is due to the intervention being evaluated, which is done through proper implementation of a research design. When researchers fail to implement research designs appropriately, it may introduce certain confounding variables, which can limit or qualify interpretation of the results of the study. Unfortunately, this is the case in the first 19 years of social story research.

For example, when researchers switch from a baseline condition to an intervention condition when the participant's behavior is trending in the wrong direction, it is impossible to know if the intervention had an effect on the behavior or if the behavior change would have occurred naturally (i.e., in the absence of the intervention). Thus, it is important for researchers not to start intervention until the participant is displaying steady behavior that is not trending in the desired direction of the intervention. When social stories are combined with other procedures, it limits the researcher's ability to know if the social story procedure was solely responsible for behavior change. It could be that the social story alone could have caused the behavior change, that the other procedure was responsible for the behavior change, or that the treatment package was needed to change the participants' behaviors.

These are two examples of how not adhering to the correct implementation of a research design can mitigate the conclusiveness

of findings. Based on these and other limitations (analyzed in this study), it is not known how effective social stories actually are in changing behavior. Yet, social stories are still commonly implemented (e.g., Reynhout & Carter, 2009) and recommended (The National Autism Center, 2009) for children and adolescents with ASD. The question has to be why is this procedure still being implemented and recommended by clinicians, teachers, and parents. One possible reason why it is being implemented is that it is a relatively easy procedure to implement as compared to more difficult procedures, such as video modeling (e.g., Charlop & Milstein, 1989), script fading (e.g., Sarokoff, Taylor, & Poulson, 2001), and the teaching interaction procedure (e.g., Leaf et al., 2012). A second possible reason is that the social story is still perceived to be an effective intervention (Reynhout & Carter, 2009); this could be due to a lack of understanding of research designs, a self-fulfilling prophecy (e.g., desire for a procedure you are implementing to be successful), or a reliance on subjective measurement.

Based on this review, other recent reviews, and comparative studies that were not included in this analysis (e.g., Leaf et al., 2012), there are several recommendations that can be made for clinicians and researchers. Although social stories is an easy procedure to implement, at this point, clinicians should apply other empirically supported procedures (e.g., video modeling, script fading, behavioral skills training, the teaching interaction procedure) when trying to teach pro-social behavior or decrease aberrant behavior for children and adolescents diagnosed with autism. Although there are a lot of studies that have examined social stories, only three studies have research rigor to show experimental support that the procedure is effective and clinicians should only implement procedures that are demonstrated to be effective in the empirical research.

Although social stories may not currently be a promising avenue for clinicians, teachers, and parents, they are a promising avenue for researchers. Future researchers should evaluate the effects of social stories using a wide variety of single subject methodology; however, these future researchers must ensure that the research methodology is implemented appropriately. In doing so, we can better assess how effective social stories truly are. If empirically sound research shows that social stories are an effective procedure, then future researchers should compare social stories to other procedures (e.g., video modeling, script fading, behavioral skills training) to determine which procedures are the most efficient and effective. In doing so, clinicians will be able to implement the most efficacious empirically proven procedures to individuals diagnosed with autism.

References

* Indicates articles that were evaluated in the review

*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.

*Agosta, E., Graetz, J. E., Mastropieria, M. A., & Scruggs, T. E. (2004). Teacher-researcher partnerships to improve social behavior through social stories. *Intervention in School and Clinic, 39,* 276–287.

Ali, S., & Frederickson, N. (2006). Investigating the evidence base of Social Stories. *Educational Psychology in Practice,* 22, 355–377.

Bailey, J. S., & Burch, M. R. (2002). Research methods in applied behavior analysis. Thousand Oaks, California: Sage Publications, Inc.

*Barry, L. M., & Burlew, S. B. (2004). Using social stories to teach choice and play skills to children with autism. *Focus on Autism and Other Developmental Disabilities, 19,* 45–51.

*Bernard-Ripoll, S. (2007). Using a self-as-model video combined with social stories™ to help a child with Asperger syndrome understand emotions. *Focus on Autism and Other Developmental Disabilities, 22,* 100–106.

*Bledsoe, R., Myles, B. S., & Simpson, R. L. (2003). Use of a social story intervention to improve mealtime skills of an adolescent with Asperger syndrome. *Autism,* 7, *289-295.*

*Brownell, M. D. (2002). Musically adapted social stories to modify behaviors in students with autism: Four case studies. *Journal of Music Therapy, 2,* 117–144.

*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.

Charlop, M. H., & Milstein, J. P. (1989). Teaching autistic children conversational speech using

video modeling *Journal of Applied Behavior Analysis, 22,* 275–285.

*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.

*Crozier, S., & Tincani, M. J. (2007). Effects of social stories on prosocial behavior of preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders,* 37, 1803–1814.

*Delano, M., & Snell, M. E. (2006). The effects of social stories on the social engagement of children with autism. *Journal of Positive Behavior Interventions, 8,* 29–42.

*Dodd, S., Hupp, S. D. A., Jewell, J. D., & Krohn, E. (2008). Using parents and siblings during a social story intervention for two children diagnosed with PDD-NOS. *Journal of Developmental and Physical Disabilities, 20,* 217–229.

*Graetz, J. E., Mastropieri, M. A., & Scrugss, T. E (2009). Decreasing inappropriate behaviors for adolescents with autism spectrum disorders using modified social stories. *Education and Training in Developmental Disabilities, 44,* 91–104.

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.

Gray, C. (1994). *Social stories.* Arlington, TX: Future Horizons.

Gray, C. (1995). Teaching children with autism to read social situations. In K. A. Quill (Ed.)., *Teaching children with autism* (pp. 219–241). New York: Delmar.

Gray, C. (2000). (Illustrated edition). *The new Social Story book.* Arlington, TX: Future Horizons Inc.

Gray, C. (2002). *My Social Stories book.* London, Englad: Jessica Kingsley.

Gray, C. (2003). Social Stories™ 10.0. *Jenison Autism Journal, 15.*

Gray, C. (2004). Social Stories 10.0: The new defining criteria. *Jenison Autism Journal,* 15, 1–21.

*Hagiwara, T., & Myles, B. S. (1999). A multimedia social story intervention: Teaching skills to children with autism. *Focus on Autism and Other Developmental Disabilities, 14,* 82–95.

*Hanley-Hocdorfer, K., Bray, M. A., Kehle, T. J., & Elinoff, M. J. (2010). Social stories to increase verbal initiation in children with autism and asprger's disorder. *School Psychology Review, 39,* 484–492.

*Hung, L. C. (2011). Autism in Taiwan: Using social stories to decrease disruptive behaviour. *The British Journal of Developmental Disabilities, 57,* 71–80.

*Hutchins, T. L., & Prelock, P. A. (2006). Using social stories and comic strip conversations to promote socially valid outcomes for children with autism. *Seminars in Speech and Language, 27,* 47–59.

*Ivey, M. L., Heflin, J., & Alberto, P. (2004). The use of social stories to promote independent behaviors in novel events for children with PDD-NOS. *Focus on Autism and Other Developmental Disabilities, 19,* 164–176.

*Klett, L. S., & Turan, Y. (2012). Generalized effects of social stories with task analysis for Teaching menstrual care to three young girls with autism. *Sexual Disabilities, 30,* 319–336.

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.

*Kuoch, H., & Mirenda, P. (2003). Social story interventions for young children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 18,* 219–227.

*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.

Leaf, J. B., Oppenheim-Leaf, M. L., Call, N. A., Sheldon, J. B., Sherman, J. A., Taubman, M., . . . Leaf, R. (2012). Comparing the teaching interaction procedure to social stories for people with autism. *Journal of Applied Behavior Analysis, 45,* 281–298.

*Lorimer, P. A., Simpson, R. L., Myles, B. S., & Ganz, J. B. (2002). The use of social stories as preventative behavioral intervention in a home setting with a child with autism. *Journal of Positive Behavior Interventions, 4,* 53–60.

*Mancil, R. G., Haydon, T., & Whitby, P. (2009). Differentiated effects of paper and computer-assisted social stories™. *Focus on Autism and Other Developmental Disabilities, 24,* 205–215.

*Moore, P. S. (2004). The use of social stories in a psychology services for children with learning disabilities: A case study of a sleep problem. *British Journal of Learning Disabilities, 32,* 133–138.

National Standards Report. (2009). The national standards project— addressing the need for evidence-based practice guidelines for autism spectrum disorders, *National Autism Center.*

*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.

*O'Connor, E. (2009). The use of social story dvds to reduce anxiety levels: A case study of a child with autism and learning disabilities. *Support of Learning, 24,* 133–136.

*Okada, S., Ohtake, Y., & Yanagihara, M. (2010). Improving the manners of a student with autism: The effects of manipulating perspective holders

in social stories™-A pilot study. *International Journal of Disability: Developmental and Education, 57,* 207–219.

*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.

*Quilty, K. M. (2007). Teaching paraprofessionals how to write and implement social stories for students with autism spectrum disorders. *Remedial and Special Education, 28,* 182–189.

*Reichow, B., & Sabornie, E. J. (2009). Brief report: Increasing verbal greetings initiations for a student with autism via a social story™ intervention. *Journal of Autism and Developmental Disorders, 39,* 1740–1743.

Reynhout, G., & Carter, M. (2009). The use of social stories by teachers and their perceived efficacy. *Research in Autism Spectrum Disorders, 3,* 232–251.

*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 Disabilities, 22,* 173–182.

*Richter, S., & Test, D. (2011). Effects of multimedia social stories on knowledge of adult outcomes and opportunities among transition-aged youth with significant cognitive disabilities. *Education and Training in Autism and Developmental Disabilities, 46,* 410–424.

*Rogers, M. F., & Myles, B. S. (2001). Using social stories and comic strip conversations to interpret social situations for an adolescents with Asperger syndrome. *Intervention in School and Clinic, 36,* 310–313.

*Samuels, R., & Stansfield, J. (2011). The effectiveness of social stories™ to develop social interactions with adults with characteristics of autism spectrum disorder. *British Journal of Learning Disabilities, 40,* 272–285.

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 Disabilities,* 19, 194–204.

*Sansosti, F. J., & Powell-Smith, K. A. (2006). Using social stories to improve the social behavior of children with Asperger syndrome. *Journal of Positive Behavior Intervention, 8,* 43–57.

*Sansosti, F. J., & Powell-Smith, K. A. (2008). Using computer-presented social stories and video models to increase the social communication skills of children with high-functioning autism spectrum disorders. *Journal of Positive Behavior Interventions, 10,* 162–178.

Sarokoff, R. A., Taylor, B. A., & Poulson, C. L. (2001). Teaching children with autism to engage in conversational exchanges: Script fading with embedded textual stimuli. *Journal of Applied Behavior Analysis, 34,* 81–84.

*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.

*Scattone, D., Tingstrom, D. H., & Wilczynski, S. M. (2006). Increasing appropriate social interactions of children with autism spectrum disorders using social stories™. *Focus on Autism and Other Developmental Disabilities, 21,* 211–222.

*Scattone, D. (2008). Enhancing the conversation skills of a boy with asperger's disorder through social stories™ and video modeling. *Journal of Autism and Developmental Disorders, 38,* 395–400.

*Schneider, N., & Goldstein, H. (2010). Using social stories and visual schedules to improve socially appropriate behaviors in children with autism. *Journal of Positive Behavior Interventions, 12,* 149–160.

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

*Thiemann, K. S., & Goldstein, H. (2001). Social stories, written text cues, and video feedback: Effects on social communication of children with autism. *Journal of Applied Behavior Analysis, 34,* 425–446.

Received: 5 November 2013
Initial Acceptance: 6 January 2014
Final Acceptance: 20 February 2014