
SOCIAL STORIES AS AN INTERVENTION METHOD FOR INDIVIDUALS WITH AUTISM: A DESCRIPTIVE REVIEW OF SEVEN META-ANALYSES

► DIMITRIS VROIKOS

Electrical Engineer, Computer Engineer
Med. Organization and Administration of Education, Educational Leadership
Director of Vocational Laboratory Centers, Eastern Thessaloniki
dvroikos@gmail.com

► IOANNA AIVATZIDOU

Kindergarten Special Education Teacher – Economist
Meds. Special Education, Educational Sciences, Planning, Management and
Tourism Policy
Director of a multi-class preschool in Eastern Thessaloniki
iaivatzidou@gmail.com

Abstract

The present paper is a descriptive literature review of meta-analytic studies, which examine the effectiveness of social stories as an intervention tool for individuals with Autism Spectrum Disorder (ASD). The aim is to synthesize and critically present seven meta-analyses, based on the main stages of their implementation. The search for meta-analyses was carried out in scientific databases using keywords such as “meta-analysis,” “social stories,” and “autism.” Although the international literature recognizes the positive impact of social stories on reducing problematic behaviour and enhancing social skills, the findings of this review indicate small to moderate effectiveness, which varies depending on participant characteristics. Furthermore, it was found that social stories are not sufficiently documented as an evidence-based practice (EBP), yet they remain for many educators a practical and accessible strategy in educational settings.

Keywords: Autism, Social stories, Meta-analysis

Introduction

The prevalence of Autism Spectrum Disorder (ASD) has increased markedly over the past two decades (Maenner et al., 2023). Teachers in both general and special education across all levels are called upon to create appropriate supportive environments for students with autism, applying specialized approaches and intervention strategies. Social stories represent one of the most widespread strategies, as they help children with autism understand social situations and adopt appropriate social behaviours (Rodrigues da Silva & Mwangi, 2025, p. 2). In contrast to many other interventions for children with autism, social stories can highlight the uniqueness that these children may have. Specifically, because children with autism often adhere strictly to

routines, a social story can serve to establish a routine or rule that the child may apply in a social situation. By utilizing this dynamic, social stories may be less intrusive for children with autism than other interventions designed to address social skill deficits. Presenting instructions in the form of written commitments (as opposed to the interactive nature of oral instructions) minimizes the social aspect of teaching and reduces the resistance that may arise toward verbal commands. Thus, a social story may allow children with autism to perceive the rules governing social behaviour in a way that maximizes the likelihood of benefiting from the proposed guidelines.

The purpose of this paper is to conduct a descriptive literature review of studies from the last decade that follow the methodology of meta-analysis and investigate the effectiveness of social stories as an intervention for individuals with ASD. The study aims to synthesize and critically present seven meta-analyses according to the main stages defined for conducting a meta-analysis: formulation of the research question, determination of inclusion criteria, search and evaluation of the literature, statistical analysis, and interpretation of findings. Additionally, emphasis is placed on highlighting the existing knowledge regarding intervention through social stories, as well as the methodological issues and writing characteristics emerging from the examined studies.

Social stories and autism

Autism is a neurodevelopmental disorder. It is a lifelong disability and belongs to the pervasive developmental disorders, or more precisely, to Autism Spectrum Disorders (ASD). Autism Spectrum Disorder (ASD) is characterized by persistent deficits in social communication and social interaction, alongside restricted, repetitive patterns of behaviour, interests, or activities. Symptoms appear early in development and cause significant impairment in functioning. This definition follows the DSM-5-TR criteria. The DSM-5-TR criteria for Autism Spectrum Disorder (ASD) include (Yu, Ozonoff, & Miller, 2023):

1. Persistent deficits in social communication and interaction, such as difficulty with social reciprocity, nonverbal communication, and forming relationships.
2. Restricted, repetitive patterns of behaviour, interests, or activities, including repetitive movements, insistence on sameness, intense interests, and unusual sensory responses.
3. Symptoms appear in early developmental period.
4. Symptoms cause significant impairment in daily functioning.
5. Symptoms are not better explained by intellectual disability or other developmental disorders.

Individuals who exhibit the disorder present significant differences among themselves, as the deficits they demonstrate may range from very mild to very severe, and for this reason, autism is considered a spectrum disorder (National Research Council, 2001). One of the most prominent deficits of children with autism relates to social skills, such as initiating and responding to a conversation, coping with a change in routine, understanding

how others may feel or think, and generally how to respond appropriately in a social situation and interact with peers and others in their environment (Chamberlain, Kasari, & Rotheram-Fuller, 2007, p.231). Deficits in social behaviours may become more evident, distinct, and critical when individuals with autism reach school age and adulthood. Since social skills are an important aspect of everyday life, improving social functioning is one of the most significant intervention goals for individuals with autism (Weiss & Harris, 2001, p.798).

Social stories are an important intervention strategy in special education and represent a method of teaching social skills. They constitute a dynamic strategy that can be used with children with autism as well as with children experiencing a range of other social dysfunctions or deficits. A social story is a short story written for a specific child, describing a social situation, a person, a skill, or an event based on expected and appropriate social responses (Scattone, Wilczynski, Edwards & Rabian, 2002, p.535). Each social story is designed to teach children with autism how to manage their own behaviour during a specific social situation by describing the place where the activity will occur, when it will happen, what will happen, who is involved, and why the child should behave in a certain way. They have been used to reduce problematic behaviours such as tantrums, self-injury, aggression, stereotypy, property destruction, noncompliance, and others. They involve individualized stories that provide individuals with specific information about skills and situations they find challenging and often include visual material to facilitate comprehension for the intended recipient. As Grandin (1988, p.6) points out, children with autism tend to learn better visually, which suggests that the effectiveness of social stories is enhanced through the use of images during the intervention.

Meta-analysis

Meta-analyses are systematic literature reviews in which data undergo statistical processing. They are secondary studies and constitute powerful tools for examining the literature and critically analysing results derived from primary research. They contribute to consolidating or deconstructing theories that emerge from independent studies of a related subject, aiming to group data and draw clearer conclusions. They are systematic quantitative reviews and represent a research approach grounded in a specific scientific methodology. They are particularly important and useful, as they can provide new knowledge to the educational community by offering answers to previously unanswered questions. Importantly, they can be conducted without the costs associated with primary research. In the field of education, they have been applied to various issues, with their results influencing practice and policy worldwide (Borenstein, Hedges, Higgins, & Rothstein, 2011, p.38).

To conduct a meta-analysis, there must be primary studies that share common basic characteristics, have strong methodology, are free from systematic errors, and present results that are not inconsistent. For a meta-analysis to be scientifically valid,

its writing and implementation must follow scientific principles and rules that minimize potential errors inherent in any research study. The key components of a meta-analysis include formulating the research question, searching the literature, selecting and evaluating studies, statistically processing the data, and interpreting the results.

Methodology

The search strategy included widely accessible and multidisciplinary electronic search engines (Google, Bing, Yahoo), electronic databases (ERIC, Heal-Link), and scientific article search engines (Google Scholar, Openarchives.gr), aiming to cover a wide range of sources, including theses, dissertations, and open-access articles. However, specialized databases such as PsycINFO and PubMed were not used, due to limited access and the intention to include broader document types. This limitation should be acknowledged as a weakness of the study, as it may have affected the comprehensiveness of the search.

The keywords used for the search were “meta-analysis,” “social stories,” and “autism,” while the publication period was limited to the last decade. The search yielded seven studies published between 2009 and 2019, which met the criteria relevant to the research topic.

To synthesize the data, a descriptive review of meta-analyses was conducted, aiming to present high-level evidence from multiple meta-analyses and provide a broader and more consolidated understanding of social stories interventions for individuals with autism. Since meta-analyses already aggregate results from individual studies, their descriptive review allows for the identification of overarching trends, methodological strengths and weaknesses, and gaps in the literature, without the need to re-analyze all primary studies. This approach is particularly useful when several meta-analyses exist on related questions, as in the present case (Ntais & Talias, 2024).

The descriptive synthesis of these studies follows the stages characteristic of a structured meta-analytic review. These stages include formulating the research question, defining inclusion and exclusion criteria, conducting the literature search, selecting and evaluating studies, performing statistical analysis of the data, and interpreting the results (Patelarou & Brokalaki, 2010, p. 124).

Descriptive review

Research questions – objectives

The first stage in every study is the clear formulation of the research question, which must be scientifically justified and which the conducted meta-analysis attempts to answer. The transformation of the research question into a research project is achieved through research objectives, which state the steps to be followed in order to derive answers and determine how the research process will be structured. The individual research questions in the selected studies can be summarized in one central research question, namely the use of social stories as an intervention tool for children with autism, along with the objectives surrounding this

question as they emerged from the examined studies presented below.

In the study by Wang & Spillane (2009), the aim was to review research studies published over the past ten years regarding interventions to improve social skills in children and adolescents with autism. Among these interventions there were social stories. Key objectives included examining the outcomes of the studies, evaluating whether these interventions met the necessary scientific criteria, and determining whether their results were based on clear evidence.

The main objectives of the Kokina & Kern (2010) meta-analysis were to examine the overall effectiveness of social stories, describe the ways in which social stories were used in research studies, and investigate the role of a set of variables involved in the intervention, including characteristics of both the facilitator and the participant.

Test, Richter, Knight, & Spooner (2011) conducted a systematic meta-analysis that included a descriptive review, qualitative analysis, and meta-analysis using the percentage of non-overlapping data (PND). The main objectives of the study were:

- a) to conduct a comprehensive review of the quality of the literature using current criteria based on the evidence standards developed by Horner et al. (2005),¹
- b) to calculate PND, and
- c) to determine whether social stories can be considered an evidence-based practice (EBP).

The study by Reynhout & Carter (2011) evaluated the effectiveness of social stories by examining single-case studies. It focused first on determining whether social stories are effective in promoting behaviour change in individuals with autism or other disabilities, and secondly on investigating whether the effectiveness of social story interventions varies depending on the study or participant characteristics.

In the study by Bozkurt & Vuran (2014), the purpose was to investigate whether the use of social stories for teaching skills was effective, according to the findings. The specific objectives serving this purpose were to address questions concerning the research context and the characteristics of the participants, the key skills taught, the reasons for selecting them, as well as the data collection and evaluation methods used. Furthermore, the study explored whether the process followed proper scientific methodology, clarifying who conducted the research, which research models were used, what sentence types were employed in the social stories, and what independent variables were applied in the social story intervention.

In the study by McGill, Baker & Busse (2015), the aim was to conduct a quantitative review of social story interventions for students diagnosed with autism. The initial hypothesis suggested that small to moderate effects would be demonstrated in reducing problematic behaviours, based on prior research. It was also hypothesized that outcome variability would be observed across indicators. The primary objective was to quantitatively examine results for various outcome measures. A secondary objective was

1. The bibliographic references contained in each study are not listed in the References section at the end of the present paper; however, they are provided as footnotes. 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.

to assess the potential impact of variable fluctuations on intervention outcomes.

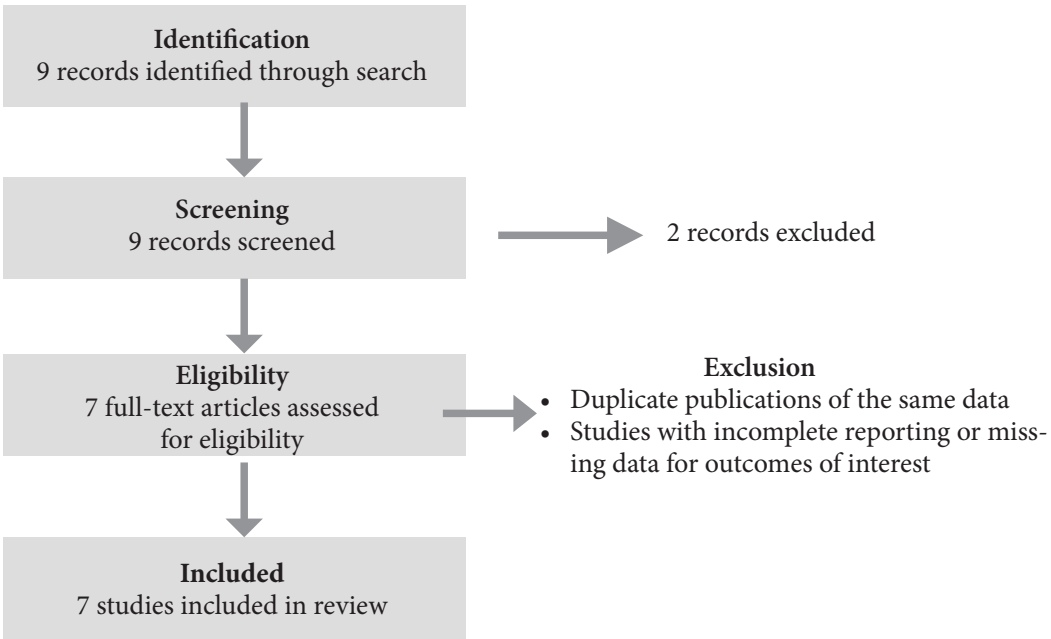
Finally, the study by Wahman, Pustejovsky, Ostrosky, & Santos (2019) evaluated the effectiveness of social stories in reducing challenging behaviour and improving prosocial skills in children under the age of five, including those with autism. The objectives of this study were to assess the quality of available evidence using WWC standards, synthesize findings using parametric effect size estimates with meta-analytic methods appropriate for behavioural outcomes, and investigate potential moderators of intervention effectiveness.

Selection and evaluation of studies – Inclusion and exclusion criteria

An important step in a meta-analysis is defining the study population-sample, as well as the inclusion and exclusion criteria for the studies. By applying the predetermined criteria, the sample is derived. These criteria may refer to participant characteristics, the type of intervention, duration of implementation, methods of comparison and analysis, as well as the number of studies. In the studies examined in this review, the selection criteria are, for the most part, common, allowing for a basis of comparison. Table 1 presents the PRISMA flow chart for study selection in the social stories interventions review. They are presented in detail per study and summarized in Table 2 that follows.

In the study of Wang and Spillane (2009), an electronic search was conducted for studies published between 1997 and 2008. The studies were selected based on the following criteria. The participants had been diagnosed with autism between birth and 21 years of age and were students receiving special education services either at home or in the school environment. They had used interventions targeting social skills (studies measuring functional

Table 1
PRISMA Flow Chart.



No.	Authors	Title	Year of Publication of the Study	Year of Publication of the Studies Reviewed	Number of Studies Reviewed	Age of Reference Groups
1	Wang, P., & Spillane, A.	Evidence-based social skills interventions for children with autism: A meta-analysis	2009	1997-2008	36	0 – 21 age
2	Kokina, A., & Kern, L.	Social Story™ interventions for students with autism spectrum disorders: A meta-analysis	2010	2002-2009	18	0 – 21 age
3	Test, D. W., Richter, S., Knight, V., & Spooner, F.	A comprehensive review and meta-analysis of the social stories literature	2011	1995-2007	18	5 – 18 age
4	Reynhout, G., & Carter, M. (2011).	Evaluation of the efficacy of Social Stories™ using three single subject metrics	2011	1995-2009	62	2 - 57 (71%, 5 to 12 age)
5	Bozkurt, S., & Vuran, S.	An Analysis of the Use of Social Stories in Teaching Social Skills to Children with Autism Spectrum Disorders	2014	1998-2010	22	0-15 age
6	McGill, R. J., Baker, D., & Busse, R. T.	Social Story™ interventions for decreasing challenging behaviours: A single-case meta-analysis 1995–2012	2015	1995–2012	27	5-12 age
7	Wahman, C. L., Pustejovsky, J. E., Ostrosky, M. M., & Santos, R. M.	Examining the Effects of Social Stories™ on Challenging Behaviour and Prosocial Skills in Young Children: A Systematic Review and Meta-Analysis	2019	2002-2012	12	3-10 age

Table 2

Summary table of the seven meta-analyses.

skills such as daily living skills, reduction of problem behaviours, or non-social communication skills were not included) and evaluated the effectiveness of social competence interventions. They had been published in a peer-reviewed journal between 1997 and 2008 and met the evaluation criteria of intervention strategy, characterized as evidence-based practice (EBP) by the Council for Exceptional Children Division for Research in 2005.

In order to be included in the meta-analysis of Kokina and Kern (2010), studies had to use social stories as the sole intervention. Furthermore, they had to apply a single-case design with experimental control and provide graphical baseline and intervention data allowing the calculation of the percentage of non-overlapping data (PND). Finally, participants had to have

a primary diagnosis of autism from an independent diagnostic center. The following were excluded: (a) studies without quantitative data (e.g., qualitative case studies) or that used group action designs, (b) studies employing non-experimental designs, (c) studies using social stories as part of a set of interventions (e.g., social stories combined with verbal prompts, video modeling, role play, etc.), (d) studies with participants with disorders other than autism (e.g., behavioural disorders, intellectual disabilities), and (e) studies where the PND index could not measure the effectiveness of the intervention.

For a study to be included in the review of Test, Richter, Knight, & Spooner (2011), social stories had to represent part of the interventions, assess the impact of the intervention on a dependent variable, report quantitative or qualitative results, and be published in a journal. Studies excluded from this review were those that used a single-case design to demonstrate the impact of social stories and those that did not examine social stories as an independent variable.

The inclusion criteria in the study of Reynhout & Carter (2011) were that the studies had been published in English, were available in databases before October 23, 2009, and implemented a social story intervention. In addition, at least one participant with a disability had to be included, and a small-N research design had to be used. The studies had to present data in graphic or text form and provide information regarding learner performance. Since social stories are applied for a wide range of purposes, no further restrictions were imposed on study outcomes. Data concerning teacher perceptions were excluded from the analysis, as well as studies concerning the use of social scripts, mutual story telling, and narrative therapy. Descriptive cases not including quantitative data were also excluded, and no geographical restriction was applied.

In the selection of articles included in the study of Bozkurt & Vuran (2014), the criteria concerned articles published in scientific journals between 1991 and 2011, in which experimental studies related to individuals with autism were examined. The studies examined dependent variables of social skills and independent variables related only to social stories. Articles concerning the effectiveness of social stories in simple single cases that presented only social stories and social stories with additional interventions were considered. In addition, studies that used a single-case design with experimental control and provided graphical data allowing the calculation of the percentage of non-overlapping data (PND) were examined.

In the study of McGill, Baker & Busse (2015), seven specific selection criteria were applied to investigate potential studies for inclusion in the meta-analysis. The participants were school-age children from preschool through Grade 12 who had been diagnosed with autism. The study used social stories as the primary component in the intervention package. The intervention program primarily aimed at reducing problem behaviours, while the studies also sought to increase pro-social behaviours; however, these results were not incorporated in the meta-analysis.

The study used a single-case research design with experimental control. It also had to follow the standards adopted by the *What Works Clearinghouse* for evaluating single-case research.

The studies used in the research of Wahman, Pustejovsky, Ostrosky, & Santos (2019) had to meet the following criteria. They had to be published in a peer-reviewed journal between 1995 and 2018 and report one or more single-case designs. Social stories had to be used as the primary intervention for reducing challenging behaviour and increasing social behaviour, and at least one participant in the intervention had to be under the age of five. They had to present outcomes for at least one measure of challenging behaviour and be conducted in the United States. The participants in the studies could be both children with challenging behaviour and children with autism.

Recording, coding, and statistical analysis of data

An important step in a meta-analysis is the recording of the main elements of the studies so that comparison and determination of the degree of similarity can be made. The degree of similarity is characterized by the term “assessment of study heterogeneity.” To minimize errors, the data recorded from the studies result from the collaboration of two researchers (Patelarou & Brokalaki, 2010, p. 126).

During the statistical analysis of the data, all study results are essentially synthesized to produce an aggregated outcome. The statistical analysis (meta-analysis) of the study data can be conducted using random effects models, fixed effects models, and regression models. These models are used to calculate the summary effect estimate. Subsequently, statistical tests are performed to check the consistency of the result as well as subgroup analyses. In addition, the study should present data in graphs depicting individual data points. Typically, single-case studies are interpreted through visual inspection of graphs; however, this may be subjective, resulting in a lack of objectivity in interpreting results. As an alternative, the calculation of PND (Percentage of Non-overlapping Data) using these graphs has been proposed as a statistical measure in a meta-analysis.

The PND is a widely used statistical index for measuring and meta-analyzing single-case experimental designs. Scruggs et al. (1987, p. 27) note that the PND is the only significant evaluation criterion that can be consistently applied to most single-case studies. The PND determines the magnitude of behavioural change from baseline to intervention phase by calculating the percentage of overlap between data in these phases. The number of intervention data points exceeding the highest or lowest baseline data point is divided by the total number of intervention data points, and the result is multiplied by 100%. The PND is usually used to measure outcomes in meta-analyses of single-case research as an alternative to traditional visual methods of analysis. A PND between 91–100 is considered a highly effective intervention, between 71–90 moderately effective, between 51–70 minimally effective, and between 0–50 ineffective (Scruggs & Mastropieri, 1998, p. 224).

In studies where statistical data are used, inter-rater reliability is also examined, representing the degree of agreement between evaluators. It estimates the extent of homogeneity among the scores given by different raters. This evaluation is useful for determining whether the tools used by raters –such as a specific scale– are appropriate for measuring a given variable. If evaluators do not agree, either the scale is flawed, or the raters need retraining.

In the 36 single-case research studies used by Wang & Spillane (2009), a coding system developed by Mastropieri and Scruggs (1985–1986)² was applied and modified according to criteria described by Horner et al. (ibid.). Each study was categorized based on characteristics such as the number of participants, diagnosis, and age, the description of the target behaviours and skills, as well as the description of the intervention and research design. Furthermore, studies were categorized according to generalized outcomes, as measured by the percentage of non-overlapping data (PND), and confirmation that the study assessed integrity and validity.

To determine inter-rater reliability for the coding and PND analysis procedures, the first author coded all single-case studies and calculated PND, while the second author randomly selected 30% (12) of the studies and independently coded and calculated PND for those. Inter-rater reliability was calculated by dividing the total number of agreements by the sum of agreements and disagreements multiplied by 100. The mean agreement between the two authors was 97% (range 78%–100%).

In the meta-analysis by Kokina & Kern (2010), each study was summarized, and information was coded into three categories concerning the study methodology, the key characteristics of the intervention, and the participants' characteristics. Specifically, coding included experimental design (e.g., ABAB design, multiple-baseline designs), assessment of skill maintenance and generalization, treatment fidelity (i.e., accuracy of implementation), social validity (i.e., acceptability of the intervention to stakeholders), and inter-observer reliability (i.e., consistency of measurement among observers).

To determine whether specific study characteristics were associated with higher intervention effectiveness, PND was calculated. To summarize intervention outcomes and estimate effectiveness for each study characteristic, median PND values were used instead of means, since the median is a positional measure less affected by outliers and more appropriate when PND data are not normally distributed. The first author served as the primary coder. Two additional independent raters –doctoral students in special education– coded a random sample of 5 out of 18 studies (28%) and calculated PND scores. The inter-rater reliability for study coding was 94% (range 89–97%). Disagreements were primarily due to difficulties interpreting participant characteristics and were resolved through discussion. Agreement for PND calculations was 100%.

In the meta-analysis by Test, Richter, Knight, & Spooner (2011), qualitative indicators for single-case research, developed

2. Mastropieri, M. A., & Scruggs, T. E. (1985-1986). Early intervention for socially withdrawn chil dren. *The Journal of Special Education*, 19, 429-442.

by Horner et al. (ibid.), were used to evaluate the methodological quality of each included study. Two authors reviewed the studies using a form developed based on these standards. The researchers calculated PND scores for all social story interventions meeting inclusion criteria and also computed PND for maintenance and generalization of skills.

Inter-rater reliability data were collected for each category included in the three analyses (i.e., descriptive analysis, alignment with quality indicators, and analysis of intervention outcomes). A second researcher randomly selected at least 27% of the articles in each analysis to determine inter-rater reliability by dividing agreements by the sum of agreements and disagreements. For descriptive findings, inter-rater reliability was calculated item-by-item, and reliability was computed for 8 of 28 studies (28.6%), ranging from 80% to 100%, with a mean of 94.9%. For coding quality indicators, two researchers independently evaluated 8 of 28 studies, yielding inter-rater reliability between 85.7% and 100% (mean 95.8%). For PND checks, inter-rater reliability was assessed point-by-point on each data graph, with five of 18 studies (27.8%) evaluated, showing 100% agreement.

In the study by Reynhout & Carter (2011), a PND statistic was calculated for each study using the combined number of non-overlapping data points across all participants and conditions, as described by Schlosser et al. (2008)³ and Scruggs et al. (ibid.). In addition, a PEM (Percent of Data Points Exceeding the Median) statistic was calculated using the grouped number of data points exceeding the median across participants and conditions. An IRD (Improvement Rate Difference) statistic was also computed as described by Parker & Hagan-Burke (2007),⁴ by measuring improvement rates from baseline to treatment phases across individuals and conditions. IRD calculation involves eliminating overlap between baseline and intervention data by removing the smaller number of data points.

In some cases, two statistical methods were applicable, and both were used with the more realistic result being selected. Inter-rater reliability checks were conducted on a randomly selected sample of 14 studies (23%) for recording study details and assessing qualitative research quality. Agreements were calculated as a percentage of agreements over agreements plus disagreements. Reliability measures for each section of the study were as follows: participant description (age, cognitive ability, autism level, diagnosis) 94%; study description (design, target behaviours, settings, construction of social story, use of additional strategies) 94%; outcome agreement 93%; and maintenance/generalization data 98%. Overall inter-rater reliability for research quality was 95%. Reliability for participant information quality was 98%, dependent variables 96%, independent variables 100%, key characteristics 89%, internal validity 100%, external validity 93%, and social validity 95%.

In the meta-analysis by Bozkurt & Vuran (2014), a total of 22 studies published between 1998 and 2010 were included. PND scores were calculated for intervention, maintenance, and generalization outcomes. Inter-rater reliability was calculated for 8

3. Schlosser, R., Lee, D. L., & Wendt, O. (2008). Application of the percentage of non-overlapping data (PND) in systematic reviews and meta-analyses: A systematic review of reporting characteristics. *Evidence-based Communication Assessment and Intervention*, 2(3), 163–187.

4. Parker, R., & Hagan-Burke, S. (2007). Median-based overlap analysis for single case data. A second study. *Behavior Modification*, 31, 919–936.

of the 22 studies (at least 30%). To ensure reliability in coding and PND analysis, two authors independently coded each study and compared results, achieving 100% agreement on study characteristics. PND scores were calculated from the graphs provided in each study and summarized for each one. For ABAB designs, PND scores were computed for baseline and intervention phases per participant (i.e., total non-overlapping points divided by total intervention data points). For multiple designs, separate PND scores were computed for each behaviour, and their mean was used as the overall PND for the study. PNDs were calculated for intervention, maintenance, and generalization outcomes. Maintenance and generalization were assessed by comparing PND between baseline and maintenance/generalization phases. PND could not be computed for 10 (45.45%) single-case studies using designs that did not allow identification of a functional relationship (e.g., AB, ABAC, ABC).

In the study by McGill, Baker & Busse (2015), two authors independently coded inclusion criteria, study information, and statistical computations. Reliability and procedural fidelity were assessed by the lead author, who had prior experience in single-case research and meta-analysis. Coding included parameters such as treatment integrity (i.e., whether studies reported integrity data for $\geq 25\%$ of intervention sessions), social validity (i.e., whether studies explicitly gathered acceptability data from stakeholders), reliability assessment (i.e., whether studies systematically assessed dependent variable accuracy for $\geq 25\%$ of sessions), and functional assessment (i.e., whether data on antecedents, consequences, duration, latency, or frequency were reported as part of intervention planning).

To compare findings with prior social story research, three single-case indicators were used to evaluate each study. The use of multiple indicators aligns with best practice guidelines (Beretvas & Chung, 2008).⁵ Intervention effectiveness was evaluated using Visual Analysis Ratings (VAR), a method in applied behaviour analysis that assesses whether intervention data demonstrate effectiveness. Researchers calculated overall PND statistics for each qualifying single-case study. For complex designs, PNDs were computed for each primary interval, then averaged per case to produce an overall estimate. Fail-safe N analyses were also conducted to assess publication bias, following Orwin's (1983)⁶ formula for standardized effect sizes, estimating the stability of single-case effect results.

In the meta-analysis by Wahman, Pustejovsky, Ostrosky, & Santos (2019), single-case data were coded based on descriptive characteristics, research design strength, and experimental control strength. Participant demographics (age, disability, gender, race/ethnicity) were coded as reported in the studies. Study design characteristics such as data type, targeted skills or behaviours, presence of maintenance/generalization phases, and procedural fidelity ($\geq 33\%$ of sessions, with $\geq 80\%$ accuracy) were also coded.

WWC (What Works Clearinghouse) standards by Kratochwill et al. (2013)⁷ were used to assess design rigour (internal validity)

5. 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.

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

7. Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34, 26–38.

and strength of experimental evidence (visual analysis). Only single-case data meeting WWC design standards –with or without reservations– were included. Average inter-rater reliability across quality indicators and studies was 93% (range 74%–100%). Both coders reviewed discrepancies and reached consensus. Effect size metrics were computed using *WebPlotDigitizer* (Rohatgi, 2018),⁸ known for producing highly reliable data. Extracted data were organized in Excel. A multilevel meta-analysis model (Pustejovsky, 2018)⁹ was used to summarize effect sizes across cases and studies, incorporating random effects for studies and nested cases. In addition to overall mean effect estimates, meta-regression analyses examined how participant and study design characteristics influenced effect size variability. All analyses were conducted in R (version 3.5.1; R Core Team, 2018),¹⁰ using the *SingleCaseES* package for effect size computation (Pustejovsky & Swan, 2018),¹¹ *metafor* for meta-analysis (Viechtbauer, 2010),¹² and *clubSandwich* for variance estimation (Pustejovsky, 2017).¹³

Results

An important stage is the correct presentation, evaluation, and interpretation of the results, which lead to the final conclusion. At this stage, the strength of the results, their practical application, and their generalizability must be assessed. Through this process, the necessity for conducting new studies to strengthen the findings also emerges.

The results of the studies by Wang et al. (2009), Kokina et al. (2010), Test et al. (2011), and Bozkurt et al. (2014) confirmed previous conclusions regarding the questionable effectiveness of social story interventions for students with autism. Most authors agree that social stories are a promising intervention; however, according to meta-analytic findings, they demonstrate questionable effectiveness. They should not be considered an evidence-based practice for teaching social skills to individuals with autism. Nevertheless, they appear to be a promising practice that requires further research (Wang et al., 2009, p. 338).

The intervention using social stories appeared to be much more effective when used to target the reduction of problematic behaviours rather than to teach appropriate social skills. A possible explanation is that social behaviours, situations, and concepts are more abstract and complex and, therefore, more difficult for students to comprehend. According to the “complexity hypothesis”, social story interventions that describe simple, discrete behaviours, rather than complex social or nonsocial routines, had higher effectiveness (Kokina et al., 2010, p. 822).

The review of the studies showed extreme variability among individual results, which was expressed in a wide range of PND scores. In the study by Kokina et al. (2010), the overall mean PND score was 60% (range 11–100%), which was below the PND threshold score of 70, proposed for effective interventions (Scruggs & Mastropieri, 1998).

The overall results for PND and IRD measures in the study by Reynhout & Carter (2011) indicated that the social story intervention was slightly effective. This result was consistent with

8. Rohatgi, A. (2018). *WebPlotDigitizer* (version 4.1). Retrieved from <https://automeris.io/Web-PlotDigitizer>.

9. Pustejovsky, J. E. (2018). Using response ratios for metaanalyzing single-case designs with behavioral outcomes. *Journal of School Psychology*, 68, 99–112.

10. R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.r-project.org/>

11. Pustejovsky, J. E., & Swan, D. M. (2018). *SingleCaseES*: A calculator for single-case effect sizes. Retrieved from <https://github.com/jepusto/SingleCaseES>

12. Viechtbauer, W. (2010). Conducting meta-analyses in R with the *metafor* package. *Journal of Statistical Software*, 36(3), 1–48.

13. Pustejovsky, J. E. (2017). *clubSandwich*: Cluster-robust (sandwich) variance estimators with small-sample corrections. Retrieved from <https://cran.r-project.org/package=clubSandwich>

other previously published social story meta-analyses, although based on much smaller samples, such as those by Kokina & Kern (2010), Reynhout & Carter (2006), and Test et al. (2010). According to the PND data, the overall IRD data were just above the lowest levels that could be considered meaningful.

Test, Richter, Knight, and Spooner (2011), through their research, found that social stories could not be considered as evidence-based practices (EBP). Despite many critiques warning practitioners about the limitations of social stories as EBPs, the reality is that social stories remain a common practice used by special education teachers. On the other hand, since social stories are not yet considered an evidence-based practice –as measured by the quality indicators of Horner et al. (ibid.)– there is concern about whether educators should use social stories.

In the study by Bozkurt & Vuran (2014), 18.75% of the studies examined the effectiveness of social stories as a stand-alone intervention, 65.62% examined the effectiveness of social stories with additional interventions, and 15.62% were comparative studies. In the descriptive studies where social stories were presented alone, most studies showed that they were effective in teaching social skills to children with autism. The mean PND for the effectiveness of social stories presented alone was 57.46%, with a range of 19.02% to 87.82%, whereas the mean PND for social stories combined with additional interventions was 65.66%, with a range of 0% to 100% across all studies. According to this study, it is an important finding that social stories alone are not as effective as social stories with additional intervention. At this point, the greater effectiveness of social stories with supplementary interventions may be attributed to the added intervention strategies.

The results of the meta-analysis by McGill, Baker, and Busse (2015) showed that social story interventions for children and adolescents diagnosed with autism were associated with small to large reductions in targeted problem behaviours, with the outcomes varying depending on intervention characteristics, the implementer of the intervention, the behaviour of the participant, the duration of the treatment, and the type of publication. In line with previous studies, some evidence was found for greater effects when individual study variables were considered. However, the current meta-analysis indicates that the effectiveness of social story interventions is limited and that they should be used cautiously as a treatment for addressing problem behaviours in children and adolescents with autism.

The results of Wahman, Pustejovsky, Ostrosky, and Santos (2019) showed variability in methodological rigour and in the effectiveness of using social stories both as a stand-alone intervention and in combination with other intervention approaches. Thus, social stories cannot be considered an evidence-based practice (EBP) for young children. Although several studies, such as Zimmerman & Ledford (2017), suggest avoiding the use of social stories as an intervention, other reviews, such as those by Reynhout & Carter (2011) and Test et al. (2011), indicate that social story interventions are slightly effective. The findings from

14. Zimmerman, K. N., & Ledford, J. R. (2017). Beyond ASD: Evidence for the effectiveness of social narratives. *Journal of Early Intervention*, 39, 199–217.

this study highlight the need for additional research to improve our understanding of the effectiveness of social story interventions, whether implemented alone or in combination with other interventions.

Conclusions

From the interpretation of the results, we can determine whether the main objectives of the studies examined were achieved and whether the key research questions were answered. In relation to the topic under investigation, these questions concerned, in short, the overall effectiveness of social stories and their contribution to behaviour change in individuals with autism or social behaviour difficulties.

In conclusion, it was found that social story interventions demonstrated small to moderate effectiveness and that their effectiveness varied depending on the study or the participants' characteristics. Furthermore, it emerged that social stories cannot be characterized as an evidence-based practice (EBP). Initial hypotheses suggesting that the effects on reducing challenging behaviours and improving pro-social skills in children are small to moderate were also confirmed.

All in all, although research findings on the effectiveness of social stories as interventions are not particularly encouraging, they are still considered by many educators to be an effective and feasible intervention (Smith, 2001, p. 344), as their ease of implementation makes them an accessible intervention option for teachers and parents who wish to improve the social behaviour of children with autism.

Social stories remain a practical intervention due to their adaptability and ease of use in educational settings. They are most effective when tailored to the individual's cognitive and communicative needs and are combined with other supports such as visual aids or peer modelling to enhance understanding and generalization of social skills. Educators should consider simplifying language and focusing on tangible behaviours for better outcomes, consistent with the complexity hypothesis. Professional development on writing and implementing social stories, alongside ongoing data-driven adaptation, is essential to improve fidelity and effectiveness. Despite modest effects, social stories offer accessible and feasible support, especially when integrated thoughtfully into broader intervention plans (Schreibman et al., 2015).

There remains a clear need for more rigorous randomized controlled trials examining the effectiveness of social stories across diverse populations and settings. Cross-cultural studies are important for validating intervention generalizability globally. Further exploration into technology-enhanced social stories, including digital and interactive formats, could improve accessibility, engagement, and outcomes. Researchers should also focus on identifying which components and characteristics of social stories maximize benefits to better inform educational practices.

References

- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.
- Bozkur, S., & Vuran, S. (2014). An Analysis of the Use of Social Stories in Teaching Social Skills to Children with Autism Spectrum Disorders. *Educational Sciences: Theory & Practice*, 14(5), 1875–1892.
- Chamberlain, B., Kasari, C., & Rotheram-Fuller, E. (2007). Involvement or isolation? The social networks of children with autism in regular classrooms. *Journal of autism and developmental disorders*, 37(2), 230–242.
- Grandin, T. (1988). Teaching tips from a recovered autistic. *Focus on Autistic Behavior*, 3, 1–8.
- Kokina, A., & Kern, L. (2010). Social Story™ interventions for students with autism spectrum disorders: A meta-analysis. *Journal of autism and developmental disorders*, 40(7), 812–826.
- Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Shaw, K. A. (2023, March 24). *Prevalence and characteristics of autism spectrum disorder among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. Morbidity and Mortality Weekly Report. Surveillance Summaries*, 72(2), 1–14.
- 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*, 31(1), 21–42.
- National Research Council, (2001). *Educating children with autism*. Washington: National Academy Press.
- Ntais, C., & Talias, M. A. (2024). Unveiling the value of meta-analysis in disease prevention and control: A comprehensive review. *Medicina*, 60(1629).
- Patelarou, E., & Brokalaki, H. (2010). *Methodology of systematic review and meta-analysis* [Μεθοδολογία της συστηματικής ανασκόπησης και μετα-ανάλυσης]. *Nosileftiki*, 49(2), 122–130.
- Reynhout, G., & Carter, M. (2011). Evaluation of the efficacy of Social Stories™ using three single subject metrics. *Research in Autism Spectrum Disorders*, 5(2), 885–900.
- Rodrigues da Silva, D., & Mwangi, G. (2025). *Exploring the effects of Social Stories intervention on communication skills and anxiety reduction in children with autism*. *PsychNexus Journal*, 3(3).
- 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(6), 535–543.
- Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., ... Halladay, A. (2015). *Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder*. *Journal of Autism and Developmental Disorders*, 45(8), 2411–2428.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special education*, 8(2), 24–33.

- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior modification*, 22(3), 221–242.
- Smith, C. (2001). Using social stories to enhance behaviour in children with autistic spectrum difficulties. *Educational psychology in Practice*, 17(4), 337–345.
- 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(1), 49–62.
- Wahman, C. L., Pustejovsky, J. E., Ostrosky, M. M., & Santos, R. M. (2019). Examining the Effects of Social Stories™ on Challenging Behavior and Prosocial Skills in Young Children: A Systematic Review and Meta-Analysis. *Topics in Early Childhood Special Education*, 1–13.
- Wang, P., & Spillane, A. (2009). Evidence-based social skills interventions for children with autism: A meta-analysis. *Education and Training in Developmental Disabilities*, 318–342.
- Weiss, M. J., & Harris, S. L. (2001). Teaching social skills to people with autism. *Behavior modification*, 25(5), 785–802.
- Yu, Y., Ozonoff, S., & Miller, M. (2023). Assessment of autism spectrum disorder. *Assessment*, 31(1), 3–20.