Case Study 1: Evidence Based Practice Report

Theme: Interventions implemented by parents.

How effective is It Takes Two to Talk®–The Hanen Program® at improving the language and interaction skills of pre-school children with special educational needs and disabilities

Summary

It Takes Two to Talk - The Hanen Program (ITTT) for Parents of Children with Language Delays is an early language intervention that supports parents to apply language facilitation strategies flexibly, to support their child's development (Weitzman et al., 2017). The intervention is based on the social-interactionist perspective of language acquisition, which views language development as an interactive process between the caregiver and their child (Nor & Rashid, 2018). The ITTT intervention teaches parents responsive interaction strategies through group and individual training sessions, delivered by a qualified speech-language pathologist/ therapist (Weitzman et al., 2017).

The first studies to investigate the effectiveness of ITTT found consistent positive effects of the programme on parental language use; however, mixed effects were found for child language and interaction outcomes (Girolametto et al., 1996a; Girolametto et al., 1996b; Girolametto et al., 1998). Thus, the present review evaluated the effectiveness of the ITTT programme for improving the language and interaction skills of pre-school children with special educational needs and disabilities, based on research published after these initial studies. Five studies were identified and evaluated. Overall, the findings provide promising evidence for the use of ITTT for improving some language and interaction outcomes and suggest that the intervention has comparable effects to clinician-directed therapy. The methodological quality and relevance of each study is discussed in relation to these findings, and areas for future research are identified.

Introduction

It Takes Two to Talk- The Hanen Program

It Takes Two to Talk - The Hanen Program (ITTT) is a family-centred model of early language intervention for children under four years with receptive and expressive language delays, who may also have additional developmental needs (Pepper & Weitzman, 2004). The intervention is part of a suite of programmes developed by The Hanen Centre, a Canadian charitable organisation that was founded in 1975 to support parents and professionals in their daily interactions with children, to foster lifelong language, literacy and social skills (The Hanen Centre, 2016a).

Hanen programmes adopt an indirect service delivery model, whereby parents and carers are taught to facilitate reciprocal interactions with their children in naturalistic settings (Weitzman et al., 2017). The ITTT programme is comprised of an orientation meeting for parents, a pre-program assessment for each child, eight group training sessions and three individual video feedback sessions, with all components delivered by a Hanen Certified speech and language pathologist or therapist (SLT). During the group sessions, parents are taught responsive interaction strategies including language modelling, child-oriented strategies and interaction promoting strategies (see Table 1; Weitzman et al., 2017). During the individual sessions, which take place at regular intervals between the group sessions, parent-child interactions are videoed and reviewed with a SLT, to support parents' use of the strategies taught in the group sessions and to engage in problem-solving (Baxendale & Hesketh, 2003).

Parents are instructed to use the strategies that best support their child's individual language goals, which are formulated by the parent with support from a SLT. The parent implements the strategies using focused stimulation, which involves frequent repetitions of a goal within an interaction, to facilitate comprehension and language production (Weitzman et al., 2017). The focused stimulation technique was included in ITTT in 1996 (Girolametto et al., 1996b).

Table 1

Responsive Interaction Strategies Taught in ITTT

Type of Strategy	Examples of Strategies	Aim
Child-oriented	Being Face to Face; Observe, Wait and Listen; Following the Child's Lead by joining in with play, imitating, interpreting and commenting	To encourage the child to initiate interactions
Interaction-promoting	Matching Turns; Cueing your Child to Take a Turn; Asking Questions that Keep the Conversation Going	To promote balanced turn-taking between children and caregivers during interactions
Language modelling	Matching What you Say to What's Happening; Repeating Important Words; Using a Variety of Words; Highlighting Your Language and Expanding your Child's Message	To increase the child's receptive and expressive language skills

Rationale and Relevance

ITTT adopts a naturalistic approach to intervention, in which strategies to support a child's language are applied to the child's everyday activities,

interactions and routines (Sheldon & Rush, 2001). It is argued that this ecological approach to language intervention, as opposed to traditional clinician-directed therapy, enables children to better generalise their language skills to new situations and allows parents to continue providing intervention on an ongoing basis (Roberts & Kaiser, 2011). Further, research suggests that, for expressive language needs, indirect treatment, involving the training of parents, is as effective as direct clinic-based treatment, with indirect treatment considered more cost-effective (Law et al., 1998).

The ITTT programme is based on the social-interactionist theory of language acquisition (Nor & Rashid, 2018), which highlights the interactive nature of parent-child interactions, that may either inhibit or encourage child language development. Within this perspective, the responsivity hypothesis posits that parent responses that are contingent with child utterances and focus on the child's interests are more easily processed by the child, as they reduce contextual ambiguity and increase the saliency of the input (Weitzman et al., 2017). This processing advantage allows the child to direct more effort to language learning, supporting language development (Dominey & Dodane, 2004). Indeed, research has shown positive correlations between the contingency of maternal input to child speech and child language outcomes (Girolametto et al., 1999; Salerni & Suttora, 2022).

A number of randomised control trials were conducted to assess the effects of ITTT during the programme's development. Girolametto et al. (1996a) found positive effects of ITTT on maternal input and child target word acquisition, early gestures and aggressive and destructive behaviours for

children with language delay; however, no significant differences were found for total vocabulary size using a standardised measure. Using a larger sample, Girolametto et al. (1996b) found similar positive effects of ITTT on maternal input and child lexical development. However, in contrast to the first study, Girolametto et al. (1996b) also found improvements in child use of multiword combinations and early morphemes using standardised language measures. Finally, ITTT was found to be effective for children with Down Syndrome, with positive effects found on maternal input and target word acquisition (Girolametto et al., 1998). However, these language improvements were not reflected in differences on standardised vocabulary measures.

The effectiveness of the ITTT programme is of significant relevance to the United Kingdom (UK) context, as the programme is widely used in National Health Service (NHS) Speech and Language Therapy Teams (e.g. Central and North West London NHS Foundation Trust, 2022; Blackpool Teaching Hospitals NHS Foundation Trust, 2023). Additionally, there are over 100 speech and language therapists certified to deliver the ITTT programme in the UK (The Hanen Centre, 2016b), making it feasible for educational psychologists (EPs) to recommend the intervention to parents. Further, speech language and communication is the most common area of need for children requiring special educational needs and disability (SEND) support in the UK (Department for Education, 2022), highlighting the importance of effective interventions for this population.

Review Question

As early randomised control trials of ITTT were conducted by one research group and found consistent effects of the programme on parental language outcomes but mixed effects on child language outcomes (Girolametto et al., 1996a; Girolametto et al., 1996b; Girolametto et al., 1998), the reviewer suggests that a systematic review of the literature since these studies were conducted, that focuses on the impact of ITTT on child language and interaction outcomes, is necessary.

Thus, the present review focuses on the question: How effective is It Takes Two to Talk–The Hanen Program at improving the language and interaction skills of pre-school children with special educational needs and disabilities.

Critical Review

Systematic Literature Search

A systematic literature search was conducted on the 20th January 2022 using the databases PsycINFO (OVID), Web of Science and ERIC (EBSCO). The search terms used and rationale are shown in Table 2. The intervention synonyms were based on a scoping search of the literature concerning It Takes Two to Talk- The Hanen Program. Language strategy search terms, such as ‘focused stimulation’, were not included due to their use in other interventions that are not the focus of the present review.

Table 2

Search Terms

Intervention Focused Search Terms	**Rationale**
"Hanen* Program*" OR “Hanen* Parent Program*” OR "It Takes Two to Talk" OR ITTT	Broader ‘Hanen Program’ terms were used to include studies that follow the ITTT procedure without explicitly stating ITTT (which is a more recently devised programme name)

Note. Quotation marks were used to search for exact phrases and groups of words, an Asterix (*) was used to conduct a truncation search to allow for different endings of words, and ‘OR’ was used to find one term or the other.

The initial search produced 70 results, of which 33 were removed as duplicates and the resulting 37 were screened by title and abstract. 15 articles were excluded at title screening and a further 12 were excluded at abstract screening, according to the exclusion criteria which were applied hierarchically (see Table 3). At this stage, an ancestral search was conducted of the ten remaining articles, which produced one relevant result, and thus 11 articles underwent screening of the full text. Six articles were excluded after full text screening, resulting in five articles eligible for review (see Table 4). All articles excluded after removing duplicates are shown in Appendix A, and Figure 1 presents a PRISMA flow diagram (Page et al., 2021) of the systematic literature search.

Table 3

Inclusion and Exclusion Criteria

Study Feature	Inclusion Criteria	Exclusion Criteria	Rationale
1. Type and Date of Publication	Peer reviewed journal articles published between 1999 and 2022	Non-peer reviewed journals, dissertations, books and grey literature or peer reviewed journals published before 1999	The present review is concerned with studies conducted after the initial research into the approach up to 1998. To ensure the quality of the articles
2. Language	Studies published in English	Studies published in any language that is not English	To ensure that the reviewer is able to access and comprehend the article, as the reviewer's first language is English
3. Intervention	Studies that use It Takes Two to Talk- The Hanen Program intervention or studies based on this intervention model	Studies that do not use It Takes Two to Talk- The Hanen Program intervention or model. Studies that specify a different Hanen approach, such as Target Word	The aim of the review is to assess the effectiveness of It Takes Two to Talk- The Hanen Program
4. Population	Children aged 0-4 years identified as having a special educational need and disability (SEND)	Adults or children older than four years, and/or without an identified special educational need and	The ITTT programme is appropriate for children under four with language and other developmental needs

Study Feature	Inclusion Criteria	Exclusion Criteria	Rationale
		disability (SEND)	(Weitzman et al., 2017)
5. Outcome Measure	Studies that collect pre- and post-measures of intervention impact on at least one language or child-focused interaction measure	Studies that do not include a child language or interaction measure or studies that do not collect pre- and post-measures	The present review is concerned with the impact of the intervention on child language and interaction outcomes
6. Research Design	Studies that employ an experimental or quasi-experimental research design	Studies that do not use an experimental or quasi-experimental method, including secondary analyses, descriptive studies and qualitative reports	This criterion is based on Petticrew and Roberts (2008) typology of suitable studies for effectiveness reviews
7. Geographic Context	Studies that were conducted in Organisation for Economic Co-operation and Development (OECD) countries	Studies conducted in countries not part of the OECD	Studies from OECD countries are more generalisable to the UK context, due to similarities between policies and principles which impact on education and culture

Figure 1

PRISMA Flow Chart of the Systematic Screening

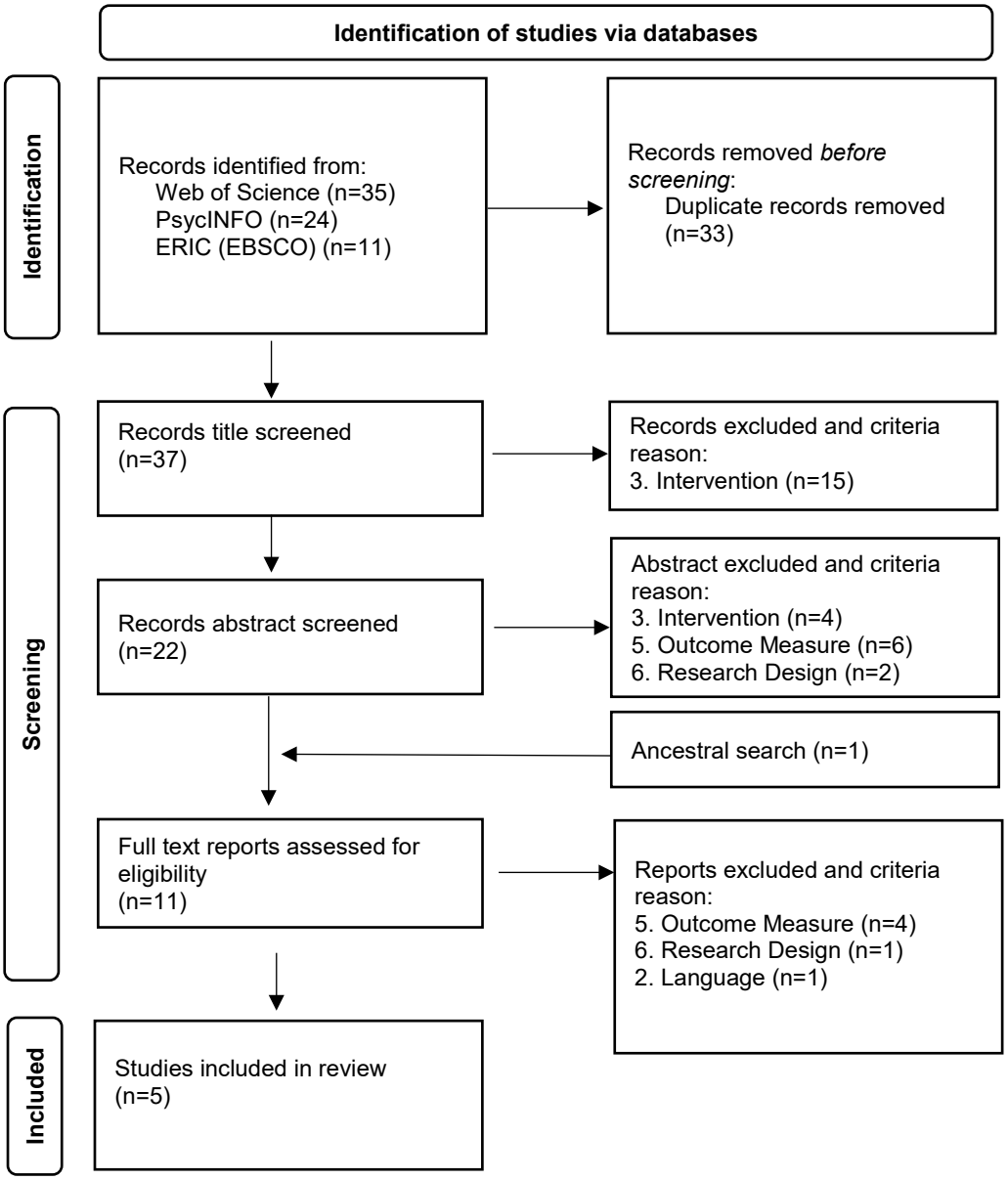

Table 4

References for Studies Included in the Review

	Reference
1.	Baxendale, J., & Hesketh, A. (2003). Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *International Journal of Language & Communication Disorders*, *38*(4), 397-415. https://doi.org/10.1080/1368282031000121651
2.	Cologon, K., Wicks, L., & Salvador, A. (2017). Supporting caregivers in developing responsive communication partnerships with their children: Extending a caregiver-led interactive language program. *Child Language Teaching and Therapy*, *33*(2), 157-169. https://doi.org/10.1177/0265659016650978
3.	Nicastri, M., Giallini, I., Ruoppolo, G., Prosperini, L., de Vincentiis, M., Lauriello, M., ... & Mancini, P. (2021). Parent training and communication empowerment of children with cochlear implant. *Journal of Early Intervention*, *43*(2), 117-134. https://doi.org/10.1177/1053815120922908
4.	Pennington, L., Thomson, K., James, P., Martin, L., & McNally, R. (2009). Effects of It Takes Two to Talk—The Hanen Program for Parents of preschool children with cerebral palsy: Findings from an exploratory study. *Journal of Speech, Language, and Hearing Research, 52*(5), 1121–1138. https://doi.org/10.1044/1092-4388(2009/07-0187)
5.	Senent-Capuz, N., Baixauli-Fortea, I., & Moret-Tatay, C. (2021). Parent-Implemented Hanen Program It Takes Two to Talk®: An Exploratory Study in Spain. *International journal of environmental research and public health*, *18*(15), 8214. https://doi.org/10.3390/ijerph18158214

Mapping the Field

The studies included in the review all used quasi-experimental research designs to investigate the effectiveness of ITTT. A detailed mapping of the field was conducted and can be found in Appendix B.

Weight of Evidence

Gough's Weight of Evidence Framework (WoE; Gough, 2007) was used to evaluate the value of each study in relation to the review question. The framework includes three dimensions: WoE A relates to the methodological quality of the study in relation to quality standards, WoE B relates to the methodological relevance of the research design for addressing the review question and WoE C relates to the relevance of the study topic to the review question.

WoE A calculations were based on an adapted version of Gersten et al.'s (2005) coding protocol, due to its relevance for appraising quasi-experimental research articles. Protocol adaptations and rationale are shown in Appendix C. Criteria for WoE B and C judgements were developed by the reviewer. Full explanations for how WoE A, B and C were calculated can be found in Appendix D. A completed coding protocol for WoE A is shown in Appendix E.

Ratings for WoE A, B and C were averaged for each study to give an overall weighting (WoE D). Thresholds for WoE D ratings were created by calculating an even interval between scores 1-3 (0.67) and then decreasing each value by (0.17) to adjust for the 0-1 interval used in WoE A judgements. A summary of the WoE ratings for each study are shown in Table 5.

Table 5

Summary of Weight of Evidence Ratings

Study	WoE A	WoE B	WoE C	WoE D
Baxendale & Hesketh (2003)	1 (low)	2.6 (high)	2.5 (high)	2 (medium)
Cologon, Wicks & Salvador (2017)	1 (low)	2.6 (high)	1.5 (medium)	1.7 (medium)
Nicastri et al. (2021)	3 (high)	2.6 (high)	2.25 (medium)	2.6 (high)
Pennington et al. (2009)	0 (very low)	2 (medium)	2.25 (medium)	1.4 (low)
Senent-Capuz, Baixauli-Fortea & Moret-Tatay (2021)	1 (low)	2.2 (medium)	3 (high)	2.1 (medium)

Note. WoE D ratings ≥ 2.5 are considered 'high', <2.5 and ≥1.5 are considered 'medium' and <1.5 are considered 'low'

Critical Review of Included Studies

Participants

In total, there were 103 child participants across the five studies included in the present review, with sample sizes ranging between ten and 37 participants. The age range of participants was provided in four of the studies, ranging from 16 to 44 months. Nicastri et al. (2021) only provided the mean ages of participants, which was 25.6 months in the intervention group and 26.2 months in the control group. As ITTT is suitable for pre-school children under age four (Weitzman et al., 2017), the intervention was appropriate for all participants.

Additionally, all studies reported the gender split of participants in the total sample, with 66 males and 37 females across the five studies. Three studies included more males than females and two studies included an equal gender split (see Table B1). The larger number of males in some studies may reflect findings that indicate a higher incidence of language delay in males (Etchell et al., 2018), however these differences were found to be small and influenced by a variety of other factors. The uneven gender split across the five studies reduces the generalisability of findings to females.

The special educational needs and disabilities (SEND) of the participants was considered important for the relevance of the intervention to the participants, as ITTT was developed to support pre-school children with language delay, who may also have other developmental needs (Pepper & Weitzman, 2004). Two studies in the review included participants with specific language difficulties (Baxendale & Hesketh, 2003; Senent-Capuz et al., 2021), and received higher WoE C ratings. The remaining three studies included participants with other primary needs (see Table B1), and thus received lower WoE C ratings. Furthermore, one study did not assess participant language level prior to inclusion in the intervention (Cologon et al., 2017), and thus received a lower WoE C rating, as this makes it unclear whether the participants' language level was appropriate for the intervention.

Geographic contexts in which the studies were conducted include: UK (Baxendale & Hesketh, 2003), UK and Australia (Pennington et al., 2009), Australia (Cologon et al., 2017), Italy (Nicastri et al., 2021) and Spain

(Senent-Capuz et al., 2021). All of these countries are OECD countries, which have similar education and government systems, thus geography was not included in WoE C ratings.

Research Design

Petticrew and Roberts (2003) consider quasi-experimental designs less suitable for assessing effectiveness than randomised control trials (RCTs), but better than non-experimental or qualitative methods. As all five studies used quasi-experimental research designs, they received an equal WoE B rating for this criterion. Cologon et al. (2017) also used a randomised control design for the follow-up phase of their study, however as the first phase was quasi-experimental, the mixed design received an equal WoE B rating to the other studies. One study did not include a follow-up assessment after the first post-test assessment (Senent-Capuz et al., 2021), which resulted in a lower WoE B rating.

Two studies used an active control condition, resulting in the highest WoE B rating for this criterion (Baxendale & Hesketh, 2003; Senent-Capuz et al., 2021). In these studies, the absence of a no treatment control group was justified for ethical reasons. Two studies used a no intervention or no further intervention control group (Nicastri et al. 2021; Cologon et al., 2017), resulting in a 'medium' WoE B rating, as any improvement following intervention compared to the no treatment control group may be attributed to other factors such as attention (Rogers & Revesz, 2019). One study did not include a control group (Pennington et al., 2009), thus receiving a 'low' WoE

B rating for that criterion. Of the studies that included a control group, one study did not use a random or matched groups procedure to allocate participants to conditions (Baxendale & Hesketh, 2003), which resulted in a lower B rating for the control group allocation criterion.

Finally, justification for using a quasi-experimental design and discussion of the potential limitations of this approach have been cited as criteria for effective quasi-experimental research designs (Eliopoulos et al., 2005; Moreno-Fernandez et al., 2008). One study did not include an explicit justification of the research design (Senent-Capuz et al., 2021), resulting in a lower WoE B rating for this criterion.

Intervention

Two studies reported implementing ITTT in the standard form, described by Weitzman et al. (2017), and thus received higher WoE C ratings (Pennington et al., 2009; Senent-Capuz et al., 2021). Two studies extended ITTT to include more individual sessions after the standard programme was completed (Cologon et al., 2017; Nicastri et al., 2021) and one study used an earlier form of the Hanen Program that did not specify ITTT, resulting in the lowest WoE C rating (Baxendale & Hesketh, 2003); however, the procedure was judged as equivalent to ITTT and thus the study was included in the present review.

With regards to replicability, only two studies clearly described and specified the strategies taught to parents in the group sessions and the types of

activities used, resulting in higher WoE A ratings (Nicastri et al., 2021; Senent-Capuz et al., 2021). However, Nicastri et al. (2021) adapted the ITTT session content to include additional information related to supporting deaf children, resulting in a lower WoE C rating for the intervention criterion. Finally, only one study assessed the fidelity of intervention implementation (Cologon et al., 2017), resulting in a higher WoE A rating.

Language and Interaction Outcome Measures

As the ITTT programme targets multiple aspects of language development, it is important that studies assessing the intervention's effectiveness include multiple reliable and valid tools to measure each construct (Gersten et al., 2005). The tools used to measure child language and interaction outcomes differed between studies (see Table B1). Two studies included more than two outcome measures and assessed both language and interaction skills, receiving higher WoE A ratings for the number of outcome measures indicator and higher WoE C ratings for the scope of outcome measures criterion (Baxendale & Hesketh, 2003; Senent-Capuz et al., 2021). One study included more than two outcome measures but only assessed language outcomes (Nicastri et al., 2021), resulting in a lower WoE C rating. Two studies included only one outcome measure (Cologon et al., 2017; Pennington et al. 2009), resulting in lower WoE A and C ratings.

Reliability and validity information for all outcome measures was established by the reviewer and is shown in Table B2. One study provided reliability and validity information for all outcome measures, resulting in a higher WoE A

rating (Senent-Capuz et al., 2021). Three studies utilised standardised measures of language skills, improving the reliability and validity of the findings (Baxendale & Hesketh, 2003; Nicastri et al., 2021; Senent-Capuz et al., 2021). Baxendale & Hesketh (2003) also utilised two additional non-standardised language measures, mean length of utterance (Brown, 1973) and proportional number of utterances. These measures had high inter-rater reliability in the study and MLU has been shown to have concurrent validity with other measures of clausal development (Rice et al., 2006), suggesting these measures have acceptable reliability and validity. Nicastri et al. (2021) used an unspecified test of speech recognition, for which no reliability and validity information is available, and thus the findings of this test should be interpreted with caution.

Cologon et al. (2017) employed the parent-child interaction checklist (Stewart, 2006) and did not provide any reliability and validity information for this measure, resulting in a lower WoE A rating. The percentage of inter-rater agreement for the checklist in the study was 67%, however Miles and Huberman (1994) propose an acceptable standard of 80% agreement on 95% of all codes, suggesting the parent-interaction checklist had inadequate inter-rater reliability in this study. Further, as no validity information is available for tool, the findings should be interpreted with caution.

Pennington et al. (2009) also failed to report reliability and validity information for the interaction coding scheme utilised (Pennington & McConachie, 2001), resulting in a lower WoE A rating. Kappa reliability coefficients of 0.77 and

0.78 for inter-rater agreement may be interpreted as moderate (McHugh, 2012), indicating acceptable inter-rater reliability for this tool. However, as no information regarding the validity of the measure is available, the findings should be interpreted with caution.

All studies also included additional parent outcome measures. However, as the focus of the present review is the effectiveness of the intervention on child outcomes, these measures will not be discussed in this review, but are shown in Table B1.

Findings and Effect Sizes

The descriptors used to interpret the effect sizes of findings are shown in Table 6 and the effect sizes are summarised in Table 7. Two studies reported effect sizes and received higher WoE A ratings (Nicastri et al., 2021; Pennington et al., 2009). Where partial eta squared was reported or no effect sizes were reported, the website Psychometrica (Lenhard & Lenhard, 2016) was used to calculate Cohen's *d* effect sizes.

Table 6

Descriptors for Partial Eta Squared and Cohen's d Effect Sizes (Cohen, 1992)

Partial Eta Squared η_p^2	**Cohen's *d***	**Descriptor**
<0.01	<0.2	Negligible
0.01	0.2	Small
0.06	0.5	Medium
0.14	0.8	Large

Table 7

Effect Sizes for Studies Included in the Review

Study	**Child language and/or Interaction Outcomes**	**Type of Comparison**	***p*-value**		**Reported Effect Size and Descriptor**		**Converted Effect Size (Cohen's *d*) and Descriptor**	
			Post-test	Follow-up	Post-test	Follow-up	Post-test	Follow-up
Baxendale & Hesketh (2003) N=37	Pre-school Language Scale-3 UK (PLS3-UK)	Between-subjects	.161	.248	N/A	N/A	d=0.472 (small)	d=0.392 (small)
	Menth Length of Utterance (MLU)		.353	.882	N/A	N/A	d=0.328[a] (small)	d=0.06[a] (negligible)
WoE D: 2 (medium)	Number of children with improved MLU at follow-up		N/A	.402	N/A	N/A	N/A	d=0.278 (small)
	Proportional number of utterances (PROPUTT) of parent and child		.927	.778	N/A	N/A	d=0.032 (negligible)	d=0.183[a] (negligible)

Study	**Child language and/or Interaction Outcomes**	**Type of Comparison**	***p*-value**		**Reported Effect Size and Descriptor**		**Converted Effect Size (Cohen's *d*) and Descriptor**	
			Post-test	Follow-up	Post-test	Follow-up	Post-test	Follow-up
Cologon, Wicks & Salvador (2017) N=10 WoE D: 1.7 (medium)	Parent- child interaction checklist (PIC) (control group)	Within-subjects	<.044*	<.91	N/A	N/A	d=1.654 (large)	d=0.067 (negligible)
	Parent- child interaction checklist (PIC) (experimental group)		<.016*	<.028*	N/A	N/A	d=2.336 (large)	d=1.925 (large)
Nicastri et al. (2021) N=28 WoE D: 2.6 (high)	Word comprehension (time by group effect)	Between-subjects	.001*	N/A	ηp2=0.34 (large)	N/A	d=1.436 (large)	N/A
	Word production (time by group effect)		.028*	N/A	ηp2=0.17 (large)	N/A	d=0.905 (large)	N/A
	Sentence comprehension (time by group effect)		.001*	N/A	ηp2=0.69 (large)	N/A	d=2.984 (large)	N/A
	Bisyllabic word recognition		N/A	.88	N/A	d=0.01 (negligible)	N/A	N/A

Study	**Child language and/or Interaction Outcomes**	**Type of Comparison**	***p*-value**		**Reported Effect Size and Descriptor**		**Converted Effect Size (Cohen's *d*) and Descriptor**	
			Post-test	Follow-up	Post-test	Follow-up	Post-test	Follow-up
	Sentence recognition		N/A	.68	N/A	d=0.00 (negligible)	N/A	N/A
	Bisyllabic word recognition in noise		N/A	.34	N/A	d=0.03 (negligible)	N/A	N/A
	Peabody Picture Vocabulary Test for lexical comprehension		N/A	.012*	N/A	d=0.34 (small)	N/A	N/A
	Test for Reception of Grammar		N/A	.019*	N/A	d=0.93 (large)	N/A	N/A
	Boston Naming Test		N/A	.007*	N/A	d=0.31 (small)	N/A	N/A
Pennington et al. (2009)	Child initiation moves	Within-subjects	.005*	p=.887	d=0.95 (large)	d=0.046 (negligible)	N/A	N/A
	Child response moves		.062	p=.058	d=-0.507 (medium)	d=0.685 (medium)	N/A	N/A
N=11	Child follow-up moves		.309	N/A	d=-0.155 (negligible)	N/A	N/A	N/A
WoE D: 1.4 (low)	Child 'no responses'		.015*	N/A	d=-0.764 (medium)	N/A	N/A	N/A
	Child control function		.007*	p=.496	d=0.901 (large)	d=0.224 (small)	N/A	N/A

Study	**Child language and/or Interaction Outcomes**	**Type of Comparison**	***p*-value**		**Reported Effect Size and Descriptor**		**Converted Effect Size (Cohen's *d*) and Descriptor**	
			Post-test	Follow-up	Post-test	Follow-up	Post-test	Follow-up
	Child comply function		.233	p=.109	d=-0.228 (small)	d=-0.563 (medium)	N/A	N/A
Senent-Capuz, Baixauli-Fortea & Moret-Tatay (2021)[b]	Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP) Communication	Between-subjects	.80	N/A	N/A	N/A	d=0.011 (negligible)	N/A
N=17	CSBS DP Gestures		.05	N/A	N/A	N/A	d=1.1 (large)	N/A
WoE D: 2.1 (medium)	Reynell Developmental Language Scales (RDLS-III) Comprehension		.46	N/A	N/A	N/A	d=-0.245 (medium)	N/A
	RDLS-III Expression		.76	N/A	N/A	N/A	d=0.047 (negligible)	N/A
	RDLS-III Total		.67	N/A	N/A	N/A	d=-0.149 (negilible)	N/A
	MacArthur-Bates Communicative Development		.96	N/A	N/A	N/A	d=-0.081 (negligible)	N/A

Study	Child language and/or Interaction Outcomes	Type of Comparison	*p*-value		Reported Effect Size and Descriptor		Converted Effect Size (Cohen's *d*) and Descriptor	
			Post-test	Follow-up	Post-test	Follow-up	Post-test	Follow-up
	Inventories (MCDI) Vocabulary							
	MCDI Word Endings		.13	N/A	N/A	N/A	d=-0.905 (large)	N/A
	MCDI Morphosyntactic Complexity		.61	N/A	N/A	N/A	d=0.032 (negligible)	N/A

[a]= equal attrition rates assumed
[b]= CSBS DP Emotion scale was not included due to lack of relevance to the present review
*= the results reached statistical significance at $p \leq 0.05$

When considering the effectiveness of ITTT on child interaction outcomes, both Cologon et al. (2017) and Pennington et al. (2009) found large effects of ITTT on child interaction skills, using a within-groups pre-post-test design. Specifically, Pennington et al. (2009) found a significant increase in child initiation moves and control functions, with large effect sizes, and a significant decrease in the child not responding, with a medium effect size. Cologon et al. (2017) received a 'medium' WoE D rating, which was affected by a 'low' overall WoE A rating, and Pennington et al. (2009) received a 'low' WoE D rating, which reflects its 'very low' overall WoE A rating, impacted by the lack of a control condition.

For child language outcomes, Nicastri et al. (2021) found that the ITTT group achieved better outcomes compared to the no treatment control group, with significant time by group interaction effects for all language outcomes post-intervention, with large effect sizes. The study's large effect sizes and 'high' WoE D rating suggest that considerable weight should be given to this study. However, the ITTT procedure was adapted to include six additional individual sessions over a period of six months, as a reinforcement and maintenance phase. Similarly, Cologon et al. (2017) also found a large effect size for the impact of four additional monthly individual sessions after standard ITTT was completed. These findings have practical implications for intervention implementation.

When assessing the effectiveness of ITTT compared to more traditional clinician-directed therapy, both Baxendale and Hesketh (2003) and Senent-

Capuz et al., (2021) found no significant differences between the ITTT group and the clinician-directed therapy control group for any child language or interaction measures. Both studies received a 'medium' WoE D rating, suggesting their results should be given weight when drawing conclusions, despite some methodological limitations.

However, it is important to note that Senent-Capuz et al. (2021) found large effect sizes for gesture and word ending measures, which indicate that with a larger sample size, these differences may have been significant. Inspection of the descriptive statistics indicate that the ITTT group scored higher on the gestures measure post-intervention, however the clinician-directed therapy group scored higher on the word endings measure post-intervention.

Finally, four studies included a follow-up assessment post-intervention. For within subjects designs, Cologon et al. (2017) and Pennington et al. (2009) found no significant differences between child interaction scores at post-intervention and four months post-intervention assessment points. Similarly, Baxendale and Hesketh (2003) found no significant differences six months post-intervention between the ITTT group and clinician-directed therapy group for any measure. Furthermore, Nicastri et al. (2021) conducted a three-year follow-up assessment, finding large effect sizes for child language comprehension skills. Small effect sizes for child lexical skills were found and no significant long-term effects of ITTT were found for word or sentence recognition.

Conclusions and Recommendations

Conclusions

The aim of the present review was to investigate the effectiveness of It Takes Two to Talk–The Hanen Program at improving the language and interaction skills of pre-school children with special educational needs and disabilities. Five studies met the inclusion criteria; based on overall WoE criteria (Gough, 2007), one study received a rating of 'high', three were rated 'medium', and one was rated 'low'.

For the highest rated study included in the review, ITTT was found to improve child language outcomes including word comprehension, word production and sentence recognition (Nicastri et al., 2021), suggesting that ITTT is effective for improving these child language skills.

Further, two studies with 'medium' WoE D ratings found that the effects of ITTT on child language and interaction outcomes were not significantly different to traditional clinician-directed therapy (Baxendale & Hesketh, 2003; Senent-Capuz et al., 2021). However, large effect sizes were found by Senent-Capuz et al. (2021) for gesture and word ending measures, with ITTT leading to larger scores on a gesture measure post-intervention, and clinician-directed therapy leading to larger scores on a word ending measure post-intervention, suggesting that the two types of therapy may be more suited for improving different language skills.

The finding of no significant differences between the two therapies for child language and communication outcomes is consistent with earlier reviews of

indirect language interventions (Law et al., 1998), suggesting that for families that cannot access or do not wish to access traditional therapies for their child, ITTT is an effective alternative option. However, it has been argued that parent training programmes are often more time intensive, which has been shown to influence parental decisions to participate in ITTT (Pennington et al., 2007).

ITTT was also found to have large effects on child interaction outcomes, however the two studies on which this conclusion is based had the lowest ratings of the five studies (Cologon et al., 2017; Pennington et al., 2009). Therefore, further research is needed to determine the effects of ITTT on child interaction skills, using more methodologically rigorous procedures, such as using a randomised control design, including active control groups and utilising more reliable and valid outcome measures (Gersten et al., 2015).

Finally, the positive effects of ITTT on some child language and interaction skills appear to be maintained for at least four to six months (Baxendale & Hesketh, 2003; Cologon et al., 2017; Pennington et al., 2009), but do not continue to improve. ITTT was also associated with more lasting gains for language comprehension skills (Nicastri et al., 2021).

Therefore, this review finds some evidence for the effectiveness of ITTT for improving child language outcomes and concludes that the effects of ITTT on language and interaction outcomes are not significantly different to the effects of clinician-directed therapy. There is weak evidence for the

effectiveness of ITTT for improving child interaction outcomes, due to the methodological limitations of studies assessing this domain.

Limitations and Recommendations

The present review has a number of limitations which should be considered alongside these findings. There was substantial variability in how ITTT was delivered between studies, with adaptations being made to the number of individual sessions and session content, making it hard to draw conclusions about the intervention's overall effectiveness. The findings suggest that increasing the number of individual sessions continues to improve outcomes (Cologon et al., 2017), whereas follow-up measures conducted after four months of no further intervention find that child skills are maintained but do not continue to improve (Cologon et al., 2017; Pennington et al., 2009). These findings have practical implications for providers of the ITTT programme, as an intended outcome of the programme is that parents continue to use to the strategies after programme completion, resulting in continued improvement in child language and interaction skills (Weitzman et al., 2017). Thus, future research should determine the number of additional sessions needed to optimise child outcomes, whilst considering cost and time implications, and factors affecting continued improvement in child language and interactions skills should be explored.

In addition, each study included small sample sizes, which limits the generalisability of the findings. Further, as all studies were quasi-experimental in design, biases may have been introduced when recruiting

participants and assigning groups, reducing the validity of between-group comparisons. Thus, further research into the effectiveness of ITTT, using larger samples and more rigorous methodology is needed.

Finally, the focus of the present review did not include the wider effects of ITTT on parent and child outcomes; for example, research has identified secondary benefits of ITTT including improvements in child behaviour and accompanying reductions in parental stress (Konza et al., 2010; Pennington et al., 2010; Whittingham et al., 2011). Thus, for some families, ITTT may have benefits beyond primary language and interaction outcomes which were not considered in the present review.

Overall, the literature published over the last twenty years suggests that ITTT is an effective early intervention for supporting some language skills of pre-school children with special educational needs and disabilities; however, future research is needed to investigate the effects of ITTT on interaction skills, due to methodological limitations of studies assessing this outcome.

References

Axelrod, B. N., Ricker, J. H., & Cherry, S. A. (1994). Concurrent validity of the MAE visual naming test. *Archives of Clinical Neuropsychology*, *9*(4), 317-321. https://doi.org/10.1093/arclin/9.4.317

Baxendale, J., & Hesketh, A. (2003). Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *International Journal of Language & Communication Disorders*, *38*(4), 397-415. https://doi.org/10.1080/1368282031000121651

Benton, A. L., & Hamsher, K. (1978). Multilingual aphasia examination manual. *Iowa City: University of Iowa*.

Bishop, D. V. M. (1982). TROG Test for reception of grammar. *Published by the author and available from Age and Cognitive Performance Research Centre, University of Manchester, UK*.

Bishop, D. V., & Garsell, M. (2003). Test for reception of grammar: Version 2: TROG-2 manual.

Boucher, J., & Lewis, V. (1997). Pre-school Language Scale-3 (UK Adaptation).

Blackpool Teaching Hospitals NHS Foundation Trust. (2023). *Groups and Workshops.* https://www.bfwh.nhs.uk/our-services/speech-and-language-therapy/childrens-speech-and-language-therapy-service/our-services/blackpool-fylde-wyre/groups-workshops/

Brown, R. (1973). A first language: The early stages. *Harvard University Press*. https://doi.org/10.1017/s030500090000074x

Caselli, M. C., & Casadio, P. (2007). Il Primo Vocabolario del Bambino. *Consiglio Nazionale Delle Ricerche, Istituto di Psicologia e Fondazione MacArthur.*

Central and North West London NHS Foundation Trust. (2022). Milton Keynes Early Years Speech and Language Therapy Team. https://www.cnwl.nhs.uk/services/community-services/milton-keynes-childrens-speech-and-language-therapy-services/milton-keynes-early-years-speech-and-language-therapy-team

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin*, *112*, 1155-1159. https://doi.org/10.1037/0033-2909.112.1.155

Cologon, K., Wicks, L., & Salvador, A. (2017). Supporting caregivers in developing responsive communication partnerships with their children: Extending a caregiver-led interactive language program. *Child Language Teaching and Therapy*, *33*(2), 157-169. https://doi.org/10.1177/0265659016650978

Cutugno, F., Prosser, S., & Turrini, M. (2000). Audiometria Vocale, vol IV. *Padova, GN Resound Italia*.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*, *34*(3), 565-571. https://doi.org/10.1044/jshr.3403.565

Department for Education. (2022). *Special educational needs and disability: an analysis and summary of data sources.* https://assets.publishing.service.gov.uk/government/uploads/system/u

ploads/attachment_data/file/1082518/Special_educational_needs_publication_June_2022.pdf

Dominey, P. F., & Dodane, C. (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. *Journal of Neurolinguistics*, *17*(2-3), 121-145. https://doi.org/10.1016/s0911-6044(03)00056-3

Dunn, L. M., & Dunn, L. M. (1981). Peabody picture vocabulary test-revised. *American guidance service, Incorporated.*

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). British picture vocabulary scale II. *Windsor. UK: Nfer-Nelson*.

Edwards, S., Garman, M., Hughes, A., Letts, C., & Sinka, I. (1999). Assessing the comprehension and production of language in young children: an account of the Reynell Developmental Language Scales III. *International Journal of Language & Communication Disorders*, *34*(2), 151-171. https://doi.org/10.1080/136828299247487

Eliopoulos, G. M., Harris, A. D., Lautenbach, E., & Perencevich, E. (2005). A systematic review of quasi-experimental study designs in the fields of infection control and antibiotic resistance. *Clinical Infectious Diseases*, *41*(1), 77-82. https://doi.org/10.1086/430713

Etchell, A., Adhikari, A., Weinberg, L. S., Choo, A. L., Garnett, E. O., Chow, H. M., & Chang, S. E. (2018). A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia*, *114*, 19-31. https://doi.org/10.1016/j.neuropsychologia.2018.04.011

Fenson, L., Marchman, V., & Thal, D. (1993). MacArthur-bates communicative development inventories: words and sentences form MCDI. *Baltimore (MD): Brookes*.

Gardner, M. F. (1981). Expressive One- Word Picture Vocabulary Test. *Novato, CA: Academic Therapy Publications*

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. https://doi.org/10.1177/001440290507100202

Girolametto, L., Pearce, P., & Weitzman, E. (1996a). The effects of focused stimulation for promoting vocabulary in children with delays: A pilot study. *Journal of Childhood Communication Development, 17*(2), 39–49. https://doi.org/10.1177/152574019501700205

Girolametto, L., Pearce, P., & Weitzman, E. (1996b). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech and Hearing Research, 39*(6), 1274–1283. https://doi.org/10.1044/jshr.3906.1274

Girolametto, L., Weitzman, E., & Clements-Baartman, J. (1998). Vocabulary intervention for children with Down syndrome: Parent training using focused stimulation. *Infant-Toddler Intervention: A Transdisciplinary Journal, 8*(2), 109–126. Retrieved from https://eric.ed.gov/?id=EJ568724

Girolametto, L., Weitzman, E., Wiigs, M., & Pearce, P. S. (1999). The relationship between maternal language measures and language development in toddlers with expressive vocabulary delays. *American Journal of Speech-Language Pathology*, *8*(4), 364-374. https://doi.org/10.1044/1058-0360.0804.364

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. https://doi.org/10.1080/02671520701296189

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). Boston Naming Test. *Philadelphia: Lea & Febiger.*

Konza, D., Maloney, C., & Grafton, P. (2010). It Takes Two to Talk: A Focused Intervention Program for Parents and Children with Language Delays. *International Journal of Interdisciplinary Social Sciences*, *5*(6), 225-236. https://doi.org/10.18848/1833-1882/cgp/v05i06/59301

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for primary speech and language delay: a systematic review of the literature. *International journal of language & communication disorders*, *33*(sup1), 21-23. https://doi.org/10.3109/13682829809179388

Lee, L. L., Koenigsknecht, R. A., & Mulhern, S. T. (1974). Developmental sentence scoring. *North Western University.*

Lenhard, W. & Lenhard, A. (2016). *Computation of effect sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Psychometrica. DOI: 10.13140/RG.2.2.17823.92329

López Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005). MacArthur: Inventario de desarrollo comunicativo. Manual y Cuadernillos. *Madrid, TEA Ediciones.* ISBN: 84-7174-820-7

Mariscal, S., Nieva, S. & López-Ornat, S. (2010). Observar y medir el desarrollo gramatical temprano: un estudio de validez concurrente. *Psicothema, 22*(1), 50-56. ISSN: 0214-991

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276-282. https://doi.org/10.11613/bm.2012.031

Miles, M & Huberman A 1994, Qualitative data analysis: An expanded source book, 2nd edn, *Sage, Thousand Oaks, California*

Moreno-Fernández, M. M., Ramos-Álvarez, M. M., Valdés-Conroy, B., & Catena, A. (2008). Criteria of the peer review process for publication of experimental and quasi-experimental research in Psychology: A guide for creating research papers. *International Journal of Clinical and Health Psychology*, *8*(3), 751-764. Retrieved from https://www.redalyc.org/articulo.oa?id=33712016009

Nicastri, M., Giallini, I., Ruoppolo, G., Prosperini, L., de Vincentiis, M., Lauriello, M., ... & Mancini, P. (2021). Parent training and communication empowerment of children with cochlear implant. *Journal of Early Intervention*, *43*(2), 117-134.

https://doi.org/10.1177/1053815120922908

Nor, N. M., & Ab Rashid, R. (2018). A review of theoretical perspectives on language learning and acquisition. *Kasetsart Journal of Social Sciences, 39*(1), 161-167. https://doi-org.libproxy.ucl.ac.uk/10.1016/j.kjss.2017.12.012

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pennington, L., & McConachie, H. (2001). Predicting patterns of interaction between children with cerebral palsy and their mothers. *Developmental Medicine and Child Neurology*, *43*(2), 83-90. https://doi.org/10.1111/j.1469-8749.2001.tb00720.x

Pennington, L., & Noble, E. (2010). Acceptability and usefulness of the group interaction training programme It Takes Two to Talk to parents of pre-school children with motor disorders. *Child: care, health and development*, *36*(2), 285-296. https://doi.org/10.1111/j.1365-2214.2009.01054.x

Pennington, L., & Thomson, K. (2007). It Takes Two to Talk–The Hanen Program® and families of children with motor disorders: a UK perspective. *Child: care, health and development*, *33*(6), 691-702.

https://doi.org/10.1111/j.1365-2214.2007.00800.x

Pennington, L., Thomson, K., James, P., Martin, L., & McNally, R. (2009). Effects of It Takes Two to Talk—The Hanen Program for Parents of preschool children with cerebral palsy: Findings from an exploratory study. *Journal of Speech, Language, and Hearing Research, 52*(5), 1121–1138. https://doi.org/10.1044/1092-4388(2009/07-0187)

Pepper, J., & Weitzman, E. (2004). *It takes two to talk: A practical guide for parents of children with language delays*. The Hanen Centre.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology & Community Health*, *57*(7), 527-529. https://doi.org/10.1136/jech.57.7.527

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Faculty Publications, Department of Psychology. 435.* https://doi.org/10.1044/1092-4388(2006/056)

Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and language*, *71*(2), 267-284. https://doi.org/10.1006/brln.1999.2166

Roberts, M.Y., & Kaiser, A.P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology, 20*(3), 180–199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Rogers, J., & Revesz, A. (2019). Experimental and quasi-experimental designs. In *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge. https://doi.org/10.4324/9780367824471-12

Roth, C. (2011). Boston naming test. *Encyclopedia Clin Neuropsychol*, *2*, 430-433. https://doi.org/10.1007/978-0-387-79948-3_869

Salerni, N., & Suttora, C. (2022). Semantic Contingency of Maternal Verbal Input Directed at Very Preterm and Full-Term Children. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.800568

Scarborough, H. S. (1990). Index of productive syntax. *Applied psycholinguistics*, *11*(1), 1-22. https://doi.org/10.1017/s0142716400008262

Semel, E. M., Wiig, E. H., & Secord, W. (1987). CELF-R: Clinical evaluation of language fundamentals--revised. *Psychological Corporation, Harcourt Brace Jovanovich.*

Senent-Capuz, N., Baixauli-Fortea, I., & Moret-Tatay, C. (2021). Parent-Implemented Hanen Program It Takes Two to Talk®: An Exploratory Study in Spain. *International journal of environmental research and public health*, *18*(15), 8214. https://doi.org/10.3390/ijerph18158214

Sheldon, M.L., & Rush, D.D. (2001). The ten myths about providing early intervention services in natural environments. *Infants and Young Children, 14,* 1–13. https://doi.org/10.1097/00001163-200114010-00004

Stella, G., Pizzoli, C., & Tressoldi, P. (2000). Peabody. Test Psicolinguistico.

Stewart, B. (2006). Parent-child interaction checklist. *Cairns City, Australia: Queensland Health*.

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: matching methods for causal inference. *Best practices in quantitative methods,* 155-176. https://doi.org/10.4135/9781412995627.d14

Suraniti, S., Ferri, R., & Neri, V. (2009). TROG 2–test for reception of grammar Version 2 (Edizione Italiana). *Firenze (IT): Giunti OS*.

The Hanen Centre. (2016a). *Not-For-Profit Charity.* https://www.hanen.org/About-Us/Not-For-Profit-Charity.aspx

The Hanen Centre. (2016b). *Find a Hanen Speech Language Pathologist.* https://www.hanen.org/Special-Pages/Members-Search.aspx

Vance, B., Kutsick, K., & West, R. (1987). Concurrent validity of the Peabody Picture Vocabulary Test-R and the Expressive One-Word Picture Vocabulary Test for language-delayed and non-language-delayed young children. *Diagnostique*, *13*(1), 3-9. https://doi.org/10.1177/073724778701300101

Weitzman, E., Girolametto, L. & Drake, L. (2017). Hanen Programs® for Parents: Parent Implemented Early Language Intervention. In *Treatment of Language Disorders in Children*, 2nd ed.; McCauley, R.J., Fey, M.E., Gillam, R.B., Paul, H., Eds.; Brookes Publishing Co.: Baltimore, MD, USA, 2017; pp. 27–56.

Wetherby, A. M., Allen, L., Cleary, J., Kublin, K., & Goldstein, H. (2002). Validity and reliability of the communication and symbolic behavior

scales developmental profile with very young children. *Journal of Speech, Language, and Hearing Research, 45*(6), 1202-1218. https://doi.org/10.1044/1092-4388(2002/097)

Wetherby, A. M., & Prizant, B. M. (2002). Communication and Symbolic Behavior Scales: Developmental Profile, (1st normed ed.). *Paul H Brookes Publishing Co.* https://doi.org/10.1037/t11529-000

Whittingham, K., Wee, D., & Boyd, R. (2011). Systematic review of the efficacy of parenting interventions for children with cerebral palsy. *Child: care, health and development*, *37*(4), 475-483. https://doi.org/10.1111/j.1365-2214.2011.01212.x

Wiig, E. H., Secord, W., & Semel, E. (1992). CELF-Preschool. Clinical Evaluation of Language Fundamentals—Preschool. Examiner's Manual. *The Psychological Corporation, San Antonio*.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). The Preschool Language Scale-3. *San Antonio: The Psychological Corporation Harcourt Brace Jovanovich.*

Appendices

Appendix A: List of Excluded Studies

Table A1.

References of Excluded Studies and Exclusion Criteria Met

Study Reference	Stage of Screening Excluded	Exclusion Criteria
Baxendale, J., Frankham, J., & Hesketh, A. (2001). The Hanen Parent Programme: a parent's perspective. *International Journal of Language & Communication Disorders*, *36*(sup1), 511-516. https://doi.org/10.3109/13682820109177938	Abstract Screening	5 Outcome Measure
Bizikova, L., Jungcurt, S., McDougal, K., & Tyler, S. (2020). How can agricultural interventions enhance contribution to food security and SDG 2.1?. *Global Food Security*, *26*, 100450. https://doi.org/10.1016/j.gfs.2020.100450	Title Screening	3 Intervention
Chaney, D., Brown, R. J., & Shelton, T. S. (1999, October). SIRTF prototype telescope. In *Advanced Telescope Design, Fabrication, and Control* (Vol. 3785, pp. 48-55). SPIE. https://doi.org/10.1117/12.367620	Title Screening	3 Intervention
Costa, M. D., Bergmann, J. A. G., Resende, A. S. C., Fonseca, C. G., & Faria, F. J. C. (2005). Study on genetic subdivision of the Mangalarga Marchador horse breed. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, *57*, 272-280. https://doi.org/10.1590/s0102-09352005000200021	Title Screening	3 Intervention
Cunningham, B. J., Kwok, E., Earle, C., & Oram Cardy, J. (2019). Exploring participation and impairment-based outcomes for Target Word™: A parent-implemented intervention for preschoolers identified as late-to-talk. *Child Language Teaching and Therapy*, *35*(2), 145-164. https://doi.org/10.1177/0265659019846931	Title Screening	3 Intervention
de Carlos Isla, M., & Baixauli Fortea, I. (2016). Parent-implemented Hanen program More than words in Angelman syndrome: A case study. *Child Language Teaching and Therapy*, *32*(1), 35-51. https://doi.org/10.1177/0265659014567784	Title Screening	3 Intervention

Study Reference	Stage of Screening Excluded	Exclusion Criteria
Evans, J., & Bricker, D. (1982). *Differences in Language and Cognitive Development: ECI-3. Early Childhood Intervention Catalog Module*.	Title Screening	3 Intervention
Friedman, D. B., Thomas, T. L., Owens, O. L., & Hébert, J. R. (2012). It takes two to talk about prostate cancer: A qualitative assessment of African American men's and women's cancer communication practices and recommendations. *American journal of men's health*, *6*(6), 472-484. https://doi.org/10.1177/1557988312453478	Title Screening	3 Intervention
Girolametto, L., & Weitzman, E. (2006). It takes two to talk—The Hanen program for parents: Early language intervention through caregiver training. *Treatment of language disorders in children*, 77-103. In McCauley, R. J., Fey, M. E., & Gillam, R. B. (Eds.). (2006). *Treatment of language disorders in children*. Paul H. Brookes Pub..	Absract Screening	5 Outcome Measure
Girolametto, L., & Weitzman, E. (2007). Promoting peer interaction skills: Professional development for early childhood educators and preschool teachers. *Topics in language disorders*, *27*(2), 93-110. https://doi.org/10.1097/01.tld.0000269927.96009.b7	Title Screening	3 Intervention
Hayes, N., & Rooney, T. (2019). 'I do it all the time! My mam does it!' Leveraging the familiar to enhance communication skills in early years educators. *Early Child Development and Care*, *189*(5), 707-717. https://doi.org/10.1080/03004430.2017.1339277	Title Screening	3 Intervention
Hettiarachchi, S. (2022). The effectiveness of the adapted Box Clever language intervention programme in the development of vocabulary and narrative skills of deaf and hard of hearing children. *Deafness & Education International*, *24*(1), 65-91. https://doi.org/10.1080/14643154.2020.1721158	Title Screening	3 Intervention
Hsiao, C., Koren-Karie, N., Bailey, H., & Moran, G. (2015). It takes two to talk: Longitudinal associations among infant–mother attachment, maternal attachment representations, and mother–child emotion dialogues. *Attachment & human development*, *17*(1), 43-64. https://doi.org/10.1080/14616734.2014.981671	Abstract Screening	3 Intervention
Johnsen, L., Lyckegaard, N. B., Khanal, P., Quistorff, B., Raun, K., & Nielsen, M. O. (2018). Fetal over-and undernutrition differentially program thyroid axis adaptability in adult sheep. *Endocrine*	Title Screening	3 Intervention

Study Reference	Stage of Screening Excluded	Exclusion Criteria
connections, *7*(5), 777. https://doi.org/10.1530/ec-18-0014		
Kamp, W. M., Sellers, C. M., Stein, S., Lim, J. K., & Kim, H. S. (2020). Direct-acting antivirals improve overall survival in interventional oncology patients with hepatitis C and hepatocellular carcinoma. *Journal of Vascular and Interventional Radiology*, *31*(6), 953-960. https://doi.org/10.1016/j.jvir.2019.12.809	Title Screening	3 Intervention
Konza, D., Maloney, C., & Grafton, P. (2010). It Takes Two to Talk: A Focused Intervention Program for Parents and Children with Language Delays. *International Journal of Interdisciplinary Social Sciences*, *5*(6). https://doi.org/10.18848/18331882/cgp/v05i06/59301	Full Text Screening (included at Ancestral Search)	5 Outcome Measure
Kwok, E. Y., Jane Cunningham, B., & Oram Cardy, J. (2020). Effectiveness of a parent-implemented language intervention for late-to-talk children: a real-world retrospective clinical chart review. *International journal of speech-language pathology*, *22*(1), 48-58. https://doi.org/10.31219/osf.io/k9p8e	Abstract Screening	3 Intervention
Lowman, A. E., Macenka, S. A., Redding, D. C., & Basinger, S. A. (1998, August). Infrared Telescope Technology Testbed primary mirror test results. In *Space Telescopes and Instruments V* (Vol. 3356, pp. 727-734). SPIE. https://doi.org/10.1117/12.324491	Title Screening	3 Intervention
McCauley, R. J., Fey, M. E., & Gillam, R. B. (Eds.). (2006). *Treatment of language disorders in children*. Paul H. Brookes Pub..	Abstract Screening	3 Intervention
Moeller, D., Probst, P., & Hess, M. (2008). Implementation and evaluation of a parent training for language delayed children. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *57*(3), 197-215. https://doi.org/10.13109/prkk.2008.57.3.197	Full Text Screening	2 Language (full text published in German)
Newbury, J., & Sutherland, D. (2020). Measurement of child-directed speech: A survey of clinical practice. *International Journal of Speech-Language Pathology*, *22*(4), 399-413. https://doi.org/10.1080/17549507.2019.1650111	Absract Screening	3 Intervention
Pennington, L., & Noble, E. (2010). Acceptability and usefulness of the group interaction training programme It Takes Two to Talk to parents of pre-	Full Text Screening	5 Outcome Measure

Study Reference	Stage of Screening Excluded	Exclusion Criteria
school children with motor disorders. *Child: care, health and development*, *36*(2), 285-296. https://doi.org/10.1111/j.1365-2214.2009.01054.x		
Pennington, L., & Thomson, K. (2007). It Takes Two to Talk–The Hanen Program® and families of children with motor disorders: a UK perspective. *Child: care, health and development*, *33*(6), 691-702. https://doi.org/10.1111/j.1365-2214.2007.00800.x	Full Text Screening	5 Outcome Measure
Rose, T., Scarinci, N., Meyer, C., Harris, S., Forsingdal, S., Anger, N., & Webb, K. (2020). The It Takes Two to Talk®–The Hanen Program® for Parents: impacts on child behaviour and social-emotional functioning. *Speech, Language and Hearing*, *23*(3), 180-188. https://doi.org/10.1080/2050571x.2019.1622832	Full Text Screening	5 Outcome Measure
Scarinci, N., Rose, T., Cronan, A., & Lambertz, K. (2018). Speech pathology student experiences and perceptions of working with parents in a Hanen It Takes Two to Talk family-centred clinical placement. *Speech, Language and Hearing*, *21*(3), 132-141. https://doi.org/10.1080/2050571x.2017.1329890	Abstract Screening	5 Outcome Measure
Senent-Capuz, N., Fortea, I. B., & Perales, M. J. (2022). Social Validity Evaluation of the Hanen Program It Takes Two to Talk® in Spain. *Communication Disorders Quarterly*, *43*(4), 224-233. https://doi.org/10.1177/1525740120967013	Abstract Screening	5 Outcome Measure
Stahl, H. P., Radacsi, D., Heydenburg, T. J., Gehan, A., Bourgeois, R. P., Radomski, B., ... & Macenka, S. A. (1997, November). Fabrication and testing of the ITTT beryllium secondary mirror. In *Optical Manufacturing and Testing II* (Vol. 3134, pp. 62-71). SPIE. https://doi.org/10.1117/12.295156	Title Screening	3 Intervention
Stier, M. T., Crout, R. R., Hansen, D. A., Krim, M. H., Nonnenmacher, A. L., Paquin, R. A., ... & Vollaro, J. (1998, August). Telescope design for the infrared telescope technology testbed. In *Space Telescopes and Instruments V* (Vol. 3356, pp. 712-726). SPIE. https://doi.org/10.1117/12.324490	Title Screening	3 Intervention
Syal, S., & Anderson, A. K. (2013). It takes two to talk: A second-person neuroscience approach to language learning. *Behavioral and Brain Sciences*, *36*(4), 439. https://doi.org/10.1017/s0140525x12002130	Abstract Screening	5 Outcome Measure

Study Reference	Stage of Screening Excluded	Exclusion Criteria
Te Kaat-van den Os, D. J., Jongmans, M. J., Volman, M. C. J., & Lauteslager, P. E. (2017). Parent-implemented language interventions for children with a developmental delay: A systematic review. *Journal of Policy and Practice in Intellectual Disabilities*, *14*(2), 129-137. https://doi.org/10.1111/jppi.12181	Abstract Screening	6 Research Design
Weitzman, E. (1994). The Hanen Program for Early Childhood Educators: Inservice Training for Child Care Providers on How to Facilitate Children's Social, Language, and Literacy Development. *Infant-Toddler Intervention: The Transdisciplinary Journal*, *4*(3), 173-202.	Absract Screening	5 Outcome Measure
Whittingham, K., Wee, D., & Boyd, R. (2011). Systematic review of the efficacy of parenting interventions for children with cerebral palsy. *Child: care, health and development*, *37*(4), 475-483. https://doi.org/10.1111/j.1365-2214.2011.01212.x	Absract Screening	6 Research Design
Zulkifli, S., Short, K., Kleiman, C., Kidd, J. C., Earley, J., Beckett, S., ... & McCabe, P. (2022). Evaluating the dose–response relationship of the number of sessions of “It Takes Two to Talk®” in young children with language delay. *International Journal of Speech-Language Pathology*, 1-11. https://doi.org/10.1080/17549507.2022.2080270	Full Text Screening	6 Research Design

Appendix B: Mapping the Field

Table B1.

Description of Studies Included in Review

Author (date), Country	Sample: Child Participants	Study Design	Participants (Age, gender, need)	Intervention	Outcome Measures	Assessment Points
Baxendale & Hesketh (2003) United Kingdom	37 Intervention group:19	Quasi-experimental pre/post-test design Clinician-directed therapy comparison group Allocation to group by geographical location	Intervention group: Age M=31.5 months Gender: 14 males, 5 females Need: 7 expressive language delay, 12 receptive and expressive language delay Comparison group Age: M=34.3 months Gender: 16 males, 2 females Need: 8 expressive	Hanen Parent Programme: 11-week programme including 8 weekly group sessions for parents and three home visits delivered by speech and language therapists	Child language measures: Pre-school Language Scale-3 (UK) (PLS-3-UK) Child interaction measures: Language output of child analysed from audio-taped interactions by mean length of utterance (MLU) and proportional number of utterances (PROPUTT) Parent measures: Parent language-modelling techniques (PARESP)	Assessment 1: pre-intervention Assessment 2: 6 months later (post-intervention) Assessment 3: 12 months after assessment 1

Author (date), Country	Sample: Child Participants	Study Design	Participants (Age, gender, need)	Intervention	Outcome Measures	Assessment Points
			language delay, 10 receptive and expressive language delay			
Cologon, Wicks & Salvador (2017) Australia	10 Experimental group in phase 2:5	Phase 1: Single group pre/post-test design Phase 2: Randomised control design (no further treatment control group) Convenience sample	Age: M=33.4 months Gender: 5 males, 5 females Need: 5 down syndrome, 3 global developmental delay, 2 Rubenstein Taybi syndrome	Phase 1: ITTT program, 8 group sessions with parents only and 3 home visits, facilitated by certified SLT Phase 2: four additional monthly individual sessions for experimental group only	Child interaction measures: Parent-child interaction checklist (PIC) assessed using video-footage of parent-child interactions. Parent measures: Caregiver evaluations	Pre-intervention and post-intervention (after 10-12 weeks) Follow-up assessment four months after post-intervention assessment
Nicastri et al. (2021) Italy	28 Intervention group:14	Quasi-experimental pre/post-test design	Intervention group: Age: M=25.6 months Gender: 7 males, 7 females	ITTT program with additional information relevant to	Child language measures: MacArthur-Bates Communicative Development Inventory (MCDI) and Gestures and Words Form	Pre-intervention (within 1 month of start of study)

Author (date), Country	Sample: Child Participants	Study Design	Participants (Age, gender, need)	Intervention	Outcome Measures	Assessment Points
		No treatment control group (matched for: parental SES, education, child age, hearing age, pre-implant tone average and language level) Allocation to groups by arrival order at center	Need: children who were profoundly deaf and received Cochlear Implants Control group: Age: M=26.2 months Gender: 7 males, 7 females Need: children who were profoundly deaf and received Cochlear Implants	families of DHH children Phase 1: 9 parent groups sessions, 3 individual sessions Phase 2: 6 additional individual sessions	Parent Measures: The Communication-Promoting Behaviours Checklist for Caregivers Parent Stress Index-Short Form (PSI-SF) 3 years post-intervention child language measures: Boston Naming Test, Peabody Picture Vocabulary Test, Test for Reception of Grammar- Version 2, Speech recognition Test	Post-intervention (within one month of the end of PT) Child assessment at 3-year follow-up
Pennington et al. (2009) United Kingdom and Australia	11	Single group repeated measures design No control group	Age: M=26 months Gender: 8 males, 3 females Need: 10 with cerebral palsy and 1 with myopathy	ITTT program: 7-8 parent group sessions and 3 home visits over approx.13 weeks	Child interaction measures: Child moves and pragmatic functions coded from video-taped parent-child interactions (coding system based on discourse analysis method) Parent measures:	Pre-intervention (4-months pre and 1-month pre) Post-intervention (1-month post

Author (date), Country	Sample: Child Participants	Study Design	Participants (Age, gender, need)	Intervention	Outcome Measures	Assessment Points
					Coding of maternal language input (Systematic Analysis of Language Transcripts)	and 4-months post)
Senent-Capuz, Baixauli-Fortea & Moret-Tatay, (2021) Spain	17 Intervention group: 10	Quasi-experimental pre-post-test design Clinician-directed therapy comparison group Groups matched for gender, age, cognitive level and expressive language Allocation to group by parental preference	Intervention group: Age: M=29 months Gender: 5 males, 5 females Need: met the criteria for late-talking children Control Group: Age: M=29 months Gender: 4 males, 3 females Need: met the criteria for late-talking children	ITTT program: 8 parent group sessions, 3 home visits	Child language measures: MacArthur-Bates Communicative Development Inventory (MCDI), Reynell Developmental Language Scales (RDLS-III) Child interaction measures: Communication and Symbolic Behaviour Scales Developmental Profile (CSBS DP) Parent measures: Parenting Stress Index-Short Form (PSI-SF) Parent Perception of Language development (PPOLD)	Pre-intervention (1-month pre-) Post-intervention (6 months post)

Table B2.

Reliability and Validity Information for Outcome Measures

Study Author (date)	Outcome Measure	Used to Measure	Reliability	Validity
Baxendale & Hesketh (2003)	Pre-school Language Scale-3 (UK) (PLS-3-UK) (Zimmerman et al., 1992; Boucher & Lewis, 1997)	Standardised measure of receptive and expressive language	For children over one year of age, internal consistency α=.73-.94 across scales Test-retest stability coefficient r=.82-.94 across scales Percentage of inter-rater agreement 89%, r=.98 (Boucher & Lewis, 1997)	Construct validity measured through correlations between scales, r=.64. Interpreted as evidence for distinct scales Concurrent validity with the Clinical Evaluation of Language Fundamentals-Revised (CELF-R, Semel et al., 1987), r=.69-.82 across scales (Boucher & Lewis, 1997)
	Mean length of utterance (MLU) (Brown, 1973)	Index of linguistic productivity	Inter-rater agreement in Baxendale & Hesketh (2003), r_s=.998	Concurrent validity with other measures of clausal development including Developmental Sentence Scoring (Lee

				et al., 1974), r=.56-.70, and Index of Productive Syntax (Scarborough, 1990), r=.70-.80 (Rice et al., 2006)
	Proportional number of utterances (PROPUTT) (author developed)	Ratio of child to parent utterances	Inter-rater agreement in Baxendale & Hesketh (2003), r_s=.943	This measure is described in the paper as a 'new tool', and thus research in not available regarding the measure's validity
Cologon, Wicks & Salvador (2017)	Parent-child interaction checklist (PIC) (Stewart, 2006)	Observational rating scale of child and caregiver interactions	Percentage of inter-rater agreement in Cologon et al. (2017) 67%	No validity data available
Nicastri et al. (2021)	MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993; Caselli & Casadio, 2007)	Standardised measure of children's early language development, including vocabulary comprehension and expression and use of gestures and phrases	Internal consistency, α=.39 (infant form Gesture scale)-.96 (vocabulary scales) (Fenson et al., 1993) Test-retest reliability, r=.8-.9 across scales, except for 12-month group where r=.6, perhaps reflecting rapid language development	Concurrent validity with the Expressive One Word Picture Vocabulary Test (Gardner, 1981), r=.73 (Dale, 1991)

		at this age (Fenson et al., 1993)	
Boston Naming Test (Kaplan et al., 1983; Riva et al., 2000)	Standardised measure of object naming ability	Internal consistency α=.78-.96 (Roth, 2011)	Concurrent validity with the Visual Naming Test of the Multilingual Aphasia Examination (Benton & Hamsher, 1989), r=.86 (Axelrod et al., 1994)
Peabody Picture Vocabulary Test (Dunn & Dunn, 1981; Stella et al., 2000)	Standardised test of receptive vocabulary	Internal consistency α=.61-.88 (Dunn & Dunn, 1981)	Concurrent validity with the Expressive One-Word Picture Vocabulary Test (Gardner, 1981) r=.74 (Vance et al., 1987).
Test for Reception of Grammar- Version 2 (Bishop & Garsell, 2003; Suraniti et al., 2009)	Standardised test measuring understanding of grammatical contrasts	Internal consistency α=.88 (Bishop & Garsell, 2003)	Concurrent validity with CELF-Preschool (Wiig et al., 1992), r=.58. (Bishop & Garsell, 2003)
Speech recognition Test (unspecified) Sentences sourced from Cutugno et al. (2000)	Unspecified test of speech recognition	This measure is unspecified and thus no reliability information is available	This measure is unspecified and thus no validity information is available

Pennington et al. (2009)	Interaction coding scheme based on discourse analysis (developed by Pennington & McConachie, 2001)	Measure of child moves and pragmatic functions during parent-child interactions	Inter-rater agreement in Pennington et al. (2009) k=.78, p<.01 for children's moves and k=.77, p<.01 for children's functions	This measure is author developed and thus no validity information is available
Senent-Capuz et al. (2021)	MacArthur-Bates Communicative Development Inventories- Spanish Adaptation (López Ornat et al., 2005)	Standardised measure of children's early language development, including vocabulary comprehension and expression and use of gestures and phrases	Internal consistency α=.99 for vocabulary and morphosyntactic complexity (López Ornat et al., 2005)	Concurrent validity with observational and MLU language measures, r=.65-.89 (Mariscal et al., 2010)
	Reynell Developmental Language Scales (RDLS-III) (Edwards et al., 1999)	Standardised measure of expressive and receptive language	Internal consistency α=.97 for receptive scale and α=.96 for expressive scale (Edwards et al., 1999)	Concurrent validity with British Picture Vocabulary Scale II (Dunn et al., 1997) r=.68-.75, and Test of Reception of Grammar (Bishop, 1982) r= .67-.70 (Edwards et al., 1999)

	Communication and Symbolic Behaviour Scales Developmental Profile (CSBS DP) (Wetherby & Prizant, 2002)	Standardised measure of social, communication and symbolic abilities	Internal consistency α=.73-.95 across subscales (Wetherby & Prizant, 2002)	Construct validity assessed through correlations between CSBS DP measures, r=.59-.71 (Wetherby et al., 2002)

Appendix C: Adaptations to Gersten et al. (2005) Coding Protocol

The Gersten et al. (2005) coding protocol for group experimental and quasi-experimental research in special education was adapted for the current review. The option of 'Not applicable/ Unknown' was added for all quality indicators, to allow for complete coding. For example, essential quality indicator two refers to comparison across conditions, however one of the studies in the review (Pennington et al., 2009) does not include a control condition, therefore the code 'not applicable' is most appropriate. It was judged appropriate to included indicators that involve comparison across conditions, despite not all of the studies in the review including a control condition, as studies that include a control condition should receive a higher methodological quality score.

Appendix D: Weight of Evidence Rating Criteria

WoE A: Methodological Quality

Table D1 shows the WoE A threshold criteria as outlined by Gersten et al. (2005). Only high quality and acceptable thresholds are outlined by Gersten et al. (2005), thus the reviewer created thresholds for low and very low ratings. Table D2 shows the WoE A ratings by criteria category for each study.

Table D1.

WoE A Criteria with Rationale

WOE A Rating	Criteria	Rationale
3 (high)	≥ 9 essential criteria and ≥ 4 desirable criteria	These thresholds are outlined in Gersten et al. (2005)
2 (medium)	≥ 9 essential criteria and ≥ 1 desirable criteria	
1 (low)	≥ 6 essential criteria but <9	Threshold created by reviewer
0 (very low)	<6 essential criteria	Threshold created by reviewer

Table D2.

WoE A Ratings

Quality Indicator Category	Baxendale & Hesketh (2003)	Cologon, Wicks & Salvador (2017)	Nicastri et al. (2021)	Pennington et al. (2009)	Senent-Capuz, Baixauli-Fortea & Moret-Tatay (2021)
Describing participants (/3)	3	2	3	2	3
Intervention implementation (/3)	1	2	2	0	2
Outcome measures (/2)	2	1	2	1	2
Data analysis (/2)	1	1	2	2	1
Total essential (/10)	7	6	9	5	8
Total desirable (/8)	4	4	4	3	5
WoE A rating	1 (low)	1 (low)	3 (high)	0 (very low)	1 (low)

WoE B: Methodological Relevance

WoE B is a review-specific judgement related to the quality and relevance of the methodology of each study for addressing the review question. Thus, for the present review, WoE B assessed the methodology for evaluating the effectiveness of It Takes Two to Talk- The Hanen Program for improving the language and interaction skills of pre-school children with SEND.

WoE B criteria were created by the reviewer and are shown in Table D3. The five criteria were study design, control condition, control group allocation, data collection and justification of design. The score for each study was averaged across WoE B criteria to produce an overall WoE B rating. A summary of scores is presented in Table D4.

Table D3.

WoE B Criteria with Rationale

Criteria	Weighting	Rationale
Study design	3 Randomised control trials (RCTs) as the only design used 2 Quasi-experimental studies, mixed designs and cohort studies 1 Designs that are not experimental e.g. qualitative research, surveys, non-experimental evaluations	Petticrew and Roberts (2003) outlined this hierarchy of study designs with regards to their appropriateness for effectiveness research questions

Criteria	Weighting	Rationale
Control condition	3 Active control group (alternative intervention or therapy) 2 Wait-list or no intervention control 1 No control group	Active control groups allow the effects of the intervention to be isolated from the effects of similar interventions or therapies
Control group allocation	3 Control groups were randomly allocated or matched 2 Control groups were not randomly allocated or matched 1 No control group	Selection bias is reduced through random allocation. (Rogers & Revesz, 2019). However, a matched groups procedure has been argued to be more appropriate for some educational research where randomisation is not possible (Stuart & Rubin, 2008).
Data collection	3 Outcomes are measured pre and post-intervention and at a follow-up time point 2 Outcomes are measured pre and post-intervention 1 Outcomes are measured post-intervention	Pre-post measures allow measurement of intervention effects over time, with follow-up measures showing lasting intervention effects (Gersten et al., 2005)
Justification of research design and discussion of limitations	3 Authors justify their research design and reference potential limitations of that design 2 Authors justify their research design 1 No justification of the research design is presented	Justification for using a quasi-experimental design and discussion of limitations of the approach are cited as criteria for effective quasi-experimental research designs (Eliopoulos et al., 2005; Moreno-Fernandez et al., 2008)

Table D4.

Summary of WoE B Ratings

Study	Study design	Control condition	Control group allocation	Data collection	Justification of design	Overall WoE B
Baxendale & Hesketh (2003)	2	3	2	3	3	2.6 (high)
Cologon, Wicks and Salvador (2017)[c]	2	2	3	3	3	2.6 (high)
Nicastri et al. (2021)	2	2	3	3	3	2.6 (high)
Pennington et al. (2009)	2	1	1	3	3	2 (medium)
Senent-Capuz, Baixauli-Fortea & Moret-Tatay (2021)	2	3	3	2	1	2.2 (medium)

Note. WoE B ratings ≥ 2.5 are considered 'high', <2.5 and ≥1.5 are considered 'medium' and <1.5 are considered 'low'

[c] As the study included multiple phases, WoE A criteria related to the use of a control group refers only to phase 2 of the study.

WoE C: Topic Relevance

WoE C assesses the relevance of each study's topic to the present review question. The four criteria created by the reviewer included the intervention implemented, description of participant need, assessment of participant need and the scope of language and interaction outcome measures used, as shown in Table D5. Table D6 presents a summary of WoE C ratings.

Table D5.

WoE C Ratings Criteria

Criteria	Ratings	Rationale
Intervention	3 ITTT is the only intervention implemented 2 an adapted version of ITTT is the main intervention condition (e.g. increased number of sessions, additional session content not specified in the Hanen Program) 1 ITTT is not the main intervention, or The Hanen Program is specified without specifying ITTT	Studies that use the standard ITTT intervention are most relevant to the review which aim to assess the effectiveness of the standard programme
Participant need described	3 Participants were described as having a specific language or communication need 2 Participants were described as having a need, but a specific language or communication need was not specified 1 Participants were not identified as having any additional need	The ITTT programme was developed to support pre-school children with language delay (Pepper & Weitzman, 2004), thus studies that do not identify their participants as having a language delay or need are less relevant to the review
Participant need assessed	3 Child language level was assessed by the study's researchers pre-intervention 2 Child language level was assessed pre-intervention, not by the study's researchers 1 Child language level was not assessed prior to inclusion in the intervention	Gersten et al. (2005) describe the importance of researchers assessing the needs of participants, rather than relying on previous diagnostic labels, to ensure accurate description of need and to support the generalisability of the results to other

Criteria	Ratings	Rationale
		individuals with similar needs
Scope of child language and interaction outcome measures	3 More than two child outcomes have been measured, that assess both language and interaction skills 2 More than one child language or interaction outcome has been measured 1 Only one child language or interaction outcome has been measured	As the review focuses on two separate skills, language and interaction, separate measures are needed to assess these constructs. Gersten et al. (2005) describe the importance of studies including multiple tools to measure each construct

Table D6

Summary of WoE C Ratings

Study	Intervention	Need Described	Need Assessed	Outcome Measure	Overall WoE C
Baxendale & Hesketh (2003)	1	3	3	3	2.5 (high)
Cologon, Wicks and Salvador (2017)	2	2	1	1	1.5 (medium)
Nicastri et al. (2021)	2	2	3	2	2.25 (medium)
Pennington et al. (2009)	3	2	3	1	2.25 (medium)
Senent-Capuz, Baixauli-Fortea & Moret-Tatay (2021)	3	3	3	3	3 (high)

Note. WoE B ratings ≥ 2.5 are considered 'high', <2.5 and ≥1.5 are considered 'medium' and <1.5 are considered 'low'

Appendix E: Completed Coding Protocol

Baxendale, J., & Hesketh, A. (2003). Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. International Journal of Language & Communication Disorders, 38(4), 397-415.

Essential Quality Indicators

Quality Indicators for Describing Participants

1. Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown
2. Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown
3. Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

4. Was the intervention clearly described and specified?
   ☐Yes
   ☒No
   ☐Not applicable/ Unknown
5. Was the fidelity of implementation described and assessed?
   ☐Yes
   ☒No
   ☐Not applicable/ Unknown
6. Was the nature of services provided in comparison conditions described?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown

Quality Indicators for Outcome Measures

7. Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown
8. Were outcomes for capturing the interventions effect measured at the appropriate times?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown

Quality Indicators for Data Analysis

9. 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
   ☐Not applicable/ Unknown
10. Did the research report include not only inferential statistics but also effect size calculations?
    ☐Yes
    ☒No
    ☐Not applicable/ Unknown

Desirable Quality Indicators

1. 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
   ☐Not applicable/ Unknown
2. 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
   ☐Not applicable/ Unknown
3. Were outcomes for capturing the intervention's effect measured beyond an immediate posttest?

☒Yes
☐No
☐Not applicable/ Unknown

4. Was evidence of the criterion-related validity and construct validity of the measures provided?
   ☐Yes
   ☒No
   ☐Not applicable/ Unknown
5. 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
   ☐Not applicable/ Unknown
6. Was any documentation of the nature of instruction or series provided in comparison conditions?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown
7. Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?
   ☐Yes
   ☒No
   ☐Not applicable/ Unknown
8. Were results presented in a clear, coherent fashion?
   ☒Yes
   ☐No
   ☐Not applicable/ Unknown

Total Essential = 7
Total Desirable = 4
WoE A Rating = 1 (low)