

Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for SEND

How Effective are Computer-based Social Stories Intervention for Primary School Aged Children diagnosed with Autistic Spectrum Disorder to improve social skills?

Summary

This systematic review aims to explore the effectiveness of computer-based social stories (SS) for children with Autistic Spectrum Disorder (ASD). SS were created by Gray (1998) as an intervention for children with ASD to improve their social skills. Gray (2000) described SS as a combination of tailored short stories to support children with ASD to understand the world around them and what behaviours are expected. Computer-versions have increased in popularity for children with ASD but there is not yet a review of the evidence.

Five studies met the inclusion criteria and were reviewed following Gough's (2007) weight of evidence framework using Gersten et al.'s (2005) and Horner et al.'s (2005) protocols for methodological quality. Findings suggest that computer-assisted SS are effective in improving target behaviours associated with social skills for children with ASD. These effects were also maintained at follow-up. Interestingly, paper-based SS were very effective too suggesting that preference for a particular format may be down to practicalities and resources. Differences in interventions are discussed along with implications for use and future directions.

Introduction

The need for schools to support children with ASD

Autistic spectrum disorder (ASD) is diagnosed when a person has persistent difficulties with social communication and interaction and has restricted and repetitive patterns of behaviour, as specified in the DSM-V (APA, 2013). It is widely acknowledged that people with ASD share these specific difficulties, but their diagnosis will affect them all in very different ways and impair their everyday functioning (Rogers & Ozonoff, 2005). The diagnostic rate of ASD has rapidly increased in recent years and as such it is suggested that 1 in 68 children are currently being diagnosed, although this figure is widely debated (CDC, 2014). There has been consideration within the literature as to whether this is caused by an increased awareness or the broadening of the criteria rather than an increase in prevalence (Gillberg, 1998; Wing & Potter, 2002). Nonetheless, Neitzel (2010) found that children with ASD are more likely to present with challenging behaviours that will prevent them from learning in school and will often feel isolated because of their social communication difficulties. Whilst Myers and Johnson (2007) state that there is not currently a cure for ASD, school interventions can improve their skills and as such improve their quality of life.

Social stories and its evidence base

Social Stories (SS) were created by Gray (1998) as an intervention for children with ASD to improve their social skills (Green et al. 2006); and prosocial behaviour (Crozier & Tincani, 2007). Gray (2000) describes SS as a combination of tailored short stories to support children with ASD to understand the world around them and what behaviours are expected. Relevant information about the specific scenario needs to be included such as where and when it happens, who is involved,

what they are doing and why (Gray, 2004). Green et al. (2006) found that SS were widely used in schools for children with ASD because of its cost-effectiveness, ease of administration and ability to be personalised to support real-world problems for the individual child. Gray (1999) developed key characteristics of SS to ensure that intervention quality was upheld. In 2010, Gray restructured the original characteristics and developed 10 criteria for creating SS which is widely used today (see appendix A).

There has been considerable research to provide evidence of SS as an effective intervention to support social skills and pro-social behaviour. Overall, much of the literature on SS has tended to be single subject study designs and therefore systematic reviews have attempted to draw conclusions from these studies. Karkhaneh et al. (2010) conducted a qualitative analysis of 6 controlled trials and all but one of the trials revealed good effectiveness for improving social interactions. This review highlights that SS has promising impact for children with ASD but did reveal that more rigorous research is needed and an exploration into the maintenance and generalisability of the intervention.

Kokina and Kern (2010) conducted a meta-analysis comprising 18 single subject studies. The results of the studies varied widely, likely because of the variability in how the SS intervention was implemented (i.e. frequency and length of session) and the research methodology chosen (i.e. single subject or group experimental design). Interestingly, the results showed that SS were more effective when preventing poor behaviour than when attempting to encourage pro-social behaviour which is important to consider when designing SS. Ali and Frederickson (2006) conducted a systematic review for SS and found overall significant outcomes across studies, although most were administered alongside other interventions too

which made it difficult to draw conclusions. These systematic reviews provide some promising evidence for SS, although findings have been mixed. Up to 72% of the research on SS were conducted on children aged between 6-10 years which suggests a need for further research on a range of ages beyond this.

Social stories in computer-format

Over the last 20 years there have been rapid advances in technology and as such computers are widely available in schools and children's homes (Barron, Harmes, & Kemker, 2006). Because of this, SS have recently been presented in digital format. Gray (2010) recommends a presentation style that meets the needs of the child with Ploog et al. (2013) finding that most children with ASD show a preference towards computer-based instruction. This suggests there could be a greater impact compared to paper-versions. Golan and Baron-Cohen (2006) argue that computer interventions are highly standardised and yet can be individualised so that children can work at their own pace and ability. This is useful in the classroom where teachers have limited time; although each child would need access to their own device which could be costly.

Goldsmith and Leblank (2004) found that computer-based activities increased motivation, attention and learning for children with autism which suggests computer-based SS could have a greater impact compared to the original version. However, children could be distracted by other apps on the device. Hagiwara and Smith Myles (1999) introduced the first SS in a computer format for 3 males aged 7-9 years. Results were varied and depended on the duration of the intervention, individual differences, target behaviours and consistency of the environment. Since then,

studies have started to explore the effectiveness of computer-based SS. Despite this, Gray (2012) found that up to 85% are still presented in paper format.

Psychological basis

Kokina and Kern (2010) found within their systematic review that SS area useful intervention that complements the strengths and supports the difficulties associated with ASD. The weak central coherence theory (Happe, 2005) argues that people with ASD have a specific perceptual cognitive style and therefore struggle to understand context as they focus on details rather than 'seeing the bigger picture'. SS builds upon these respective difficulties: the step by step story of a situation emphasises the minor details that occur so the child with ASD can understand what is important. Quill (1997) also argues that children with ASD have a preference for visual stimuli because of their difficulties with language. The pictorial SS scenarios build upon these preferences.

Furthermore, it is widely acknowledged that children with ASD have difficulties with their 'theory of mind' which is the ability to understand what others are thinking and feeling (Korkmaz, 2011). Gray (1994) created SS with this difficulty in mind and provided children with autism with the missing information to help them understand what is happening in the scenario and how others might be thinking. Because of a lack of theory of mind, the world can seem unpredictable and anxiety provoking (Lipsky, 2011). Moore and Sosa (2013) argue that computer programs generate clear routines and reduce distractions whilst Charlop, et al. (2000) found that the repetitive and predictable routine associated with computers reduces anxiety for children with ASD who typically struggle with change. This therefore suggests that

computer-based SS have potential to be an effective intervention for children with ASD.

Rationale and Relevance to Educational Psychology

Given the significant increase in ASD diagnosis, schools will be under greater pressure to be inclusive and provide support to cater for the needs of children with ASD. EPs must support in this process by recommending evidence-based interventions (Rosenweig, 2009). Whilst research for children with ASD has grown over the last 30 years including the use of SS, Burgess and Gutstein (2007) found that social outcomes for children with ASD remain poor. Therefore, as technological advancements have increased the use of technology in schools, research must be available to determine its effectiveness for use as an intervention, specifically combined with SS. Therefore, a systematic review of computer-based SS will allow EPs to be equipped with the knowledge of its effectiveness to recommend in their practice.

Review question

Therefore, the current systematic review will aim to address the following research question *“How effective are computer-based SS intervention for primary school aged children with Autistic Spectrum Disorder to improve social skills?”*

Critical review of the evidence

Literature searching

On 15th December 2017, a literature search was carried out in Psych Info, SCOPUS, and Eric databases in order to ensure access to published articles involving psychology and education. The following search terms were used:

"social stor**"

AND

(ASD or "autis* spectrum disorder*" or autism)

AND

(Computer or internet or online or web-based)

The search returned 31 text results (PsychInfo, 15; SCOPUS, 13; and Eric, 3) including 9 duplicates (see appendix B for reference list). The title and abstracts of the remaining 22 were reviewed against the inclusion and exclusion criteria, and full text reviewed if needed (See table 1).

Table 1
Inclusion and Exclusion Criteria

No	Factor	Inclusion criteria	Exclusion Criteria	Rationale
1	Participants	Sample must include children aged between 5-11 years	Sample does not include children aged between 5-11 years.	The majority of the research is tested on this age group and thus useful to provide a summary of the majority of the literature on this topic. Additionally, this intervention is suitable for primary-aged children. Studies may be included if they also include outside this age range, as long as they are predominantly assessing primary-aged children.
2	Participants	Diagnosis of ASD	No diagnosis	SS intervention was designed for children with ASD. The outcomes are typically improving social skills.
3	Intervention	Computer based SS	Not used computer-based SS	The research question is specifically looking at computer-based SS because the research suggests children have a preference for computers
4	Intervention	SS follows Gray's (2000) criteria	Not referenced Gray's criteria	As SS are created by individual teachers, it is important to follow a criterion to ensure that the intervention is similar to the evidence base. This also ensures consistency. Gray's criteria is highly regarded in the literature.
5	Setting	Conducted in school	Delivered in any other setting	The research question is specifically exploring the use of a school-based intervention to support ASD children. In school, it is also easier to measure fidelity of the intervention.
6	Study design	Quantitative studies with pre and post measures	Qualitative studies	Need to measure effectiveness of intervention and easier to compare quantitative studies against each other.
7	Date of publication	Studies since 2007	Any studies before 2007	Wanted to look at the research over the last 10 years.
8	Publication type	Published journal article or thesis	Literature that is not published in these forms	These publications would have undergone a rigorous ethics procedure whereas other studies may have not had the same level of ethics approval and supervision.
9	Language	Written in English	Not written in English	The researcher is monolingual and cost and time constrains do not allow for translation.

17 of the articles were removed because they did not meet the inclusion and exclusion criteria (see appendix C). See figure 1 for breakdown of the exclusion process.

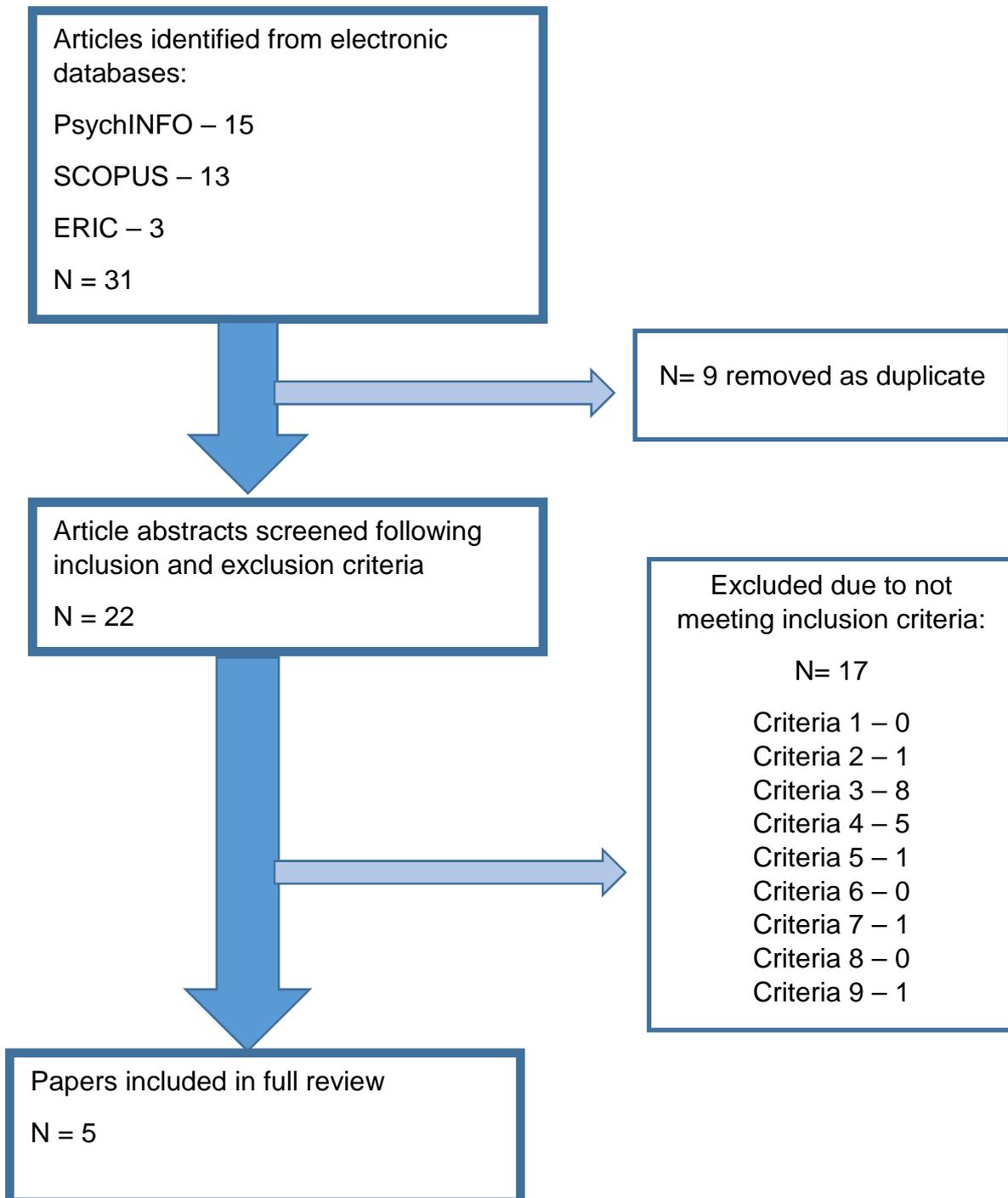


Figure 1: *Flow chart of the exclusion process*

The remaining 5 studies (see table 2) were fully reviewed and included in the review.

Table 2: *List of References of Articles That Have Been Fully Reviewed*

No	Reference
1	Chan, J. M. (2009). <i>Pre-service teacher-implemented social stories' intervention for students with autism spectrum disorders in general education settings</i> (Doctoral dissertation, The University of Texas at Austin).
2	Mancil, G. R., Haydon, T., & Whitby, P. (2009). Differentiated effects of paper and computer-assisted Social Stories™ on inappropriate behaviour in children with autism. <i>Focus on Autism and Other Developmental Disabilities</i> , 24(4), 205-215.
3	Pop, C. A., Simut, R. E., Pinte, S., Saldien, J., Rusu, A. S., Vanderfaeillie, J., ... & Vanderborght, B. (2013). Social robots vs. computer display: does the way social stories are delivered make a difference for their effectiveness on ASD children?. <i>Journal of Educational Computing Research</i> , 49(3), 381-401.
4	Thompson, R. (2016). <i>Social Stories: An Intervention Tool to Help Decrease Undesired Behaviors in Children with Characteristics of Autism Spectrum Disorders</i> (Doctoral dissertation, Department of Special Education, University of Utah).
5	Vandermeer, J., Beamish, W., Milford, T., & Lang, W. (2015). iPad-presented social stories for young children with autism. <i>Developmental Neurorehabilitation</i> , 18(2), 75-81.

Weight of evidence

The final 5 studies were evaluated using Gough's (2007) Weight of Evidence (WoE) framework. This comprises three sections: WoE A assesses methodological quality; WoE B assesses methodological relevance and WoE C assesses topic relevance. Each section was given a score out of 3 and then averaged for the WoE D which is the overall WoE score.

The WoE A score was calculated using one of two protocols depending on the study design. Pop et al. (2013) is a group experimental design and was critically evaluated using Gersten et al.'s (2005) protocol (see appendix D). The other four studies were single subject designs and thus critically evaluated using Horner et al.'s (2005) protocol (see appendix E). WoE B and WoE C scores were calculated by a

checklist created solely based on the relevance to the research question (see appendix F-J).

The overall WoE D was the average of the WoE A, B and C scores. It was decided to give each of the scores equal weighting because it was equally important to have a robust study design as it was to be relevant to the research question. The WoE D scores were given a description of the weighting: 2.5 or above = high; 1.5 – 2.4 = medium; 1.4 or below = low. This criteria ensured that an article would only be given a high score if they scored high in at least 2 of the WoE sections. See table 3 for weight of evidence scores.

Table 3
WoE Scores

Study	WoE A - Methodological Quality	WoE B - Methodological Relevance	WoE C - Topic Relevance	WoE D – Overall score
Chan (2009)	2.28 (medium)	1.6 (low)	2.25 (medium)	2.06 (medium)
Mancil et al.(2009)	2.85 (high)	2 (medium)	2.75 (high)	2.48 (high)
Pop et al. (2013).	3 (high)	2.6 (high)	2 (medium)	2.53 (high)
Thompson (2016)	2.42 (medium)	2.3 (medium)	2 (medium)	2.24 (medium)
Vandermeer et al. (2015)	2.5 (high)	1.3 (low)	2.25 (medium)	2.01 (medium)

Note: 1.4 or less = low; 1.5-2.4 = medium; 2.5 or above = high.

Participant details

Participants were aged 0-13 years (based on the inclusion criteria). Mancil et al. (2009) was the only study that only included children at primary-age (5-11 years) which was reflected in a higher WoE-C rating. This suggested that future research could broaden the age range within the literature if computer-versions were deemed effective, especially for children aged over 13 years. Four of the five studies included

more boys than girls which is unsurprising as the population of children with ASD is typically much larger for boys, with a ratio of 4:1 (Werling & Gershwind, 2013). Pop et al. (2013) did not state gender in specific numbers but did state that groups did not differ in terms of gender. None of the studies commented on the social economic status of the participants thus it is difficult to draw a conclusion on this demographic.

All children across the studies had a diagnosis of ASD (as referenced in the inclusion/ exclusion criteria). All of the studies selected participants based on the children's level of functioning assessed through a range of questionnaires. One study used the Social Communication Questionnaire and the Autism Diagnostic interview (Pop et al. 2013); Vandermeer et al. (2015) used the Vineland Adaptive Behaviour Scale, Mullen Scales of Early Learning and the Preschool Language scale and two of the studies assessed through the Children's Autism Rating scale and the Gilliam Autism Rating scale (Mancil et al. 2009; and Chan, 2009). Thompson (2016) used functional assessment. All studies stated that the children chosen were significantly below age appropriate levels in their social skills and thus were not scores differently in the weightings.

Sample size

36 children participated in the SS intervention across the studies. All of the studies except Pop et al. (2013) had very small samples because they were single subject designs and therefore Pop et al. (2013) received a higher WoE-B score. The four single subject designs were evaluated for their effectiveness by calculating a percentage of non-overlapping data (PND). Cohen (1992) was used to assess for the power of Pop et al. (2013) as this was the only group experimental design. As

this study only had 20 participants, it was underpowered which needs to be considered in the analysis.

Setting

Three of the studies were conducted in the USA (Chan, 2009; Mancil et al. 2009; and Thompson, 2016) and thus given a higher score on WoE C as it is highly relevant to the UK education system. Vandermeer et al. (2015) was conducted in Australia and Pop et al. (2013) was conducted in Romania which is slightly different to the UK education system and thus scored lower on WoE C. However, all countries were part of the Organisation for Economic Co-operation and Development (OECD) organisation which is considered broadly similar to the UK system. Therefore, positive effects across the studies would suggest that the intervention can be generalised to different countries and as such would be promising to the UK education system. There was a mixture of provisions that the intervention was implemented in across the studies. Mancil et al. (2009) and Chan (2009) were both set in mainstream schools whilst Pop et al. (2013), Vandermeer et al. (2015) and Thompson (2016) were set in special educational needs school. Thompson (2016) and Pop et al. (2013) also conducted the intervention outside of the classroom; these studies scored lower on the WoE C as they were less inclusive and required more resources. Reviewing studies from a range of settings increases external validity and shows whether the intervention is effective in the different provisions.

There was a mixture of experience and training given to the teachers that were delivering the interventions across the studies. Vandermeer et al. (2015), Mancil et al., (2009) and Thompson et al. (2016) used the children's current teacher whom had some experience with autism and/or special educational needs and had

received some training on social stories. Pop et al. (2013) used experienced psychotherapists and Chan (2009) utilised undergraduate students with experience in Applied Behaviour Analysis (ABA) training and ASD. These two studies scored lower on the WoE C because they were a more expensive resource.

Research design

All of the studies used a quantitative method. Four out of the five studies were single subject designs. Whilst this design is regarded as less superior in the hierarchy of evidence (Wolpert et al. 2004) and thus given a lower score on WoE A, single case designs are useful as they integrate both research and practice providing practical and conceptual advantages (Barker, Pistrang, & Elliott, 2002). Chan et al. (2009), Vandermeer et al. (2015) and Thompson (2010) had a multiple baseline single subject design whilst Mancil et al. (2009) administered an ABABCBC design. Pop et al. (2013) was a group experimental design that randomly allocated participants to one of three groups: computer-based SS, robot-based SS or control group and therefore given a higher rating in the WoE A score to reflect these differences. Pop et al. (2013) and Thompson (2016) also conducted a follow-up to assess whether impact had maintained after the intervention. These two studies were given a higher score in the WoE B criteria as they explored maintenance.

Treatment fidelity

All studies used SS as the intervention but varied significantly in terms of content, frequency, implementation and target behaviour. However, each SS was written based on Gray's (2010) criteria (as identified in the inclusion criteria). Assessment of treatment fidelity is important to ensure that the intervention is implemented accurately and consistently. Chan (2009), Mancil et al. (2009) and Pop

et al. (2013) used a checklist to mark whether each step was satisfactory. Thompson (2016) did not measure treatment fidelity specifically but used the same researcher to deliver the intervention who was highly experienced to ensure accuracy and consistency. Finally, Vandermeer et al. (2015) made subtle changes to their intervention to ensure treatment integrity i.e. providing clear and transparent details of implementation and making adjustment to seating conditions. As all of the studies considered treatment fidelity, none needed to be weighted differently.

Social validity:

Social validity is important within research to ensure that the intervention is relevant, important and useable. Three out of the five studies commented on social validity. Chan (2009) administered a social validity questionnaire to teachers which addressed issues such as relevance of target behaviours, ease of implementation, and motivation to use the intervention in the future and found positive results. Mancil et al. (2009) also administered a social validity scale which found that teachers felt SS did not use up too much time, that they would use it in the future and that the computer-version was better than the paper versions. Finally, Thompson (2016) used a combination of scaling, video ratings and teacher surveys to elicit social validity and found overall positive results. The other two studies did not comment on social validity which was reflected in their lower WoE-A score.

Measures:

There was a range of target behaviours that were being measured. Chan (2009) and Pop et al. (2013) assessed target behaviours that were aiming to improve social communication skills i.e. appropriate social responses and eye contact, and

therefore received a high WoE-C Score. Vandermeer et al. (2015) and Mancil et al. (2009) assessed target behaviours that were relevant to the classroom but not directly related to social communication i.e. on task behaviour and level of pushing and therefore received a slightly lower WoE-C score. Thompson (2016) was not clear on each individual's target behaviour and thus was scored lower on the WoE-C. There were a range of measures used to assess for the target behaviours. Four of the studies used a time sampling observation to measure the individual target behaviours. Observation assessment measures have construct and ecological validity because it is representative of the participant's typical everyday behaviour (Gardner, 2000). However, observations could lack reliability due to observation bias; although studies commented on their inter-rater reliability. It is therefore useful to use a second measure to capture the results which two of the studies did and consequently received a higher score on the WoE-B: Thompson (2016) used functional assessment too and Chan (2009) used comprehension questions. Alternatively, Pop et al. (2013) used a 7 point Likert scale developed from the 5 point scale of Barry and Burlew (2004). This is a useful quantitative measure although as only this measure was used, it received a lower score on the WoE-B. Variation in measures across the studies is useful to show that SS can impact a variety of target behaviours. However, this makes replication and comparison across studies difficult.

Table 4: PNS scores

Study	PP	Mean: baseline	Mean: comp	PND	PND desc	Mean: paper	PND	PND desc	Mean follow-up	PND	PND desc	Outcome
Chan (2009)	C1	41	61	100%	VE	-	-	-	-	-	-	Appr sitting
	C2	29	41	100%	VE	-	-	-	-	-	-	Attention
	C3	29	38	100%	VE	-	-	-	-	-	-	Appr comment
	C4	11	6	88.5%	E	-	-	-	-	-	-	Appr comment
	C5	58	100	100%	VE	-	-	-	-	-	-	Eye contact
	C6	55	-	-	-	80	100%	VE	-	-	-	Attention
	mean	37.1	49.2	97.7%	VE	80	100%	VE	-	-	-	
Mancil et al. (2009)	C1 trans	9.5	1.25	75%	E	5.5	66%	VE	1.5	100%	VE	Level of pushing
	C1 clas	23.3	7.25	100%	VE	9.33	100%	VE	11.66	100%	VE	Level of pushing
	C2 trans	18.5	1.75	100%	VE	5.25	100%	VE	1,5	100%	VE	Level of pushing
	C2 clas	30	9.33	100%	VE	7.66	100%	VE	13	100%	VE	Level of pushing
	C3 trans	19.5	0	58%	QU	3.5	77.7%	E	0.25	100%	VE	Level of pushing
	C3 clas	17.66	8.6	100%	VE	8	100%	VE	9.3	100%	VE	Level of pushing
	mean	23.65	8.39	88%	E	8.33	90.6%	VE	11.32	100%	VE	
Thompson (2016)	C1	0.57	0.108	92.8%	VE	0.093	95.2%	VE	0.119	85.7%	E	Engaging in a task
	C2	0.508	0.159	92.3%	VE	0.148	90%	VE	0.184	91.6%	VE	Engaging in a task
	C3	0.762	0.15	100%	VE	0.118	100%	VE	0.074	100%	VE	Eating food socially appropriately
	C4	0.64	0.099	100%	VE	0.193	100%	VE	0.102	100%	VE	Raising hand for teacher attention
	mean	0.62	0.129	96.3%	VE	0.138	96.3%	VE	0.119	94.3%	VE	

Vandermeer et al. (2015)	C1	23.5	24.9	0%	IE	-	-	-	-	-	-	on-task
	C2	22.8	27.8	70%	QU	-	-	-	-	-	-	on-task
	C3	18.5	27	0%	IE	-	-	-	-	-	-	on-task
	mean	21.6	26.5	23.3	IE	-	-	-	-	-	-	on-task
	n											
			Mean overall	82%	E		93.5%	VE		97.7%	VE	
			IPDN									

Percentage of non-overlapping data (PND) determines to what extent the study has found that the intervention is effective (MacArthur, Graham, & Fitzgerald, 2006). See table 5 for breakdown of scores and descriptors.

Table 5:
PND Descriptors

Score	Descriptor	Initial
91% and above	Very Effective treatment	VE
71-90%	Effective treatment	E
51-70%	Questionable effectiveness	QU
Below 50%	Ineffective	IE

The four single subject design studies revealed that overall the computer-assisted SS intervention for children with ASD was effective in improving target behaviours. The PND overall mean was 82% with 2 out of the 4 single case studies showing very effective results overall. Vandermeer et al. (2015) results showed ineffective scores but explained that this was because the baseline scores were very high to start with. Also, only 1 of the 3 children showed substantial improvements following intervention and explained that this could have been because of the limited time within the intervention phase. This suggested that the measurements were not sensitive and the intervention was too short and thus this study should be treated with caution. Pop et al. (2013) was the only group experimental design. The results found that the mean at baseline for level of prompt needed to provide an expected social response for children with ASD was 2.41 and had reduced to 1.65 following the computer-

assisted SS. The effectiveness was calculated by Cohen's D which compares the mean difference between two groups in standard deviation units. A small effect would be 0.2; a medium effect would be 0.5; and a large effect would be 0.8. The post-test showed a medium effect between the computer-assisted SS and control group ($d=.33$). It is therefore important to note that the two studies with the highest weight of evidence (Pop et al. 2013; Mancil et al. 2009) show promising results for computer assisted SS.

Three of the five studies measured the effectiveness of paper-format SS too. Although there was relatively little difference in means between the two versions, the paper-version received an overall very effective PND score. This suggested that paper-versions might be more effective than computer-versions. Two of the five studies recorded follow-up data for the intervention which was reflected in their WoE-B score. Both studies showed that computer-format and paper-format SS effectiveness can be maintained. It must be noted that PND scores should be treated with caution because they do not measure the magnitude of the effect and therefore a very effective result does not necessarily depict that there were large gains following intervention.

Conclusion and Recommendations

This review evaluated five studies to depict the effectiveness of computer-assisted SS for children with ASD. Four of the five studies were single subject designs and one was a group experimental design. Overall, the studies showed that computer assisted SS has promising evidence in supporting children with ASD. The single subject design studies showed an overall effective result and the group experimental showed a small to medium effect. Of the three studies that assessed paper-format SS, this intervention also appeared to be highly effective. Both versions of SS showed positive long term outcomes (although only two studies assessed for this) which has promising implications for implementation in schools.

All of the studies had very small sample sizes which makes it difficult to generalise the results to determine if other children with ASD would find it equally effective. This is particularly pertinent as one study found that two of the three children found it to be ineffective. Furthermore, the majority of the sample consisted of boys and therefore the findings may not be able to be generalised to girls, especially as the research has found that there is large variability in characteristics of ASD across gender (Dworzynski, Ronald, Bolton, & Happé, 2012). The variance in results across the studies seemed to occur because of the large variance in how it was delivered such as type of school (mainstream or special); who delivered the intervention (psychologist or teacher) and how long it ran for and the frequency. Two of the studies were set in mainstream schools and showed effective to very effective results whilst two were set in special schools with only one showing to be effective. Future research could explore the reasons for this predicting that those in mainstream might be higher functioning. This may suggest that SS may be more

effective in children with higher functioning ASD. There seemed to be no relationships between the type of qualification of person who delivered the program and effectiveness which suggests that SS would be suitable to be implemented by their current teacher to save on resources. Interestingly, interventions that ran for longer tended to have better outcomes which is important to consider in future implementation. All of the studies adhered to Gray's criteria.

Therefore, the computer-based SS intervention has potential to be used in practice in schools to support children with ASD. As computer-based SS did not appear to be any more effective than paper-versions, choice of format should depend on preference, feasibility and resources. Computer-formats have the added benefits of being cost effective as multiple children can be using the intervention simultaneously on separate computers in class, with only one adult monitoring. The children can also work at their own pace and ability. However, the initial outlay may be more expensive than paper-versions. Therefore, EPs may wish to explore the use of either format to support children with ASD.

Despite its promising findings across the current literature, future research is needed to build upon the existing evidence base. Foremost, there needs to be an increase in large scale empirical research with a group experimental design to develop robust evidence and answer questions regarding generalisability. Furthermore, there needs to be an increase in studies that explore different age ranges to determine whether SS is as effective across different age ranges.

References

- Ali, S., & Frederickson, N. (2006). Investigating the evidence base of social stories. *Educational Psychology in Practice, 22*(4), 355-377.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
- Barker, C., Pistrang, N., & Elliott, R. (2002). Foundations of qualitative methods. *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners, Third Edition*, 73-95.
- Barron, A. E., Harmes, J. C., & Kemker, K. J. (2006). Technology as a classroom tool: Learning with laptop computers. *Handbook of Research on Literacy in Technology at the K-12 level*, 271-286.
- Barry, L. M., & Burlew, S. B. (2004). Using social stories to teach choice and play skills to children with autism. *Focus on Autism and Other Developmental Disabilities, 19*(1), 45-51.
- Burgess, A. F., & Gutstein, S. E. (2007). Quality of life for people with autism: Raising the standard for evaluating successful outcomes. *Child and Adolescent Mental Health, 12*(2), 80-86.
- Centers for Disease Control and Prevention. (2014). CDC (2014). *About the National Health and Nutrition Examination Survey*.
- Chan, J. M. (2009). *Pre-service teacher-implemented social stories' intervention for students with autism spectrum disorders in general education settings* (Doctoral dissertation, The University of Texas at Austin).
- Charlop-Christy, M. H., Le, L., & Freeman, K. A. (2000). A comparison of video modeling with in vivo modeling for teaching children with autism. *Journal of Autism and Developmental Disorders, 30*(6), 537-552.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-156.
- 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.
- Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders?. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(8), 788-797.
- Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review, 3*(3), 185-198.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164.
- Gillberg, C. (1998). Asperger syndrome and high-functioning autism. *The British Journal of Psychiatry, 172*(3), 200-209.
- Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology, 18*(2), 591-617.
- Goldsmith, T. R., & LeBlanc, L. A. (2004). Use of technology in interventions for children with autism. *Journal of Early and Intensive Behavior Intervention, 1*(2), 166- 172.

- Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213-228.
- Gray, C. (1994). *Comic strip conversations: Illustrated interactions that teach conversation skills to students with autism and related disorders*. Future Horizons.
- Gray, C. (2000). The new social stories book. *Arlington, TX: Future Horizons*, 5.
- Gray, C. (2004). Social stories 10.0: The new defining criteria and guidelines. *Jenison Autism Journal*, 15(4), 2-21.
- Gray, C. (2012). What are Social Stories™. *Using storytelling to support children and adults with special needs: Transforming lives through telling tales*, 95.
- Gray, C. A. (1998). Social stories and comic strip conversations with students with Asperger syndrome and high-functioning autism. *Asperger Syndrome or High-functioning Autism?*, 167-198.
- 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. (2015). The new social story book: Revised and expanded 15th anniversary edition. Arlington, Texas: Future Horizons.)
- Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities*, 27(1), 70-84.
- 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.
- Happé, F. (2005). The weak central coherence account of autism. *Handbook of Autism and Pervasive Developmental Disorders, Volume 1, Third Edition*, 640-649.
- 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(2), 165-179.
- Karkhaneh, M., Clark, B., Ospina, M. B., Seida, J. C., Smith, V., & Hartling, L. (2010). Social Stories™ to improve social skills in children with autism spectrum disorder: A systematic review. *Autism*, 14(6), 641-662.
- 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.
- Korkmaz, B. (2011). *Theory of mind and neurodevelopmental disorders of childhood*. *Paediatrician. Research*, 69, 101R-8R.
- Lipsky, D. (2011). *From anxiety to meltdown: How individuals on the autism spectrum deal with anxiety, experience meltdowns, manifest tantrums, and how you can intervene effectively*. Jessica Kingsley Publishers.
- Mancil, G. R., Haydon, T., & Whitby, P. (2009). Differentiated effects of paper and computer-assisted Social Stories™ on inappropriate behavior in children with autism. *Focus on Autism and Other Developmental Disabilities*, 24(4), 205-215.
- Moore, S., & Sosa, J. (2013). *U.S. Patent Application No. 14/069,933*.
- Myers, S. M., & Johnson, C. P. (2007). Management of children with autism spectrum disorders. *Pediatrics*, 120(5), 1162-1182.

- Neitzel, J. (2010). Positive behavior supports for children and youth with autism spectrum disorders. *Preventing School Failure: Alternative Education for Children and Youth*, 54(4), 247-2
- Ploog, B. O., Scharf, A., Nelson, D., & Brooks, P. J. (2013). Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43(2), 301-322.
- Pop, C. A., Simut, R. E., Pinte, S., Saldien, J., Rusu, A. S., Vanderfaeillie, J., ... & Vanderborght, B. (2013). Social robots vs. computer display: does the way social stories are delivered make a difference for their effectiveness on ASD children? *Journal of Educational Computing Research*, 49(3), 381-401.
- Quill, K. A. (1997). Instructional considerations for young children with autism: The rationale for visually cued instruction. *Journal of Autism and Developmental Disorders*, 27(6), 697-714.
- Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, 46(12), 1255-1268.
- Rosenzweig, K. (2009). Are Today's General Education Teachers Prepared to Meet the Needs of Their Inclusive Students?
- Thompson, R. (2015). *Social Stories: An Intervention Tool to Help Decrease Undesired Behaviors in Children with Characteristics of Autism Spectrum Disorders* (Doctoral dissertation, Department of Special Education, University of Utah).
- Vandermeer, J., Beamish, W., Milford, T., & Lang, W. (2015). iPad-presented social stories for young children with autism. *Developmental Neurorehabilitation*, 18(2), 75-81
- Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current Opinion in Neurology*, 26(2), 146-153.
- Wing, L., & Potter, D. (2002). The epidemiology of autistic spectrum disorders: is the prevalence rising?. *Developmental Disabilities Research Reviews*, 8(3), 151-161.
- Wolpert, R. L., & Mengersen, K. L. (2004). Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. *Statistical Science*, 450-471.

Appendix

Appendix A

Table 6:
Gray (2005) Criteria for Writing a Social Story

No	Criteria	Detail
1	The Goal	Share accurate information using content, format, and voice that is descriptive, meaningful, and physically, socially, and emotionally safe for the intended audience
2	TWO-step discovery	Authors gather relevant information to 1) improve their understanding of the audience in relation to a situation, skill, or concept, and/or 2) identify the specific topic(s) and type(s) of information to share in the Story.
3	THREE parts and a title	“A Social Story has a title and introduction that clearly identifies the topic, a body that adds detail, and a conclusion that reinforces and summarizes the information
4	FOURmat makes it mine	The Social Story format is tailored to the individual abilities, attention span, learning style and – whenever possible – talents and/or interests of the Audience.
5	FIVE factors define voice and vocabulary	Social Story has a patient and supportive voice and vocabulary that is defined by five factors. They are: (1) 1st or 3rd person perspective; (2) Positive & patient tone; (3) Past, present, or future tense; (4) Literally accurate; and (5) Accurate meaning
6	SIX questions guide story development	A Social Story answers relevant “wh” questions, describing the context (where), time-related information (when), relevant people (who), important cues (what) basic activities, behaviours, or statements (how) and the reasons or rationale behind them (why).
7	SEVEN is about sentences	A Social Story is comprised of Descriptive Sentences, and may also have one or more Coaching Sentences. Sentences adhere to all applicable Social Story Criteria.
8	A GR-EIGHT formula	“One Formula ensures that every Social Story describes more than directs.”
9	NINE make is mine	“Every Social Story is reviewed and revised until it meets all applicable Social Story Criteria.”
10	TEN guides to implementation	The 10 guides to implementation 1) Plan for comprehension, 2) Plan story support, 3) Develop a story schedule, 4) Plan a positive introduction, 5) Monitor!, 6) Organize the stories, 7) Mix and match stories to build concepts, 8) Story reruns and sequels, 9) Recycle instruction into applause, 10) Stay current

Appendix B

Table 7:

References and Rationale for Excluded Studies

Excluded References	Excl Crite ria	Rationale
Bader, R. (2006) <i>Using social stories to increase emotion recognition and labeling in school-age children with autism [dissertation]</i> San Diego Alliant International University	3	Intervention not computer-based
Bauminger-Zvieli, Nirit; Kugelmass, Dana Shoham, (2013) Mother-stranger comparisons of social attention in jealousy context and attachment in hfasd and typical pre-schoolers. <i>Journal of Abnormal Child Psychology</i> 41(2), 253-264.	3	Examining attachment and mother jealousy and intervention not online.
Bellini, S., & Peters, J. K. (2008). Social skills training for youth with autism spectrum disorders. <i>Child and adolescent psychiatric clinics of North America</i> , 17(4), 857-873.	3	Not computer based SS
Chan, J. M., O'Reilly, M. F., Lang, R. B., Boutot, E. A., White, P. J., Pierce, N., & Baker, S. (2011). Evaluation of a Social Stories™ intervention implemented by pre-service teachers for students with autism in general education settings. <i>Research in Autism Spectrum Disorders</i> , 5(2), 715-721.	3	Survey of types of assessment that EPS use. Not evaluating effectiveness of computer social stories.
Deyro, M. C., Simon, E. W., & Guay, J. (2016). Parental awareness of empirically established treatments for autism spectrum disorders. <i>Focus on Autism and Other Developmental Disabilities</i> , 31(3), 184-195.	4	Not computer social stories just general interventions for ASD
Eason, L. R. (2014). <i>Computer assisted instruction to improve theory of mind in children with autism</i> . University of North Texas.	5	Completed at home
Edeiken-Cooperman, N. (2014). The Use of Computer-Assisted Instruction as an Instructional Tool to Teach Social Stories to Individuals Who Have Been Diagnosed on the Autism Spectrum. <i>SAGE Open</i> , 4(4), 2158244014554205.	4	Computer assisted instruction not intervention
Hagiwara, T., & Smith Myles, B. (1999). A multimedia social story intervention: Teaching skills to children with autism. <i>Focus on Autism and other developmental disabilities</i> , 14(2), 82-95.	7	Too old
Hess, K. L., Morrier, M. J., Heflin, L. J., & Ivey, M. L. (2008). Autism treatment survey: Services received by children with autism spectrum disorders in public school classrooms. <i>Journal of autism and developmental disorders</i> , 38(5), 961-971.	4	General classroom strategies
Jeekratok, K., Chanchalor, S., & Murphy, E. (2014). Web-Based Social Stories and Games for Children with Autism. <i>International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)</i> , 9(4), 33-49.	9	Thailand children and emphasis on learning

Excluded References

Excluded References	Excl	Rationale
	Crite	
	ria	
Ku, J. (2013). <i>A psychoeducational and support workbook for siblings of children with autism</i> . Azusa Pacific University.	2	Siblings of children with ASD
Long, S. K. (2012). <i>Social Skills Interventions for Children with Autism Spectrum Disorders: Teachers' Acceptability and Likelihood to Keep and Use Progress Monitoring Data</i> (Doctoral dissertation, University of Florida).	3	Teacher's perceptions and not computer based.
Marion, I. W., Nelson, T. M., Sheller, B., McKinney, C. M., & Scott, J. M. (2016). Dental stories for children with autism. <i>Special Care in Dentistry</i> , 36(4), 181-186.	4	Dental stories
Pinto, M. (2015). Viewing and exploring the subject area of information literacy assessment in higher education (2000–2011). <i>Scientometrics</i> , 102(1), 227-245.	4	Teacher's perspectives. Literacy
Sansosti, F. J., & Powell-Smith, K. A. (2008). Using computer-presented social stories and video models to increase the social communication skills of children with high-functioning autism spectrum disorders. <i>Journal of Positive Behavior Interventions</i> , 10(3), 162-178.	3	Not on computer alone. Also included video modelling
Taghizadeh, N., Davidson, A., Williams, K., & Story, D. (2015). Autism spectrum disorder (ASD) and its perioperative management. <i>Pediatric Anesthesia</i> , 25(11), 1076-1084.	3	Not computer social stories
Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pinte, S., ... & David, D. O. (2012). Using the social robot probo as a social story telling agent for children with ASD. <i>Interaction Studies</i> , 13(3), 348-372.	3	Robot not computer

Appendix C

Table 7: *Mapping the Studies*

Author	Location	Participant details	Study design	Outcome measures	Outcome variable	Intervention	Key findings
Chan (2009)	USA Mainstream school	6 participants aged 6-13 years (5 boys and 1 girl) diagnosed with ASD.	Single subject multiple baseline across participants design.	30 minute observation at 10 second intervals and comprehension questions	target behaviour (appropriate sitting, eye contact, attention, appropriate comments, working independently)	Social stories presented on a computer (PowerPoint) for 5 participants and bound book for the 6 th . It was conducted in class once a day by trained teachers with ABA experience	There was improved performance in the immediate condition for one, no differentiation between conditions for the second, and better performance in the delay condition for the third participant. There was also a treatment effect for two other participants but not for the final one.
Mancil et al. (2009)	USA Mainstream school	3 participants aged 6-9 years (2 male and 1 female) diagnosed with ASD.	ABABCBC single-subject design comparing a social story in paper format and power-point.	5 minute intervals during 10 minute transition time and 15 minute recess.	Level of pushing.	SS paper and computer versions daily before lunch	Reduction of pushing in both groups. Outcomes were slightly better for the power-point format.
Pop et al., (2013).	Romania Autism centre	20 children aged 4-9 years diagnosed with ASD.	Randomly allocated to: computer SS (7) robot SS (6), and control group (7).	7 point Likert scale developed from 5 point scale of Barry and Burlew (2004)	Level of prompt (verbal, gestural and physical) to provide social response.	Computer or robot SS Intervention conducted by psychotherapists. 6 sessions lasting 10-15 minutes	Significant difference between control group and robot but not between control group and computer-format.

Author	Location	Participant details	Study design	Outcome measures	Outcome variable	Intervention	Key findings
Thompson (2016)	USA, 4 SEND schools.	4 children aged 0-12 years (all male) diagnosed with ASD.	2 Groups: Paper based and no intervention group and Tablet based and no intervention group	Behavioural observations/ functional assessment to measure rate of undesired behaviours per minute. Pre, during and post.	Target behaviours depended on child.	For each child, 2 stories were written either on ipad or paper. Daily sessions ranging from 4-10 minutes outside the classroom in a quiet space.	SS in paper and ipad version were effective. There was no difference in effectiveness between both groups.
Vandermeer et al. (2015)	Australia SEND school	3 participants aged 4-6 years (2 boys and 1 girl) diagnosis with ASD.	A single-subject with multiple baseline across participants design	5 minute interval observation at 10 second intervals (baseline, intervention and follow-up).	On-task behaviour	Personalised Ipad-assisted social stories (Stories2Learnapp) embedded in class routine for 5 minutes. 18 sessions over 4 weeks.	Combination of social story with ipad was an effective intervention for one of the three participants.

Appendix D

Horner et al., (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education

Article Reference: Chan, J. M. (2009). *Pre-service teacher-implemented social stories' intervention for students with autism spectrum disorders in general education settings* (Doctoral dissertation, The University of Texas at Austin)

Description of Participants and Setting

Participants are described with sufficient detail to allow others to select individuals with similar characteristics; (e.g., age, gender, disability, diagnosis).

Yes

No

N/A

Unknown / unable to code

The process for selecting participants is described with operational precision.

Yes

No

N/A

Unknown / unable to code

Critical features of the physical setting are described with sufficient precision to allow replication.

Yes

No

N/A

Unknown / unable to code

Overall Rating of Evidence: X 3 2 1 0

Dependent Variable

Dependent variables are described with operational precision.

Yes

No

N/A

Unknown / unable to code

Each dependent variable is measured with a procedure that generates a quantifiable index.

Yes

No

N/A

Unknown / unable to code

Measurement of the dependent variable is valid and described with replicable precision.

Yes

No

N/A

Unknown / unable to code

Dependent variables are measured repeatedly over time.

Yes

No

N/A

Unknown / unable to code

Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards

Yes

No

N/A

Unknown / unable to code

Overall Rating of Evidence: 3 2 1 0

Independent Variable

Independent variable is described with replicable precision.

Yes

No

N/A

Unknown / unable to code

Independent variable is systematically manipulated and under the control of the experimenter.

Yes

No

N/A

Unknown / unable to code

Overt measurement of the fidelity of implementation for the independent variable is highly desirable.

Yes

No

N/A

Unknown / unable to code

Overall Rating of Evidence: 3 2 1 0

Baseline

The majority of single-subject research studies will include a baseline phase that provides repeated measurement of a dependent variable and establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur.

- Yes
- No
- N/A
- Unknown / unable to code

Baseline conditions are described with replicable precision.

- Yes
- No
- N/A
- Unknown / unable to code

Overall Rating of Evidence: 3 2 1 0

Experimental Control/internal Validity

The design provides at least three demonstrations of experimental effect at three different points in time.

- Yes
- No
- N/A
- Unknown / unable to code

The design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses).

- Yes
- No
- N/A
- Unknown / unable to code

The results document a pattern that demonstrates experimental control.

- Yes
- No
- N/A
- Unknown / unable to code

Overall Rating of Evidence: 3 2 1 0

External Validity

Experimental effects are replicated across participants, settings, or materials to establish external validity.

Yes

No

N/A

Unknown / unable to code

Overall Rating of Evidence: 3 2 1 0

Social Validity

The dependent variable is socially important.

Yes

No

N/A

Unknown / unable to code

The magnitude of change in the dependent variable resulting from the intervention is socially important.

Yes

No

N/A

Unknown / unable to code

Implementation of the independent variable is practical and cost effective

Yes

No

N/A

Unknown / unable to code

Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.

Yes

No

N/A

Unknown / unable to code

Overall Rating of Evidence: X 3 2 1 0

Average WoE A across the 7 judgement areas:

Sum of X / N = 9 / 7 = 1.3

X = individual quality rating for each judgement area

N = number of judgement areas

Overall Rating of Evidence: X 3 2 1 0

Appendix E

Coding protocol: Gersten et al., (2005). *Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education*

Full Study Reference: Pop, C. A., Simut, R. E., Pinteá, S., Saldien, J., Rusu, A. S., Vanderfaeillie, J., ... & Vanderborght, B. (2013). Social robots vs. computer display: does the way social stories are delivered make a difference for their effectiveness on ASD children?. *Journal of Educational Computing Research*, 49(3), 381-401.

Research design: Group experimental

Type of Publication: Journal

Essential Quality Indicators

Describing Participants

Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?

Yes

No

N/A

Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

Yes

No

N/A

Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Yes

No

N/A

Unknown/Unable to Code

Implementation of the Intervention and Description of Comparison Condition

Was the intervention clearly described and specified?

Yes

No

N/A

Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

Yes

- No
- N/A
- Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

Yes

- No
- N/A
- Unknown/Unable to Code

Outcome Measures

Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

Yes

- No
- N/A
- Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

Yes

- No
- N/A
- Unknown/Unable to Code

Data Analysis

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?

Yes

- No
- N/A
- Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

Yes

- No
- N/A
- Unknown/Unable to Code

Desirable Quality Indicators

Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

- Yes
- No
- N/A
- Unknown/Unable to Code

Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

- Yes
- No
- N/A
- Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured beyond an immediate posttest?

- Yes
- No
- N/A
- Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

- Yes
- No
- N/A
- Unknown/Unable to Code

Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

- Yes
- No
- N/A
- Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

- Yes
- No
- N/A
- Unknown/Unable to Code

Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

- Yes
- No
- N/A

Appendix F

Table 8:
WoE A Criteria, Weightings and Rationale

Protocol	Weightings	Rationale
A Horner et al., (2005)	3 Mean score across the seven key areas between 2.1- 3	Mean scores range from 0-3 have been mapped onto the weightings from 1-3.
	2 Mean score across the seven key areas between 1.1- 2	
	1 Mean score across the seven key areas between 0- 1	
B Gersten et al., (2005)	3 Scored 9 or more on the essential criteria and 4 or more on the desirable criteria	Gersten et al., (2005) provided these weighting recommendations
	2 Scored 9 or more on the essential criteria and 2 or more on the desirable criteria	
	1 Scored less than 9 on the essential criteria and less than 2 on the desirable criteria	

Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

Yes

No

N/A

Unknown/Unable to Code

Overall Rating of Evidence: 3 2 1

Appendix G

Table 9:
WOE B Criteria, Weightings and Rationale

Criteria	Weightings	Rationale
A Study Design	3 Randomised control trial studies	Apart from systematic reviews, these are different types of evidence suited for research questions around effectiveness (Petticrew and Roberts, 2003)
	2 Cohort studies and experimental designs	
	1 Qualitative research, surveys, single case studies, case-control studies, non-experimental evaluations	
B Appropriateness of measurement	3 Pre-post and follow ups for all participants and used two different measures	Pre and post measurements are useful to measure effectiveness. Post-test immediately after intervention is useful to reduce likelihood of other effects interplaying. Follow up measures allow for measurement of long term effects. More than 1 measure is useful to increase reliability.
	2 Pre and post taken immediately after and used 1 measure.	
	1 Pre and post not taken immediately after and only used 1 measure	
C Comparison	3 The study measures the impact of computer version, paper version and a control.	It is useful to compare all three versions to see whether computer-version SS is more effective than no intervention and/or paper-version SS.
	2 The study measures the impact of computer version and paper version	
	1 The study only measures the impact of computer version.	

Appendix H

Table 10:
Scores for Each Criteria of WoE B and Overall Score.

Study	Criteria A	Criteria B	Criteria C	WOE B
Chan (2009)	1	2	2	1.6
Mancil et al. (2009)	1	2	3	2
Pop et al. (2013)	3	2	3	2.6
Thompson (2016)	1	3	3	2.3
Vandermeer et al. (2015)	1	2	1	1.3

Appendix I

Table 11:
WOE C Criteria, Weightings and Rationale

Criteria	Weightings	Rationale
A Setting	3 Schools in UK or America	The more the setting matches UK education system, the more valid the results and more the results can be generalised.
	2 Schools in OECD countries	
	1 Schools outside of OECD countries	
B Intervention implementation	3 Implemented by school staff in class with little additional cost.	Implementation needs to be cost effective to increase likelihood to be used in the future.
	2 Implemented outside the classroom.	
	1 Implemented outside the class room and / or by specialist staff	
C Intention of intervention	3 The dependent variable has a social communication impact	Social stories are intended to help with social skills and ultimately improve behaviours in the school classroom.
	2 The dependent variable will have an impact on behaviours in the classroom.	
	1 It is not clear what the impact of the dependent variable will be.	
D Age of participants	3 Children are aged between 5-11 years only	5-11 years is the primary-age and thus meets the research question.
	2 Children are aged between 0 -16 years	
	1 Not specified age range	

Appendix J

Table 12:
WoE C Breakdown of Scores for Each Study and Overall WoE C Score

Studies	A	B	C	D	WOE C
Chan (2009)	3	1	3	2	2.25
Mancil et al. (2009)	3	3	2	3	2.75
Pop et al. (2013)	2	1	3	2	2
Thompson (2016)	3	2	1	2	2
Vandermeer et al. (2015)	2	3	2	2	2.25