Ankara University Faculty of Educational Sciences Journal of Special Education

Advance Online Publication

SYSTEMATIC REVIEW

Recieved Date: 08.07.20
Accepted Date: 25.03.21
OnlineFirst: 14.04.21

Is Social Story™ an Evidence-Based Practice? A Meta-Analysis and Comprehensive Descriptive Analysis Study*

Seray Olçay[1] **Üzeyir Emre Kıyak**[2] **Özlem Toper**[3]

Abstract

Introduction: Debates about the evidence-base for social stories™ (SSs) have been ongoing and meta-analysis studies show that SSs are not an evidence-based practice.

Method: We conducted a descriptive and meta-analysis research by evaluating studies in which SSs were used both to increase appropriate behaviors and decrease inappropriate behaviors published until June 2020. We reviewed seven studies, which used single-case designs and published in internationally peer-reviewed journals and also met criteria for quality indicators for single-subject designs. We calculated the effect sizes of these seven studies by using the improvement rate difference (IRD).

Findings: The results of our study show that the effect sizes of the included studies are large and SSs meet 5-3-20 rule by scarcely. The findings of our study also demonstrate that dependent variables of research examined the effectiveness of social story were increasing appropriate behaviors. Moreover, the overwhelming majority of the participants of the research were primary school age children and had autism spectrum disorders diagnosis.

Discussion: Our study demonstrate that social story is an evidence-based practice. However, future research is required to examine the effectiveness of social story on individuals with different developmental disabilites, in different age groups. Besides, there is still a need for qualified research to strengthen the evidence-base of SSs.

Keywords: Evidence-based practices, social stories, systematic review, meta-analysis, single-subject research design.

To cite: Olçay, S., Kıyak, Ü. E., & Toper, Ö. (2021). Is Social Story™ an evidence-based practice? a meta-analysis and comprehensive descriptive analysis study. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi, Erken Görünüm.* https://doi.org/10.21565/ozelegitimdergisi.766765

*This study was presented at the 29th National Special Education Conference in Kuşadası, Aydın, Turkey.

[1]Assoc. Prof., Hacettepe University, E-mail: solcaygul@hacettepe.edu.tr, https://orcid.org/0000-0002-5007-7466

[2]***Corresponding Author:*** Ph. D., Bursa Uludag University, E-mail: uzeyiremre@gmail.com, https://orcid.org/0000-0002-2681-9407

[3]Assist. Prof., Bursa Uludag University, E-mail: otoper@gmail.com, https://orcid.org/0000-0002-0243-4844

Introduction

Social Story (SS) was first developed by Carol A. Gray in 1991 for the purpose of telling about new routines and changes, explaining the behaviors of other people, and teaching academic and socials skills to individuals with Autism Spectrum Disorders (ASD) (Gray & Garand, 1993). As a technique, it has become widespread overtime for teaching different skills such as self-care skills and safety skills for children with different diagnoses including intellectual disabilities and learning disorders (Hagiwara & Smith-Myles, 1999; Kurt & Kutlu, 2019). Many studies have shown that SSs have positive effects on decreasing problem behaviors as well as on teaching new skills such as academic, social, and safety skills (e.g., Agosta et al., 2004; Scattone, et al., 2002). While SSs have been a sole intervention as an independent variable in some studies (e.g., Sansosti & Powell-Smith, 2006; Scattone et al., 2006), other researchers examined the effectiveness of SSs combined with various techniques and methods such as video modeling, response cost, and prompting (e.g., Bernad-Ripoll, 2007).

SSs, regarded as easy and practicable by different individuals such as teachers, families, peers and support staff (Test et al., 2011), were mentioned as an evidence-based practice (EBP) in a National Standards Report published by the National Autism Center (NAC) in 2009 and reports published by the National Professional Development Center (NPDC) in 2014 and Steinbrenner et al. (2020). On the other hand, descriptive and meta-analysis studies related to SSs have questioned whether SSs are an EBP or not (Ali & Frederickson, 2006; Kuoch & Mirenda, 2003; Nichols et al., 2005; Sansosti et al., 2004). In a descriptive analysis of SSs, Kuoch and Mirenda (2003) analyzed 10 SSs research articles published up to 2002 in which SSs were used, including case studies. They reported that SSs were used together with other techniques in seven of the 10 studies; five of them did not meet the criteria specified by Carol A. Gray and six were limited by case studies or an AB design. Similarly, descriptive studies conducted between 1994-2015 emphasize that SSs research has limitations in terms of methodological characteristics, primarily experimental control (Ali & Frederickson, 2006; Leaf et al., 2015; Nichols et al., 2005; Sansosti et al., 2004), and there is a need for more research, especially research which is rigorously designed to be able to count SS intervention as an EBP (Ali & Frederickson, 2006; Leaf et al., 2015). The findings of the meta-analysis studies on the effectiveness of SSs are that SSs are in fact lacking an evidence-base (EB); SS can be shown among promising practices, but more research is needed to establish an EB (Kokina & Kern, 2010; Leaf et al., 2015; Reynhout & Carter, 2006; Sani-Bozkurt & Vuran, 2014; Test et al., 2011). These studies emphasized that SSs are devoid of EB and qualitative indicators are regarded as a criterion only in three meta-analysis studies, which examined whether they are met for research findings, which have been analyzed for meta-analysis (Leaf et al., 2015; Qi et al., 2018; Test et al., 2011). Test et al. (2011) used the quality indicators suggested by Horner et al. (2005), Leaf et al. (2015) and Qi et al. (2018) used the quality indicators suggested by Kratochwill et al. (2013) for assessing the quality of studies conducted with single-subject research designs as inclusion criteria for meta-analysis. In the other three meta-analysis studies (Kokina & Kern, 2010; Reynhout & Carter, 2006; Sani-Bozkurt & Vuran, 2014), it can be seen that the studies included in the analysis were not only studies in which experimental control was established, but were also single-subject studies and/or case studies in which experimental control was weak and which had limitations in terms of qualitative indicators were examined, so it is not possible to discuss the EB of such studies. Therefore, studies which meet the quality indicators for meta-analysis, which makes it possible to discuss an EB of a practice. Horner et al.'s (2005) study was the first to use quality indicators for single-subject research in 2005; it was followed by Reichow et al. (2008) and Kratochwill et al. (2013). Although the common ground is provided in terms of qualitative indicators by groups of researchers, some differences can also be observed in these studies. One of these differences concerns the collection of treatment fidelity data in the research. Treatment fidelity takes place in the quality indicators determined by Horner et al. (2005) and Reichow et al. (2008). However, efforts to establish EB are based on improving reliability and credibility by increasing reproducibility, transparency, and openness in the exposition of the entire research process (Cook et al., 2018). The collection of treatment fidelity data is therefore an important indicator. Reynhout and Carter (2006), who stated that no treatment fidelity was reported in 12 of the 16 studies which they examined, expressed this as a limitation in their study. It is therefore evident that indicators related to treatment fidelity should be added to the indicators determined by Kratochwill et al. (2013).

Regarding to publication years of the meta-analysis studies of SS, it is seen that studies conducted up to 2011 were included in the analyses, but approximately 20 SS studies have been carried out over the following eight years (e.g., Kim et al., 2014). Because of the need to consider studies to establish an EB of SSs in descriptive analysis and meta-analysis studies, it is thought that the inclusion of studies conducted between 2012 and 2020 in the analysis is an important part of deciding whether SS is an EBP or not. Additionally, considering the updates of the quality indicators proposed by Kratochwill et al. (2013) and SSs writing criteria by Gray (2010, 2015) are

considered, it can be said that analysis of SSs studies carried out particularly in the years after 2010 is necessary for both experimental and methodological rigor. As to effect size calculation techniques in the analysis of SS are examined, it is seen that the percentage of non-overlapping data (PND) technique was used in all studies. There is no consensus on the most appropriate non-parametric technique for analyzing single-subject research designs in the relevant literature (Parker et al., 2011; Rakap, 2015; Tekin-İftar et al., 2018), but PND is often preferred because it is a more highly consistent technique than other techniques and is easy to calculate. Meta-analysis studies using different non-parametric techniques are needed to calculate the effect sizes of SSs interventions in the literature. Based on all these points and the limitations of the previous meta-analysis studies, this current study is conducted to provide a comprehensive descriptive and meta-analysis of the SSs studies.

For this purpose, we (a) analyzed studies which met the inclusion criteria by refining indicators related to treatment fidelity on those set out by Kratochwill et al. (2013), (b) conducted a comprehensive descriptive analysis of the studies, which met the quality indicators in terms of demographic, methodological, and outcome characteristics, (c) calculated the effect sizes of those studies by using the improvement rate difference (IRD), and (d) assessed studies to determine whether or not SSs intervention met the 5-3-20 criterion determined for EBP. It is hoped that the findings of this study will provide an important response to the question of whether there is an EB for SS, and will contribute to the points which have to be taken into consideration in SSs research.

Method

Preparation Process

Prior to the literature search, we met to decide how to collect data reliably and authentically and devised the coding tables to be used for the analysis. We created five tables to use on (a) search procedures, (b) evaluation procedures of "Single-Case Intervention Research Design Standards", (c) procedures for conducting a comprehensive descriptive analysis, (d) use of the software program UnGraph5, and (e) the intervention effect calculation phases. We reviewed each item in the tables and reached an agreement on how to code the tables. Two of us then independently coded three randomly selected studies. The reliability was calculated by using the formula of "number of agreements of between coders / (number of agreements + number of disagreements between coders) x 100" formula. We continued this process until our agreement reached 100% and then we began to collect and code data.

Search Procedures

We aimed to identify all the studies related to SS, so we did not set a start date and we attempted to identify all the studies published up to December 2018. We also wanted to evaluate the effectiveness of SS on teaching skills and on decreasing the disruptive behaviors of individuals with different diagnoses, so we located studies using *Academic Search Complete*, *Article First*, *EBSCOhost*, *JSTOR*, *PsycINFO*, *ScienceDirect*, *Worldcat.org* and *Web of Science* using the keywords *social stories, social skills, problem behaviors*, *intervention, developmental disabilities*, *autism, autism spectrum disorder, intellectual disability, mental retardation, developmental delay,* and *learning disability*. Additionally, we reviewed the reference lists of the identified research and meta-analysis studies which had examined the effectiveness of SSs to identify articles that we had not identified on a prior search, but this did not add any unidentified studies. As a result of this search procedure, 92 studies were identified; 9 descriptive studies, 3 case studies, 16 experimental studies, 7 review studies, 4 meta-analysis studies, and 53 single-subject studies. We evaluated the identified studies according to the inclusion and exclusion criteria using the coding table.

We included in the review studies which met the following criteria; (a) published in English and Turkish in national and international peer-reviewed journals, (b) had participants who had at least a diagnosis, (c) examined the effectiveness of SSs on teaching different skills or decreasing disruptive behaviors, and (d) used the single-case research method. We excluded studies which (a) used other research methods (e.g., narrative research and group experimental research), (b) used SSs combined with other procedures, (c) compared the effectiveness of different procedures, and (d) did not present graphed data for visual analysis. 39 studies that did not use a single-subject research methodology were eliminated because they did not meet the inclusion criteria. Since the full text of one of the 53 studies conducted with a single-subject research methodology (Hanley-Hochdorfer et al., 2010) could not be reached, it was not taken into further analysis in this study. Of 52 studies, 20 were not included in the further analysis because fifteen had used SSs in conjunction with other interventions (e.g., Litras et al., 2010), four had compared SSs with different interventions (e.g., Acar et al., 2017), one (Burke et al., 2004) had been conducted

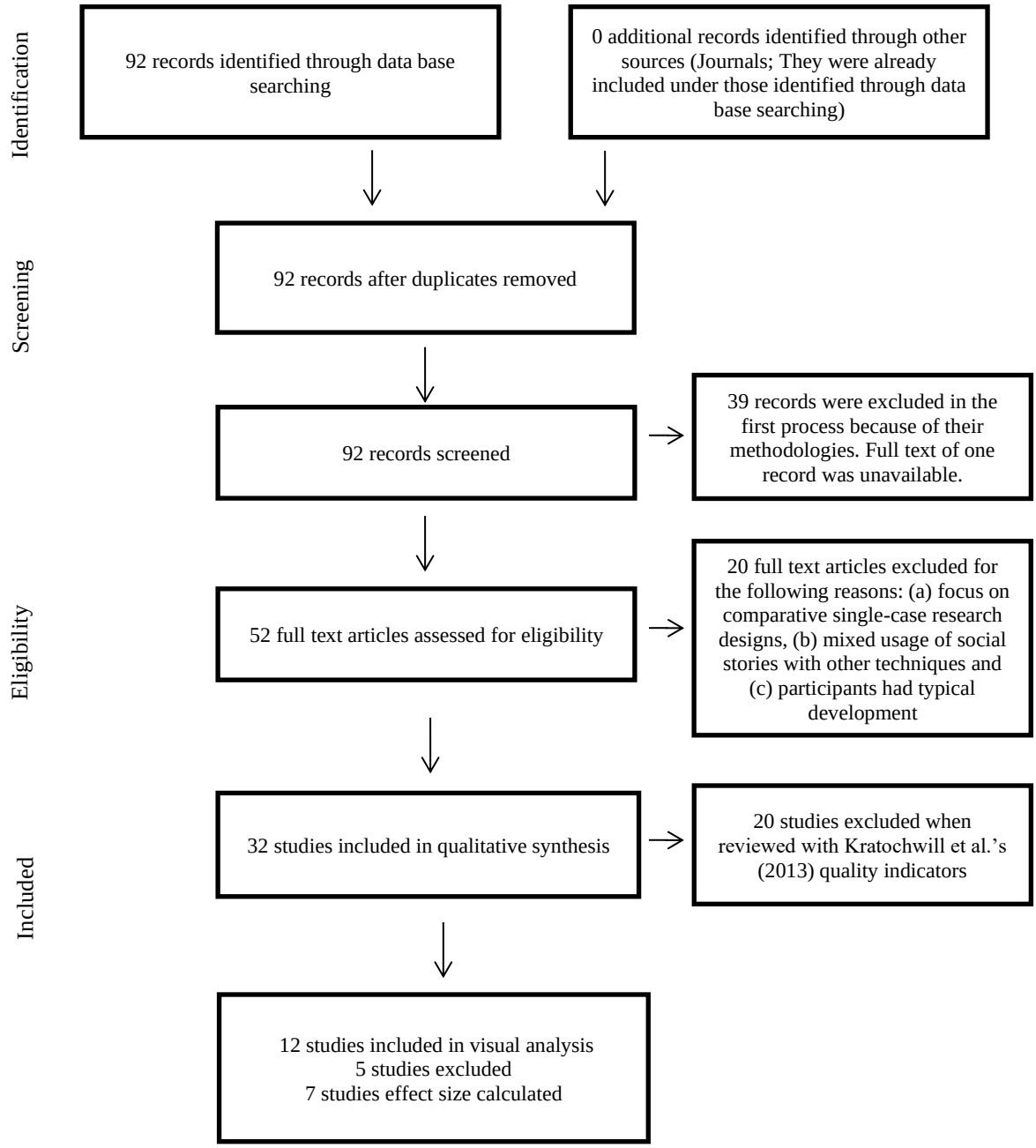
with participants without disabilities. The remaining 32 studies were included in further analysis. The whole process is shown in Figure 1.

Evaluation Procedures

We evaluated the quality indicators and demographic parameters of the selected studies which met the inclusion criteria in this phase.

Figure 1

Procedures Followed During the Search



Evaluating Studies by Single-Case Intervention Research Design Standards

We first evaluated the studies according to the inclusion and exclusion criteria for further analysis. We used the quality indicators recommended by Kratochwill et al. (2013) with the addition of two further criteria, 'collection of procedural fidelity data' and 'procedural fidelity index' located in other rubrics (Odom et al., 2005; Reichow et al., 2008). We used a final form of the coding table which comprised the quality indicators and the additional criteria. There were 10 criteria: (a) systematic manipulation of the independent variable(s), (b) collection of inter-observer data for at least 20% of all sessions, (c) inter-observer agreement on at least 80% of all sessions, (d) collection of procedural fidelity data for at least 20% of all sessions, (e) procedural fidelity of at least 80% of all sessions, (f) at least three demonstrations of effect, (g) at least five data points per condition, (h) at least three data points per condition, (i) clarification of design standards, and (j) clarification of evidence for effectiveness in the coding table. We coded the criteria (a) to (i) as 'yes' (Y) or 'no' (N) and we used trio decoding for the (i) criterion, or more clearly 'clarification of design standards.' We coded those studies which met all of the (a) to (h) criteria as 'meet standards' (MS) and those which met the (a) to (h) criteria but not the (g) criterion as 'meet standards with reservation' (MS-R). Studies that did not meet any of the (a) to (h) criteria except for (g) were coded as 'do not meet standards' (n-MS). We coded the (j) criterion ('clarification of evidence for effectiveness') based on the results of visual analysis of the studies coded as MS and MS-R. We analyzed six features of the graphics of the studies in terms of their *level, trend, variability, the immediacy of effect*, *overlap,* and *consistency across similar phases* (Kratochwill et al., 2013).

Studies which met all of the characteristics listed above were coded as 'strong effect' (evidence); studies which presented at least three demonstration of effect but did not present in one situation as 'moderate effect' and other studies were coded as 'no effect' (no evidence). We calculated the intervention effect of studies coded as strong effect and moderate effect after visual analyses.

Conducting the Comprehensive Descriptive Analysis

We conducted comprehensive descriptive analysis of the studies coded as MS or MS-R using the single-case intervention research design standards. We evaluated each study in terms of five demographic characteristics: (a) characteristics of participants (e.g., number, age, gender, and so on), (b) target behaviors (skills taught or decreased behaviors), (c) measurement of behaviors, (d) the criterion for target behaviors, and (e) instructional setting, and in terms of eight methodological characteristics and results: (a) research design, (b) characteristics of SS, (c) instruction, (d) implementer, (e) social validity, (f) maintenance, (g) generalization, and (h) general effect. We used the descriptive analysis coding table developed by us for this purpose.

Intervention Effect Calculations

We calculated the intervention effect for studies coded as MS or MS-R and 'strong effect' or 'moderate effect' using the single-case intervention research design standards. We used the "improvement rate difference (IRD)" technique for calculating the intervention effect, which distinguishes our study from other meta-analyses conducted on studies of the effectiveness of SSs. IRD is the feed ratio between baseline and intervention. IRD has the following advantages: (a) it can easily be calculated manually, (b) it corresponds to visual analysis, and (c) it has a confidence interval because of the existence of a known sampling distribution (Parker et al., 2009). We calculated the intervention effect using the calculation tool available in the link http://www.singlecaseresearch.org (Vannest et al., 2011). We considered the intervention effect as 'small' when the effect size scored under 50%, as 'moderate' when the effect size scored between 50% and 70%, and as 'large' when effect size scored 70% or above (Parker et al., 2009). We calculated the intervention effects separately for each tier in the study. For this purpose, we converted the data in each tier into digitized data using the UnGraph5 software which is effective for digitalizing graphical data. One of us exported the digitized data in each tier into a Microsoft Excel file for further analysis (Tekin-İftar et al., 2018).

Determination of Evidence Base for Using SSs

We used the three criteria known as the 5-3-20 rule (Kratochwill et al., 2013) as the base for evaluating whether SSs interventions were EB or not. These criteria were determined as (a) five studies classified as 'strong' or 'moderate' effect and MS or MS-R, (b) practice conducted by at least three groups of researchers with no overlapping authorship from three different geographic regions, and (c) the total number of participants in all the included studied equaled at least to 20.

Reliability

We made reliability calculations for each phase in the study: (a) inclusion and exclusion procedures, (b) procedures for evaluating studies by single-case intervention research design standards, (c) procedures for conducting comprehensive descriptive analysis, (d) digitalizing data using UnGraph5, and (e) IRD calculations. For this purpose, two of us coded all the studies independently. As already explained, consistency between the coders was achieved by using the formula 'number of agreements of between coders / (number of agreements + number of disagreements between coders) x 100' (Tekin-İftar et al., 2018). We set an acceptable criterion for reliability between coders as at least 80%. For the inclusion and exclusion criteria, the inter-coder reliability was set at 95.74% (range = 85.71-100); reliability for the descriptive analysis of the single-subject interventions model standards was 96.10% (range = 90.9-100); the reliability of the descriptive analysis was 96.10% (range = 90.9-100), the reliability of the digitized data by UnGraph 5 was 99.52% (range = 97.2-100) and the reliability of the IRD calculations was 100%.

Results

Quality Indicators of Single-Case Studies

As already described and as shown in Figure 1, we found 32 studies that met the inclusion criteria for our meta-analysis. As can be seen in Table 1, of these 32 studies, 20 (62.5%) did not meet the criteria for quality indicators. Of these 20 studies, we coded one as nMS because the inter-observer reliability ratio was under 80% (Dodd et al., 2008), procedural fidelity data were not collected or were under 80% in eight studies (e.g., Khantreejitranon, 2018; Moudry-Quilty, 2007), three demonstrations of experimental effect could not be shown in two studies (e.g., Hung & Smith, 2011) and eight studies did not meet more than one criterion. We did not include these 20 studies in the visual analysis. Of the 12 studies which met the criteria for quality indicators, we coded one study (3.57%) as MS and 11(28.57%) as MS-R and included them in the visual analysis. Of these 12 studies, we coded seven (58.33%) which met all the criteria determined for visual analysis as 'strong effect' (evidence), and the remaining five (41.67%) as 'no effect'. We calculated effect size for the seven studies coded as 'strong effect'.

Descriptive Analysis of SSs Studies

Demographic Features of SSs Studies

Numbers and Ages of Participants. As can be seen in Table 2, each of the 12 studies analyzed had been conducted with three participants, so the total number of participants was 36. The ages of these 36 participants ranged from 2 to 17, and we examined three age categories: pre-school, aged between birth and 5; primary school, aged between 6 and 11; and secondary school, aged 12 or over. Five participants (13.8%) in three studies (Boşnak & Turan, 2020; Kurt & Kutlu, 2019; Thompson & Johnston, 2013) were of pre-school age, 19 (52.7 %) were primary school age (e.g., Delano & Snell, 2006; Özdemir, 2008b), and 12 (33.3%) were secondary school age (e.g., Graetz et al., 2009).

Gender and Disability. Of the 36 participants, three were female (8.4%) and 33 were male (91.6%). Of the 12 studies, participants were diagnosed with ASD (N = 26, 72.2%) in nine studies (e.g., Olçay-Gül &Tekin-İftar, 2016; Özdemir, 2008b;), with Asperger's syndrome in two studies (Sansosti & Powell-Smith, 2006; Scattone et al., 2006), with severe intellectual disability in only one study (Kim et al., 2014) and with mild intellectual disability in one study (Bıçakçı & Olçay-Gül, 2019).

Target Behaviors. The dependent variables of the 12 studies could be separated into two groups, increasing appropriate behaviors and decreasing inappropriate behaviors. The dependent variables of nine studies (75%) were increasing appropriate behaviors (e.g., Delano & Snell, 2006; Sansosti & Powell-Smith, 2006), and two studies (16.6%) were decreasing inappropriate behaviors (Özdemir, 2008b; Scattone et al., 2002). The dependent variables of the one remaining study (%8.4) were both decreasing the inappropriate behaviors of the participants and increasing their academic engagement level (Kim et al. 2014)

When the dependent variables of studies were examined separately, it could be seen that especially increasing social interaction and social skills within the context of increasing appropriate behaviors were targeted. These target behaviors were sharing toys with peers (Boşnak & Turhan, 2020); social interaction with peers, seeking attention, initiating comments, initiating requests and making contingent responses (Delano & Snell, 2006); following direction independently within 15 seconds when given verbal direction ('stand up' or 'get up'), using appropriate vocal pitch when talking with peers or adults, keeping the hands down and keeping materials away from the lips when seated in the classroom (Graetz et al., 2009); greeting skills, expressing happiness or needs appropriately in response to others' behaviors or comments and asking permission to have snacks or various items (Olçay-Gül & Tekin-İftar, 2016); teaching kind behaviors, holding conversations and engaging in group activities (Sansosti & Powell-Smith, 2006); staying seated during circle time and snack-time activities, engaging in tactile play activities with no more than one initial verbal prompt (Thompson & Johnston, 2013) were determined as target behaviors. Besides, target behaviors were verbal, physical, or gestural initiation or responsing to a peer, commenting, or asking questions related to the activity or conversation; continued engagement in the same activity as the peer; a response to a peer's comment or question with a comment related to the conversation; an initiated comment or question related to the conversation; and a physical gesture such as nodding to indicate approval or disagreement in one study (Scattone et al., 2006). Additionally, safety skills were dependent variables in two studies as crossing skills (Bıçakçı & Olçay-Gül, 2019) and abduction-prevention skills (Kurt & Kutlu, 2019).

Table 1

Coding Studies According to Kratochwill et al. (2013) Criteria

Studies	Systematic man. of IV	IOA for 20% of sessions	IOA at or above 80 %	PF for 20% of sessions	PF at or above 80%	At least three demons. of effect	Five data points per condition	Three data points per condition	Classification of standards	Classification of evidence of effectiveness
Adams et al. (2004)	Y	Y	Y	N	N	Y	Y	Y	nMS	-
Almutlag & Martella (2018)	Y	N	N	Y	Y	N	Y	Y	nMS	-
Beh-Pajooh et al. (2011)	Y	Y	Y	N	N	N	Y	Y	nMS	-
Bıçakçı & Olçay-Gül (2019)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Bledsoe et al. (2003)	Y	Y	Y	N	N	N	Y	Y	nMS	-
Boşnak & Turhan (2020)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	No evidence
Crozier & Tincani (2007)	Y	Y	Y	Y	Y	N	N	N	nMS	-
Delano & Snell (2006)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	No evidence
Dodd et al. (2007)	Y	Y	N	Y	Y	Y	N	Y	nMS	-
Graetz et al. (2009)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Hagiwara & Myles (1999)	Y	Y	Y	N	N	Y	Y	Y	nMS	-
Hung & Smith (2011)	Y	Y	Y	Y	N	N	Y	Y	nMS	-
Ivey et al. (2004)	Y	Y	Y	Y	N	Y	N	Y	nMS	-
Karayazi et al. (2014)	Y	Y	Y	Y	N	N	N	Y	nMS	-
Khantreejitranon (2018)	Y	Y	Y	N	N	Y	Y	Y	nMS	-
Kim et al. (2014)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Klett & Turan (2012)	Y	N	N	N	N	Y	N	Y	nMS	-
Kurt & Kutlu (2019)	Y	Y	Y	N	N	Y	Y	Y	nMS	-
Kuttler et al. (1998)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Lorimer et al. (2002)	Y	Y	Y	N	N	N	N	Y	nMS	-
Moudry-Quilty (2007)	Y	Y	Y	Y	N	Y	Y	Y	nMS	-
Norris & Dattilo (1998)	Y	Y	Y	Y	Y	N	Y	Y	nMS	-
Olçay-Gül & Tekin-İftar (2016)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Özdemir (2008a)	Y	Y	Y	N	N	Y	N	Y	nMS	-
Özdemir (2008b)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Reichow & Sabornie (2009)	Y	Y	Y	N	N	N	N	Y	nMS	-
Sansosti & Powell-Smith (2006)	Y	Y	Y	Y	Y	Y	Y	Y	MS	No evidence
Scattone et al. (2006)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	No evidence
Scattone et al. (2002)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Thompson & Johnston (2013)	Y	Y	Y	Y	Y	Y	N	Y	MS-R	Strong
Vandermeer et al. (2015)	Y	Y	Y	N	N	Y	N	Y	nMS	-
Wright & McCathren (2012)	Y	Y	Y	Y	N	Y	Y	Y	nMS	-

Note: Demonst. = demonstration; IOA = inter observer agreement; IV = independent variable; Man.= manipulation; MS = meeting standards; MS-R = meeting standards with reservations; N = no; nMS = not meeting standards; PF = procedural fidelity; Y = yes.

The dependent variables of studies in which the aim was to decrease inappropriate behaviors were off-task behaviors (such as leaving seat or classroom without permission, turning the head away from the teacher, looking at peers or engaging in motor movements such as nail-biting); disturbing others and making disruptive noises (Kim et al., 2014); using a loud voice, such as raising the voice above the typical level in class, chair tipping, cutting in the lunch line and disturbing other students nearby by pushing them (Özdemir, 2008b); chair tipping, shouting and staring at a female for three consecutive seconds or more (Scattone et al., 2002).

Measurement of Behaviors. Researchers collected data related to dependent variables by using continuous and sampling recording methods selected according to behavior characteristics. Partial interval recording was used in all of the studies to examine the effects of SSs on decreasing inappropriate behaviors (Kim et al., 2014, Özdemir, 2008b; Sansosti & Powell-Smith, 2006; Scattone et al., 2002). In studies which sought to increase social interaction, social skills, and appropriate behaviors, event and duration recording (Delano & Snell, 2006), controlled event recording (Boşnak & Turhan, 2020; Olçay-Gül & Tekin-İftar, 2016), momentary time sampling (Thompson & Johnston, 2013) and partial interval recording (Scattone et al., 2006) were used. Interval recording and duration recording were used together in one study (Graetz et al., 2009). Task analysis was used in two studies (Bıçakçı & Olçay-Gül, 2019; Kurt & Kutlu, 2019).

Criteria for Target Behaviors. Delano and Snell (2006) determined criteria for instruction sessions as (a) an increase in level, (b) an accelerated slope, (c) an increase in the median, and (d) little or no overlap with baseline data, and they started instruction sessions with the next participant when each participant had met these conditions. The criteria was performing target behaviors 100% at least three consecutive sessions in two studies (Bıçakçı & Olçay-Gül, 2019; Kurt & Kutlu, 2019). There were no data about teaching criteria in the remaining studies analyzed (e.g., Graetz et al., 2009).

Instructional Settings and Teaching Format. The settings were different areas in schools in all of the studies except for two. Olçay-Gül and Tekin-İftar (2016), conducted their study in the participants' homes. Kurt and Kutlu (2019) conducted instruction sessions in home for one of three participants. Training and probe trials in all of the studies were conducted in a 1:1 instructional arrangement.

Methodological Characteristics

Research Design. Studies predominantly used a single-case multiple baseline design across participants (N = 6; 50%; e.g., Graetz et al., 2009; Özdemir, 2008b). Three studies (25%) used a multiple probe design across participants (Boşnak & Turhan, 2020; Delano & Snell, 2006; Kim et al., 2014) and in two studies (16.6%) a multiple probe design across dyads was used (Bıçakçı & Olçay-Gül, 2019; Olçay-Gül & Tekin-İftar, 2016). One study used concurrent multiple baseline design across participants (Kurt & Kutlu, 2019).

Characteristics of SSs. We regarded SSs in two parts, the format and the design of the SS. As it is shown in Table 3, the hard-copy book format was used in nine studies (e.g., Graetz et al., 2009) and a tablet-assisted SSs format was used in three studies (Boşnak & Turhan, 2020; Kim et al., 2014; Kurt & Kutlu, 2019). SSs were written individually regarding the target behaviors for the participant. Symbolic pictures were used combined with texts in five studies (e.g., Delano & Snell, 2006). Four studies (e.g., Graetz et al., 2009) used participants' photographs in stories. Kurt and Kutlu (2019) used peer's modeled photographs. Delano and Snell (2006) used text-only SSs for one participant. There was no information about the characteristics of the SSs in the remaining three studies (e.g., Scattone et al., 2006).

There was no information about whether the researchers had used any criteria for designing SSs in three studies (Kurt & Kutlu, 2019; Olçay-Gül & Tekin-İftar, 2016; Sansosti & Powell-Smith, 2006), whereas the researchers in the other nine studies stated that they had used the criteria for designing the SSs developed by Gray. Gray and Garand's (1993) criteria were used in one study (Graetz et al., 2009) and Gray's (1998) criteria were used in four studies (Kim et al., 2014; Özdemir, 2008b; Scattone et al., 2006; Scattone et al., 2002). Bıçakçı and Olçay-Gül (2019) and Delano and Snell (2006) used Gray's (2000) criteria; Boşnak and Turhan (2020) used Gray's (2002) criteria and Thompson and Johnston (2013) used SSs 10.0 proposed by Gray (2004).

Instruction. We collected data about instructional procedures. This included the presentation of SS, frequency of reading SS, comprehension questions, whether feedback on participants' answers after comprehension questions was given or not, and the fading procedures of SSs under the title 'implementation of SS'.

Table 2

Demographic and Methodological Characteristics of Studies

Studies	Participant characteristics: Age, number, gender, label, cognitive level	Settings/ teaching format	Target behavior/ criteria for target behavior/ assessment procedures	Research model	Intervention: Presentation of social story, prompt, reinforcement, reinforcement schedule	Interventionist	Social validity	Main.	Gen.	General effect
Bıçakçı & Olçay-Gül (2019)	7-9, 3M, ASD, ID, 110[1],120[2]	School 1:1	IAB; %100	MPDAD	HC, MP, SOR, ND	Peers	Interview peers & participants	+	+	+++
Boşnak & Turhan (2020)	5-6, 2M 1F, ASD, ND	School 1:1	IAB, ND, FR	MPDAP	TA, ND, SOR, ND	SEP	OA, teachers	+	-	+++
Delano & Snell (2006)	6-9, 3 M, ASD, ND	School 1:1	IAB + DR-FR	MPDAP	HC, ND, ND, ND	Experimenter	SC	+	+	+++
Graetz et al. (2009)	12-13, 2M 1F, ASD, ND[1-3]- 42[2]	School 1:1	IAB, ND, IR-DR	MBDAP	HC, ND, ND, ND	SEP	OA-SEP, SET, SLT	+	+	+++
Kim et al. (2014)	17; 2M 1F; ID, 49[1]- 73[2]- 76[3]	School 1:1	DIB-IAB, ND, PIR	MPDAP	TA, MP, TS, ND	SET	OA, parents, SET, LSC	+	+	+++
Kurt & Kutlu (2019)	7-4-12, 3M, ASD, 81[1], 114[2], 90[3]	School and home	IAB, %100	NonMBDAP	TA, MP, SOR, ND	Experimenter	OA, parents and participants	+	-	+++
Olçay-Gül & Tekin-İftar (2016)	12-16, 3M, ASD, ND	Home 1:1	IAB, ND, CER	MPDAD	HC, VP, SOR, ND	Mothers and sister	SC, OA, parents	+	+	+++
Özdemir (2008b)	7-9; 3M, ASD, 62[1]-86[1].74[3]	School 1:1	DIB + PIR	MBDAP	HC, ND, ND, ND	Teachers	OA, teachers	+	ND	+++
Sansosti & Powell-Smith (2006)	9-11, 3M, AS, ND	School 1:1	IAB, ND, PIR	MBDAP	HC, ND, ND, ND	Caregivers	SC	+	ND	++-
Scattone et al. (2006)	8-13, 3M, AS, 40[1]-ND- 82[3]	School 1: 1	IAB, ND, PIR	MBDAP	HC, ND, ND, ND	SET, SEP	IRP, SET-SEP	ND	-ND	++-

Table 2 (*continued*)

Studies	Participant characteristics: Age, number, gender, label, cognitive level	Settings/teaching format	Target behavior/criteria for target behavior/assessment procedures	Research model	Intervention: Presentation of Social Story, prompt, reinforcement, reinforcement schedule	Interventionist	Social validity	Main.	Gen.	General effect
Scattone et al. (2002)	7-15, 3M, ASD, 67[1]-74[2]-95[3]	School 1:1	DIB, ND, PIR	MBDAP	HC, ND, ND, ND	SET, SEP	SVC SET-SEP	ND	ND	+++
Thompson & Johnston (2013)	3-5, 3M, ASD, ND	School 1:1	IAB + MTS	MBDAP	HC, ND, ND, ND	OT	OA	+	+	IAB: +++ SR: ++-

Note: ASD = autism spectrum disorder; AS = asperger syndrome; CER = controlled event recording; DIB = decreasing inappropriate behavior; DR = duration recording; FR = frequency recording; Gen = generalization; HC = hardcopy; IAB = increasing appropriate behavior; ID = intellectual disability; IRP = intervention rating profile; LSC = licensed school counselor; Main = maintenance; MBDAP = multiple baseline design across participants; MP = model prompting; MPDAD = multiple probe design across dyads; MTS = momentary time sampling; ND = no data; OA = objective assessment; OT = occupational therapist; PIR = partial interval recording; SC = social comparison; SEP = special education personnel; SET = special education teacher; SLP = speech and language therapist; SOR = social reinforcement; SR = self-regulation; TA= tablet assisted; TS = token system; VP = verbal prompt.

The researchers read SSs to participants in two study (Delano & Snell, 2006; Kurt & Kutlu, 2019), while family members read SSs to participants in two studies (Olçay-Gül & Tekin-İftar, 2016; Sansosti & Powell-Smith, 2006). It was stated that interventionists or staff read SSs in three studies (Boşnak & Turhan, 2020; Graetz et al., 2009; Thompson & Johnston, 2013). In four of the studies, instruction sessions were arranged for the participants who were literate read stories after instructors had read them first, whereas only instructors had read stories in all sessions for illiterate participants (Kim et al., 2014; Özdemir, 2008b; Scattone et al., 2002; Scattone et al., 2006). Peers read SSs only in one study (Bıçakçı & Olçay-Gül, 2019). There was no information about the frequency of reading SSs in five studies (e.g., Delano & Snell, 2006; Kim et al., 2014); SSs were read to each participant at least twice in one study (Graetz et al., 2009). Stories were read twice a day for each participant in two studies (Bıçakçı & Olçay-Gül, 2019; Sansosti & Powell-Smith, 2006) and once a day for each participant in another three studies (Boşnak & Turhan, 2020; Kurt & Kutlu, 2019; Scattone et al., 2006). In one study, stories were read twice a day for two participants, and once a day for one participant (Özdemir, 2008b). Although there was no information about the frequency of reading stories, it was indicated in one study that the participants had opportunities to access the SSs on off-time (Scattone et al., 2002).

Although interventionists asked comprehension questions about stories in most of the studies, only one study provided no information about this (Özdemir, 2008b). It was stated in one study that stories and strategies were discussed with participants at the end of the story (Thompson & Johnston, 2013). Four studies (Bıçakçı & Olçay-Gül, 2020; Kim et al., 2014; Kurt & Kutlu, 2019; Olçay-Gül & Tekin-İftar, 2016) presented feedback information, stating that participants' correct responses were reinforced and incorrect responses resulted in re-reading stories and error-correction. In one study correct responses were reinforced but there was no information about feedback for the incorrect responses (Boşnak & Turhan, 2020). However, six studies did not present information about feedback. Researchers did not mention about feedback but indicated that stories were read until the participants responded to comprehension questions with 100% accuracy in 1 study (Scattone et al., 2006).

Fading procedures arranged in six studies (e.g., Delano & Snell, 2006). Delano and Snell (2006) specified a fading criterion as performing 40% of target behaviors on four of the last six data points after 15 instruction sessions compared with a baseline for the first fading and they read SSs once in every two sessions. They began a second fading procedure after participants had achieved 40% of the target behaviors on four of the last six data points over the baseline and read SSs once in every three sessions. When participants reached the criterion, researchers began a 'no story phase'. Özdemir (2008b) used the same fading criterion as Delano and Snell (2006) but it was conducted with a slight difference. They continued reading the story without directive sentences after the first criterion for fading and she read stories once in every two sessions through the second fading until the same criterion was met. Then she began a 'no story phase' and collected data for five sessions. Olçay-Gül and Tekin-İftar (2016) made arrangements for both content of stories and reading time. Arrangements on the content of SSs were made in three phases: (a) directive sentences were excluded from stories; (b) only the title, first and last sentences were read, and (c) the SSs book was shown to participants before entering the setting in which the target behavior was observed. They also extended the time between reading time and performing time of the target behavior. Similarly, Kurt & Kutlu (2019) extended the time between reading SSs time and performing time. Bıçakçı and Olçay-Gül (2019) removed the directive sentences in the first phase of fading process and reminded participants by saying "Don't forget the story" in the second phase. Sansosti and Powell-Smith (2006) faded stories two weeks after instruction sessions; SSs were read once a day for the first fading week and once every other day in the second fading week, unlike the instruction sessions in which SSs were read twice a day.

Implementer. The first authors (researcher, special education teacher, and occupational therapist) of three studies had conducted instruction sessions (e.g., Delano & Snell, 2006). The second author conducted instruction sessions in one study (Kurt & Kutlu, 2019). Instructional sessions were conducted by special education teachers or teaching assistants in three studies and by special education staff member (paraprofessional) in two studies (e.g., Scattone et al., 2006). Families conducted instruction sessions in one study (e.g. Olçay-Gül & Tekin-İftar, 2016) and peers conducted in one study (Bıçakçı & Olçay-Gül, 2019).

Social Validity. Social validity data were collected in all of the studies. Subjective evaluation was used in nine studies by interviewing or using rating scales with parents, specialists or peers (e.g., Kim et al., 2014). Data were collected also from participants in two studies (Kurt & Kutlu, 2019; Bıçakçı & Olçay-Gül, 2019). The social comparison was used in two studies (Delano & Snell, 2006; Sansosti et al., 2006). Social validity data were collected using both social comparison and subjective evaluation in one study (Olçay-Gül & Tekin-İftar, 2016).

Maintenance and Generalization. Maintenance data about dependent variables were collected in ten studies (e.g., Delano & Snell, 2006), however it were not collected in two studies (Scattone et al., 2002; Scattone et al., 2006). Also Bıçakçı and Olçay-Gül (2019) could not collect maintenance data one of three participants. Generalization sessions were planned in six studies (e.g., Delano & Snell, 2006), but not in five studies (Özdemir, 2008b). Generalization data were collected across both settings and people in two studies (Delano & Snell, 2006; Olçay-Gül & Tekin-İftar, 2016) and across only settings in two studies (Bıçakçı & Olçay-Gül, 2019; Graetz et al., 2009) and activities in two studies (Kim et al., 2014; Thompson & Johnston, 2013). Kurt and Kutlu (2019) didn't conduct generalization sessions but authors reported that they used different lures and different strangers for each participant in order to provide generalization for them across people and settings.

Table 3

Social Story Characteristics

Studies	Gray's criteria	Pictures of SSs	Who read the story	Comprehension check questions	Prompts	Reinforcement	Fading system	Frequency of reading SSs
Bıçakçı & Olçay-Gül (2019)	Gray (2000)	+PS	Peers	Y	MP, RRSQ	SRDQ	Y	TT
Boşnak & Turhan (2020)	Gray (2002)	+PS	Interventionist	Y	ND	SRDQ	N	OT
Delano &Snell (2006)	Gray (2000)	+PS	Interventionist	Y	ND	ND	Y	ND
Graetz et al. (2009)	Gray and Garand (1993)	+PRP	Interventionist	Y	ND	ND	N	ALT+FA
Kim et al. (2014)	Gray (1998)	+PRP	Interventionist+ children	Y	OQ	SRDQ	N	ND
Kurt & Kutlu (2019)	ND	+PRP a peer	Interventionist	Y	MP, RRDQ	SRDQ	Y	OT
Olçay-Gül & Tekin-İftar (2016)	ND	+PS	Interventionist	Y	RRSQ, MP	SRDQ	Y	ND
Özdemir (2008b)	Gray (1998)	+PRP	Interventionist+ children	ND	ND	ND	Y	TT, OTFOP
Sansosti & Powell-Smith (2006)	ND	+PS	Interventionist	Y	ND	ND	Y	TT
Scattone et al. (2006)	Gray (1998)	ND	Interventionist+ children	Y	RRUCM	ND	N	OT
Scattone et al. (2002)	Gray (1998)	ND	Interventionist+ children	Y	ND	ND	N	FA
Thompson & Johnston (2013)	Gray (2004), social stories 10.0	ND	Interventionist	ND	ND	ND	N	ND

Note: ALT = at least twice; FA = free access; MP = model prompt; N = No; ND = no data; OQ = only questions; OT = one time; OTFOP = one time for one participant; PRP = participant's real photo; PS = picture symbols; RRSQ = reread the story and questions; RRUCM = reread story until criterion met; SRDQ = social reinforcement during questions; TT= two times; Y = yes.

General Effect. When the SSs effects on dependent variables were examined for all 36 participants in the 12 studies, positive outcomes were obtained for 34 participants (94.4%) but there was no effect for two participants (5.6%). The researchers in two studies (Sansosti & Powell-Smith, 2006; Scattone et al., 2006) stated that there had been no effect of SSs in one of the three participants in each study.

Effects of SSs Instruction

Of the 12 studies which met the standards or the standards with reservations, we calculated the effect size of seven studies coded as the strong effect in the visual analysis process by using IRD. We calculated the effect size separately for each tier and then estimated the mean value for each participant. From the IRD values shown in Table 4, it can be seen that the IRD values of all seven studies were greater than 70% and the effect size was, therefore determined to be large. Of the 7 studies for which we calculated effect sizes, we examined the effects of SSs on decreasing problem behaviors in 3 studies (e.g., Graetz et al., 2009); on teaching social skills in two studies (e.g., Olçay-Gül & Tekin-İftar, 2016) and safety skills in one study (Bıçakçı & Olçay-Gül, 2019) and on both decreasing problem behaviors and teaching social skills in one study (Kim et al., 2014). The findings of studies that examined the effects of SSs on both decreasing problem behaviors and teaching social skills show that the effect size was large for all participants.

Determination of EBP

We used the 5-3-20 rule described above to determine EB. The first criterion is that there must be at least five studies that meet standards and standards with reservations. The second is that studies must be conducted by at least three groups of researchers with no overlapping authorship from three different geographic regions. In our research, seven studies met this criterion. The third criterion for the determination of EB is that there must be at least 20 participants in studies that meet the first two criteria. The total number of participants in our study was 21, hence the SSs meet the 5-3-20 rule.

Table 4

Percentage of IRD Calculations for Social Stories

Studies	Intervention IRD	Number of tiers	Effect size
Graetz et al. (2009)	0.8671	3	Large effect
Kim et al. (2014)	0.9722	6	Large effect
Olçay-Gül & Bıçakçı (2019)	1.0000	3	Large effect
Olçay-Gül &Tekin-İftar (2016)	1.0000	3	Large effect
Özdemir (2008b)	1.0000	3	Large effect
Scattone et al. (2002)	0.9722	3	Large effect
Thompson & Johnston (2013)	0.7445	3	Large effect

Note: IRD = improvement rate difference.

Discussion

We carried out a meta-analysis and a comprehensive review of studies that examined the effectiveness of SSs on increasing appropriate behaviors and decreasing inappropriate behaviors. We first set inclusion criteria and then analyzed the studies which met them in terms of the quality indicators recommended by Kratochwill et al. (2013). Consequently, we identified 32 studies that sought to examine the effectiveness of SSs on increasing appropriate behaviors and decreasing inappropriate behaviors in individuals who had different types of developmental disabilities and then analyzed the findings in terms of the quality indicators recommended by Kratochwill et al. (2013). Then we conducted a visual analysis of studies that we coded as MS or MS-R on a prior process and we finally calculated the effect size of studies that we classified as 'strong effect' (evidence). As a result of the analysis, we coded nine studies as MS and MS-R. Of these 12 studies, we found that seven had a strong effect and we calculated the effect sizes of these seven studies. When the meta-analysis and descriptive analysis studies in which the effects of SSs investigated were examined, it was found that the researchers in only two studies (Test et al., 2011; Qi et al., 2018) had analyzed the quality indicators for including studies on effect size calculation. Test et al. (2011) had used the quality indicators recommended by Horner et al. (2005) and Qi et al. (2018) had used those recommended by Kratochwill et al. (2013), which they regarded as a limitation. We assessed the studies in terms of the procedural fidelity which took place in the quality indicators recommended by Horner et al. (2005) but did not take place in the one recommended by Kratochwill et al. (2013). Although it was emphasized that not collecting procedural fidelity data was a limitation and that it needed to be collected in the meta-analysis and review studies on SSs (Reynhout & Carter, 2006; Test et al., 2011), other meta-analysis and review studies did not assess studies using this particular criterion. Within this context, it can be thought that it is one of the strengths of this current study that we conducted a more detailed assessment procedure by using different criteria as quality indicators.

We analyzed studies, which examined the effects of SSs on both increasing appropriate behaviors and decreasing inappropriate behaviors and calculated the effect sizes of seven studies, which were then subjected to further analysis. Of the seven studies, SSs' effects on increasing appropriate behaviors in four studies (Bıçakçı & Olçay-Gül, 2019; Graetz et al., 2009; Olçay-Gül & Tekin-İftar, 2016; Thompson & Johnston, 2013); on decreasing inappropriate behaviors in two studies (Özdemir, 2008b; Scattone et al., 2002) and on both increasing appropriate behaviors and decreasing inappropriate behaviors in one study (Kim et al., 2014) was examined. The results of the effect size calculation showed that the effect size of SSs on both increasing appropriate behaviors and decreasing inappropriate behaviors was substantial. This finding is different from those of other studies which have shown that SSs are more effective on decreasing inappropriate behaviors than on increasing appropriate behaviors

(Kokina & Kern, 2010; Qi et al., 2018). It can be thought that this discrepancy is a consequence of using different inclusion criteria and quality indicators and of the non-parametric techniques used in the effect size calculation.

We used the 5-3-20 rule as a base for determining EB and found that there were at least five studies which had examined the effects of SSs and were coded as MS and MS-R and classified the effects as strong or moderate and that these studies had been conducted by at least three groups of researchers with no overlapping authorship from three different geographic regions, but the SSs studies did not meet the criterion of involving at least 20 participants. This finding is similar to those of meta-analysis studies of SSs which were grounded on quality indicators (Qi et al., 2018; Test et al., 2011). Differentiating some of our findings from those of Qi et al. (2018) is due to the fact that we included studies conducted with participants with different diagnoses, and from those of Test et al. (2011) is the fact that we included studies which had examined only the effects of SS. It can be said that an original and strong aspect of the current study is that we analyzed studies in which the participants had different diagnoses and which had examined only the effects of SS.

The findings of the descriptive analysis show that most of the participants in the examined studies were primary-school students (N = 19) whose ages ranged from 6 to 11, followed by participants aged 12 and over (N = 12); the group with the fewest participants was pre-school children aged under 5. The findings of the studies which were conducted specifically with primary-school students were consistent with those of Kokina and Kern (2010) and Sani-Bozkurt and Vuran (2014). The findings of the other age groups are different from those of previous studies. For instance, Kokina and Kern (2010) stated that 28 primary school students and 10 pre-school children took part in their study and Sani-Bozkurt and Vuran (2014) stated that 57.14% of their participants were primary-school students and 31.42% were pre-school children, while only 5.74% were 12 or over. This differentiation can be seen as an expected result caused by the differences in both the number of studies analyzed as a result of the inclusion criteria and the grouping of the age ranges. Even so, it can be said that SSs studies were commonly conducted predominantly with primary-school students. The need for further studies conducted particularly with pre-school children and with other age groups is obvious.

Most of the participants in the studies had been diagnosed with ASD and were male and this finding is consistent with those of previous studies (Kokina & Kern, 2010; Qi et al., 2018; Reynhout & Carter, 2006; Sani Bozkurt & Vuran, 2014; Test et al., 2011. Of all the studies analyzed, there was only one (Kim et al., 2014) in which the participants had a severe intellectual disability and one participant had a mild intellectual disability (Bıçakçı & Olçay-Gül, 2019). It can be thought that SSs had been introduced which were especially suitable for children with ASD, even though Gray (2000) emphasized that SSs can be used for children with different types of disabilities. We, therefore, recommend that the effects of SSs should be examined for individuals who have different types of developmental disabilities.

Reynhout and Carter (2006, 2009) emphasized that information about participants was limited. Therefore, the current study confirms this considering that SSs intervention requires that participants have verbal communication skills and reading skills (at least at the beginner level) because a crucial element of SSs use is asking comprehension questions. Most of the studies were conducted with participants who had higher levels of these skills than the majority of individuals with developmental disabilities. Surprisingly, the results of one study (Kim et al., 2014) which was conducted with participants who had severe intellectual disabilities with limited vocabulary were promising, in contrast with the general opinion about using SSs with high-functioning children who have a particular vocabulary and cognitive skills. Considering that three of the included studies gave no information about the cognitive level of the participants (Boşnak & Turhan, 2020; Kim et al., 2014; Olçay-Gül & Tekin-İftar, 2016) and three contained no information about the prerequisite skills of the participants (Olçay-Gül & Tekin-İftar, 2016; Scattone et al., 2002; Thompson & Johnston, 2016), we recommend that future studies should be designed in terms of replicability for both researchers and interventionists and should provide much more information about the participants.

Previous studies mainly focused on increasing appropriate behaviors, especially social skills and social communication skills (Kokina & Kern, 2010; Reyhout & Carter, 2006; Test et al., 2011, but we found that SSs were effective for decreasing inappropriate behaviors, as was also found in a previous meta-analysis (Kokina & Kern, 2010). Only two studies (Bıçakçı & Olçay-Gül, 2019; Kurt & Kutlu, 2019) focused on safety skills. Many previous studies concentrated on social skills, so there is a clear need for studies that will examine the effect of

SSs on different skills such as self-care skills, safety skills, and academic skills, as Gray (2004) stated that SSs could be effectively used for teaching different skills.

All of the included studies used the 1:1 teaching format because SSs are designed as individualized (Gray, 2004). However, SSs could also be effective when delivered to a group of students by providing observational learning. It can, therefore, be recommended that future research should focus on the effects and the efficacy of using SSs in a group format and observational learning, especially on teaching classroom rules and the need to follow routines.

The findings show that SSs had been implemented by teachers, teaching assistants and parents as well as researchers with high levels of procedural fidelity in six studies (e.g., Graetz et al., 2009; Olçay-Gül & Tekin-İftar, 2016), which shows that SSs are one of the teaching methods which can be implemented effectively by both teachers and non-specialists. Also in one study SSs was implemented by peers (Bıçakçı & Olçay-Gül, 2019). It should not be forgotten that teachers, parents, peers and paraprofessionals must receive training in how to write and implement SS, as was shown in the studies which were analyzed except for one in which it was delivered only by teachers (Scattone et al., 2002). Of all the included studies, the most detailed information about training on writing and implementing SSs was provided by Bıçakçı & Olçay-Gül (2019) and Olçay-Gül and Tekin-İftar (2016), so one recommendation is that future research should be undertaken in that way for using SSs effectively in instructional settings. In one study (Boşnak & Turhan, 2020) gave training for implementing of SSs and collecting data. Moreover, training on collecting data about target behaviors for teachers, paraprofessionals, and parents is important as specified by Test et al. (2011).

There are different ways of delivering SS: Gray (2004) listed PowerPoint presentations, textile SSs books, and puppet books. However, conventional hard-copy books were used in all of the studies analyzed except three (Boşnak & Turhan, 2020; Kurt & Kutlu, 2019; Kim et al., 2014), in which SSs were delivered by tablet. Additionally, all of the studies provided information about visually-assisted SSs except for three studies (Scattone et al., 2002; Scattone et al., 2006; Thompson & Johnston, 2013). Visual elements were included in SSs in four studies (e.g., Delano & Snell, 2006; Sansosti & Powell-Smith, 2006) and participants' photographs were used with SSs in three studies (Graetz et al., 2009; Kim et al.,2004; Özdemir, 2008b), whereas text-only SSs were used in only one study (Delano & Snell, 2006). And only one study used one peer's photograph (Kurt & Kutlu, 2019). SSs are said to be more effective when supported by visual material (Kokina & Kern, 2010), so the option of using visual cues is recommended and should be continued as it is a concretizing referent for individuals with developmental disabilities, particularly ASD.

Author's Contributions

All three authors operated screening, inclusion and exclusion procedure. The first author conducted meta-analysis process with the second author, wrote the introduction part of the manuscript. The first and second author reported the method part of the manuscript. The descriptive analysis process conducted by the second author and the third author. All three authors contributed the discussion part.

References

References marked with one asterisk indicate studies reviewed for Kratochwill et al.'s (2013) criteria.

References marked with two asterisks indicate studies retained for descriptive analyses.

References marked with three asterisks indicate studies retained for effect size analyses.

Acar, C., Tekin-İftar, E., & Yıkmış, A. (2017). Effects of mother-delivered social stories and video modeling in teaching social skills to children with autism spectrum disorders. *The Journal of Special Education, 50*(4), 215-226. https://doi.org/10.1177/0022466916649164

*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*(2), 87-94. https://doi.org/10.1177/10883576040190020301

Agosta, E., Graetz, J. E., Mastropieri, M. A., & Scruggs, T. E. (2004). Teacher-researcher partnerships to improve social behavior through Social Stories. *Intervention in School and Clinic*, *39*(5), 276-287. https://doi.org/10.1177%2F10534512040390050401

*Ali, S., & Frederickson, N. (2006). Investigating the evidence base of social stories. *Educational Psychology in Practice*, *22*(4), 355-377. https://doi.org/10.1080/02667360600999500

Almutlaq, H., & Martella, R. C. (2018). Teaching elementary-aged students with autism spectrum disorder to give compliments using a social story delivered through an ipad application. *International Journal of Special Education*, *33*(2), 482-492.

*Beh-Pajooh, A., Ahmadi, A., Shokoohi-Yekta, M., & Asgary, A. (2011). The effect of social stories on reduction of challenging behaviours in autistic children. *Procedia-Social and Behavioral Sciences*, *15*(1), 351-355. https://doi.org/10.1016/j.sbspro.2011.03.100

Bernad-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*(2), 100-106. https://doi.org/10.1177%2F10883576070220020101

***Bıçakcı, M., & Olçay-Gül, S. (2019). Effect of peer delivered social stories on the crossing skills of primary school students with developmental disabilities. *Eğitim ve Bilim*, *44*(199), 1-22. https://doi.org/10.15390/EB.2019.8168

*Bledsoe, R., Smith, B., & Simpson, R. L. (2003). Use of a social story intervention to improve mealtime skills of an adolescent with Asperger syndrome. *Autism*, *7*(3), 289-295. https://doi.org/10.1177%2F1362361303007003005

**Boşnak Ö., & Turhan, C. (2020). Presentation of social stories with tablet computers in social skill instruction for students with autism spectrum disorder. *Elementary Education Online*, *19*(4), 2161-2170. https://doi.org/10.17051/ilkonline.2020.763781

Burke, R. V., Kuhn, B. R., & Peterson, J. L. (2004). Brief report: A "storybook" ending to children's bedtime problems-the use of a rewarding social story to reduce bedtime resistance and frequent night waking. *Journal of Pediatric Psychology*, *29*(5), 389-396. https://doi.org/10.1093/jpepsy/jsh042

Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*, *85*(1), 104-118. https://doi.org/10.1177%2F0014402918793138

*Crozier, S., & Tincani, M. (2007). Effects of social stories on prosocial behavior of preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(9), 1803-1814. https://doi.org/10.1007/s10803-006-0315-7

**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*(1), 29-42. https://doi.org/10.1177%2F10983007060080010501

*Dodd, S., Hupp, S. D., 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*(3), 217-229. https://doi.org/10.1007/s10882-007-9090-4

***Graetz, J. E., Mastropieri, M. A., & Scruggs, T. E. (2009). Decreasing inappropriate behaviors for adolescents with autism spectrum disorders using modified social stories. *Education and Training in Developmental Disabilities*, *44*(1), 91-104. https://www.jstor.org/stable/24233466

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), 1-10.

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

Gray, C. A. (1998). Social stories and comic strip conversations with students with Asperger syndrome and high-functioning autism. In E. Schopler, G. B. Mesibov, & L. J. Kunce (Eds.), *Asperger syndrome or high-functioning autism?* (pp. 167-198). Springer.

Gray, C. A. (2000). *Writing Social Stories with Carol Gray.* Future Horizons.

Gray, C. (2004). Social Stories™ 10.0: The new defining criteria and guidelines. *Autism Journal*, *15*(4), 2-21.

Gray, C. (2010). *The new Social Story™ book: Revised and expanded 10th anniversary edition.* Future Horizons.

Gray, C. (2015). *The new Social Story™ book: Revised and expanded 15th anniversary edition*. Future Horizons.

*Hagiwara, T., & Smith-Myles, B. (1999). A multimedia social story intervention: Teaching skills to children with autism. *Focus on Autism and Other Developmental Disabilities*, *14*(2), 82-95. https://doi.org/10.1177/108835769901400203

Hanley-Hochdorfer, K., Bray, M. A., Kehle, T. J., & Elinoff, M. J. (2010). Social stories to increase verbal initiation in children with autism and Asperger's disorder. *School Psychology Review*, *39*(3), 484-492. https://doi.org/10.1080/02796015.2010.12087767

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. https://doi.org/10.1177/001440290507100203

*Hung, L. C., & Smith, C. S. (2011). Autism in Taiwan: Using social stories to decrease disruptive behaviour. *The British Journal of Development Disabilities*, *57*(112), 71-80. https://doi.org/10.1179/096979511798967197

*Ivey, M. L., Juane Heflin, L., & 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*(3), 164-176. https://doi.org/10.1177/10883576040190030401

*Karayazi, S., Kohler Evans, P., & Filer, J. (2014). The effects of a Social Story™ intervention on the pro-social behaviors of a young adult with autism spectrum disorder. *International Journal of Special Education*, *29*(3), 126-133.

*Khantreejitranon, A. (2018). Using a social story intervention to decrease inappropriate behavior of preschool children with autism. *Kasetsart Journal of Social Sciences*, *39*(1), 90-97. https://doi.org/10.1016/j.kjss.2017.12.019

***Kim, M. S., Blair, K. S. C., & Lim, K. W. (2014). Using tablet assisted Social Stories™ to improve classroom behavior for adolescents with intellectual disabilities. *Research in Developmental Disabilities*, *35*(9), 2241-2251. https://doi.org/10.1016/j.ridd.2014.05.011

*Klett, L. S., & Turan, Y. (2012). Generalized effects of social stories with task analysis for teaching menstrual care to three young girls with autism. *Sexuality and Disability*, *30*(3), 319-336. https://doi.org/10.1007/s11195-011-9244-2

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. https://doi.org/10.1007/s10803-009-0931-0

Kratochwill, T. R., Hitchcock, J. H., 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*(1), 26-38. https://doi.org/10.1177/0741932512452794

Kuoch, H., & Mirenda, P. (2003). Social story interventions for young children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *18*(4), 219-227. https:/doi.org/10.1177/10883576030180040301

**Kurt, O., & Kutlu, M. (2019). Effectiveness of social stories in teaching abduction-prevention skills to children with autism. *Journal of Autism and Developmental Disorders*, *49*(9), 3807-3818. https://doi.org/10.1007/s10803-019-04096-9

*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*(3), 176-182. https://doi.org/10.1177/108835769801300306

Leaf, J. B., Oppenheim-Leaf, M. L., Leaf, R. B., Taubman, M., McEachin, J., Parker, T., Waks, A. B., & Mountjoy, T. (2015). What is the proof? A methodological review of studies that have utilized social stories. *Education and Training in Autism and Developmental Disabilities*, *50*(2), 127-141. https://www.jstor.org/stable/24827530

Litras, S., Moore, D. W., & Anderson, A. (2010). Using video self-modelled social stories to teach social skills to a young child with autism. *Autism Research and Treatment*, *2010*, 1-9. https://doi.org/10.1155/2010/834979

*Lorimer, P. A., Simpson, R. L., Smith Myles, B., & Ganz, J. B. (2002). The use of social stories as a preventative behavioral intervention in a home setting with a child with autism. *Journal of Positive Behavior Interventions*, *4*(1), 53-60. https://doi.org/10.1177/109830070200400109

*Moudry-Quilty, K. (2007). Teaching paraprofessionals how to write and implement social stories for students with autism spectrum disorders. *Remedial and Special Education*, *28*(3), 182-189. https://doi.org/10.1177/07419325070280030701

National Autism Center. (2009). *The National Autism Center's national standarts report*. https://www.nationalautismcenter.org/reports/

National Professional Development Center on Autism. (2014). *Evidence-based practices*. http://autismpdc.fpg.unc.edu/evidence-based-practices

Nichols, S. L., Hupp, S. D., Jewell, J. D., & Zeigler, C. S. (2005). Review of social story interventions for children diagnosed with autism spectrum disorders. *Journal of Evidence-Based Practices for Schools*, *6*(1), 90-120. https://doi.org/10.1007/978-1-4614-4788-7_138

*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*(3), 180-186. https:/doi.org/10.1177/108835769901400307

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*(2), 137-148. https//doi.org/10.1177/001440290507100201

***Olçay-Gül, S., & Tekin-İftar, E. (2016). Family generated and delivered social story intervention: Acquisition, maintenance, and generalization of social skills in youths with ASD. *Education and Training in Autism and Developmental Disabilities*, *51*(1), 67-78. https://www.jstor.org/stable/26420365

*Özdemir, S. (2008a). Using multimedia social stories to increase appropriate social engagement in young children with autism. *Turkish Online Journal of Educational Technology-TOJET*, *7*(3), 80-88.

***Özdemir, S. (2008b). The effectiveness of social stories on decreasing disruptive behaviors of children with autism: Three case studies. *Journal of Autism and Developmental Disorders*, *38*(9), 1689-1696. https://doi.org/10.1007/s10803-008-0551-0

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, *75*(2), 135-150. https://doi.org/10.1177/001440290907500201

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*(4), 303-322. https://doi.org/10.1177/0145445511399147

Qi, C. H., Barton, E. E., Collier, M., Lin, Y. L., & Montoya, C. (2018). A systematic review of effects of social stories interventions for individuals with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, *33*(1), 25-34. https://doi.org/10.1177/1088357615613516

Rakap, S. (2015). Effect sizes as result interpretation aides in single-subject experimental research: Description and application of four nonoverlap methods. *British Journal of Special Education*, *42*(1), 11-33, 3. https://doi.org/10.1111/1467-8578.12091

*Reichow, B., & Sabornie, E. J. (2009). Brief report: Increasing verbal greeting initiations for a student with autism via a Social Story™ intervention. *Journal of Autism and Developmental Disorders*, *39*(12), 1740. https://doi.org/10.1007/s10803-009-0814-4

Reichow, B., Volkmar, F. R., & Cichetti, D. V. (2008). Development of an evaluative method for determining the strength of research evidence in autism. *Journal of Autism and Developmental Disorders*, *38*(7), 1311-1319. https://doi.org/10.1007/s10803-007-0517-7

Reynhout, G., & Carter, M. (2006). Social Stories™ for children with disabilities. *Journal of Autism and Developmental Disorders*, *36*(4), 445-469. https://doi.org/10.1007/s10803-006-0086-1

Reynhout, G., & Carter, M. (2009). The use of social stories by teachers and their perceived efficacy. *Research in Autism Spectrum Disorders*, *3*(1), 232-251. https://doi.org/10.1016/j.rasd.2008.06.003

Sani-Bozkurt, 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 and Practice*, *14*(5), 1875-1892. https://doi.org/10.12738/estp.2014.5.1952

**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 Interventions*, *8*(1), 43-57. https://doi.org/10.1177/10983007060080010601

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*(4), 194-204. https://doi.org/10.1177/10883576040190040101

**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*(4), 211-222. https://doi.org/10.1177/10883576060210040201

***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. https://doi.org/10.1023/A:1021250813367

Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yucesoy-Ozkan, S., & Savage, M. N. (2020*). Evidence-based practices for children, youth, and young adults with autism*. The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice Review Team. https://ncaep.fpg.unc.edu/sites/ncaep.fpg.unc.edu/files/imce/documents/EBP%20Report%202020.pdf

Tekin-İftar, E., Olçay-Gul, S., & Collins, B. C. (2018). Descriptive analysis and meta analysis of studies ınvestigating the effectiveness of simultaneous prompting procedure. *Exceptional Children*, *85*(3), 309-328. https://doi.org/10.1177/0014402918795702

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. https://doi.org/10.1177/1088357609351573

***Thompson, R. M., & Johnston, S. (2013). Use of social stories to improve self-regulation in children with autism spectrum disorders. *Physical & Occupational Therapy in Pediatrics*, *33*(3), 271-284. https://doi.org/10.3109/01942638.2013.768322

*Vandermeer, J., Beamish, W., Milford, T., & Lang, W. (2015). iPad-presented social stories for young children with autism. *Developmental Neurorehabilitation*, *18*(2), 75-81. https://doi.org/10.3109/17518423.2013.809811

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2011). *Single case research: Web based calculators for SCR analysis (Version 1.0) [Web-based application]*. College Station: Texas A&M University. singlecaseresearch. org

*Wright, L. A., & McCathren, R. B. (2012). Utilizing social stories to increase prosocial behavior and reduce problem behavior in young children with autism. *Child Development Research*, *13*(1), 1-13. https://doi.org/10.3109/17518423.2013.809811

Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi

Erken Görünüm

SİSTEMATİK DERLEME

Gönderim Tarihi: 08.07.20
Kabul Tarihi: 25.03.21
Erken Görünüm: 14.04.21

Sosyal Öykü Bilimsel-Dayanaklı Bir Uygulama mıdır? Bir Meta-Analiz ve Kapsamlı Betimsel Analiz Çalışması*

Seray Olçay[1] **Üzeyir Emre Kıyak**[2] **Özlem Toper**[3]

Öz

Giriş: Sosyal öykülerin bilimsel dayanağına ilişkin tartışmalar devam etmektedir ve alanyazında gerçekleştirilmiş olan meta-analiz çalışmaları sosyal öykülerin bilimsel dayanağının henüz oluşmadığını göstermektedir.

Yöntem: Bu çalışmada hem uygun davranışların artırılmasında hem de uygun olmayan davranışların azaltılmasında sosyal öykülerin kullanıldığı Haziran 2020 tarihine kadar yayımlanmış araştırmalar değerlendirilerek betimsel ve meta-analiz gerçekleştirilmiştir. Değerlendirmeler sonucunda, tek-denekli araştırma metodolojisiyle tasarlanmış, uluslararası hakemli dergilerde yayımlanmış ve tek-denekli araştırmalar için belirlenmiş olan niteliksel göstergeleri karşılayan yedi çalışmaya ulaşılmıştır. Ardından, bu yedi çalışmanın 'İyileşme Oran Farkı (İOF [Improvement Rate Difference-IRD])' tekniği kullanarak etki büyüklükleri hesaplanmıştır.

Bulgular: Çalışma sonuçları, analiz edilen araştırmaların etki büyüklüklerinin yüksek olduğunu ve sosyal öykülerin 5-3-20 kuralını karşıladığını göstermektedir. Ayrıca, çalışma sonuçları sosyal öykülerin genellikle uygun davranışların arttırılmasında kullanıldığını göstermektedir. Bununla birlikte çalışmada, katılımcıların büyük çoğunluğunun otizm spektrum bozukluğu tanısına sahip olduğu ve ilkokul çağındaki bireyler olduğu bulgularına ulaşılmıştır.

Tartışma: Bu çalışma, sosyal öyküleri bilimsel-dayanaklı uygulamalar arasında olduğunu göstermektedir. Bununla birlikte, sosyal öykülerin farklı gelişimsel yetersizliklere sahip, farklı yaş gruplarındaki bireylerle etkililiğin sınanması gerekmektedir. Ayrıca, sosyal öykülerin bilimsel-dayanaklı olma ölçütlerini karşılamış olsa dahi halen konuya ilişkin nitelikli araştırma gereksinimi devam etmektedir.

Anahtar sözcükler: Bilimsel-dayanaklı uygulamalar, sosyal öyküler, sistematik derleme, meta-analiz, tek denekli araştırmalar.

Atıf için: Olçay, S., Kıyak, Ü. E., & Toper, Ö. (2021). Sosyal öykü bilimsel-dayanaklı bir uygulama mıdır? Bir meta-analiz ve kapsamlı betimsel analiz çalışması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğititim Dergisi, Erken Görünüm.* https://doi.org/10.21565/ozelegitimdergisi.766765

*Bu çalışma 29. Ulusal Özel Eğitim Kongresi'nde sözlü bildiri olarak sunulmuştur.

[1]Doç. Dr., Hacettepe Üniversitesi, E-posta: solcaygul@hacettepe.edu.tr, https://orcid.org/0000-0002-5007-7466

[2]***Sorumlu Yazar***: Araş. Gör. Dr., Bursa Uludağ Üniversitesi, E-posta: uzeyiremre@gmail.com, https://orcid.org/0000-0002-2681-9407

[3]Dr. Öğr. Üyesi, Bursa Uludağ Üniversitesi, E-posta: otoper@gmail.com, https://orcid.org/0000-0002-0243-4844

Giriş

Sosyal öyküler ilk kez 1991 yılında Carol A. Gray tarafından Otizm Spektrum Bozukluğu (OSB) tanısı olan bireylere, değişikliklerin ve yeni rutinlerin anlatılması, başka kişilerin davranışlarının açıklanması, akademik ve sosyal becerilerin öğretimi amacıyla geliştirilmiş (Gray & Garand, 1993); zamanla bu öykülerin zihin yetersizliği, öğrenme güçlüğü gibi farklı tanılara sahip çocuklar için öz bakım, güvenlik becerileri gibi farklı becerilerin öğretiminde de kullanımı giderek yaygınlaşmıştır (Hagiwara & Smith-Myles, 1999; Kurt & Kutlu, 2019). Akademik beceriler, sosyal beceriler, güvenlik becerileri gibi yeni becerilerin öğretiminin yanı sıra sosyal öykülerin problem davranışların azaltılmasında da olumlu etkileri olduğunu gösteren araştırma örnekleri bulunmaktadır (ör., Agosta, vd., 2004). Bu araştırmaların bazılarında sosyal öykülerin etkisi tek başına incelenirken, (ör., Sansosti & Powell-Smith; 2006; Scattone vd., 2006), bazı araştırmalarda sosyal öykülerin video model, tepkinin bedeli, ipuçları gibi farklı stratejilerle birlikte kullanıldığı görülmektedir (ör., Bernad-Ripoll, 2007).

Aileler, öğretmenler, akranlar ve destek personeli gibi farklı bireyler tarafından kullanımı kolay ve pratik bulunan sosyal öykülerin (Test vd., 2011), Amerika Birleşik Devletleri'ndeki kuruluşlar tarafından (ör., National Autism Center [NAC], 2009; The National Clearinghouse on Autism Evidence & Practice [NCAEP], 2020; National Professional Development Center on Autism [NPDC], 2014) bilimsel dayanaklı (Türkiye'de bilimsel dayanaklı uygulamalar terimi için kanıt temelli uygulamalar, kanıta dayalı uygulamalar, delile dayalı uygulamalar gibi farklı ifadeler de kullanılagelmektedir.) olduğu ifade edilmiştir. Öte yandan sosyal öykülerin etkililiğine yönelik meta-analiz araştırmalarının bulguları da sosyal öykülerin aslında bilimsel dayanaktan yoksun olduğu, umut vadeden uygulamalar arasında gösterilebileceği ve bilimsel dayanağının oluşması için daha fazla araştırma gereksinimi bulunduğu şeklindedir (Kokina & Kern, 2010; Reynhout & Carter, 2006; Sani-Bozkurt & Vuran, 2014; Test vd., 2011). Sosyal öykülerin bilimsel dayanaktan yoksun olduğunun vurgulandığı bu araştırmalar incelendiğinde yalnızca iki meta-analiz çalışmasında meta-analize dahil edilen araştırmalar için niteliksel göstergelerin karşılanıp karşılanmadığı ölçüt olarak alınmıştır (Qi vd., 2018; Test vd., 2011). Test ve diğerleri (2011) tarafından yürütülen araştırmada meta-analize dâhil etme ölçütleri arasında Horner ve diğerlerinin (2005); Qi ve diğerleri (2018) tarafından yürütülen araştırmada ise Kratochwill ve diğerlerinin (2013) nitelikli araştırmalar için yayımlamış oldukları kalite göstergeleri dikkate alınmıştır. Diğer üç meta-analiz çalışmasında ise (Kokina & Kern, 2010; Reynhout & Carter, 2006; Sani-Bozkurt & Vuran, 2014) analiz sürecine dâhil edilen araştırmaların sadece deneysel kontrolün kurulduğu çalışmalar değil, deneysel kontrolün zayıf olduğu tek denekli araştırmalar ve/veya vaka çalışmalarını da kapsadığı, niteliksel göstergeler açısından sınırlılıkları olan araştırmaların analize alındığı ve bu tür çalışmalar için ise bilimsel dayanaktan söz etmenin mümkün olmadığı görülmektedir. Bir uygulamanın bilimsel dayanağından söz edebilmek amacıyla gerçekleştirilecek olan meta-analiz araştırmalarına niteliksel göstergeleri karşılayan araştırmaların dahil edilmesinin önemli bir gereklilik olduğu açıktır. Tek denekli araştırmalar için niteliksel göstergelere ilişkin ilk çalışma 2005 yılında Horner ve diğerleri tarafından yayımlanmıştır. Bu çalışmayı Reichow ve diğerleri (2008) ve Kratochwill ve diğerleri (2013) tarafından yayımlanan çalışmalar izlemiştir. Bu çalışmalarda niteliksel göstergeler açısından bir görü birliği sağlanmasına karşın bazı farklılıklar da gözlenmektedir. Bu farklılardan biri de araştırmalarda uygulama güvenirliği verilerinin toplanmasına ilişkindir. Uygulama güvenirliği Horner ve diğerleri (2005) ve Reichow ve diğerleri (2008) tarafından geliştirilmiş olan niteliksel ölçütlerde bir gösterge olarak yer almamıştır. Oysa bilimsel-dayanak oluşturma çabaları, tüm araştırma sürecinin aktarımında tekrarlanabilirliği, şeffaflığı ve açıklığı arttırarak güvenirliği ve inandırıcılığı geliştirme üzerine kuruludur (Cook vd., 2018). Bu nedenle araştırmalarda uygulama güvenirliği verilerinin toplanmış olması önemli bir göstergedir. Günümüze kadar olan süreçte belirlenen göstergelerin iyileştirilmesi ve eksikliklerinin giderilerek kullanılmasıyla niteliksel göstergeler daha sağlıklı bir şekilde işletilebilir. Bu durum alanyazında gerçekleştirilen 16 araştırmadan 12'sinde uygulama güvenirliğinin rapor edilmediğini belirten Reynhout ve Carter (2006) tarafından yürütülen araştırmada da bir sınırlılık olarak dile getirilmiştir. Dolayısıyla niteliksel göstergelere ilişkin son kaynak olan Kratochwill ve diğerleri (2013) tarafından belirlenen göstergelere uygulama güvenirliği ile ilgili göstergelerin eklenmesi gerekliliği açıktır.

Sosyal öykülere yönelik meta-analiz çalışmaları yayın yılları açısından incelendiğinde en son 2011 yılına kadar gerçekleştirilmiş araştırmaların analizlere dâhil edildiği görülmektedir. Oysaki günümüze kadar olan ortalama sekiz yıllık süre zarfında yaklaşık 20 sosyal öykü araştırması daha yürütülmüştür (ör., Kim vd., 2014; Olçay-Gül & Tekin-İftar, 2016). Konu ile ilgili betimsel analiz ve meta analiz araştırmalarında sosyal öykülerin bilimsel dayanağının oluşması için araştırma gereksinimi olduğu vurgusu göz önüne alındığında, 2012-2018 yılları arasında yürütülen araştırmaların analizlere dâhil edilmesinin sosyal öykülerin bilimsel dayanaklı bir uygulama

olup olmadığına karar verilmesinde önemli bir yere sahip olduğu düşünülmektedir. Ek olarak özellikle Kratochwill ve diğerleri (2013) tarafından yayımlanan tek denekli araştırmalara ilişkin kalite göstergelerinin, diğer yandan Gray (2010) tarafından son şekli verilen sosyal öykü yazım ölçütlerinin tarihleri göz önünde bulundurulduğunda, özellikle 2010 ve sonrası yıllarda gerçekleştirilmiş olan sosyal öykü araştırmalarının gerek deneysel gerekse yöntemsel açıdan niteliklerinin incelenmesinin bir gereksinim olduğunu söylemek mümkündür.

Sosyal öykülere yönelik meta analiz araştırmalarında etki büyüklüklerinin hesaplanmasında kullanılan teknikler incelendiğinde ise tüm araştırmalarda etki büyüklüklerinin hesaplanmasında örtüşmeyen veri yüzdesinin (ÖMVY) kullanıldığı görülmüştür. Alanyazında tek denekli araştırma desenlerinin analizinde en uygun parametrik olmayan tekniğin hangisi olduğu konusunda henüz bir uzlaşma bulunmamaktadır (Tekin-İftar vd., 2018). ÖMVY kolay hesaplanabilir, diğer tekniklerle yüksek düzeyde tutarlılık gösteren bir teknik olması nedeniyle sıklıkla tercih edilmektedir. Alanyazında sosyal öykü uygulamalarının etki büyüklüklerinin hesaplanmasında farklı parametrik olmayan tekniklerin kullanıldığı meta-analiz araştırmalarına gereksinim duyulmaktadır. Sıralanan bu noktalardan ve daha önce gerçekleştirilen meta-analiz çalışmalarının sınırlılıklarından yola çıkılarak bu araştırmada, sosyal öykülerin tek başına uygulandığı araştırmaların kapsamlı betimsel ve meta analizi amaçlanmıştır. Bu amaçla (a) dâhil etme ölçütlerini karşılayan araştırmalar Kratochwill ve diğerleri (2013) tarafından belirlenen göstergelere uygulama güvenirliğine ilişkin göstergelerin de eklendiği niteliksel göstergeler kullanılarak analiz edilmiş, (b) niteliksel göstergeleri karşılayan araştırmaların demografik, yöntemsel ve sonuç özellikleri açısından kapsamlı betimsel analizi yapılmış ve (c) bu araştırmaların etki büyüklükleri İyileşme Oran Farkı (İOF [Improvement Rate Difference-IRD]) kullanılarak analiz edilmiş, (d) sosyal öykülerin bilimsel dayanaklı uygulamalar için belirlenen 5-3-20 kuralını karşılayıp karşılamadığı değerlendirilmiştir. Bu araştırma sonucunda elde edilen bulguların, sosyal öykülerin bilimsel dayanağı olup olmadığı sorusuna yanıt olacağı ve bundan sonra gerçekleştirilecek olan sosyal öykü araştırmalarında dikkat edilmesi gereken noktalara katkı sağlayacağı düşünülmektedir**.**

Yöntem

Hazırlık Süreci

Araştırmada verilerin sağlıklı ve güvenilir bir şekilde toplanabilmesi için ilk olarak araştırmacılar bir araya gelerek araştırma sürecinde kullanacakları kodlama tablolarını hazırlamışlardır. Bu amaçla (a) tarama süreci, (b) "Tek Denekli Müdahaleler Model Standartları" değerlendirme süreci, (c) kapsamlı betimsel analiz değerlendirme süreci, (d) UnGraph5 programının kullanımı, (e) etki büyüklüklerinin hesaplanması aşamalarında kullanılmak üzere beş kodlama tablosu hazırlanmıştır. Araştırmacılar kodlama tablolarında yer alan her bir maddeyi gözden geçirerek bu maddelere ilişkin kodlamaları nasıl yapacakları konusunda görüş birliğine varmışlardır. Ardından iki araştırmacı yansız olarak belirlenen üç araştırma için hazırlanan tabloları kullanarak kodlama yapmışlardır. Kodlamalar arasındaki tutarlılık "*Kodlamacılar arası görüş birliği / Kodlamacılar arası Görüş birliği + Görüş ayrılığı X 100*" formülü kullanılarak analiz edilmiştir. Bu süreç kodlamalar arasındaki tutarlılık %100 oluncaya değin devam etmiştir. Araştırmacıların kodlamaları arasındaki tutarlılık %100'e ulaştıktan sonra kodlama ve veri toplama sürecine geçilmiştir.

Tarama Süreci

Araştırmada sosyal öykülerin etkililiğini inceleyen araştırmalar gözden geçirilmiştir. Araştırmada sosyal öykülerle ilgili tüm araştırmalara ulaşılması amaçlandığından taramalar için bir başlangıç tarihi belirlenmemiş; Aralık 2018'e kadar olan araştırmalara ulaşılmaya çalışılmıştır. Farklı tanıları olan öğrencilere uygun davranış ve becerilerin kazandırılmasında ya da uygun olmayan davranışların azaltılmasında sosyal öykülerin etkililiğini inceleyen araştırmaların belirlenmesi için taramalar gerçekleştirilmiştir. Ek olarak ulaşılan makalelerin kaynakçaları taranarak ve sosyal öykülerin etkililiğini konu alan araştırmalara yönelik meta analiz çalışmaları gözden geçirilerek taramalarda çıkmayan araştırmalara da ulaşılması amaçlanmış; ancak taramalar sırasında ulaşılan araştırmalar dışında bir araştırmaya ulaşılamamıştır. Tarama sonucunda dokuz betimsel araştırma, üç vaka çalışması, beş deneysel araştırma, dört gözden geçirme çalışması, beş meta-analiz çalışması ve 49 tek-denekli araştırma olmak üzere toplam 75 araştırmaya ulaşılmıştır. Tarama sonucu ulaşılan araştırmalar hazırlanan kodlama tablosu kullanılarak dâhil etme ve dışlama ölçütleri açısından değerlendirilmiştir.

Araştırmada dâhil etme ölçütleri (a) ulusal ve uluslararası hakemli dergilerde yayımlanmış olma, (b) İngilizce dilinde yazılmış olma, (c) en az bir tanıya sahip olan çocuklarla çalışılmış olma, (d) farklı davranış ya da becerilerin öğretiminde ya da uygun olmayan davranışların azaltılmasında sosyal öykülerin etkililiğini inceleme

ve (e) tek-denekli araştırma metodolojisiyle tasarlanmış olma olarak belirlenmiştir. Dışlama ölçütleri ise, (a) diğer araştırma metodolojileri ile tasarlanmış olma (ör., grup deneysel araştırma, betimsel araştırma), (b) sosyal öykülerle birlikte diğer yöntemlerin kullanıldığı uygulamaların etkililiğini inceleme, (c) farklı öğretim uygulamalarını karşılaştırıyor olma ve (d) görsel analize uygun veri sunmama olarak belirlenmiştir. Tek denekli araştırma metodolojisiyle yürütülmeyen 26 araştırma dâhil etme ölçütlerini karşılamadığı için elenmiştir. Tek denekli araştırma metodolojisiyle yürütülen 49 araştırmadan birinin tam metnine (Hanley-Hochdorfer vd., 2010) ulaşılamadığı için bu araştırmada ileri analizlere alınmamıştır. 48 araştırmadan 15'i sosyal öykülerin diğer öğretim uygulamaları ile birlikte kullanılması (ör., Litras, vd., 2010), dördü sosyal öykülerle farklı öğretim uygulamalarının karşılaştırılması (ör., Acar vd., 2017), biri ise normal gelişim gösteren katılımcılarla yürütülmesi (Burke vd., 2004) nedeniyle ileri analizlere dahil edilmemiştir. Geriye kalan 32 araştırma ileri analizlere alınmıştır.

Değerlendirme Süreci

Bu süreçte dâhil etme ölçütlerini karşılayan araştırmaların metodolojik niteliklerinin ve betimsel özelliklerinin belirlenmesi amacıyla değerlendirme ölçütleri belirlenerek değerlendirmeler yapılmıştır.

Tek Denekli Müdahaleler Model Standartlarına İlişkin Değerlendirme

Dâhil etme ve dışlama ölçütleri açısından analiz edilen ve bu analiz sonucunda ileri analizlere dâhil edilmesine karar verilen 32 araştırma öncelikle metodolojik nitelikleri açısından değerlendirmeye alınmıştır. Alanyazında tek denekli araştırma metodolojisiyle yürütülen araştırmaların niteliksel olarak değerlendirilmesi amacıyla farklı araştırmacılar tarafından geliştirilmiş göstergeler/rubrikler yer almaktadır (Horner vd., 2005; Kratochwill vd., 2013; Reichow vd., 2008). Bu araştırma kapsamında Kratochwill ve diğerleri (2013) tarafından geliştirilen "Tek Denekli Müdahaleler Model Standartları" yönergesi kullanılmış; yönergeye diğer rubriklerde (Horner vd., 2005; Reichow vd., 2008) yer alan ve araştırmacılar tarafından önemli görülen uygulama güvenirliği verilerinin toplanması ve uygulama güvenirlik katsayısına ilişkin iki ölçüt eklenmiştir. Araştırmaların değerlendirilmesinde bu ölçütlerin de eklenmesiyle son şekli verilen niteliksel ölçütlerin yer aldığı bir kodlama tablosu kullanılmıştır. Kodlama tablosunda (a) bağımsız değişkenin sistematik manipülasyonu, (b) oturumların %20'sinden gözlemciler arası güvenirlik verisi toplanması, (c) gözlemciler arası güvenirlik katsayısının %80 ve üzerinde olması, (d) oturumların %20'sinden uygulama güvenirliği verisinin toplanması, (e) uygulama güvenirliği katsayısının %80 ve üzerinde olması, (f) deneysel etkinin en az üç gösterimi, (g) her koşulda (evrede) beş veri noktası olması, (h) her koşulda üç veri noktası olması, (i) model standardının sınıflandırılması ve (j) etkililik için dayanağın sınıflandırılması olmak üzere 10 ölçüt yer almıştır.

Kodlama tablosunda "a-h" maddeleri için "evet" ya da "hayır" biçiminde kodlama yapılmıştır. "I" maddesi, başka bir ifadeyle model standardının sınıflandırılması, için üçlü bir kodlama yapılmış; "a-h" maddelerinde yer alan ölçütlerin tamamını karşılayan araştırmalar *model standartlarını karşılayan* araştırmalar olarak değerlendirilerek tabloda standartları karşılayan şeklinde kodlama yapılmıştır. Bu ölçütlerden "g" maddesini karşılayamayıp "h" maddesini karşılayan araştırmalar *model standartlarını koşullu karşılayan* araştırmalar olarak değerlendirilerek tabloda *standartları koşullu karşılayan* olarak kodlama yapılmıştır. "G" maddesi dışında "a-h" arasında yer alan ölçütlerden herhangi birini karşılayamayan araştırmalar ise *model standartlarını karşılayamayan* araştırmalar olarak değerlendirilerek tabloda *standartları karşılamayan* olarak kodlama yapılmıştır. Kodlama tablosunda yer alan "j" maddesi, başka bir ifadeyle etkililik için dayanağın sınıflandırılması, için *model standartlarını karşılayan* ya da *model standartlarını koşullu olarak karşılayan* araştırmaların görsel analiz sonuçlarına dayalı olarak kodlama yapılmıştır. Görsel analizde düzey, eğilim, kararlılık, acil etki, örtüşme ve benzer evreler arasında verilerin tutarlılık göstermesi olmak üzere altı özelliğe ilişkin değerlendirme yapılmıştır (Kratochwill vd., 2013). Araştırmada görsel analiz dört aşamalı olarak yürütülmüştür: (a) başlama düzeyi evresinde görsel analiz gerçekleştirilmiş, (b) uygulama evresinde görsel analiz gerçekleştirilmiş, (c) ardışık evrelerde karşılaştırma yapılmış ve (d) ilk üç aşamada yapılan değerlendirmeler birleştirilerek deneysel etkinin üç farklı zamanda ortaya konulup konulmadığı incelenmiştir. Sıralanan özelliklerin tümünün tüm durumlarda karşılanmış olması koşulunda tabloda "*güçlü etki (dayanak)"*, etkinin en az üç durumda gösterildiği, bir durumda gösterilmediği durumlarda "*orta düzeyde etki"*, bunun dışındaki durumlarda ise "etkisiz (*dayanak yok)"* şeklinde kodlama yapılmıştır. Araştırmada görsel analiz sonucunda tabloda güçlü etki ve orta düzeyde etki şeklinde kodlama yapılan araştırmalar için etki büyüklüğü hesaplanmıştır.

Kapsamlı Betimsel Analiz Değerlendirmesi

"Tek Denekli Müdahaleler Model Standartları" değerlendirmesinde model standartlarını karşılayan ya da koşullu karşılayan ve güçlü etki ya da orta düzeyde etkiye sahip olan araştırmalar için kapsamlı betimsel analiz yapılmıştır. Kapsamlı betimsel analizde her bir araştırma (a) katılımcı sayısı ve yaşı, (b) cinsiyet, (c) tanı, (d) beceri, (e) becerinin ölçümü, (f) öğretim ölçütü, (g) öğretim ortamı ve (h) öğretim düzenlemesi olmak üzere sekiz demografik özellik; (a) araştırma modeli, (b) sosyal öykülerin özellikleri, (c) sosyal öykülerin uygulanması, (d) uygulamacı (uygulamanın kim tarafından yürütüldüğü), (e) sosyal geçerlik, (f) kalıcılık, (g) genelleme ve (h) genel etki olmak üzere sekiz yöntemsel özellik ve araştırma sonuçları açısından değerlendirilmiştir. Bu amaçla araştırmacılar tarafından hazırlanan betimsel analiz kodlama tablosu kullanılmıştır.

Etki Büyüklüklerinin Hesaplanması

Araştırmada "Tek Denekli Müdahaleler Model Standartları" değerlendirmesinde *model standartlarını karşılayan* ya da *koşullu karşılayan* ve güçlü etki ya da orta düzeyde etkiye sahip olan araştırmalar için etki büyüklüğü hesaplanmıştır. Alanyazında tek denekli araştırmalarda etki büyüklüğünün hesaplanmasında kullanılması önerilen farklı nonparametrik teknikler bulunmakta; hangi tekniğin uygun olduğu konusunda ise bir görüş birliği bulunmamaktadır (Parker vd., 2011; Rakap, 2015; Tekin-İftar vd., 2018). Bu araştırmada sosyal öykülerin etkililiğini konu alan araştırmalara ilişkin yürütülen meta-analiz araştırmalarında kullanılan parametrik olmayan tekniklerden farklı olarak *iyileşme oran farkı* (İOF [improvement rate difference-IRD]) tekniği kullanılmıştır. İOF, başlama düzeyi ve uygulama evresi arasındaki ilerleme oranı farkıdır. İOF'un (a) basit bir şekilde elle hesaplanabilir olma, (b) görsel analizle uyumlu olma ve (c) bilinen bir örnekleme dağılımının varlığı nedeniyle güven aralığına sahip olma gibi olumlu özellikleri bulunmaktadır (Parker vd., 2009). Araştırmada etki büyüklükleri http://www.singlecaseresearch.org adresindeki İOF hesaplama aracı kullanılarak hesaplanmıştır (Vannest vd., 2011). Etki büyüklüğünün %50 ve altında olması durumunda uygulama etkisinin küçük, %50-70 arasında olması durumunda orta büyüklükte, %70 ve üzerinde olması durumunda ise büyük olduğuna karar verilmiştir (Parker vd., 2009).

Araştırmada her bir katman (tier) için etki büyüklükleri ayrı ayrı hesaplanmıştır. Bu amaçla grafiklerdeki her bir katmanda yer alan veriler, grafiksel verilerin dijitalleştirilmesinde etkililiği deneysel olarak ortaya konmuş bir yazılım programı olan UnGraph5 kullanılarak dijital veriye dönüştürülmüştür. UnGraph5 ile dijitalleştirilen veriler etki büyüklüğü hesaplaması yapmak üzere ileri analizler yapabilmek için Microsoft Excel dosyasına aktarılmıştır (Tekin-İftar vd., 2018).

Bilimsel-Dayanağın Belirlenmesi

Etki büyüklüğü hesaplamalarının ardından sosyal öykülerin bilimsel dayanaklarının oluşup oluşmadığını değerlendirebilmek için Kratochwill ve diğerlerinin (2013) önerdiği "*5-3-20 kuralı*" olarak da bilinen üç ölçüt temel alınmıştır. Bu ölçütler (a) *model standartlarını karşılayan* ve *model standartlarını koşullu karşılayan* güçlü ya da orta düzey etkiye sahip en az beş araştırmanın olması, (b) uygulamaların farklı üç coğrafi bölgeden, yazarlarının örtüşmediği en az üç araştırmacı grubu tarafından yürütülmüş olması ve (c) bu araştırmalardaki toplam katılımcı sayısının en az 20 olması olarak belirlenmiştir.

Güvenirliklerin Hesaplanması

Araştırmada her bir aşama için güvenirlik hesaplamaları yapılmıştır: (a) Dâhil etme ve dışlama süreci, (b) Tek Denekli Müdahaleler Model Standartlarına ilişkin analiz, (c) Betimsel analiz, (d) UnGraph 5 ile dijitalleştirilmiş veriler ve (e) İOF hesaplamaları. Bu amaçla her bir aşamada iki araştırmacı tüm araştırmaları birbirinden bağımsız olarak kodlamışlardır. Araştırmada kodlayıcılar arasındaki tutarlılığı belirlemek üzere "Kodlayıcılar arası görüş birliği / (Kodlayıcılar arası görüş birliği + Kodlayıcılar arası görüş ayrılığı) x 100" (Tekin-İftar vd., 2018) formülü kullanılmıştır. Kodlayıcılar arası güvenirlik için kabul edilebilir ölçüt %80 ve üzeri olarak belirlenmiştir. Dâhil etme ve dışlama ölçütleri için kodlayıcılar arası güvenirlik %95.74 (ranj = 85.71-100), Tek Denekli Müdahaleler Model Standartlarına ilişkin kodlamalara yönelik güvenirlik %100, betimsel analize ilişkin güvenirlik %96.10 (ranj = 90.9-100), UnGraph 5 ile dijitalleştirilmiş verilere ilişkin güvenirlik %99.52 (ranj = 97.2-100), İOF hesaplamalarına ilişkin güvenirlik ise %100 olarak bulunmuştur.

Bulgular

Tek-Denekli Araştırmaların Niteliksel Göstergeleri

Dâhil etme ölçütlerini karşılayan 32 araştırma meta analize alınmıştır. Bu araştırmaların 19'u (%67.86) niteliksel göstergeler için belirlenen ölçütleri karşılayamamıştır. Niteliksel göstergeleri karşılayamayan araştırmalardan biri gözlemciler arası güvenirlik katsayısı %80'nin altında olduğu için (Dodd vd, 2008), sekizi uygulama güvenirliği verileri toplanmadığı ya da uygulama güvenirliği katsayısı %80'nin altında olduğu için (ör., Khantreejitranon, 2018; Moudry-Quilty, 2007), ikisi deneysel etkinin en az üç gösterimi sağlanamadığı için (ör., Hung & Smith, 2011), sekizi ise birden fazla ölçütü karşılayamadığı için standartları karşılamaz olarak kodlanmıştır. Bu 19 araştırma görsel analize alınmamıştır. Niteliksel göstergeleri karşılayan araştırmalardan biri (%3.57) MS, sekizi ise (%28.57) standartları koşullu karşılayan olarak kodlanarak görsel analize dâhil edilmiştir. Görsel analize alınan araştırmalardan altısı (%66.67) görsel analiz için belirlenen tüm ölçütleri karşıladığından "güçlü etki (dayanak)", üç araştırma (%33.33) da etkisiz olarak kodlanmıştır. Güçlü etki olarak kodlanan altı araştırma için etki büyüklüğü hesaplanmıştır.

Kapsamlı Betimsel Analiz Bulguları

Demografik Özellikler

Katılımcıların Sayısı ve Yaşı. Analizi gerçekleştirilen 12 araştırmanın her birinin üç katılımcı ile gerçekleştirildiği dolayısıyla tüm katılımcı sayısının 36 olduğu görülmüştür. Yaşları 2 ile 17 arasında değişen 36 katılımcı, 0-5 yaş arası okul öncesi, 6-11 yaş arası ilköğretim birinci kademe ve 12 yaş üstü ikinci kademe olarak üç grupta incelenmiştir.

Cinsiyet ve Tanı. Katılımcı özellikleri cinsiyet açısından incelendiğinde ise katılımcıların üçünün kadın (%8.4), 33'nün erkek olduğu görülmüştür (%91.6). 12 araştırmanın dokuzunda (ör., Olçay-Gül & Tekin-İftar, 2016) OSB tanılı bireylerin katılımcı olduğu görülürken (N = 26, %72.2), iki araştırma (Sansosti & Powell-Smith, 2006; Scattone vd., 2006) Asperger Sendromu tanılı bireylerin katılımı ile gerçekleştirilmiştir (N = 6, %16.6). Sadece bir araştırmada ise (Kim vd., 2014) katılımcıların ağır düzeyde zihinsel yetersizlik tanıları bulunduğu belirtilirken, (N = 3, %8.3); bir araştırmanın bir katılımcısının da orta düzeyde zihinsel yetersizlik tanısının bulunduğu ifade edilmiştir (Bıçakçı & Olçay-Gül, 2019).

Beceri. Analiz edilen araştırmaların bağımlı değişkenleri genel olarak uygun davranışların artırılması ve uygun olmayan davranışların azaltılması olmak üzere iki gruba ayrılabilmektedir. 12 araştırmanın dokuzunda (%75) uygun davranışların artırılması bağımlı değişken olarak ele alınmışken, iki araştırmada (%16.6) uygun olmayan davranışların azaltılmasının hedeflendiği görülmüştür (Özdemir, 2008b, Scattone vd., 2002). Bir araştırma ise (Kim vd., 2014) uygun olmayan davranışların azaltılmasını hedef almakla birlikte aynı zamanda katılımcıların akademik çalışmalarla meşgul olma düzeylerinde de artış hedeflemiştir.

Öğretim Ortamı. Araştırmaların tümünde öğretim ortamı okul içinde farklı alanlar olarak belirtilmişken sadece bir araştırmada (Olçay-Gül & Tekin-İftar, 2016) ev ortamında öğretim sunulurken; bir araştırmada ise tek bir katılımcının öğretiminin ev ortamında gerçekleştirildiği belirtilmiştir (Kurt & Kutlu, 2019).

Öğretim Düzenlemesi. Tüm araştırmalarda öğretim oturumları her bir katılımcı ile bire bir öğretim düzenlemesi biçiminde gerçekleştirilmiştir.

Yöntemsel Özellikler

Araştırma Modeli. Araştırmaların altısı (%50) katılımcılar arası çoklu başlama düzeyi modeli ile desenlenmişken (ör., Graetz vd., 2009), üç araştırmada (%25) katılımcılar arası çoklu yoklama modeli (Delano & Snell, 2006; Kim vd., 2014) ve iki araştırmada ise (%16.6) çiftler arası çoklu yoklama modeli kullanılmıştır (Olçay-Gül & Tekin-İftar, 2016). Bir çalışmada ise katılımcılar arası eş zamanlı olmayan çoklu başlama düzeyi modeli kullanılmıştır (Kurt & Kutlu, 2019).

Uygulamacı. Sosyal öykü öğretim oturumlarının üç araştırmada (ör., Delano & Snell, 2006), birinci araştırmacılar tarafından gerçekleştirildiği belirtilmiştir (araştırmacı, özel eğitim öğretmeni ve iş-uğraşı terapisti). Bir araştırmada öğretimi ikinci araştırmacı yürütürken (Kurt & Kutlu, 2019), dört araştırmanın üçünde öğretim oturumlarının öğretmen ya da yardımcıları tarafından, ikisinde ise özel eğitim personeli tarafından gerçekleştirildiği görülmektedir (ör., Scattone vd., 2006) Bir araştırmada öğretim oturumları akranlar tarafından

gerçekleştirilirken (Bıçakçı & Olçay-Gül, 2019), iki araştırmada ise öğretim oturumları ebeveynler tarafından yürütülmüştür (Olçay-Gül & Tekin-İftar, 2016; Sansosti & Powell-Smith, 2006).

Sosyal Geçerlik. Araştırmaların tümünde sosyal geçerliğe ilişkin veri toplandığı görülmektedir. Dokuz araştırmada görüşme ya da derecelendirme ölçekleri kullanılarak ailelerden ve/veya eğitimcilerden görüş alınırken (ör., Kim vd., 2014) üç araştırmada (ör., Delano & Snell, 2006) sosyal geçerlik verileri sosyal karşılaştırma aracılığıyla toplanmıştır.

İzleme Oturumları. Araştırmaların 10'unda bağımlı değişkenlerin kalıcılığına ilişkin izleme verileri toplanmışken, iki araştırmada (ör., Scattone vd., 2006) izleme oturumlarına yönelik herhangi bir bilgi verilmemiştir.

Genelleme Oturumları. Araştırmaların altısında genelleme oturumları planlanmış ancak altı araştırmada (ör., Özdemir, 2008b) genelleme oturumlarına yer verilmemiştir. Genelleme oturumlarına yer veren araştırmaların ikisinde hem ortamlar hem de kişiler arası genellemeye yer verilmiştir (ör., Delano & Snell, 2006). Bir araştırmada (Graetz vd., 2009) sadece ortamlar arası genelleme verisi toplanırken, iki araştırmada ise (ör., Kim vd., 2014) farklı etkinlikler arası genellemeye ilişkin veri toplanmıştır.

Genel Etki. Sosyal öykülerin bağımlı değişken üzerindeki etkililiği 36 katılımcı açısından incelendiğinde, toplam 34 katılımcıda (%94.4) sosyal öykülerin etkili olduğu bulgusu elde edilmişken, iki katılımcıda (%5.5) sosyal öyküler etkili görülmemiştir. Etkililik bulguları araştırma bazında incelendiğinde ise, Sansosti ve Powell-Smith (2006) üç katılımcının birinde ve benzer şekilde Scattone ve diğerleri (2006) de üç katılımcıdan birinde sosyal öykü uygulamalarının hedef davranış üzerinde bir etkisi olmadığını ifade etmişlerdir.

Sosyal Öykü Uygulamasının Etkisi

Niteliksel göstergeleri karşılayan ve koşullu olarak karşılayan dokuz araştırmadan görsel analiz sonucunda güçlü etki olarak kodlanan altı araştırma için İOF kullanılarak etki büyüklüğü hesaplanmıştır. Her bir katman (tier) için etki büyüklüğü ayrı ayrı hesaplanmış, ardından her bir katılımcı için ortalama değere ulaşılmıştır. Tablo 4'te yer alan İOF değerleri incelendiğinde tüm araştırmalarda her bir katılımcı için ortalama İOF değerinin %70'in üzerinde olduğu görülmüş ve etki büyüklüğü yüksek olarak belirlenmiştir. Etki büyüklüğü hesaplaması yapılan altı araştırmadan üçünde sosyal öykülerin problem davranışların azaltılması (ör., Graetz vd., 2009), ikisinde sosyal becerilerin kazandırılması (ör., Olçay-Gül & Tekin-İftar, 2016), birinde ise hem problem davranışların azaltılması hem de sosyal becerilerin kazandırılması (Kim vd., 2014) üzerindeki etkililiği incelenmiştir. Sosyal öykülerin hem problem davranışların azaltılması hem de sosyal becerilerin kazandırılması üzerindeki etkililiğini inceleyen araştırma bulguları etki büyüklüğünün tüm katılımcılarda yüksek olduğunu göstermiştir.

Bilimsel Dayanağın Belirlenmesi

Bilimsel dayanağın belirlenmesinde 5-3-20 kuralı temel alınmıştır. Ölçütlerden biri güçlü ya da orta düzey etkili ile sınıflandırılan model standartlarını karşılayan ve model standartlarını koşullu karşılayan büyük ve orta düzey etkiye sahip en az beş araştırmanın olmasıdır. Araştırmada model standartlarını karşılayan ya da koşullu karşılayan altı araştırma bulunmaktadır. Ölçütlerden diğeri uygulamaların farklı üç coğrafi bölgeden, yazarlarının örtüşmediği en az üç araştırmacı grubu tarafından yürütülmüş olmasıdır. Model standartlarını karşılayan ya da koşullu karşılayan üç coğrafi bölgeden yazarları örtüşmeyen en az üç araştırmacı grubu tarafından yürütülmüş altı araştırma bulunmaktadır. Bu araştırmalardaki toplam katılımcı sayısının en az 20 olması ise diğer bir ölçüttür. Bu araştırmalarda yer alan toplam katılımcı sayısı 21'dir. Dolayısıyla sosyal öyküler 5-3-20 kuralını karşılamaktadır.

Tartışma

Bu araştırmada sosyal öykülerin uygun davranışların kazandırılması ve uygun olmayan davranışların azaltılması üzerindeki etkililiğini inceleyen araştırmaların kapsamlı betimsel ve meta-analizi amaçlanmıştır. Bu amaçla öncelikle dâhil etme ölçütleri belirlenmiş, bu ölçütleri karşılayan araştırmalar Kratochwill ve diğerleri (2013) tarafından belirlenen niteliksel göstergeler açısından analiz edilmiş, bu analiz sonucunda standartları koşullu karşılayan ve standartları karşılayan olarak kodlanan araştırmalar görsel analize alınmış, son olarak ise "güçlü etki (dayanak)" olarak sınıflanan araştırmalar için etki büyüklüğü hesaplamaları yapılmıştır. Sonuç olarak sosyal öykülerin farklı tanılara sahip bireylere uygun davranışların kazandırılması ve uygun olmayan davranışların azaltılmasındaki etkililiğini inceleyen ve dâhil etme ölçütlerini karşılayan 32 araştırma belirlenerek Kratochwill

ve diğerleri (2013) tarafından belirlenen niteliksel göstergeler açısından analiz edilmiştir. Niteliksel göstergelere yönelik yapılan analizler sonucunda dokuz araştırma standartları karşılayan ve standartları koşullu karşılayan olarak kodlanmış; bu araştırmalardan görsel analiz sonucunda güçlü dayanak olarak belirlenen altı araştırma için etki büyüklüğü hesaplaması yapılmıştır. Alanyazında sosyal öykülerin etkililiğini inceleyen araştırmalara yönelik yürütülen betimsel analiz ve meta analiz araştırmaları incelendiğinde sadece iki araştırmada etki büyüklüğü için analize dâhil edilecek araştırmaların belirlenmesinde niteliksel göstergelere başvurulduğu görülmüştür. Bu araştırmalardan birinde Test ve diğerleri (2011) Horner ve diğerleri (2005) tarafından belirlenen niteliksel göstergeleri temel almışlardır. Qi ve diğerleri (2018) tarafından yürütülen diğer araştırmada ise bu araştırmada da olduğu gibi Kratochwill ve diğerleri (2013) tarafından belirlenen ölçütler temel alınmıştır. Sözü edilen bu araştırmada yalnızca Kratochwill ve diğerleri (2013) tarafından belirlenen niteliksel göstergelere göre değerlendirme yapılması ise bir sınırlılık olarak belirtilmiştir. Bu araştırmada Horner ve diğerleri (2005) tarafından belirlenen göstergeler arasında yer alan ancak Kratochwill ve diğerleri tarafından belirlenen göstergeler arasında yer almayan bir ölçüt olarak uygulama güvenirliğine ilişkin değerlendirmelere yer verilmiştir. Konu ile ilgili meta analiz araştırmalarında sosyal öykü araştırmalarında uygulama güvenirliği verilerinin toplanılmamış olmasının bir sınırlılık olduğu ve toplanılması gerektiği dile getirilmiş olmasına (Reynhout & Carter, 2006; Test vd., 2011) rağmen araştırmalar bu ölçüt açısından değerlendirmeye alınmamıştır. Bu bağlamada araştırmada niteliksel göstergeler açısından farklı ölçütleri bir araya getiren daha ayrıntılı bir değerlendirme sürecinin yürütülmüş olmasının bu araştırmanın güçlü yönlerinden biri olduğu düşünülmektedir.

Araştırmada sosyal öykülerin hem uygun davranışların kazandırılması hem de uygun olmayan davranışların azaltılması üzerindeki etkililiğini inceleyen araştırmalar analize alınmış; araştırma sonucunda ileri analizlere dâhil edilmesine karar verilen yedi araştırma için etki büyüklüğü hesaplaması yapılmıştır. İOF ile etki büyüklüğü hesaplaması yapılan yedi araştırmanın dördünde sosyal öykülerin uygun davranışların kazandırılması (Graetz vd., 2009; Olçay-Gül & Tekin-İftar, 2016; Thompson & Johnston, 2013), ikisinde uygun olmayan davranışların azaltılması (Özdemir, 2008b; Scattone vd., 2002), birinde ise hem uygun davranışların kazandırılması hem de uygun olmayan davranışların azaltılması (Kim vd., 2014) üzerindeki etkililiği incelenmiştir. Etki büyüklüğü sonuçları hem uygun davranışların kazandırılması hem de uygun olmayan davranışların azaltılmasında sosyal öykülerin etki büyüklüğünün yüksek olduğunu göstermiştir. Bu bulgu sosyal öykülerin uygun olmayan davranışların azaltılmasında sosyal becerilerin kazandırılmasına kıyasla daha etkili olduğunu ortaya koyan meta analiz araştırmalarının bulgularından farklılık göstermektedir (Kokina & Kern, 2010; Qi vd., 2018). Bu durumun nedeninin ise araştırmalarda hem farklı dâhil etme ölçütleri ve niteliksel göstergelerin belirlenmesinden hem de etki büyüklüğü hesaplamasında kullanılan parametrik olmayan tekniklerden kaynaklanabileceği düşünülmektedir.

Araştırmada bilimsel dayanağın belirlenmesinde 5-3-20 kuralı temel alınmıştır. 5-3-20 kuralı temel alınarak yapılan değerlendirmede standartları karşılayan ve standartları koşullu karşılayan olarak kodlanan ve güçlü ya da orta düzey etki ile sınıflandırılan sosyal öykülerin etkililiğini inceleyen en az beş araştırmanın olduğu, bu araştırmaların farklı üç coğrafi bölgeden, yazarlarının örtüşmediği en az üç araştırmacı grubu tarafından yürütüldüğü ve sosyal öykü araştırmalarının toplam katılımcı sayısının 21 olduğu görülmüştür. Dolayısıyla sosyal öykülerin bilimsel dayanağının oluştuğundan ve çalışmanın alanyazınındaki raporlarla (NAC, 2009; NCAEP, 2020; NPDC, 2014) tutarlılık gösteren ilk meta analiz çalışması olduğundan söz edilebilir. Bu bulgu sosyal öykülere yönelik olarak yürütülen ve niteliksel göstergelerin temel alındığı meta analiz araştırmalarının bulgularından farklılık göstermektedir. (Qi vd., 2018; Test vd., 2011). Bu araştırmanın Qi ve diğerleri (2018) tarafından yürütülen meta analiz çalışmasından farklılaşan yanı farklı tanılara sahip bireylerle yürütülmüş olması, Test ve diğerleri (2011) tarafından yürütülen araştırmadan farklılaşan yanı ise sosyal öykülerin tek başına etkililiğini inceleyen araştırmalara yer verilmiş olmasıdır. Araştırmada OSB tanısı dışında farklı tanılara sahip olan bireylerle yürütülen araştırmaların analize alınmasının ve sosyal öykülerin salt etkisini ortaya koyan araştırmalara odaklanılmasının bu araştırmanın özgün ve güçlü yönünü oluşturduğu söylenebilir. Sonuç olarak, sosyal öyküler bilimsel dayanaklı uygulamalar arasında yer almasına karşın, bu konuya ilişkin nitelikli tasarlanmış araştırma gereksinimi devam etmektedir.

Yazarların Katkı Düzeyleri

Araştırmanın tarama ve dahil etme-dışlama süreçlerinde her üç araştırmacı da rol almıştır. Birinci araştırmacı çalışmanın giriş bölümünü yazmıştır ve ikinci araştırmacıyla birlikte yöntem bölümünü yazmışlar ve meta-analiz sürecini yürütmüşlerdir. Üçüncü araştırmacı ikinci araştırmacıyla beraber betimsel analiz sürecini yürütmüştür ve betimsel analiz bulgularını raporlamıştır. Her üç araştırmacı da tartışma bölümüne katkı sağlamıştır.