

RESEARCH REVIEW

A systematic review and meta-analysis of nonpharmacological interventions for children and adolescents with selective mutism

Gino Hipolito^{1,2,3}  | Emma Pagnamenta¹  | Helen Stacey¹  | Emily Wright¹  |
Victoria Joffe⁴  | Kou Murayama^{1,5}  | Cathy Creswell^{3,6} 

¹School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

²Paediatric Speech and Language Therapy Department, St George's University Hospitals NHS Foundation Trust, London, UK

³Department of Experimental Psychology, University of Oxford, Oxford, UK

⁴School of Health and Social Care, University of Essex, Colchester, UK

⁵Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany

⁶Department of Psychiatry, University of Oxford, Oxford, UK

Correspondence

Gino Hipolito, Paediatric Speech and Language Therapy Department, St George's University Hospitals NHS Foundation Trust, London, UK.
Email: gino.hipolito@stgeorges.nhs.uk

Funding information

Alexander von Humboldt-Stiftung; Programme Grants for Applied Research; NIHR Predoctoral Clinical Academic Fellowship, Grant/Award Number: NIHR-ICA-PCAF-2018-01-020; NIHR Research Professorship, Grant/Award Number: RP 2014-04-018; Oxford and Thames Valley NIHR Applied Research Collaboration

Abstract

Background: Selective mutism (SM) is an anxiety disorder that often starts in early years with serious and lasting consequences. Nonpharmacological interventions are commonly seen as the preferred first treatment. This systematic review identifies outcome measures used and outcomes achieved for nonpharmacological interventions for children and adolescents with SM.

Methods: Systematic searches were conducted using 13 electronic databases and hand searches, including peer-reviewed and grey literature since 1992.

Results: Twenty-five studies were identified. While specific measures varied, all studies reported an outcome measure for speaking behaviour and 18 used a measure of anxiety. Few studies reported measures of SM remission ($k = 6$), well-being ($k = 6$), academic impact ($k = 2$), or quality of life ($k = 1$). Within subject outcomes for nonpharmacological interventions were variable for improvements in speaking behaviours (very small to large positive effects) and reduction in anxiety symptoms (very small negative to large positive effects). Only five randomised controlled trials (RCTs) were included in the meta-analysis. Three studies compared a combined systems/behavioural approach with waitlist controls indicating a significant and large effect (Hedges $g = 1.06$, $p < .0001$, 95% CI: 0.57–1.56) on improved speaking behaviour. Two of these RCTs showed a large effect for SM remission favouring the intervention (Risk Ratio = 4.25, $p = .1774$, 95% CI: 0.52–34.84) but this did not reach statistical significance. Non-significant outcomes for two RCTs with active controls (Hedges $g = 0.55$, $p < .2885$, 95% CI: –0.47 to 1.57) showed considerable heterogeneity in approach and outcomes, one with large and one with negligible effects.

Conclusion: Despite the considerable impairment caused by SM, there has been little systematic evaluation of non-pharmacological interventions. Although combined systems/behavioural interventions are promising, further systematic evaluations are urgently needed to inform treatment approaches. Cross-study measurement harmonisation is required to promote learning from all studies, including wider clinical and economic outcomes.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. JCPP Advances published by John Wiley & Sons Ltd on behalf of Association for Child and Adolescent Mental Health.

Clinical Trial Registration: Not applicable.

KEYWORDS

anxiety, intervention, outcomes, selective mutism, speaking behaviour, systematic review

INTRODUCTION

Selective mutism (SM) is a condition in which a person is unable to speak in certain situations where speech is expected (e.g. school or in public) but is able to speak in other situations (e.g. home) (DSM5; American Psychiatric Associations, 2013). Selective mutism is an anxiety disorder that often starts between the ages of 2 and 5 years with a prevalence of 1 in 140 children under 8 years of age (Bergman et al., 2002; Elizur & Perednik, 2003). It has serious consequences during childhood and poor long-term outcomes. The inability to talk can have a negative impact on the child's education and social interactions with others in the short term (Bergman et al., 2002; Kumpulainen et al., 1998). There is also a high rate of emotional and behavioural problems in this group of children (Kristensen, 2001; Steinhausen & Juzi, 1996). Furthermore, two long-term follow up studies (Remschmidt et al., 2001; Steinhausen et al., 2006) assessed adults who accessed a range of interventions for SM in childhood (average age of 8.5 years) and found 42%–61% continued to have SM in adulthood hindering their communication in education, leisure activities, and at work, with many continuing to have emotional and behaviour problems. Clearly, there is a need for effective interventions to address SM and mitigate the short- and long-term negative effects of the condition.

To understand which interventions are effective, it is crucial for studies to use appropriate outcomes measures specific to the condition and that capture the impact and change of the disorder (Coster, 2013). In the first systematic review of SM interventions, Stone et al. (2002) found that 'most' studies used a narrow range of outcome measures, particularly focussed on speaking behaviour. In a more recent systematic review of assessment tools to screen and diagnose the core symptomology of SM, Rodrigues Pereira et al. (2021) found that the clinician administered Anxiety Disorders Interview Schedule (ADIS, Silverman & Albano, 1996) was most often used for diagnosing SM and comorbid anxiety disorders ($k = 23$), while the parent reported SM Questionnaire (Bergman et al., 2008) ($k = 32$) and teacher reported School Speech Questionnaire (Bergman et al., 2002) ($k = 11$) were the most common speaking behaviour outcomes in SM studies ($k = 56$) using a standardised measure in the last 10 years. However, overall there was a lack of consistency in the specific measures used for SM. This lack of consistency is a problem, as highlighted by Creswell et al. (2021), with reference to the general paediatric anxiety disorders literature, who argued that wide variation in outcome measures makes it difficult to compare and combine studies and risks faulty conclusions from meta-analyses. Furthermore, Stone et al. (2002) raised concerns that few intervention studies report on the *impact* of SM on children or the wider implications, for example, educational outcomes; and no recent systematic reviews have investigated these outcomes. As a first step it is important to capture the range of measures used across interventions studies for SM in order to make recommendations to promote greater consistency going forwards.

Key points

- This broad systematic review found an improvement in study designs and manualised treatments in the SM literature; however, few fully powered experimental studies, inconsistent use of measures, sampling gaps, insufficient reporting, and absence of health economic analyses prevent definitive answers to what works for these children and young people.
- Outcomes for SM remission and improving speaking behaviour for children (3–9 years) appear promising when using a combined systems and behavioural intervention approach.
- Future treatment studies are recommended to consistently report on intervention components, co-morbidity, onset, duration, and severity of SM.
- Regional and international research collaborations and an agreed measurement harmonisation for SM treatment studies are needed to enable more fully powered experimental studies to improve the evidence base.

In terms of treatment outcomes, although there is some evidence for efficacy of medication (specifically SSRIs) for children with SM, they are not typically seen as a first line treatment due to parental concerns about potential pharmacological side effects, particularly among young children (Manassis et al., 2016; Østergaard, 2018). On the other hand, nonpharmacological interventions, such as behavioural therapy, are perceived as effective and acceptable by parents of young children with SM (Bergman et al., 2013). Narrative reviews in the last 30 years have described nonpharmacological interventions and their outcomes (Anstendig, 1998; Cohan et al., 2006; Viana et al., 2009; Zakszeski & DuPaul, 2017) but have not systematically investigated effectiveness. Stone et al. (2002) provided the first synthesis of SM intervention research and showed that behaviour therapy was more effective in improving speaking behaviour than no treatment with a median effect size (standardised mean difference) of 1.63. However, the meta-analysis consisted of single case design studies only with small samples of up to 4 participants and so was limited to within group effect sizes which carry a high risk of bias. At that point in time, there were no randomised controlled trials (RCTs) in SM intervention research. Advances in SM research have led to larger studies and RCTs, prompting Steains et al. (2021) more recently to carry out a systematic review and meta-analysis of five RCTs. They found that psychological interventions were more effective than no treatment with a large treatment effect (overall weighted effect size Hedges' $g = 0.87$). However, this effect size combined a broad range of outcomes including speaking behaviour, SM severity rating, anxiety, global functioning and social anxiety

severity. Although a sensitivity analysis found that the means did not differ significantly for SM specific compared to non-SM specific outcomes, there was extensive heterogeneity within these two categories of outcomes. Consequently, it is not clear what the effect sizes were for the specific outcomes and whether they differed according to type of intervention. Moreover, no systematic review has investigated the rate of SM remission after nonpharmacological intervention further limiting our understanding of effectiveness.

The aim of this paper is to carry out a broad systematic review of nonpharmacological interventions for children and adolescents with SM. The systematic review will address the following questions:

1. What are the outcome measures used in nonpharmacological intervention studies for children and adolescents with SM?
2. What are the outcomes and remission rates in non-pharmacological interventions for children and adolescents with SM?
3. Which nonpharmacological interventions for children and adolescents are effective in increasing SM remission and improving speaking behaviour (the primary symptom of SM)?

METHODOLOGY

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009) and was registered on the PROSPERO International prospective register of systematic reviews on the 2 September 2019, registration number CRD42019147573. As we anticipated that there would be only a small number of RCTs, we opted to conduct a broad review including both RCTs and non-RCTs (including grey literature) to reduce publication bias and provide a comprehensive and balanced view of the evidence (Paez, 2017).

Searches in 13 databases were initially carried out between 20 September and 14 October 2019. Searches were replicated on the 29 January 2021 and 7 January 2022 to add further relevant studies. Double coding was used for the screening, full text eligibility and quality appraisal sections of the review with any disagreements resolved by an independent reviewer. Details of the search strategy, eligibility criteria, data extraction, selection procedures, data management and quality appraisal are given in Appendix S1 and Table S1.

Strategy for data synthesis

Due to the anticipated range of study designs in this broad review and to allow for comparison in the narrative synthesis, within case effect sizes of the intervention were calculated in each study using Hedges g standardised mean differences, which is useful for bias correction in small sample sizes (Borenstein et al., 2009), for continuous outcome variables. The Hedges g within case effect size was calculated for pre and post data using the open source 'effsize' package (Torchiano, 2020) in R if the raw data was available and a Hedges' g^* unbiased formula (Borenstein et al., 2009) in Microsoft Excel (2016) if only the pre and post mean and standard deviation were available. In the latter situation, a pretest-posttest correlation of 0.5 has been imputed in the formula. The direction of the within

case effect sizes was corrected to be in line with the between subject effect sizes for comparability.

The Hedges g between subject effect size (Hedges et al., 2013) was calculated for multiple baseline single case experimental design using the 'scdhlrm' package (Pustejovsky et al., 2020) in R to be in the same metric to compare with RCTs (Shadish et al., 2014). The Hedges g between subject effect size for RCTs were calculated using the 'esc' Package (Ludecke, 2019) in R. Positive values in effect sizes for speaking behaviour indicated that participants in the treatment group improved (relative to the control group where applicable). Negative values in effect sizes for anxiety measures indicated that the treatment group had lower anxiety levels after treatment (in comparison to the control group where applicable).

To measure the effectiveness of the interventions, random effects meta-analyses were conducted using 'metafor' package (Viechtbauer, 2010) in R where there were two or more eligible RCTs (Valentine et al., 2010). The speaking behaviour outcomes (primary SM symptom) were calculated as Hedges' g standardised mean difference while SM remission outcomes were converted into risk ratios. The Q statistic was used as a test of significance of heterogeneity (Borenstein et al., 2009), while the I^2 index evaluated the extent of the heterogeneity between studies (Deeks et al., 2008).

RESULTS

The search strategy yielded 810 references after duplicates and pre-1992 studies were removed. The references were screened by title and abstract with exclusion codes recorded. The reasons for exclusion at this stage were due to: non-intervention study; not specific to SM; $n < 3$; pharmacological study; age > 18 years or non-English abstract. Fifty-five papers remained for the full text eligibility assessment, of which a further 28 were excluded (see Figure 1). Twenty-five studies (described in 27 papers) were included in the qualitative synthesis, five of which underwent quantitative synthesis. Complete data of Selective Mutism Questionnaire (SMQ) scores were requested and received from Oerbeck et al. (2014) and Cornacchio et al. (2019). Characteristics of the included studies are provided in Table 1, organised by study design, of which six were RCTs (24%), five were single case experimental designs (20%), eight were pre-experimental one group pretest posttest design (32%) and six were pre-experimental single subject designs (24%).

Quality appraisal

Of the 25 studies, eight (32%) were considered to be of high quality (low risk of bias) while the rest (68%) were rated as low quality (high risk of bias) (see Table S2). The weak aspects of the studies (see Table S3) related to the method of subject selection, lack of randomisation, lack of blinding, inappropriate sample sizes, analytic methods insufficiently described or justified, estimate of variance reported for main results and lack of control for confounding variable. Only 12 studies (48%) calculated and reported effect sizes of the intervention varying considerably from Cohen's d ($k = 8$), Reliable Change Index ($k = 3$), Partial eta squared ($k = 2$), Hedges g ($k = 1$), Percentage of All Non-overlapping Data ($k = 1$), Kendall's W ($k = 1$), r ($k = 1$) to Tau- U ($k = 1$).

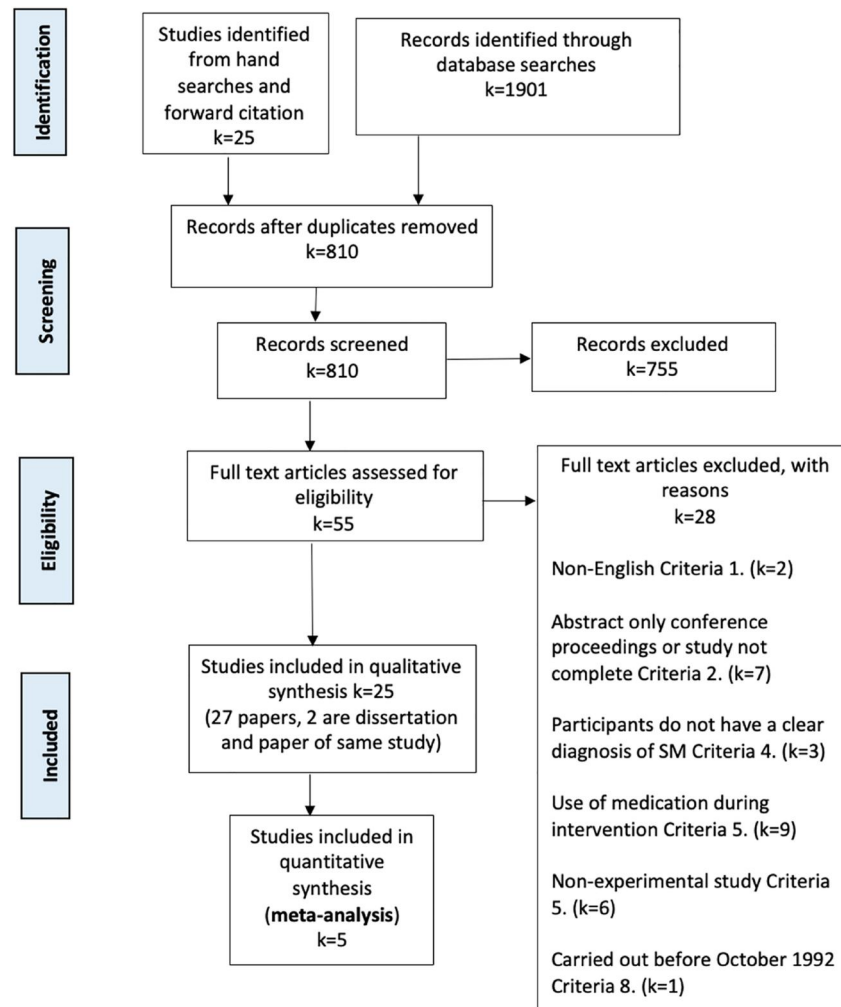


FIGURE 1 PRISMA flowchart.

Participant characteristics

All of the studies ($k = 25$) included children aged 9 years or under. Only nine studies (35%) also involved children 10 years or older, out of which only three (12%) included teenagers. Regarding SM characteristics, seven studies reported the age of onset (ranging between 2 and 5 years) and nine reported the duration (ranging between 1 and 5.5 years) of the condition prior to receiving treatment. Seven studies (28%) reported the frequency of bilingualism or multilingualism (representing between 25% and 57% of the samples); 19 studies (76%) reported on comorbidities, with 16 listing anxiety disorders, eight speech and language delays or impairments, and nine other developmental/neurodevelopmental problems (e.g. motor delay, enuresis, attention deficit hyperactivity disorder).

Types of interventions

Six studies (24%) considered behavioural therapy only. Eighteen studies (72%) used a combined behavioural and systems approach while one study (4%) incorporated a combined behavioural, systems and body-oriented therapy closely linked to play therapy (psychomotor therapy). As shown in Figure 2 (and Table S4) the interventions

varied in their use of components. The most common intervention components were exposure activities ($k = 24$) (e.g. graded exposure, stimulus fading, shaping, video self-modelling, prompting strategies during speaking situations, see Zakszeski & DuPaul, 2017 for behavioural strategy definitions) and reward systems ($k = 24$) (e.g. contingency management, reinforcement, contingent and specific verbal praise) used in 96% of interventions, followed by psychoeducation for parents and/or teachers ($k = 19$, 76%) (e.g. understanding the condition, treatment, managing anxiety and maintenance of SM), and rapport building or adapting interaction style strategies ($k = 17$, 68%) (e.g. defocused communication, nonchalant communication, child-directed interaction, verbal-directed interaction). Less common intervention components focussed on coping strategies ($k = 10$, 40%) (e.g. relaxation, emotional regulation, breathing exercises), transfer of control ($k = 10$, 40%), social skills ($k = 8$, 32%), cognitive ($k = 6$, 24%) (e.g. cognitive restructuring), problem solving ($k = 2$, 8%) and play therapy ($k = 1$, 4%) (See Table S5 for component definitions).

Characteristics of the interventions

Four studies (16%) delivered the intervention via a group format while the remaining studies (84%) took an individual approach.

TABLE 1 Study characteristics.

ID	Authors	Study design	N	Gender (male)	Age range	Average onset of SM (years)	Average duration of SM (years)	Bi or multi-lingual	Co-morbidity	Intervention				
										Type of Tx (broad)	Type of Tx (specific)	Who provided Tx	Location of Tx	Dosage of Tx (hours)
Experimental: Randomised control trial (RCT) versus active control														
1	Ooi et al. (2016)	RCT	21	8	6–12	NR	NR	NR	5 SP, 2 SeAD, 2 SAD	BT	Online CBT versus computer games	Clinician	Clinic	14 weekly sessions (14)
2	Esposito et al. (2017)	RCT	166	90	6–18	NR	NR	NR	NR	BSeoT	Psychomotor therapy versus Behavioural & educational counselling	Clinician, parent	Home	24 weeks, 45 min sessions × 3 weeks (54)
Experimental: RCT versus waiting list (WL)														
3	Bergman et al. (2013)	RCT	21	11	4–8	3.38	2.05	NR	18 SAD	BST	Integrated behaviour therapy for SM (IBTSM)	Clinician, parent, teacher	Clinic, public places, school	24 weeks, 20 sessions (20)
4	Oerbeck et al. (2014)	RCT	24	8	3–9	NR	NR	6	15 MD or LaD, 24 SAD, 7 SeAD, 6 SP, 2 GAD, 2 OCD, 2 Tics, 6 Enu, 1 Enc	BST	Home and school CBT	Clinician, parent, teacher	Home, school	12 weeks, 21 sessions (12)
5	Cornacchio et al. (2019)	RCT	29	7	5–9	NR	NR	NR	21 SAD, 8 SeAD, 7 GAD, 3 SP, 2 OCD, 2 Enu, 2 ADHD	BST	Intensive group behavioural treatment (IGBT) based on PCIT-SM	Clinician or Trained volunteer, parent	Simulated classroom, camp, public places	1 week, 5 days of 6–8 h (35)
Experimental: RCT comparing adapted treatments														
6	Bunnell et al. (2018)	RCT	15	5	5–17	NR	NR	NR	15 SAD, 2 SeAD, 1 GAD, 2 Enu	BT	Shaping with apps versus Shaping with other tools versus Shaping with reinforcement only	Clinician	Clinic	2 sessions of up to 55 min (2)
Experimental: Single case experimental design (SCED)														
7	Stone (2000)	SCED (matched pair)	6	2	4–9	2.66	3.16	NR	1 Enu, 1 LaD	BST	Videotape Training (VT) versus Video self modelling (VSM)	Parent, teacher guided by clinician	School	Both Tx: 12 weeks, 12–24 sessions, VT (27.5), VSM (20)
8	Vecchio (2008); Vecchio and Kearney (2009)	SCED (alternating treatment design)	9	2	4–9	Not clear (range 2–5)	Not clear (range 2–4)	NR	9 SAD, 2 SeAD, 2 SP, 1 ADHD, 1 Enu, 1 GAD, 1 ODD, 2 SpD	BT	Exposure practices and contingency management	Clinician, parent	Clinic, home, public places, school	8–20 weeks, 2 sessions a week (NR)

(Continues)

TABLE 1 (Continued)

ID	Authors	Study design	Participants				Intervention								
			N	Gender (male)	Age range	Average onset of SM (years)	Average duration of SM (years)	Bi or multilingual	Co-morbidity	Type of Tx (broad)	Type of Tx (specific)	Who provided Tx	Location of Tx	Dosage of Tx (hours)	
9	Mitchell and Kratochwill (2013)	SCED (multiple baseline)	4	2	5–10	NR	3	NR	NR	BST	Conjoint behavioural consultation	Parent, teacher guided by clinician	Clinic, school	10 weeks, 3–7 sessions, (NR)	
10	Solz (2015)	SCED (multiple baseline)	3	1	5–9	NR	NR	NR	1 SAD, 1 ASD, 1 SpD	BST	Video self-modeling with and without contingency management	Clinician, parent, teacher	Clinic, school, home	3–4 weeks, 12 sessions (2)	
11	Siroky (2019)	SCED (multiple baseline)	5	3	4–8	NR	NR	0	3 SAD, 2 GAD, 1 SeAD	BST	Condensed IBTSM	Clinician, parent	Clinic	16–22 weeks, 16 sessions, (13)	
Pre-experimental: One-group pretest posttest design															
12	Woodcock et al. (2007)	One-group pre-post test design	15	6	4–8	2.5	3.3	7	7 SAD, 3 SeAD, 2 GAD	BST	School based CBT	Parent, teacher guided by clinician	School, public places	12 months, 12 training sessions, parents & teachers 20 min × 3 weeks for 6 months (24)	
13	Sharkey et al. (2008)	One-group pre-post test design	5	1	5–8	3.6	2.4	NR	1 SeAD	BST	Parent group (systems) and child group therapy (CBT)	Clinician, parent	Clinic, with homework in public places and school	8 weeks, 8 sessions (12)	
14	Oerbeck et al. (2011)	One-group pre-post test design	7	2	3–5	NR	1.6	4	1 LaD	BST	Home and school CBT	Clinician, parent, teacher	Home, school	14 weeks, (12)	
15	Oerbeck et al. (2015)	One-group pre-post test design	24	8	3–9	NR	NR	6	15 MD or LaD, 24 SAD, 7 SeAD, 6 SP, 2 GAD, 2 OCD, 2 Tics, 6 Enu, 1 Enc	BST	Home and school CBT	Clinician, parent, teacher	Home, school	12 weeks, 21 sessions (12)	
16	Klein et al. (2017)	One-group pre-post test design	40	NR	5–12	2.81	3.96	NR	27 LaD	BST	Social communication anxiety treatment (S-CAT)	Clinician, parent, teacher	Clinic, school, public places	9 weeks, 3 sessions every 3 weeks, (NR)	
17	Oerbeck et al. (2018)	One-group pre-post test design	32	10	3–9	NR	NR	9	16 MD or LaD, 24 SAD, 7 SeAD, 6 SP, 2 GAD, 2 OCD, 2 Tics, 6 Enu, 1 Enc	BST	Home and school CBT	Clinician, parent, teacher	Home, school	12 weeks, 21 sessions (12)	
18	Aldrich et al. (2021)	One-group pre-post test design	112	40, 1 trans male	3–14	NR	NR	35	NR	BST	Multi-disciplinary SM treatment group based on PCIT-SM	Clinician	Clinic, locations in hospital	8 weeks, 8 sessions (12)	

TABLE 1 (Continued)

ID	Authors	Study design	Participants			Intervention									
			N	Gender (male)	Age range	Average onset of SM (years)	Average duration of SM (years)	Bi or multi-lingual	Co-morbidity	Type of Tx (broad)	Type of Tx (specific)	Who provided Tx	Location of Tx	Dosage of Tx (hours)	
19	Tan et al. (2021)	One-group pre-post test design	24	11	6–12	NR	NR	NR	7 SAD, 7SP, 3 SeAD, 1 GAD	BT	Virtual reality exposure therapy (VRET)	Clinician	Clinic, virtual classroom	10 weeks, 6 sessions (6)	
Pre-experimental: Single subject design															
20	Ortega (2011)	ABC	4	2	6–7	NR	NR	NR	NR	BST	Totally anxiety-free communication treatment (CBT)	Clinician, parent, teacher	Clinic, school	18–22 sessions, 2–3 sessions a week (NR)	
21	Paasivirta (2012)	ABA	4	1	5–7	NR	NR	2	3 AD-NOS, 1SP, 1 Enu, 1 Enc	BST	Brief teacher training programme	Teacher guided by clinician	School	3–4 weeks, 1 h training, 2–3 weekly sessions for 20 min, (2)	
22	Roslin (2013)	AB	5	2	4–7	NR	NR	NR	2 SAD, 1 SP	BST	Parent child interaction therapy (PCIT-SM)	Clinician, parent	Clinic, public places	12 weekly sessions, (12)	
23	Bunnell et al. (2016)	AB	4	1	7–10	4.25	4	NR	NR	BT	Shaping with apps	Clinician	Clinic	2 sessions of up to 55 min (2)	
24	Bork (2016); Bork and Bennett (2020)	AB	3	0	8	3.2	2.8	NR	NR	BT	Video self-modeling and stimulus fading	Parent, teacher guided by clinician	School	8 weeks, 2 sessions a week, (1–2)	
25	Haggerty (2020)	AB	25	5	4–11	NR	NR	0	20 SAD, 5 GAD	BST	Intensive summer day camp for SM	Clinician, parent	Simulated classroom, camp, public places	5 days, (40)	

Note: AB refers to A (baseline) phase and the B (intervention) phase; while ABA refers to A (baseline) phase, B (intervention) phase, and A (withdrawal/follow up) phase; ABC refers to A (baseline) phase, B (intervention) phase, B (intervention) phase, and C (other intervention) phase. Abbreviations: ADHD, attention deficit hyperactivity disorder; AD-NOS, anxiety disorder not otherwise specified; ASD, autism; BSBoT, behavioural systems body-oriented therapy approach; BST, behavioural systems therapy approach; BT, behavioural therapy approach; Enc, enuresis; Enu, enuresis; GAD, generalised anxiety disorder; LaD, language difficulties; MD, motor delay; NR, not reported; OCD, obsessive compulsive disorder; ODD, oppositional defiant disorder; SeAD, separation anxiety disorder; SAD, social anxiety disorder; SP, specific phobia; SpD, speech difficulties; Tics, tic disorders; Trans, transgender.

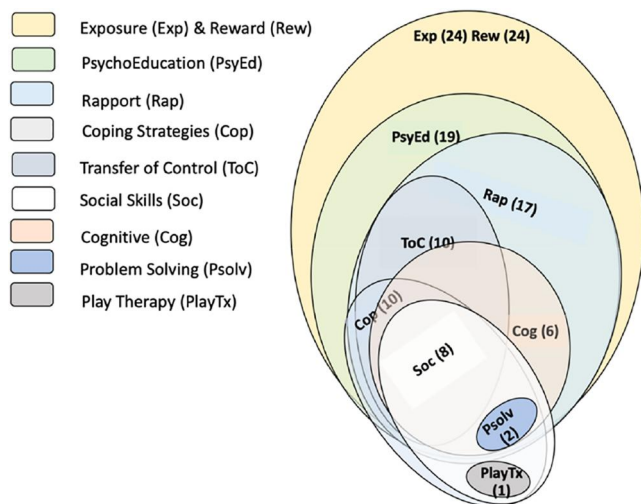


FIGURE 2 Treatment components in nonpharmacological selective mutism interventions.

Twenty-three studies (92%) had a protocol or manual for the intervention while two (8%) did not report any. Regarding who delivered the intervention with the child, the most common combination involved the clinician, parent, and teacher ($k = 12$, 48%). In four of these, the clinician guided the parent and teacher without directly working with the child. Seven studies (28%) involved both the clinician and parent, five studies (20%) used the clinician only, and one study involved the teacher only being guided by the clinician. The interventions were delivered in various combinations of locations including the school/simulated classroom, in a clinical setting, in the community, and/or the child's home. Four (16%) interventions were brief and/or intensive treatments lasting less than a week ranging between 2 and 40 h of therapist contact, eight (32%) were between 2 and 9 weeks with 2–12 h of therapist contact, ten (40%) lasted between 10 and 19 weeks varying between 12 and 27.5 h of therapist contact while three (12%) continued for 20 or more weeks requiring 20–54 h of the clinician's time.

Outcomes assessed

A wide range of outcomes were used to assess speaking behaviours across the studies (See Table S6). While the SM Questionnaire (76%) and/or the School Speech Questionnaire (48%) were consistently used, 20 other outcome measures were employed by studies which varied in measurement mode (e.g. rating scales, counts or time), respondents (e.g. independent observers, clinician, teachers, parents or child), and type of communication (e.g. words spoken, whispered, mouthed, non-verbal).

The ADIS for Children (ADIS-C/P) was consistently used to measure SM remission for all six studies with this data, however outcomes were reported in different ways. For example, in two of the studies, the authors used three categories: ongoing SM, partial remission and full remission, that is, where children no longer fulfilled the diagnostic criteria *and* were speaking freely in school (Oerbeck et al., 2015, 2018). Five studies assessed comorbid remission of social anxiety disorder (SAD); three used the ADIS-C/P and two used the

Schedule for affective disorders and schizophrenia for school aged children (K-SADS-PL, Kaufman et al., 1997).

Eighteen studies (72%) measured anxiety symptoms with 18 different measures. Ten studies (40%) measured condition severity and/or improvement however, again these varied in measures used and timings of assessments (see Table S6). Only a few studies measured broader impacts of SM on the child's life: well-being (social, emotional and behavioural problems) ($k = 6$; 24%); functional impairment or interference ($k = 4$; 16%); communication progress ($k = 2$; 8%); academic impairment/competence ($k = 2$; 8%); and quality of life ($k = 1$, 4%). Three studies (12%) measured changes in adult interactions with the child.

Outcomes

Within case effect sizes and percentage of remission

A detailed description of findings is provided in Appendix S2 and Table S3 for within case effect sizes for the study outcomes and percentage remission rates. Overall a mixed pattern of outcomes was found across the outcome variables. The few studies that reported remission rates (SM, $k = 6$; SAD, $k = 5$) displayed positive results. Most studies reported improvements in speaking behaviours ($k = 21$) and wider outcomes when reported such as well-being ($k = 3$), functional impairment or interference ($k = 4$), and academic impairment ($k = 2$) with varying within case effect sizes from slight to large. Notably, for anxiety outcomes, some studies reported positive ($k = 12$), mixed ($k = 4$) and negative ($k = 1$) changes following intervention ranging between large effect sizes decreasing and small effect sizes increasing anxiety.

Between subject effect size

Improvements in speaking behaviours in single case experimental designs

Mixed results were also found based on speaking behaviour and other communication outcomes in Single Case Experimental Designs. Mitchell and Kratochwill (2013) showed an increase in frequency of words spoken per minute with a large between subject effect size following a combined behavioural and systems approach (Hedges' $g = 2.08$, 95% CI: 0.78–3.38). In contrast, Solz (2015) found a negligible between subject effect size in verbalisations (Hedges' $g = 0.05$, 95% CI: –0.36 to 0.46), yet a large effect size in increased meaningful communication (Hedges' $g = 1.46$, 95% CI: 0.90–2.02) following a combined behavioural and systems approach. Siroky (2019) also found a small between subject effect size of observed speaking behaviours (Hedges' $g = 0.36$, 95% CI: –0.17 to 0.89) following a combined behavioural and systems approach.

Meta-analysis

Five RCTs were suitable for meta-analysis of change in speaking behaviour (all used the SMQ) of children after an intervention/

control period displayed in two forest plots that differed in controls (see Figures 3 and 4). Two of these RCTs had appropriate data for a meta-analysis of SM remission (ADIS) shown in a third forest plot (see Figure 5).

The first forest plot (see Figure 3) consisted of three non-pharmacological intervention studies involving 38 treatment participants compared to 36 waiting list (WL) control participants. Non-pharmacological interventions had a large positive effect on speaking behaviour compared to WL controls (Hedges $g = 1.06$, $p < .0001$, 95% CI: 0.57–1.56). The results indicated that there was not significant variation between effect sizes ($Q = 0.93$, $df = 2$, $p = .63$) with minimal statistical heterogeneity ($I^2 = 0.00\%$, $\tau^2 = 0$ with $SE = 0.19$) (Borenstein et al., 2009; Deeks et al., 2008). Due to the minimal heterogeneity and low number of studies a moderator analysis was not carried out. It is important to note that the effect size from Bergman et al. (2013) was calculated using the SMQ scores at week 12 in order to directly compare with the WL period, however it was only half-way through the treatment so the effect size may have been greater by week 24 (12 weeks SMQ mean = 1.32, $SD = 0.49$; 24 weeks SMQ mean = 1.74, $SD = 0.54$).

The second forest plot (see Figure 4) comprised the two non-pharmacological intervention studies with a total of 88 treatment

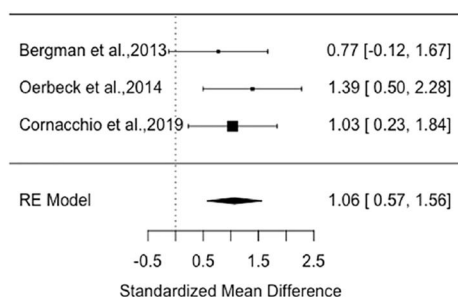


FIGURE 3 Speaking behaviour: hedges g forest plot for treatment versus waiting list control.

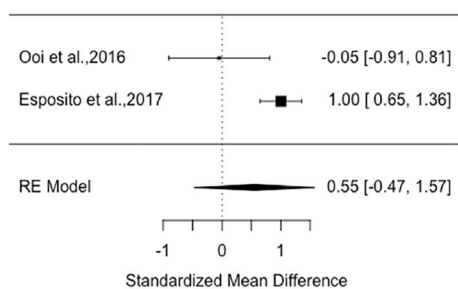


FIGURE 4 Speaking behaviour: hedges g forest plot for treatment versus active controls.

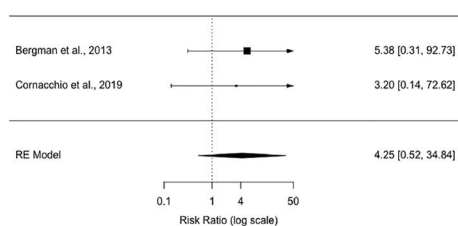


FIGURE 5 Selective mutism remission: risk ratio (log scale) forest plot for treatment versus waiting list control.

participants compared to 99 active control participants (computer games, behavioural and educational counselling to parents). There was a moderate sized improvement in speaking behaviour in the non-pharmacological interventions compared to active controls, however this did not reach statistical significance (Hedges $g = 0.55$, $p < .29$, 95% CI: -0.47 to 1.57). The results indicated that there was significant variation between effect sizes ($Q = 4.93$, $df = 1$, $p = .03$) with considerable statistical heterogeneity ($I^2 = 79.72\%$, $\tau^2 = 0.44$ with $SE = 0.78$) (Borenstein et al., 2009; Deeks et al., 2008), which reflects the disparity between Esposito et al.'s (2017) large treatment effect (Hedges $g = 1.00$, $p < .001$, 95% CI: 0.64–1.36) and Ooi et al.'s (2016) negligible treatment effect (Hedges $g = -0.05$, $p < .994$, 95% CI: -0.91 to 0.81). The clinical heterogeneity may also explain the different results between these two studies with contrasting intervention approach, active control, dosage, location, and provider of the intervention (see Table 1). Due to the small number of studies, differences in control conditions, and the varying effects, it is difficult to know what the true dispersion (τ^2) looks like (Borenstein et al., 2009). Therefore, these results must be treated with caution.

The third forest plot (see Figure 5) incorporate two non-pharmacological intervention studies (Bergman et al., 2013; Cornacchio et al., 2019) with a total of 26 treatment participants compared to 24 control participants. There was a large effect for SM remission favouring the intervention; however, this did not reach statistical significance due to the small number of studies and participants (Risk Ratio = 4.25, $p = .18$, 95% CI: 0.52–34.84). The results indicate that there was no significant variation between effect sizes ($Q = 0.06$, $df = 1$, $p = .81$) with minimal statistical heterogeneity ($I^2 = 0.00\%$, $\tau^2 = 0$ with $SE = 3.29$). However, due to the small number of studies these results must be treated with caution.

DISCUSSION

This systematic review identified 25 studies that evaluated non-pharmacological interventions for children and adolescents with SM. Although there has been progress since Stone et al. (2002) systematic review, with notably more experimental studies (44%) and manualised treatments (91%), only six RCTs have systematically tested nonpharmacological interventions for SM which together included only 276 children. Five of the six RCTs were suitable for meta-analysis, and the outcomes suggested that combined systems/behavioural approaches are promising treatments for helping children (3–9 years) with SM speak to more people in the home, school and community. A meta-analysis of two of the RCTs also suggested that combined systems/behavioural interventions have promising outcomes in terms of SM remission. However, overall, the few systematic evaluations limit conclusions about the effectiveness of nonpharmacological interventions.

Only six studies measured actual remission of SM, and these had different assessment timings and classification of remission. This is likely to be an important outcome given it takes into account both symptoms and impairment, which have been highlighted as meaningful to families in the wider child anxiety literature (Creswell et al., 2021). Furthermore, of these six studies, only two were RCTs. Although, the results indicated a large effect for SM remission,

favouring the intervention ($RR = 4.5$), the small number of studies and participants meant the results lacked precision and were not statistically significant leading to cautious conclusions. An agreed standard approach to reporting SM remission outcomes would help with consistency and enable meaningful comparisons in future analyses.

The lack of consistency of measures used also means that, at best, only tentative conclusions about outcomes for children and adolescents with SM can be drawn. The most consistent measure for speaking behaviour was the SM Questionnaire used in all of the RCTs included in the meta-analyses. However, the single case design studies measured speaking behaviours in broader ways (including words spoken, whispered, mouthed, non-verbal communication), with varying lengths of time (per session, per day, per minute) from different assessors (the clinician, parent, teacher, or child). Consequently, there is limited meaningful comparison for speaking behaviour outcomes between the identified Single Case Experimental Designs and RCTs.

While the SMQ has played a significant part in international SM research (Rodrigues Pereira et al., 2021), including the RCTs in this systematic review and meta-analysis, there are still some gaps in its utility which need to be addressed before it is recommended as the speaking behaviour outcome for SM intervention research with children and adolescents. There are many positives in using this measure. It is standardised and psychometrically sound (Cronbach's $\alpha = .97$); widely used; adaptable, having been translated in Norwegian (Oerbeck et al., 2014), Italian (Esposito et al., 2017), Spanish (Olivares-Olivares et al., 2021), Dutch (Rodrigues Pereira et al., 2022) and Japanese (Kakuta et al., 2022) and also into a teacher (School Speech Questionnaire, Bergman et al., 2002) and child reported measure (SMQ-C, Milic et al., 2020). However, the SMQ's psychometric properties have only been validated for 3- to 11-year-olds limiting its use for the adolescent population. As suggested by Bergman et al. (2008) themselves, there is a need for further research to establish whether the SMQ is suitable for adolescents or needs to be adapted. Rodrigues Pereira et al. (2022) has started that process by including adolescents in their sample (3–17 years) when validating the Dutch translation. An alternative outcome measure has been developed that covers the age ranges of 3–18 years, the Frankfurt Scale of Selective Mutism (FSSM, Gensthaler et al., 2020) with promising psychometric properties in both a diagnostic scale (Cronbach's $\alpha = .90$) and severity scale (Cronbach's $\alpha = .98$). It is a parent reported questionnaire, similar to the SMQ, which could be useful for future SM intervention research due to its age range and dual function (diagnostic and symptom severity outcome). This may also facilitate improved reporting of remission rates.

From the 18 studies that measured anxiety outcomes, the within subject effects were highly variable, from small negative to large positive effects with no clear pattern based on the type of treatment or who rated anxiety (child, parent, or teacher). Only one study used objective physiological anxiety measures (Bunnell et al., 2018). The multifarious ways in which anxiety was measured, and the different timings of measurements, may have contributed to the heterogeneous results. Again, there is a need for consensus on which anxiety outcomes to include in SM intervention research to promote consistency for future analyses. A possible starting point for this conversation could consider the wider guidelines for assessment in treatment trials

for anxiety disorders more generally (Creswell et al., 2021). In these guidelines, the authors made recommendations on the reporting of diagnostic outcomes, symptom-based and functional interference measures along with sample and treatment characteristics. More specific to anxiety outcomes, the authors recommended researchers consistently include: a multidimensional measure of anxiety symptoms, a psychometrically reliable and valid measure of the targeted symptoms, and an interference/impact measure.

Few studies measured the broader impact of the intervention on the child's life. This is concerning given the social, educational and vocational impairment associated with SM. The importance of capturing these impacts alongside remission of SM is highlighted by Remschmidt et al.'s (2001) long-term follow up study in which young people who no longer had SM 12 years post-intervention continued to struggle with independence, academic achievement, confidence and health when compared to a non-SM reference group. Although, notably, Oerbeck et al.'s (2018) longitudinal study suggested that children and adolescents with SM had comparable quality of life to other Norwegian children 5 years after the combined systems and behavioural intervention. Two decades after Stone et al.'s (2002) initial systematic review there is still scarce evidence of changes in the broader impact of SM after intervention.

Besides the overall lack of available research, limitations in the quality of the evidence, variable intervention approaches, sampling gaps, insufficient reporting, and absence of health economic analyses in the current intervention literature all prevent definitive answers to questions about what works for children with SM. Indeed the majority of studies were rated as low quality indicating a high risk of bias. The main areas of weakness were the lack of randomisation, lack of blinding, small sample sizes and lack of control for confounding results either via study design (e.g. randomised controlled trial) or through statistical methods (e.g. partial correlation). These areas of weakness reflect the few sufficiently powered RCTs and lack of collaborative work across regions to enable larger samples sizes in SM intervention research. By pooling resources, there will also be greater opportunities to ensure that treatment outcome studies are generalisable to children with SM with more diverse backgrounds and characteristics. Of particular note, only three studies included adolescents 13 years and above which may be impacted by the lack of appropriate speaking behaviour outcome measures for this age range. Consequently, the results in this review are most applicable to children (3–9 years) with SM. Furthermore few studies included or reported on bilingual children, comorbid communication disorders (e.g. speech disorders, stammers, language disorders) and neurodevelopmental conditions (e.g. autism). This is concerning since these are all common among children with SM (Kristensen, 2000; Muris & Ollendick, 2021; Steffenberg et al., 2018; Toppelberg et al., 2005). Few studies reported variables such as participants' age of onset and duration of SM (Stone et al., 2002) and severity and familial SM (Remschmidt et al., 2001) which are associated with persistent long-term SM. It will be critical that future studies consistently report these variables as potential moderators to understand the impact they may have on treatment outcomes. Going forward it will also be important to evaluate interventions in the settings in which they are routinely applied. Only just over half of the studies administered the intervention in routine clinical settings or in schools. Although there is one notable example of a research team that documented progress

of their SM intervention from development (Oerbeck et al., 2011) to feasibility/pre-evaluation (Oerbeck et al., 2014) to implementation (Oerbeck et al., 2015, 2018), finding large treatment effects on children's speaking behaviour when delivered by local therapists in community health clinics (Oerbeck et al., 2014).

Currently in SM intervention research, it can be difficult to blindly assess the treatment effects since most of the outcomes are based on parent or teacher report who are both often involved in the intervention. A potential solution to this problem is the use of an objective outcome measure for speaking behaviour. One example, that has come from recent advances in wearable technology, is a passive audio vocal measurement that appears to have the capabilities to quantify child vocalisations, vocal volume, and other conversational measures (Xu et al., 2018). While wearable technology is still developing and refining, it has potential to significantly improve the quality of SM research in the future.

In the future, it will be important for the field to explore wider questions about what works for whom, how, and overall economic viability. A current barrier to this includes the wide variation in intervention approaches, with limited descriptions of the components delivered within these interventions. As a first step we have proposed a set of definitions that can guide consistent reporting of treatment components going forward (see Table S5). The absence of health economic analyses in the literature limits our understanding of which SM interventions are worth investing in (Coast et al., 2018). Indeed, the wide variability in the amount of clinician time across the RCT interventions (from 12 to 54 h of clinician time) is likely to substantially influence their cost effectiveness, as are changes in wider health and societal costs, such as disruption to education, employment, use of health and social care services.

Strengths of this systematic review include the broad approach. The inclusion of grey literature minimises selection bias and provides a comprehensive picture of the research landscape. Robust procedures were followed, including the use of double raters for screening, full text eligibility and quality appraisal. Where possible we also conducted separate meta-analyses for distinct outcomes (i.e. speaking behaviour improvement and SM remission). However, several limitations should also be noted. Firstly, the review only included studies written in English which risks the exclusion of relevant studies. Secondly, due to few trials, the effect sizes of all the studies' outcomes were converted into the within case metric for comparison which poses a risk of over inflation or an artefact of regression to the mean. Thirdly, since the review focussed on non-pharmacological interventions, studies with participants taking medication and where they did not explicitly state they had been on a stable dose for at least 1 month prior to and during the intervention were excluded. Consequently, nine potentially relevant studies were not included.

In conclusion, the outcomes for SM remission and improving speaking behaviour for children (3–9 years) with SM appear promising when using a combined systems/behavioural intervention. However, limited fully powered systematic evaluations and the inconsistent use of outcome measures only allow for tentative conclusions to be drawn. There is an urgent need for more fully powered experimental studies which use consistent approaches to outcome measurement and include more representative and diverse samples, with consideration of moderators of treatment outcomes and of treatment cost-

effectiveness. To achieve this, regional and international research collaborations and measurement harmonisation (National Quality Forum, 2010) are crucial. While an international consensus on standard sets of outcome measures has already been achieved in other areas of child mental health (e.g. Krause et al., 2021), a similar approach is needed for SM. This will ensure the most can be made of future research through data sharing between regional and international groups to enable far greater clarity about how to achieve optimal outcomes for children and adolescents with SM.

AUTHOR CONTRIBUTIONS

Gino Hipolito: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing—original draft; Writing—review & editing. **Emma Pagnamenta:** Conceptualization; Formal analysis; Methodology; Resources; Supervision; Validation; Writing—review & editing. **Helen Stacey:** Formal analysis; Validation; Writing—review & editing. **Emily Wright:** Formal analysis; Validation; Writing—review & editing. **Victoria Joffe:** Conceptualization; Methodology; Supervision; Writing—review & editing. **Kou Murayama:** Methodology; Supervision; Validation; Writing—review & editing. **Cathy Creswell:** Conceptualization; Funding acquisition; Methodology; Supervision; Validation; Visualization; Writing—review & editing.

ACKNOWLEDGMENTS

GH is funded by an NIHR Predoctoral Clinical Academic Fellowship (NIHR-ICA-PCAF-2018-01-020). CC was funded by NIHR Research Professorship (RP 2014-04-018). CC is supported by the Oxford and Thames Valley NIHR Applied Research Collaboration. KM was partly funded by the Alexander von Humboldt Foundation (the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research). The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

CONFLICT OF INTEREST STATEMENT

The authors have declared that they have no competing or potential conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author.

ETHICAL CONSIDERATIONS

No ethical approval or patient consent was required for this research review.

PERMISSION TO PRODUCE MATERIAL FROM OTHER SOURCES

Not presenting material from other sources apart from data extracted from journals, dissertations and requested from Danielle Cornacchio and Beate Oerbeck regarding their research.

ORCID

Gino Hipolito  <https://orcid.org/0000-0003-1450-8457>

Emma Pagnamenta  <https://orcid.org/0000-0002-4703-3163>

Helen Stacey  <https://orcid.org/0000-0002-6634-4725>
 Emily Wright  <https://orcid.org/0000-0002-6619-8473>
 Victoria Joffe  <https://orcid.org/0000-0001-9132-2889>
 Kou Murayama  <https://orcid.org/0000-0003-2902-9600>
 Cathy Creswell  <https://orcid.org/0000-0003-1889-0956>

REFERENCES

- Aldrich, J. T., Blossom, J. B., Moss, A., Ray, B., Couckuyt, M., Ward, T., Fox, A. R., & Read, K. L. (2021). Effectiveness of an eight-week multidisciplinary selective mutism treatment group. *Evidence-Based Practice in Child and Adolescent Mental Health*, 8(1), 105–119. <https://doi.org/10.1080/23794925.2021.2007818>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
- Anstendig, K. (1998). Selective mutism: A review of the treatment literature by modality from 1980-1996. *Psychotherapy: Theory, Research, Practice, Training*, 35(3), 381–391. <https://doi.org/10.1037/h0087851>
- Bergman, R. L., Gonzalez, A., Piacentini, J., & Keller, M. L. (2013). Integrated behavior therapy for selective mutism: A randomized controlled pilot study. *Behaviour Research and Therapy*, 51(10), 680–689. <https://doi.org/10.1016/j.brat.2013.07.003>
- Bergman, R. L., Keller, M. L., Piacentini, J., & Bergman, A. J. (2008). The development and psychometric properties of the selective mutism questionnaire. *Journal of Clinical Child and Adolescent Psychology*, 37(2), 456–464. <https://doi.org/10.1080/15374410801955805>
- Bergman, R. L., Piacentini, J., & McCracken, J. T. (2002). Prevalence and description of selective mutism in a school-based sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(8), 938–946. <https://doi.org/10.1097/00004583-200208000-00012>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons Ltd.
- Bork, P. (2016). Augmented video self-modeling as an intervention technique for young children with selective mutism: An explanatory sequential study. [Doctoral Dissertation, Brock University]. Brock University Repository. <https://dr.library.brocku.ca/handle/10464/10427>
- Bork, P., & Bennett, S. (2020). Video self-modeling fading, and reinforcement: An effective treatment package for children with selective mutism. *Clinical Child Psychology and Psychiatry*, 25(2), 446–455. <https://doi.org/10.1177/1359104519855110>
- Bunnell, B. E., Mesa, F., & Beidel, D. C. (2018). A two-session hierarchy for shaping successive approximations of speech in selective mutism: Pilot study of mobile apps and mechanisms of behaviour change. *Behavior Therapy*, 49(6), 966–980. <https://doi.org/10.1016/j.beth.2018.02.003>
- Bunnell, B. E., Procci, K., Beidel, D. C., & Bowers, C. A. (2016). Gamification of therapy: Treating selective mutism. In D. Novak, B. Tulu, & H. Brendryen (Eds.), *Handbook of research on holistic perspectives in gamification for clinical practice* (pp. 390–410). IGI Global. <https://doi.org/10.4018/978-1-4666-9522-1.ch018>
- Coast, J., Bailey, C., & Kinghorn, P. (2018). Patient centered outcome measurement in health economics: Beyond EQ-5D and the quality-adjusted life -year-where are we now? *Annals of Palliative Medicine*, 7(Suppl 3), S249–S252. <https://doi.org/10.21037/apm.2018.03.18>
- Cohan, S. L., Chavira, D. A., & Stein, M. B. (2006). Practitioner review: Psychosocial interventions for children with selective mutism: A critical evaluation of the literature from 1990-2005. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 47(11), 1085–1097. <https://doi.org/10.1111/j.1469-7610.2006.01662.x>
- Cornacchio, D., Furr, J. M., Sanchez, A. L., Hong, N., Feinberg, L. K., Tenenbaum, R., Del Busto, C., Bry, L. J., Poznanski, B., Miguel, E., Ollendick, T. H., Kurtz, S. M. S., & Comer, J. S. (2019). Intensive group behavioural treatment (IGBT) for children with selective mutism: A preliminary randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 87(8), 720–733. <https://doi.org/10.1037/ccp0000422>
- Coster, W. J. (2013). Making the best match: Selecting outcome measures for clinical trials and outcome studies. *American Journal of Occupational Therapy*, 67(2), 162–169. <https://doi.org/10.5014/ajot.2013.006015>
- Creswell, C., Nauta, M. H., Hudson, J. L., March, S., Reardon, T., Arendt, K., Boddien, D., Cobham, V. E., Donovan, C., Halldorsson, B., In-Albon, T., Ishikawa, S., Bach Johnsen, D., Jolstedt, M., de Jong, R., Kreuze, L., Mobach, L., Rapee, R., Spence, S. H., ..., & Kendall, P. C. (2021). Research Review: Recommendations for reporting on treatment trials for child and adolescent anxiety disorders-an international consensus statement. *Journal of Child Psychology and Psychiatry*, 62(3), 255–269. <https://doi.org/10.1111/jcpp.13283>
- Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2008). Analysing data and undertaking meta-analyses. In J. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 243–296). John Wiley & Sons Ltd.
- Elizur, Y., & Perednik, R. (2003). Prevalence and description of selective mutism in immigrant and native families: A controlled study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(12), 1451–1459. <https://doi.org/10.1097/00004583-200312000-00012>
- Esposito, M., Gimigliano, F., Barillari, M. R., Precenzano, F., Ruberto, M., Sepe, J., Barillari, U., Gimigliano, R., Militerni, R., Messina, G., & Carotenuto, M. (2017). Pediatric selective mutism therapy: A randomized controlled trial. *European Journal of Physical and Rehabilitation Medicine*, 53(5), 643–650. <https://doi.org/10.23736/s1973-9087.16.04037-5>
- Gensthaler, A., Dieter, J., Raisig, S., Hartmann, B., Ligges, M., Kaess, M., Freitag, C. M., & Schwenck, C. (2020). Evaluation of a novel parent-rated scale for selective mutism. *Assessment*, 27(5), 1007–1015. <https://doi.org/10.1177/1073191118787328>
- Haggerty, D. (2020). A pilot feasibility study of an intensive summer day camp intervention for children with selective mutism. (Publication no. 27958732) [Doctoral Dissertation, Michigan State University]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/pilot-feasibility-study-intensive-summer-day-camp/docview/2407626880/se-2?accountid=13460>
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, 4(4), 324–341. <https://doi.org/10.1002/jrsm.1086>
- Kakuta, K., Takagi, J., Usui, N., Tomioka, N., Kaji, M., Kanehara, Y., & Hirose, S. (2022). Reliability and validity of the Japanese version of the selective mutism questionnaire. *Fuansho Kenkyu*, 14(1), 47–55. https://doi.org/10.14389/jasad.14.1_47
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of American Academy of Child and Adolescent Psychiatry*, 36(7), 980–998. <https://doi.org/10.1097/00004583-199707000-00021>
- Klein, E. R., Armstrong, S. L., Skira, K., & Gordon, J. (2017). Social communication anxiety treatment (S-CAT) for children and families with selective mutism: A pilot study. *Clinical Child Psychology and Psychiatry*, 22(1), 90–108. <https://doi.org/10.1177/1359104516633497>
- Krause, K. R., Chung, S., Adewuya, A. O., Albano, A. M., Rabins-Wagner, R., Birkinshaw, L., Brann, P., Creswell, C., Delaney, K., Falissard, B., Forrest, C. B., Hudson, J. L., Ishikawa, S., Khatwani, M., Kieling, C., Krause, J., Malik, K., Martinez, V., Mughal, F., ..., & Wolpert, M. (2021). International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder. *Lancet*, 8(1), 76–86. [https://doi.org/10.1016/S2215-0366\(20\)30356-4](https://doi.org/10.1016/S2215-0366(20)30356-4)
- Kristensen, H. (2000). Selective mutism and comorbidity with developmental disorder/delay, anxiety disorder, and elimination disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39(2), 249–256. <https://doi.org/10.1097/00004583-200002000-00026>
- Kristensen, H. (2001). Multiple informants' report of emotional and behavioural problems in a nation-wide sample of selective mute children and controls. *European Child & Adolescent Psychiatry*, 10(2), 135–142. <https://doi.org/10.1007/s007870170037>

- Kumpulainen, K., Räsänen, E., Raaska, H., & Somppi, V. (1998). Selective mutism among second-graders in elementary school. *European Child and Adolescent Psychiatry*, 7(1), 24–29. <https://doi.org/10.1007/s007870050041>
- Ludecke, D. (2019). Esc: Effect size computation for meta-analysis. R package version 0.5.1. <https://CRAN.R-project.org/package=esc>
- Manassis, K., Oerbeck, B., & Overgaard, K. R. (2016). The use of medication in selective mutism: A systematic review. *European Child and Adolescent Psychiatry*, 25(6), 571–578. <https://doi.org/10.1007/s00787-015-0794-1>
- Milic, M. I., Carl, T., & Rapee, R. M. (2020). Similarities and differences between young children with selective mutism and social anxiety disorder. *Behaviour Research and Therapy*, 133, 103696. <https://doi.org/10.1016/j.brat.2020.103696>
- Mitchell, A. D., & Kratochwill, T. R. (2013). Treatment of selective mutism: Applications in the clinic and school through conjoint consultation. *Journal of Educational and Psychological Consultation*, 23(1), 36–62. <https://doi.org/10.1080/10474412.2013.757151>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & PRISMA Group. (2009). Preferred reporting Items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Muris, P., & Ollendick, T. H. (2021). Current challenges in the diagnosis and management of selective mutism in children. *Psychology Research and Behavior Management*, 14, 159–167. <https://doi.org/10.2147/prbm.s274538>
- National Quality Forum. (2010). *Guidance for measure harmonization: A consensus report*. National Quality Forum.
- Oerbeck, B., Johansen, J., Lundahl, K., & Kristensen, H. (2011). Selective mutism: A home-and kindergarten-based intervention for children 3-5 years: A pilot study. *Clinical Child Psychology and Psychiatry*, 17(3), 370–383. <https://doi.org/10.1177/1359104511415174>
- Oerbeck, B., Overgaard, K. R., Stein, M. B., Pripp, A. H., & Kristensen, H. (2018). Treatment of selective mutism: A 5-year follow-up study. *European Child and Adolescent Psychiatry*, 27(8), 997–1009. <https://doi.org/10.1007/s00787-018-1110-7>
- Oerbeck, B., Stein, M. B., Pripp, A. H., & Kristensen, H. (2015). Selective mutism: Follow-up study 1 year after end of treatment. *European Child and Adolescent Psychiatry*, 24(7), 757–766. <https://doi.org/10.1007/s00787-014-0620-1>
- Oerbeck, B., Stein, M. B., Wentzel-Larsen, T., Langsrud, Ø., & Kristensen, H. (2014). A randomized controlled trial of a home and school-based intervention for selective mutism - defocused communication and behavioural techniques. *Child and Adolescent Mental Health*, 19(3), 192–198. <https://doi.org/10.1111/camh.12045>
- Olivares-Olivares, P. J., Rosa-Alcázar, Á., Nuñez, R., & Olivares, J. (2021). Psychometric properties of the selective mutism questionnaire in Spanish children. *International Journal of Clinical and Health Psychology*, 21(3), 100249. <https://doi.org/10.1016/j.ijchp.2021.100249>
- Ooi, Y. P., Sung, S. C., Raja, M., Kwan, C. H., Koh, J. B. K., & Fung, D. S. S. (2016). Web-based CBT for the treatment of selective mutism: Results from a pilot randomized controlled trial in Singapore. *Journal of Speech Pathology & Therapy*, 1(2), 1–7.
- Ortega, M. L. (2011). The generalization of verbal speech across multiple settings for children with selective mutism: A multiple-baseline pilot study. (Publication no. 3460734) [Doctoral Dissertation, Rutgers The State University of New Jersey]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/generalization-verbal-speech-across-multiple/docview/880570929/se-2?accountid=13460>
- Østergaard, K. R. (2018). Treatment of selective mutism based on cognitive behavioural therapy, psychopharmacology and combination therapy – A systematic review. *Nordic Journal of Psychiatry*, 72(4), 240–250. <https://doi.org/10.1080/08039488.2018.1439530>
- Paasivirta, M. (2012). Effects of a brief behavioural training program for teachers of children with selective mutism. (Publication no. 3522894) [Doctoral Dissertation, Hofstra University]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/effects-brief-behavioral-training-program/docview/1040874443/se-2?accountid=13460>
- Paez, A. (2017). Grey literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3), 233–240. <https://doi.org/10.1111/jebm.12265>
- Pustejovsky, J., Chen, M., & Hamilton, B. (2020). *Scdhlm: Estimate hierarchical linear models for single-case designs*. R package version 0.5.0. University of Wisconsin-Madison. <https://jepusto.github.io/scdhlm/>
- Remschmidt, H., Poller, M., Herpertz-Dahlmann, B., Hennighausen, K., & Gutenbrunner, C. (2001). A follow-up study of 45 patients with elective mutism. *European Archives of Psychiatry and Clinical Neuroscience*, 251(6), 284–296. <https://doi.org/10.1007/pl00007547>
- Rodrigues Pereira, C., Ensink, J. B. M., Güldner, M. G., Kan, K. J., De Jonge, M. V., Lindauer, R. J. L., & Utens, E. M. W. J. (2022). The validation of the selective mutism questionnaire for use in the Dutch population. *Child Psychiatry and Human Development*. <https://doi.org/10.1007/s10578-022-01387-8>
- Rodrigues Pereira, C., Ensink, J. B. M., Güldner, M. G., Lindauer, R. J. L., De Jonge, M. V., & Utens, E. M. W. J. (2021). Diagnosing selective mutism: A critical review of measures for clinical practice and research. *European Child & Adolescent Psychiatry*. <https://doi.org/10.1007/s00787-021-01907-2>
- Roslin, B. D. (2013). Adapting parent-child interaction therapy for selective mutism: A test of generalization. (Publication no. 3588028) [Doctoral Dissertation, Hofstra University]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/adapting-parent-child-interaction-therapy/docview/1427862021/se-2?accountid=13460>
- Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardised mean difference statistic: A primer and applications. *Journal of School Psychology*, 52(2), 123–147. <https://doi.org/10.1016/j.jsp.2013.11.005>
- Sharkey, L., McNicholas, F., Barry, E., Begley, M., & Ahern, S. (2008). Group therapy for selective mutism-a parents' and children's treatment group. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(4), 538–545. <https://doi.org/10.1016/j.jbtep.2007.12.002>
- Silverman, W. K., & Albano, A. (1996). *Anxiety disorders interview schedule for DSM-IV: Child version*. Graywind Publications Incorporated.
- Siroky, A. K. (2019). Integrated behaviour therapy for exclusively-anxious selective mutism: A nonconcurrent multiple baseline design across five participants. (Publication no. 22583866) [Doctoral Dissertation, Michigan State University]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/integrated-behavior-therapy-exclusively-anxious/docview/2310237433/se-2?accountid=13460>
- Solz, J. (2015). Video self-modeling as exposure with and without contingency management in treating youth with selective mutism. (Publication no. 3727067) [Doctoral Dissertation, Hofstra University]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/video-self-modeling-as-exposure-with-without/docview/1733644409/se-2?accountid=13460>
- Steains, S. Y., Malouff, J. M., & Schutte, N. S. (2021). Efficacy of psychological interventions for selective mutism in children: A meta-analysis of randomized controlled trials. *Child: Care, Health and Development*, 47(6), 771–781. <https://doi.org/10.1111/cch.12895>
- Steffenberg, H., Steffenburg, S., Gillberg, C., & Billstedt, E. (2018). Children with autism spectrum disorders and selective mutism. *Neuropsychiatric Disease and Treatment*, 14, 1163–1169. <https://doi.org/10.2147/ndt.s154966>
- Steinhausen, H., Wachter, M., Laimböck, K., & Winkler Metzke, C. (2006). A long-term outcome study of selective mutism in childhood. *Journal of Child Psychology and Psychiatry*, 47(7), 751–756. <https://doi.org/10.1111/j.1469-7610.2005.01560.x>
- Steinhausen, H. C., & Juzi, C. (1996). Elective mutism: An analysis of 100 cases. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(5), 606–614. <https://doi.org/10.1097/00004583-199605000-00015>
- Stone, B. P. (2000). Videotape modelling interventions for selective mutism: A comparative research investigation. (Publication no. 9972835) [Doctoral Dissertation, University of Wisconsin-Madison]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/videotape-modeling-interventions-selective-mutism/docview/304637914/se-2?accountid=13460>

- Stone, B. P., Kratochwill, T., Sladeczek, I., & Serlin, R. C. (2002). Treatment of selective mutism: A best-evidence synthesis. *School Psychology Quarterly*, *17*(2), 168–190. <https://doi.org/10.1521/scpq.17.2.168.20857>
- Tan, Y. R., Ooi, Y. P., Ang, R. P., Goh, D. H., Kwan, C., Fung, D. S. S., & Sung, S. C. (2021). Feasibility trial of virtual reality exposure therapy for selective mutism. *Clinical Child Psychology and Psychiatry*, *27*(2), 351–368. <https://doi.org/10.1177/135910452111056920>
- Toppelberg, C. O., Tabors, P., Coggins, A., Lum, K., & Burger, C. (2005). Differential diagnosis of selective mutism in bilingual children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(6), 592–595. <https://doi.org/10.1097/01.chi.0000157549.87078.f8>
- Torchiano, M. (2020). Effsize: Efficient effect size computation. R package version 0.8.1. <https://CRAN.R-project.org/package=effsize>
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. <https://doi.org/10.3102/1076998609346961>
- Vecchio, J., & Kearney, C. A. (2009). Treating youths with selective mutism with an alternating design of exposure-based practice and contingency management. *Behavior Therapy*, *40*(4), 380–392. <https://doi.org/10.1016/j.beth.2008.10.005>
- Vecchio, J. L. (2008). The treatment of selective mutism: A case control alternating treatments design. (Publication no. 3338188) [Doctoral Dissertation, University of Nevada]. ProQuest One Academic. <https://www.proquest.com/dissertations-theses/treatment-selective-mutism-case-control/docview/304381483/se-2?accountid=13460>
- Viana, A. G., Beidel, D. C., & Rabian, B. (2009). Selective mutism: A review and integration of the last 15 years. *Clinical Psychology Review*, *29*(1), 57–67. <https://doi.org/10.1016/j.cpr.2008.09.009>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Woodcock, E. A., Milic, M. I., & Johnson, S. G. (2007). Treatment programs for children with selective mutism. In D. A. Einstein (Ed.), *Innovations and Advances in cognitive behaviour therapy* (pp. 69–81). Australian Academic Press.
- Xu, H. Y., Stroud, J., Jozanovic, R. K., Clucas, J., Son, J. J., Koo, B., Schwarz, J., Klein, A., Busman, R., & Milham, M. P. (2018). Clinical perspective on passive audio vocal measurement in the evaluation of selective mutism. *Frontiers in Psychiatry*, *9*, 443. <https://doi.org/10.3389/fpsy.2018.00443>
- Zakszeski, B. N., & DuPaul, G. J. (2017). Reinforce, shape, expose, and fade: A review of treatments for selective mutism (2005–2015). *School Mental Health*, *9*(1), 1–15. <https://doi.org/10.1007/s12310-016-9198-8>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hipolito, G., Pagnamenta, E., Stacey, H., Wright, E., Joffe, V., Murayama, K., & Creswell, C. (2023). A systematic review and meta-analysis of nonpharmacological interventions for children and adolescents with selective mutism. *JCPP Advances*, *3*(3), e12166. <https://doi.org/10.1002/jcv2.12166>