

Case study 1: An Evidence-based practice review report.

Theme: School/Setting Based Interventions for Social, Emotional and Mental Health.

Is school presentation of selective mutism improved by intervention?

A brief systematic literature review.

Summary

A brief systematic literature review identified five studies comparing pre and post outcomes from psychological interventions targeting selective mutism. Outcomes related to school-relevant behaviours, such as speaking more at school, were used as an indirect indicator that the associated anxiety had been reduced by the intervention. Papers that met the inclusion and exclusion criteria set for this review range from 2016 – 2019, no case-studies were included. The five papers were reviewed using a weight of evidence framework (Gough,2007). Papers included a range of intervention styles, including: group intervention; psychoeducation; cognitive behavioural therapy (CBT); parent training; a summer school approach and covered different time scales. A common feature across all five studies was the use of the selective mutism Questionnaire (SMQ) as an outcome measure, which allowed for a comparison of school relevant improvements, even where interventions took place within a clinical domain. Effect sizes were identified or calculated from the intervention data. All five papers described interventions that increased social speaking and speaking with the teacher with effect sizes that ranged from medium to large for SMQ outcomes. A consideration of the common features across these effective interventions is described. Conclusions about the evidence are discussed

alongside relevance to teachers and educational psychologists (EPs) and ideas for future research.

Introduction

Selective mutism.

Selective mutism (SM) is classified by the Diagnostic and Statistical Manual of Mental Disorders (DMS 5) as an anxiety disorder (American Psychological Association, APA, 2013). It is recommended to consider SM as a diagnosis:

“If a young person consistently fails to speak in specific social situations for at least 1 month, interfering with educational or occupational achievement” (Hilt et al., 2016, p107.)

Prevalence rates are low, estimated as between 0.3 to 1% in the United States (APA, 2013), and interconnected with each schools’ awareness of SM, due to the likelihood that assessment and diagnosis will be triggered by school concerns in the first instance rather than those of home (Kovac & Furr, 2019; Stone et al., 2002).

Selective mutism is recognised to overlap with social anxiety disorder (APA, 2013). Social anxiety in children, in the United states for example, prevalent at around 4% rising to 9% in adolescence (Pickard et al., 2017) meaning that every school will be working with anxious pupils. The link between social anxiety and SM was shown in the development of the selective mutism questionnaire (SMQ) that recorded concurrent validity between the SM and social anxiety but not between SM and other anxiety disorders (Bergman et al., 2008). Due to this overlap, the small incidence of diagnosed SM seems unlikely to represent all children who experience anxiety around speaking in the classroom.

Interventions for selective mutism.

Guidance for anxiety disorders generally (NICE, 2013) recommends group or individual CBT as the psychological intervention with an evidence base, suggesting it should include between 8-12 sessions and include psychoeducation.

Psychoeducation is seen as key to supporting the environmental elements of any anxiety disorder for example, parent-training elements. The importance of a supportive environmental context extends to schools, where a large proportion of the SM is seen (Stone et al., 2002). School staff may not be therapeutically competent to deliver appropriate CBT sessions however, with support they can make use of psychological theory that underpins effective treatments to support young people in their care in making verbal contributions in the classroom. In the context of SM, altering thought processes around talking behaviour (cognitive change), is achieved through gradual exposure to experiences that reinforce positive associations with speech (behavioural element). Psychoeducation for SM could be delivered to teachers by EPs. Other skills, transferable to a school context, are gradual exposure to speaking tasks to avoid teachers inadvertently participating in reinforcing the problem behaviours.

Treatments discussed in previous reviews (Cohan et al., 2006; Ostergaard, 2018; Zakszeski & DuPaul 2017) and meta analyses (Stone et al., 2002) encompass drug treatment and a range of behavioural, cognitive, family systems and psychodynamic techniques and multi-modal techniques. The most recent review (Ostergaard, 2018) elicits a clinical perspective by comparing drug and behavioural treatments.

Cognitive behavioural approaches (Cohan et al., 2006) and behavioural approaches (Stone et al., 2002) were described as effective, despite overall evidence being evaluated as weak.

Rationale

Due to the low prevalence rate of SM the research field is relatively small compared to other anxiety disorders with the majority of school-based research being case studies (Kern et al., 2007; O'Reilly et al., 2008; Sharkey et al., 2008). Previously, Stone et al. (2002), attempted to elicit which psychological therapies were more effective and found the methodological differences and varied treatment protocols prevented like for like comparison of effect sizes.

There are two reasons reviewing evidence for non-pharmaceutical interventions is important for people working in schools. The first is that waiting lists for child and adolescent mental health services (CAMHs) are long, the Young Minds website suggests some young people are waiting up to 200 days for treatment (Young Minds, n.d). Experimental research can help identify strategies that are appropriate for use in the classroom to prevent symptomology from deteriorating while pupils wait for more specialist support. Psychologically informed interventions and those with psychoeducation can provide generalisable strategies to teaching pedagogy in primary and secondary classrooms supported by educational psychologists. In addition, in response to the publication of the Green Paper on the mental health of children and young people (Department of Health and Social Care; Department for Education, 2018) it became clear that we could expect a new generation of mental health support in schools in the UK, in the form of Mental Health Support teams by 2020. This new role within schools would be well suited to deliver evidence-based practice to young people with selective mutism, increasing capacity to deliver support.

The second reason is that links between SM and social anxiety suggest that skills and good practice developed through a consideration of what works well for SM will

have transferable skills for teachers in their work with students that are not of concern clinically but who also experience anxiety with speaking out in class.

A detailed description of the methodologies and terminology behind behaviourist principles in treatment methods for SM already exists (Zakszeski & DuPaul, 2017). Zakszeski and DuPaul (2017), in their review of research from 2005 to 2015, call for more experimental research to be carried out; this review intends to assess findings from experimental rather than case study research on SM conducted, focusing on what the outcome data can tell us about how the intervention impacts communication in school.

The review question was developed to be of relevance to educational psychologists and teachers seeking to develop their understanding of anxiety presentations in the class environment that are specific to increasing spoken contributions.

Review question

Is school presentation of selective mutism improved through intervention?

Critical Review of the Evidence Base

Literature search

Electronic database searches of the literature on selective mutism were conducted on the 21/1/2021. Databases searched were: Psycinfo; ERIC(ProQuest); and Web of Science. The search terms used are shown in Table 1. Appendix A shows the combined search method used to identify research in Web of Science. Whilst it was acknowledged that ‘elective mutism’ is an outdated term, it was included to ensure potential studies were identified, future reviews can be confident this term did not produce any additional research post-2000. Previous reviews (Cohan et al., 2006; Zakszeski & DuPaul, 2017) established that interventions for SM did not fall under one umbrella term and, as such, this review did not include intervention descriptors in the search to maximise studies being captured in the process.

Table 1

Search terms used for database searches

Focus	Context	Relevant to ‘treatment’
“selective mutism” Or “Elective mutism”	School Or education	Intervention Or Outcome

Note: “OR” combines terms. The concepts in each column were then combined using “AND” to include concepts from each section in the search.

Inclusion and exclusion criteria.

Search limits were publication between 2000-2021, papers written in English and peer-reviewed journal articles due to the time constraints of this review. Where the search database options allowed, books and dissertations were excluded.

Abstracts and articles were assessed for their eligibility using inclusion and exclusion criteria set out in Table 2. Table 3 lists included studies, following the search strategy (Figure 1). Studies excluded at the abstract review stage are listed in Appendix B along with the reason for exclusion. Petticrew and Roberts (2003), suggest, to answer a question about how effective an intervention is, it is necessary to consider randomised control trials, cohort studies and quasi-experimental research. All the studies included in this brief review (Abrami et al., 2010) were compared using their pre- and post- data for a SM specific outcome measure (Selective mutism questionnaire, SMQ).

Table 2

Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria	Rationale
Population	School-aged children and young people from 4 years to 18 years old.	Children younger than 4 years, older than 18 years.	The current review is looking at interventions appropriate for school.
Intervention	Interventions that aim to support behavioural, social or cognitive change.	Interventions that included drug treatments or methods that did not specify a behavioural, social or cognitive element to the intervention	This review aims to consider the effectiveness of interventions that are appropriate for delivery in schools.
Comparison	Pre and post data from RCT's, cohort studies and quasi-experimental designs.	Qualitative reports of outcome. Single case experimental designs.	This review is asking if there is a time by intervention interaction. (Petticrew and Roberts, 2003, typology of evidence)
Outcome	Report a quantitative pre and post-outcome measure that is relevant	Outcome measures that are clinically relevant only	For example, SMQ is a commonly used tool that reports on

	Inclusion criteria	Exclusion criteria	Rationale
	to school, (such as SMQ).	and do not link to behaviours observable in a school context.	school focused outcomes.
Context	Clinic, home or school	hospital	Focus of the review is application to school setting, origin of treatment in a school or clinic would be included as long as measures link to school relevant behaviour.
Publication date	2000-2021	Before 2000	Previous reviews report on studies up to 2015.
Peer reviewed journals	Search limit used where possible	Books, dissertations.	Due to time constraints, and to include articles that had been through a review process already.

Figure 1

PRISMA flow diagram search strategy.

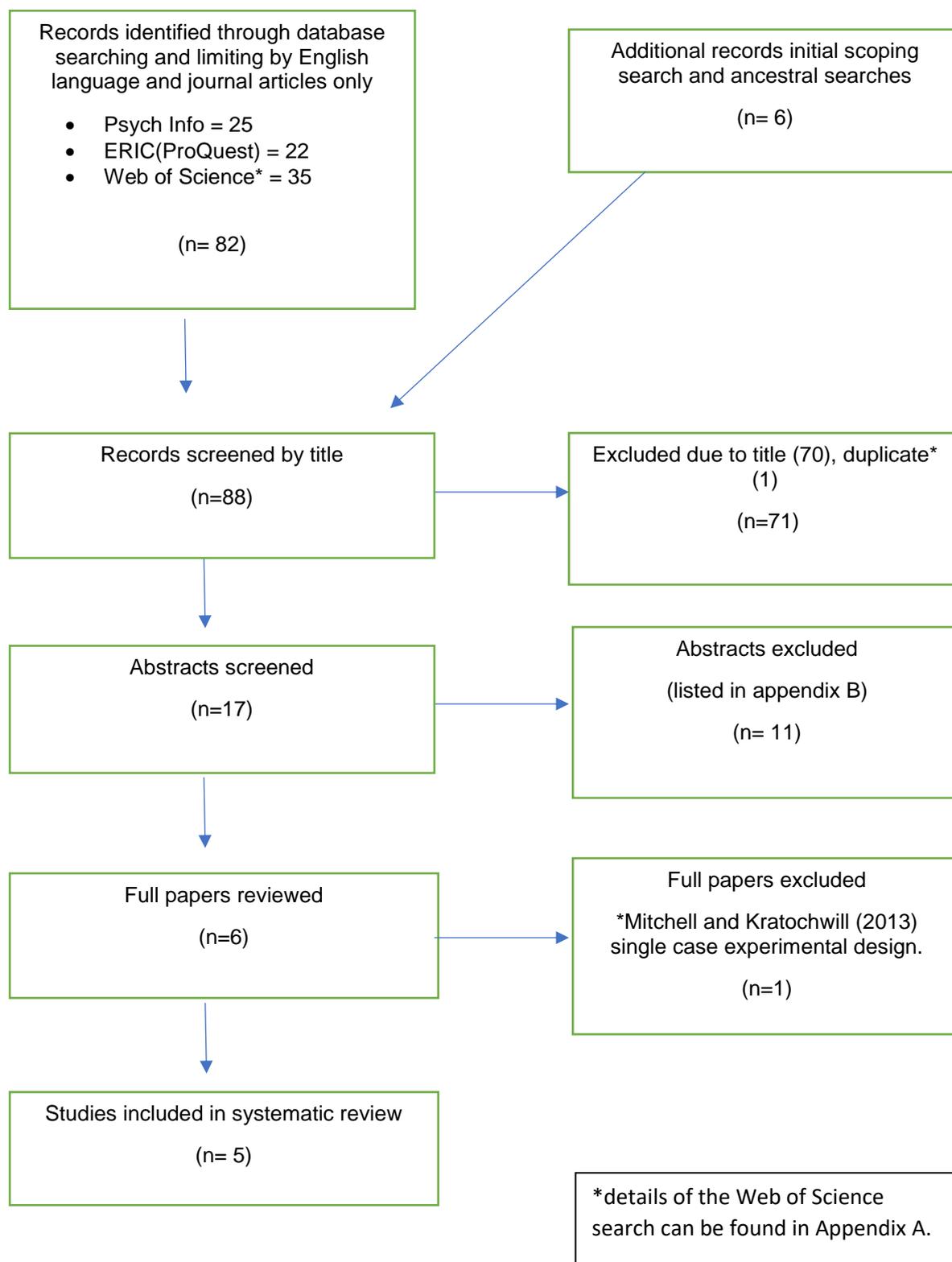


Table 3

Full references of the final 5 studies included in this review

- 1 Catchpole, R., Young, A., Baer, S., & Salih, T. (2019). Examining a novel, parent child interaction therapy-informed, behavioral treatment of selective mutism. *Journal of Anxiety Disorders*, 66. <https://doi.org/10.1016/j.janxdis.2019.102112>
 - 2 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 behavioral 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>
 - 3 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>
 - 4 Lang, C., Nir, Z., Gothelf, A., Domachevsky, S., Ginton, L., Kushnir, J., & Gothelf, D. (2016). The outcome of children with selective mutism following cognitive behavioral intervention: a follow-up study. *European Journal of Pediatrics*, 175(4), 481–487. <https://doi.org/10.1007/s00431-015-2651-0>
 - 5 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>
-

Weight of evidence

The included studies were subjected to a weight of evidence (WoE) review, using criteria described by Gough, (2007). Each study was assessed on three dimensions:

research design, WoE A; appropriateness of the evidence, WoE B; and a review of the study’s relevance to the review question, WoE C (Table 4). Values on each dimension were combined to create the overall assessment of the study as WoE D.

To assess WoE A, an adapted version (Read, et al., 2020) of Down and Black’s (1998) checklist for both randomised control trials and non-control group studies was used (Appendix C). The checklist generated a score out of 12 for each study, scores were then divided into three categories (low, medium and high) and a numerical value out of 3 was assigned to each category. WoE B was assessed based on The Oxford Levels of Evidence table (2011), using a score out of 5, recommended for review questions on intervention effectiveness (Appendix D). WoE C assessed how relevant each study was at answering the review question (Appendix E). Three levels of relevance to the review question were also assigned a number out of 3. The final WoE D assessment was made using all three values (Appendix F). The WoE review showed the studies were graded from medium to high. The lower grading of Lang et al. (2016) for WoE C reflects outcome data on school behaviour reported by home rather than school, inclusion criteria was met due to the SMQ scores being reported by sub sections for school, home, public separately allowing for a comparison of school impacted speaking behaviour.

Table 4

Overall weight of evidence (WoE) appraisal

Study	WoE A – research design	WoE B – appropriateness of evidence	WoE C – relevance to review question	WoE A-C total	WoE D – overall assessment
Catchpole et al. (2019)	3/3 High	3/5 Medium	2/3 Medium	8/11	73% High

Study	WoE A – research design	WoE B – appropriateness of evidence	WoE C – relevance to review question	WoE A-C total	WoE D – overall assessment
Cornacchio et al. (2019)	3/3 High	4/5 Medium/high	2/3 Medium	9/11	82% High
Klein et al. (2017)	3/3 High	3/5 Medium	2/3 Medium	8/11	73% High
Lang et al. (2016)	3/3 High	3/5 Medium	1/3 Low	7/11	64% Medium
Oerbeck et al. (2018)	3/3 High	3/5 Medium	3/3 High	9/11	82% High

Note: Calculations for WoE D explained in Appendix F

Participants.

Participants were school aged children from Canada, USA, Norway and Israel.

Sample sizes ranged from 24 to 40. When studying a low incidence disorder that crosses several professional contexts, pragmatic decisions had been made on how to access participants to provide group comparison data whilst retaining the external validity of the intervention. Table 5 shows information given in the papers on participant recruitment.

Table 5

Recruitment of participant information and diversity and parental context information

Study	Recruitment of participants	Diversity of participants
Catchpole et al. (2019)	An out-patients mental health department of a Children’s hospital. Met DSM V diagnostic criteria for SM. Excluded if co morbid with psychosis, mania, ASD, intellectual disability.	English language ability was a prerequisite. 32.3% of children, 51.9% of mothers and 50% of fathers had English as additional language.

	Medication was not an exclusion criterion if at least an 8-week stable dose.	Parental education recorded and covered full range of options described.
Cornacchio et al. (2019)	<p>Recruited parents who were seeking help at a SM speciality treatment centre. Met DSM V criteria for SM and where SM was dominant condition if there were other co-morbidities. Excluded if they were non-verbal with both parents due to requirements for pre-treatment baseline.</p> <p>Medication was not an exclusion criterion if stable dose has been established over 6 weeks.</p>	<p>Described as ethnically diverse sample, 1/3 Hispanic/Latino, 41.9% self-identifying as Asian. Diverse economic backgrounds of families (referred to family income levels).</p>
Klein et al. (2017)	<p>Recruited from families who had contacted specialist in SM after advertising the study online. All participants had an SM diagnosis Excluded from participation if taking medication or attending other therapies. Excluded if diagnosis was comorbid with autism or intellectual disability, uncorrected hearing or visual impairment.</p>	<p>Fluency in English was a prerequisite. 72.5% Caucasian, 2.55% African American, 15% Asian, 5% Hispanic and 5% Biracial. Mothers education level and parent anxiety is also recorded.</p>
Lang et al. (2016)	<p>All met DSMV criteria for SM in an anxiety specific clinic. 1 clinician treated all participants.</p>	<p>No further information reported.</p>
Oerbeck et al. (2018)	<p>30/32 children from the previous pilot study and RCT study and 1 child who had received the intervention previously. 3-9 years of age who met diagnostic criteria for SM and do not speak to adults in preschool or school setting. Referral from either CAMHS or school Psychology service. Excluded children had intellectual disability (IQ <50), non-Norwegian speakers. 2 children who were not participants in this follow up were reported as still symptomatic at 1 yr.</p>	<p>9 children were reported as bilingual. Familial SM reported in 11 out of 30 families</p>

A common place for recruiting participants was through medical and clinic settings. The Stone et al. (2002), review indicated the school context was more frequently used to recruit for single case designs, the focus here on non- case study designs was reflected in the higher WoE A values for these five studies. The WoE B ratings were also affected by the recruitment process, with no study allocated 5/5 by the Oxford Levels of evidence criteria due to the lack of randomised control trials (RCT's). RCT's are used less frequently in school settings, where external validity, and therefore generalisability, are prioritised over internal validity (determining causality) partly due to the complexity of systemic factors (Barker et al., 2016). Catchpole et al. (2019), Cornacchio et al. (2019) and Klein et al. (2017) collected information about the ethnic and socio-economic status of the families that suggested the samples were diverse. Collecting a diverse sample from opportunistic sampling at clinics suggests that SM is not a culturally specific issue. With symptomology of SM clearly linked to the school context by definition, studies that did not reference the school-context in either their methodology or outcomes measures would be less appropriate for this review question. WoE C was rated 'low' for Lang et al. (2016), as school relevant behaviours were reported by parents and not teaching staff, making these scores less valid.

Information gathered on participants familial background was a strength of Catchpole et al. (2019), Cornacchio et al. (2019), and Overbeck et al (2018), it supports the developing understanding around systems-based influences, for example, familiar links with SM did negatively affect outcomes (Overbeck et al., 2018). This element was missing from the WoE assessments and would be a helpful addition to the Downs and Black (1998) checklist, in order to assess how broadly findings can be applied across different groups.

Inclusion criteria had filtered studies for school-age children, ensuring the participant groups were matched well to answer the review question (assessed in WoE A scores which were high, 3/3, for all studies). Children studied ranged from 4 years old to 15 years old at the start of the intervention, with Cornacchio et al. (2019), and Oerbeck et al. (2018), finding that younger age of participants were associated with larger improvements in outcome measures.

The Catchpole et al. (2019) and Cornacchio et al. (2019) research did not use current medication as an exclusion criterion, which led to useful findings. As SM is a type of social anxiety, it may have been expected that anti-anxiety medication would improve access to treatment benefits, this was not case. Schools may assume that students receiving external services support and medication do not need further help at school, the inclusion of children with SM on medication clarifies that additional benefits can be gained from contextual interventions.

Research design

Previous reviews considered the entire body of research on SM (Stone et al., 2002) and found variability in methodological quality making outcome data hard to compare. This review defined specific research designs as part of the inclusion criteria (Table 2) to make comparison across studies more likely and included quasi-experimental designs and randomised control trials. The WoE scoring suggests that quasi-experimental methods were equally able to achieve medium and high ratings in the context of this review for 'is the intervention effective' style questions. Wait-list control ruled out maturational improvements and was a pragmatic solution when working with small prevalence disorders to build an evidence base beyond case studies. For example, Catchpole et al. (2019), and Cornacchio et al. (2019), made use of waitlist control groups to show that SM does not diminish over time,

intervention is needed. As a result of specifying research design criteria in the inclusion criteria, table 4 shows that all the research included scored 3/3 for design.

Intervention

Brief descriptions of the interventions are given (Table 6), the treatment protocols had many common features, suggesting that evidence from previous reviews (Stone et al., 2002) had fed into fine tuning intervention approaches to SM. All interventions referred to methods to generalise strategies from the intervention specific space to other settings (home, school or both) demonstrating a move from strategies that fall within one theoretical domain to those encompassing behavioural techniques, cognitive elements and systems around the child (Zakszeski & DuPaul, 2017).

Gradual exposure to speaking situations were used across interventions, as were elements of parent-training, facilitating intervention support at home. Catchpole et al. (2019), Lang et al. (2017) and Oerbeck et al. (2018) mentioned school outreach or training educational staff as part of the intervention. Cornacchio et al. (2019) and Klein et al. (2017) did not have school staff training however, they did report outcomes from school. Cornacchio et al. (2019) had a teacher reported outcome measure for return to school following intervention and Klein et al. (2017) had the most in-depth breakdown of school-based SMQ questions reported by parents (Appendix G, Table 7) and had a teacher reported Child Behaviour Checklist measure. Table 7 is included to illustrate how specific and helpful the school data from this subsection of the SMQ is, whilst generally not being reported .

Table 6

Name and brief description of each intervention

Study	Intervention name	Intervention outline and key features of protocol.
Catchpole et al. (2019)	Parent-child Interaction principles and behavioural techniques.	16 one-hour sessions conducted over max 22 weeks. Clinicians trained in Parent-child interaction therapy (PCIT-SM) and use Child directed interaction (CDI) and verbal directed interactions (VDI) 1 st session parent only training session – use of CDI and VDI. Following sessions included parent, child and therapist. Graded exposure to talking described as ‘second treatment component’. Office sessions move to variety of settings and people. Parent drive exposure sessions in the real world are used as part 3 in between sessions. Fourth component includes three sessions of school-outreach work, staff psychoeducation and support.
Cornacchio et al. (2019)	Intensive group behavioural treatment (IGBT).	5 consecutive days of 6-8 hrs of daily treatment (30-40 hours). Draws on format of Parent-child Interaction therapy - Child-directed interaction (CDI) and verbal directed interactions (VDI) used by staff. Graduated exposure to verbal communication. Parent training sessions. In vivo parent and child coaching in real life situations. Techniques mentioned and described in paper: reinforcement, prompting, shaping, stimulus fading, graduated exposure, social skills training, cognitive strategies, relaxation training and modelling.
Klein et al. (2017)	S-CAT Social Communication Anxiety Treatment	9 weeks. Sessions held in child friendly clinic setting once every 3 weeks. Main goals of S-CAT are described as reducing anxiety about speaking and reducing enabling

Study	Intervention name	Intervention outline and key features of protocol.
		<p>behaviours of parents, reduce avoidance behaviours in the child. Behavioural and cognitive strategies used. Spaced sessions provided time for parents and children to chart goal progress in communication environments that were usual for each child. This process was designed specifically for SM by Elisa Shipon-Blum – access to treatment information was given in link to website (listed in references)</p>
<p>Lang et al. (2016)</p>	<p>Modular Cognitive behavioural therapy (MCBT).</p>	<p>Length of treatment M = 12.58 months Modules used to create an individualised treatment. Included: psychoeducation; physiological training, cognitive training (cognitive restructuring), behaviour training (contingency management, exposure hierarchy, modelling, shaping, gradual desensitisation); parent training; educational or recreational staff training. Age of child was mentioned as a variable that altered programme, for example children over 10 years of age were seen separately from parents.</p>
<p>Overbeck et al. (2018)</p>	<p>School-based CBT</p>	<p>M = 21 weeks (SD = 5, range 8-24 weeks) Up to 6 months Manualised programme. Started at home using defocused communication techniques to build rapport. Psychoeducation for parents and teachers. Behavioural interventions at pre-school or school twice a week included: gradual exposure, stimulus fading, contingency management.</p>

Measures

Outcome measures (Table 8) common across studies were a version of the SMQ (Bergman et al., 2008) and the Anxiety Disorders Interview Schedule (ADIS, Silverman & Albano, 1996). The use of validated measures and the justification of them led to consistent scores on two elements of the WoE A checklist across the research field.

The call by Stone et al. (2002), for researchers to make use of standardised tools used for other anxiety disorders seems to have been answered by researchers. The ADIS was used for baseline assessment of severity of symptoms, e.g. Catchpole et al. (2019) $M = 6.48$ (range = 4.5-8, $SD = 1.23$) where a score of 4 is considered as moderate and 8, very severe (Silverman & Ollendick, 2005). ADIS was also used in waitlist control studies to demonstrate similarity of symptom severity in both groups, for example, Cornacchio et al. (2019), reported that the intervention group had ADIS score $M = 4.9$ ($SD = 0.8$), and waitlist control $M = 4.9$ ($SD = 0.7$). Klein et al. (2017) referred to a need for an ADIS score of at least 4 as an inclusion criterion. ADIS scores were used but not reported in the Lang et al. (2016) and Oerbeck et al. (2018) papers.

Since 2008 a standardised assessment for SM (SMQ, Bergman et al., 2008) has been available and was used by all researchers in this review. Internal consistency of the SMQ is .97 for both the total SMQ scores and the school subscale mentioned in this review (Table 7). The introduction of a validated measure with subscales for school, social and home is promising addition to identify specific outcomes that change for children following intervention.

A varied range of other measures were used by researchers for example Catchpole et al. (2019) and Klein et al. (2017) used both the Peabody Picture Vocabulary Test

(PPVT -2 or 4, Dunn & Dunn, 2007) and the Expressive Vocabulary Test (EVT-2, Williams, 2007) and concluded that scores on neither of these measures were predictors of outcomes on SMQ after treatment.

Klein et al. (2017) also used Family compliance rating pre-intervention and found a low score for compliance from parents associated with less positive outcomes post-intervention.

Oerbeck et al. (2018) were the only researchers to include a child self-rating of social speaking and Inventory of life quality in children and adolescents (ILC, Jozefiat, 2011).

Findings

The studies reviewed demonstrate outcomes with large effect sizes (Table 8), having been weighted positively for their methodology (Table 4).

Whilst Cornacchio et al. (2019), Oerbeck et al. (2018) found the youngest children were more likely to show improvement, Catchpole et al. (2019) did not find this to be the case. Lang et al. (2016), suggested that older children may respond to slightly different elements of the treatment, this perspective is supported elsewhere in the literature (Yeager et al., 2018). Lang et al. (2016) also expressed alarm that whilst the age of onset of SM was at 3 years, most were not seeking treatment until age 6 despite evidence that SM interventions are effective in younger children. One hypothesis for this time lag is that verbal behaviour becomes more problematic through the eyes of educationalists in Key Stage 1 and may be dismissed as a transient issue by Health professionals (Kovac & Furr, 2019).

One suggestion made was that more severe symptoms at start were associated with the least progress made (Oerbeck et al., 2018). However, Catchpole et al. (2019)

noted that 38.7% of their participants were on medication at the start of the study, indicating high levels of symptomology, but medication was not a predictor of outcome in that study and Klein et al. (2017) found no correlation between duration of SM before treatment and outcomes. More participants were female in the studies that reported gender (Table 8) however, gender was not seen as having a significant impact on outcome (Oerbeck et al., 2018).

Klein et al. (2017) reported the findings of the SMQ for each specific school relevant situations (Appendix G, Table 7), showing large effect sizes (Cohen, 1992) over time, which, in the context of a much shorter programme of treatment than previous examples, have a practical and user-friendly application to school. The Klein et al. (2017) study using S-CAT required 3 meetings to gain the large effect reported. The next shortest were 16 hours (Catchpole et al., 2019) and Oerbeck et al. (2018) that ranged from 8-24 weeks. In hindsight length of treatment as part of a feasibility analysis may have been a helpful addition to the WoE C appraisal when considering delivering interventions for school age children.

What can also be seen from Appendix G, and the reason for highlighting this level of detail from one study, was the much lower starting points for any task involving speaking to staff. These differences highlight the importance of school staff having an awareness of SM and strategies to reduce anxiety in the classroom. The involvement of school staff was a key element of the Catchpole et al. (2019) Parent-child Interaction therapy and this study reported no age-related differences in outcomes that other papers noted. Another finding to recommend the use of teacher ratings as distinct from parental views of school was from Klein et al. (2017) who found no correlation between parental and teacher ratings of school anxiety ($r(29) = .05, p = .787$).

The findings suggest intervening in the classroom with techniques based around these strategies as early as possible to lead to beneficial outcomes in relation to reducing severity of SM.

Fidelity of treatment

On reflection, an area not assessed by the WoE appraisal was an acknowledgement of fidelity of intervention protocols.

Table 9 shows that fidelity of treatment was explicitly addressed by Catchpole et al. (2019), Cornacchio et al. (2019), and Klein et al. (2017). The impressive treatment outcome for SMQ over just 3 sessions, in the Klein study has yet to be seen delivered by other therapists beyond the programme designer. Caveated against the backdrop of all studies having impressive effect sizes, intervention outcomes demonstrated across more than one therapist have increased generalisability.

Table 8

Effect sizes (ES) for SMQ (and/or SSQ) with basic study descriptors, WoE D ratings and other outcome measures used

Author and country	Study design	Sample	SMQ or SSQ outcome data reported	Effect size(s)	Overall WoE D rating	Other outcome data reported
Catchpole et al. (2019) Canada	Quasi-experimental design. Within subject waitlist controlled. 16 one-hour sessions over max 22 weeks	N = 31 (females 16, males 15) Age range = 4 to 9.75 (M = 6.47, SD = 1.68)	SMQ school speaking subscale: F (2.24,53.78) = 57.17, p<.001 SSQ teacher report: t (25) = -6.70, p <.001	$\eta^2 p = .70$ Large d = 1.07 Large	High	Anxiety disorders interview schedule (ADIS-P) The Peabody picture vocabulary test (PPVT-4) Expressive vocabulary test (EVT-2) Strong narrative assessment procedure (SNAP) The screen for social anxiety emotional disorders – parent version (SCARED-PV) Client satisfaction questionnaire (CSQ-8)
Cornacchio et al. (2019) USA	RCT with wait-list psychoeducation control group.	N = 29 (Females, 22, males,7) Age range: 5- 9yrs (M = 6.6, SD = 1.3)	Teacher report SSQ: baseline (M = 1.1, SD = 0.7), school year follow up (M = 1.6, SD = 0.7)	d = .71 Medium $\eta^2 p = .18$ Large	High	Anxiety Disorders Interview schedule for children-parent version (ADIS) Clinical global impression-improvement scale (CGI-I) Child behaviour checklist (CBCL) Children’s global assessment scale (CGAS)

Author and country	Study design	Sample	SMQ or SSQ outcome data reported	Effect size(s)	Overall WoE D rating	Other outcome data reported
			SMQ social subscale: F (1,25) = 5.35			The impairment rating scale (IRS) Client satisfaction questionnaire (CSQ) Barriers to treatment participation (BTPS) Following academic year 46% of treated children had no SM diagnosis.
Klein et al. (2017) USA	Quasi-experimental Evaluation study Longitudinal	N = 40 for 9-week trial N = 33 for follow up at 15 weeks Age range: 5-12 yrs. (M = 6.78, SD = 1.58)	SMQ school: F (3, 95.9) = 16.27, p = .53 SMQ social: F (4,128) = 22.15, p = <.001	$\eta^2 p = .34$ Large $\eta^2 p = .41$ Large	High	Behaviour assessment system for children (BASC – 2- SDH) Child behaviour checklist (CBL) Anxiety disorders interview for DSM IV (ADIS-IV) Therapist treatment fidelity Family compliance rating The Peabody picture vocabulary test (PPVT-4) The expressive vocabulary test (EVT-2) The test of narrative language (TNL)
Lang et al. (2016) Israel	Quasi-experimental Evaluation study	N = 24 Age range: 5-15 years.	SMQ school: t=19.18, p<.001	d = 2.06 Large ^a	Medium	Anxiety disorders interview for DSM IV: lifetime version (ADIS-IV-L) Clinical global impression (CGI)

Author and country	Study design	Sample	SMQ or SSQ outcome data reported	Effect size(s)	Overall WoE D rating	Other outcome data reported
	Intervention group followed up 1 year later.					84% recovery rate
Oerbeck et al. (2018) Norway	RCT and quasi-experimental pilot design. Mean 21 weeks Intervention followed up at 5 yrs.	N = 31 (follow up of pilot and RCT participants from 2015 paper) Age range at baseline:3-9yrs. Females 20, males 11)	SMQ school: 5 years post, F (4, 119) = 28.49, $p < 0.001$ At 1-year post treatment SSQ (teacher): 5-year post intervention At 1-year post intervention	d = 2.8 ^a Large ^a d = 1.52 ^a Large ^a d = 2.11 Large ^a d = 1.41 Large ^a	High	Anxiety Disorders interview schedule IV (ADIS-IV) Schedule for affective disorders and schizophrenia for school-aged children: present and lifetime version (K-SADS-PL) Inventory of life quality in children and adolescents (ILC) Child self-rating of social speaking At follow up 70% (n=21) no longer met criteria for diagnosis.

^adenotes where no ES was reported, ES was calculated using: <https://www.ai-therapy.com/psychology-statistics/effect-size-calculator>.

Table 9

Intervention fidelity

	Fidelity of treatment protocol
Catchpole et al. 2019	Clinical psychologist and child psychiatrist trained in PCIT-SM – demonstrated CDI/VDI skills to mastery criteria. Treatment adherence monitored after each session using a checklist Monthly meetings to ensure consistency and fidelity
Cornacchio et al. 2019	Staff trained to use CDI and VDI skills from PCIT and had to demonstrate proficiency Team lead: Clinical Psychologist Volunteers: undergraduate and postgraduate students All undertook 2 6-hour training led by Clinical Psychologist All staff coded at least once during week of treatment for adherence check
Klein et al. 2017	One therapist – developed the treatment Progress notes and videotaped each session, these notes formed a spreadsheet overview created by a research assistant A third of all videotapes were randomly selected and reviewed by research assistants using the spreadsheet to compare treatment delivery across patients - 80% of treatment followed the same rate.
Lang et al. 2016	One therapist – a clinical psychologist A non-manualised, modular approach, adjusted to suit individuals
Oerbeck et al. 2018	CAMHS therapists described as ‘not experts’ in SM or CBT

Conclusion and Recommendations

Conclusions

Previous reviews have identified gaps and problems within SM literature that this current review found being gradually addressed. Less variability was seen in methodological quality across the studies compared to reviews of all treatment methods (Stone et al., 2002), a consistent use of standardised measures applicable to anxiety and SM specifically (Cohan et al., 2006) and a more homogenous treatment approach despite differences in intervention labels (Zakszeski & DuPaul, 2017).

After the review by Zakszeski and DuPaul (2017), calling for larger sample research to complement the body of case study data on SM, this review was able to identify research fitting this category since 2000 that covered 2015 to 2021, a time period that includes substantial partial school closure due to COVID-19. It is positive the call for research has been responded to and will, hopefully, build as pupils return to in-school teaching.

This review found that non-pharmaceutical treatment of SM has positive outcomes for young people, with large effect sizes, and that interventions are grounded in cognitive and behavioural processes. Stone et al. (2002) predicted that behavioural techniques would have the best outcomes and intended to identify which psychological interventions were most effective, but the range of treatments at that time prevented easy comparison. Now, 19 years on, the research identified has more in common than it has to separate it.

Early intervention and working with children before symptoms are severe are positively associated with the largest improvements (Oerbeck et al., 2018). This suggests a role of educators and educational psychologists at early identification of non-verbal behaviour in the classroom to ensure that measures can be put in place to reduce speaking anxiety in the classroom early.

Interestingly in the Oerbeck 5-year study, outcomes were not affected by the CAMHS therapist's prior knowledge of SM or expertise in CBT (Table 9). This has relevance to school contexts; pastoral and supporting staff could make use of elements described in the interventions when supporting anxious pupils, with the support of an EP on how best to apply them to a classroom context.

Recommendations

Schools and educational psychologists can make use of this review by considering elements of the treatment descriptions that can be implemented into classroom routines and may benefit a range of needs, as all varieties of treatment led to significant improvement in outcomes. One example is the child-directed interaction (CDI) and verbal-directed interactions (VDI) styles of questioning that Catchpole et al. (2019) and Cornacchio et al. (2019) taken from Parent-child Interaction Therapy. This could be beneficial generally in teaching pedagogy to enhance whole-class questioning techniques, making interactions child-centred and reducing teacher input. Kovac and Furr (2019), have explained how a number of the elements within the interventions can be adopted by classroom practitioners.

A second related area is supporting identification of SM in early years setting and in the differentiation of SM from behaviours categorised as 'shy' that may reduce access to early support (Kovac & Furr, 2019) and may explain the delay from symptomatic to seeking treatment.

Overbeck et al. (2018) report the need to continue to support young people with social phobia following treatment for SM as there was evidence for the continued diagnosis of a social phobia in 23% of their participants. This research team were also the sole reporters of the child's perspective of their speaking behaviour.

Limitations and areas for future research

Whilst the ADIS was commonly used across studies to determine a baseline perspective on symptom severity, the actual data was not always reported.

Commonly reporting this data in SM research would increase the ability to compare research outcomes and postulate about effectiveness with a clearer knowledge on what level of symptomology effective outcomes were linked to.

In line with previous reviews it is still not possible to use the data collected to elicit exactly what elements of practice are effective, leaving this as an area for future research.

A suggestion for future research in this area is that it becomes the standard to report the school subsection data of the SMQ with the component elements (Table 7). This would facilitate professionals working in and with school systems to assess the domains of speaking behaviour that are most improved by the various interventions.

This review found that school presentation of SM is improved through intervention.

References and appendices.

Abrami, P. C., Borokhovski, E., Bernard, R. M., Wade, C. A., Tamim, R., Persson, T., Bethel, E. C., Hanz, K., & Surkes, M. A. (2010). Issues in conducting and disseminating brief reviews of evidence. *Evidence and Policy*, 6(3), 371–389. <https://doi.org/10.1332/174426410X524866>

American Psychiatric Association, issuing body. (2013). *Diagnostic and statistical manual of mental disorders : DSM-5*. (5th ed.).

Barker, C., Pistrang, N., & Elliott, R. R. (2005). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners: Second Edition*. In *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners: Second Edition* (3rd ed.). Chapter 8, ReWiley-Blackwell. <https://doi.org/10.1002/0470013435>

Bergman, L. R., Keller, M. L., Piacentini, J., & Bergman, A. J. (2008). The Development and Psychometric Properties of the Selective mutism Questionnaire. *Journal of Clinical Child & Adolescent Psychology*, 37(2), 456–464. <https://doi.org/10.1080/15374410801955805>

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. *Journal of Child Psychology and Psychiatry*, 47(11), 1085–1097. <https://doi.org/10.1111/j.1469-7610.2006.01662.x>

Cohen, J. (1992) "A power primer" *Psychological Bulletin* Vol.112. No.1 pp.155-159

doi:[10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155)Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk,

J. (2017). Pre-post effect sizes should be avoided in meta-analyses.

Epidemiology and Psychiatric Sciences, 26(4), 364–368.

<https://doi.org/10.1017/S2045796016000809>

Department of Health and Social Care, & Department for Education. (2018).

Government Response to the Consultation on Transforming Children and Young People 's Mental Health Provision : a Green Paper and Next Steps. July, 48.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/728892/government-response-to-consultation-on-transforming-children-and-young-peoples-mental-health.pdf

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52(6), 377–384. <https://doi.org/10.1136/jech.52.6.377>

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson assessments.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
<https://doi.org/10.3758/BF03193146>

- 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>
- Hilt, R., Nussbaum, A., & American Psychiatric Association, issuing body. (2016). *DSM-5 pocket guide for child and adolescent mental health / Robert J. Hilt, Abraham M. Nussbaum.* (First ed.) p107.
- Jozefiak T (2011) Inventory of life quality in children and adolescents. Hogrefe Psykologiförlaget AB, Stockholm
- Kern, L., Starosta, K. M., Cook, C. R., Bambara, L. M., & Gresham, F. R. (2007). Functional assessment-based intervention for selective mutism. *Behavioral Disorders*, 32(2), 94–108. <https://doi.org/10.1177/019874290703200203>
- Kovac, L. M., & Furr, J. M. (2019). What Teachers Should Know About Selective mutism in Early Childhood. *Early Childhood Education Journal*, 47(1), 107–114.
<https://doi.org/10.1007/s10643-018-0905-y>
- Kratochwill, T. R., & Shernoff, E. S. (2003). Evidence-Based Practice: Promoting Evidence-Based Interventions in School Psychology. *School Psychology Quarterly*, 18(4), 389–408. <https://doi.org/10.1521/scpq.18.4.389.27000>
- National Insitute for Health and Care Excellence (2013) General Anxiety Disorders (NICE guideline, available at <https://www.nice.org.uk/guidance/cg159/chapter/1-Recommendations#interventions-for-children-and-young-people-with-social-anxiety-disorder-2> (Accessed 1st February, 2021)

OCEBM Levels of Evidence Working Group*. "The Oxford 2011 Levels of Evidence". Oxford Centre for Evidence-Based Medicine.

<http://www.cebm.net/index.aspx?o=5653> accessed 30.01.2021. * OCEBM Table of Evidence Working Group = Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson

O'Reilly, M. O., McNally, D., Sigafos, J., Lancioni, G. E., & Green, V. (2008).

Examination of a Social problem-solving intervention to Treat Selective mutism. *Behavior Modification*, 182–195. <https://doi.org/10.1177/0145445507309018>

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>

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57(7), 527–529.

<https://doi.org/10.1136/jech.57.7.527>

Pickard, H., Rijdsdijk, F., Happé, F. & Mandy, W. (2017). Are Social and

Communication Difficulties a Risk Factor for the Development of Social Anxiety? *Journal of the American Academy of Child & Adolescent Psychiatry*, Volume 56, Issue 4, 344-351.e3. ISSN 0890-8567,

<https://doi.org/10.1016/j.jaac.2017.01.007>.

- Read, J., Sargeant, C., & Wright, S. (2020). What beliefs influence children and young people's attitudes towards the transgender population? *Educational and Child Psychology*, 37(1), 11–36.
- Sharkey, L., Mc Nicholas, 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. M. (1996). *Anxiety Disorders Interview Schedule for DSM-IV*. San Antonio, TX: Graywind.
- Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 380–411.
https://doi.org/10.1207/s15374424jccp3403_2
- Stone, B. P., Kratochwill, T. R., 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>
- Williams, K. T. (2007). *Expressive Vocabulary Test* (2nd ed.). Minneapolis, MN: Pearson Assessments.
- Yeager, D. S., Dahl, R. E., & Dweck, C. S. (2018). Why Interventions to Influence Adolescent Behavior Often Fail but Could Succeed. *Perspectives on Psychological Science*, 13(1), 101–122.
<https://doi.org/10.1177/1745691617722620>

Young Minds (n.d). *Young People Turned Away From Mental Health Support*.

Retrieved 13th February, 2021, from <https://youngminds.org.uk/blog/young-people-turned-away-from-mental-health-support/>

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). <https://doi.org/10.1007/s12310-016-9198-8>

Appendix A

Additional information for database search of Web Of Science.

Table 1a Web of science combined searches.

	Search terms used and results
Search 1	“selective mutism” or “elective mutism” AND school or education AND intervention: 82,571
Search 2	“social emotional” or social or emotional AND “selective mutism” or “elective mutism”: 934,607
Combined 1 and 2	15,552
Search 3	“selective mutism” or “elective mutism” AND “school intervention”: 210
Combined search 1 and 2 with Search 3	35 studies went to checking stage.

Appendix B

Excluded studies

Abstract and full paper exclusion	Reason for exclusion
Kern, L., Starosta, K. M., Cook, C. R., Bambara, L. M., & Gresham, F. R. (2007). Functional assessment-based intervention for selective mutism. <i>Behavioral Disorders</i> , 32(2), 94–108. https://doi.org/10.1177/019874290703200203	ERIC search. No quantifiable outcome measure of pre and post outcomes.
Manassis, K., & Tannock, R. (2008). Comparing interventions for selective mutism: A pilot study. <i>Canadian Journal of Psychiatry</i> , 53(10), 700–703. https://doi.org/10.1177/070674370805301010	Ancestral search. Drug treatment.
Martinez, Y. J., Tannock, R., Manassis, K., Garland, E. J., Clark, S., & McInnes, A. (2015). The Teachers’ Role in the Assessment of Selective mutism and Anxiety Disorders. <i>Canadian Journal of School Psychology</i> , 30(2), 83–101. https://doi.org/10.1177/0829573514566377	Web of Science search. No quantifiable outcome measure of pre and post student data. Reliability and validity assessment of teacher reporting tool.
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. <i>Child and Adolescent Mental Health</i> , 19(3), 192–198. https://doi.org/10.1111/camh.12045	Ancestral search. Data linked to 2015 and 2018 follow up studies.
Oerbeck, B., Stein, M. B., Pripp, A. H., & Kristensen, H. (2015). Selective mutism: follow-up study 1 year after end of treatment. <i>European Child and Adolescent Psychiatry</i> , 24(7), 757–766. https://doi.org/10.1007/s00787-014-0620-1	Web of science. Data linked to 2014 and 2018 follow up studies. (2018 selected as it had not been used in previous published reviews).
Oerbeck, B., Overgaard, K. R., Bergman, R. L., Pripp, A. H., & Kristensen, H. (2020). The Selective mutism Questionnaire: Data from typically developing children and children with selective mutism. <i>Clinical Child Psychology and Psychiatry</i> , 25(4), 754–765. https://doi.org/10.1177/1359104520914695	Ancestral search. Study of typically developing peers for SMQ development.
Reilly, M. O., McNally, D., Sigafos, J., Lancioni, G. E., & Green, V. (2008). Examination of a Social to Treat Selective mutism. <i>Behavior Modification</i> , 182–195.	ERIC search. 2 cases described.

<p>Remschmidt, H., Poller, M., Herpertz-Dahlmann, B., Hennighausen, K., & Gutenbrunner, C. (2001). A follow-up study of 45 patients with elective mutism. <i>European Archives of Psychiatry and Clinical Neuroscience</i>, 251(6), 284–296. https://doi.org/10.1007/PL00007547</p>	<p>Web of Science search. 15 year follow up study with a clinical focus.</p>
<p>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. (2020). Effectiveness of a behavioral treatment protocol for selective mutism in children: Design of a randomized controlled trial. <i>Contemporary Clinical Trials Communications</i>, 19. https://doi.org/10.1016/j.conctc.2020.100644</p>	<p>Ancestral search. Outline of prospective study design with no data presented. Future reviewers/researchers would be advised to read.</p>
<p>Sharkey, L., Mc Nicholas, F., Barry, E., Begley, M., & Ahern, S. (2008). Group therapy for selective mutism - A parents' and children's treatment group. <i>Journal of Behavior Therapy and Experimental Psychiatry</i>, 39(4), 538–545. https://doi.org/10.1016/j.jbtep.2007.12.002</p>	<p>Web of Science search. Case study data.</p>
<p>Stone, B. P., Kratochwill, T. R., Sladeczek, I., & Serlin, R. C. (2002). Treatment of selective mutism: A best-evidence synthesis. In <i>School Psychology Quarterly</i> (Vol. 17, Issue 2, pp. 168–190). https://doi.org/10.1521/scpq.17.2.168.20857</p>	<p>Web of Science. Review paper.</p>

Appendix C

Weight of evidence A.

The Downs and Black (1998) checklist for the assessment of methodological quality adapted by Read, Sargeant and Wright (2020) was used in this review. An example completed checklist is shown below. WoE A assessments were made for each study.

Where power was not stated in the paper Faul et al., (2007) G*Power software was used to assess power.

Catchpole, R., Young, A., Baer, S., & Salih, T. (2019). Examining a novel, parent child interaction therapy-informed, behavioral treatment of selective mutism. *Journal of Anxiety Disorders*, 66. <https://doi.org/10.1016/j.janxdis.2019.102112>

Question	Descriptor	Score
1. Focus on a specific, well-defined problem, construct or population?	YES 1 NO 0	1
2. Is the hypothesis/aim/objective of the study clearly described?	YES 1 NO 0	1
3. Are the main outcomes to be measured clearly described in the introduction or methods section? <i>If the main outcomes are first mentioned in the results section, the question should be answered no.</i>	YES 1 NO 0	1
4. Are the characteristics of the participants included in the study clearly described?	YES 1 NO 0	1
5. Were the participants in the study representative of the entire population from which they were recruited? <i>The study must identify the source population for patients and describe how the patients were selected. Participants would be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine.</i>	YES 1 NO 0 UNABLE TO DETERMINE 0	1
6. Were the main outcome measures used accurate (valid and reliable)? <i>For studies where the outcomes measures are clearly</i>	YES 1 NO 0 UNABLE TO DETERMINE 0	1

Question	Descriptor	Score
<i>described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered yes.</i>		
7. Were the statistical tests used to assess the main outcomes appropriate? <i>The statistical tests used must be appropriate to the data. For example, non-parametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described it must be assumed that the estimates used were appropriate and the question should be answered yes.</i>	YES 1 NO 0 UNABLE TO DETERMINE 0	1
8. Are the main findings of the study clearly described? <i>Simple outcome data should be reported for all major findings so that the reader can check the major analyses and conclusions.</i>	YES 1 NO 0	1
9. Does the study provide estimates of the random variability in the data for the main outcomes? <i>In non-normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.</i>	YES 1 NO 0	1
10. Have actual probability values been reported for the main outcomes except where the probability value is less than 0.001?	YES 1 NO 0	1
11. If any of the results of the study were based on 'data dredging' was this made clear? <i>Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.</i>	YES 1 NO 0 UNABLE TO DETERMINE 0	1
12. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than five per cent?	YES 1 NO 0	1
Total:		12
	9-12	High
	5-8	Medium
	0-4	Low

Appendix D

Weight of evidence B.

The Oxford Levels of Evidence Working group (2011) produced a table suggesting that research questions asking about how helpful an intervention is should use the following 5 categories of assessment for allocating a score based on the type of study. Researchers can make use of the Oxford Levels of Evidence to discern which part of the table best suits their research question focus. Below is an extract of the chart seen as suitable for this review: Does this intervention help? These values are tabulated in the main body of the paper (Table 4) for each study.

“Does this intervention help? (treatment benefits)”

	Systematic review of randomized trials or n-of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study	Case-series, case-control studies, or historic	Mechanism based reasoning
	5/5	4/5	3/5	2/5	1/5
WoE B	high	Medium/high	Medium	Medium/low	low
		Cornacchio et al., 2019.	Catchpole et al., 2019 Klein et al., 2017 Lang et al., 2016 Oerbeck et al., 2018.		

Appendix E

Weight of evidence C.

Assessment for WoE C was for relevance of the study to the review question and the context of how and where the research had taken place.

Type of sample: all studies included school age children as part of the inclusion criteria and were recruited through specialist clinics and health care provisions.

Issues of validity were assessed as part of the WoE A evaluation and all identified studies were using a common outcome measure (SMQ). A key point of difference that remained to be assessed was the issue of reliability of data reported for the school context, was school context data reported by home or school?

<p>Low 1</p>	<p>The paper is about non-pharmaceutical treatment for selective mutism but is not directly drawn from school context and/or has no outcome measures linked to school, such as teacher reporting SMQ.</p>
<p>Medium 2</p>	<p>The paper describes a non-pharmaceutical treatment for selective mutism that could be adopted for school use and/or reports on outcomes drawn from the school context.</p>
<p>High 3</p>	<p>The paper describes a non-pharmaceutical treatment for selective mutism that took place in a school and reports on school-relevant outcomes in speech and teacher reported outcomes.</p>

Appendix F

WoE D: Overall assessment of weight of evidence.

Studies were allocated a numerical value for each WoE component. Each of these component values were then added to create a score out of 11 and this value was used to calculate a percentage (Table 5). The table below shows how the percentages were allocated an overall WoE label to mirror the three-level evaluation of the individual WoE A, B and C, and were divided into three values to match. It is acknowledged that whilst the studies identified in this review do not fall across the whole value range, it may be that future updates to this review may need to make full use of the categories.

WoE A-C totals	WoE D	Percentage score of each study
0-3	Low	33% or below
4-7	Medium	34% to 66%
8-11	High	67% to 100%

Appendix G

Table 7.

Findings extracted from Klein et al. (2016) showing changes over time on the school-specific sub sections of the SMQ

	Pre-treatment	End of treatment	15 weeks after start of treatment	Effect sizes ^a Partial eta squared (η^2p)
Speaks to most peers	12.2%	36.3%	36.4%	.35 large
Speaks to selected peers	33.3%	42.4%	54.4%	.27 Large
Answers teacher when called	24.25	42.4%	45.5%	.27 Large
Asks teacher questions	6.1%	24.3%	27.3%	.30 large
Speaks to most staff/pupils	6.1%	30.3%	24.2%	.31 Large
Speaks to group in front of class	6%	18.2%	27.3%	.23 Large

^aEffect size descriptors from Cohen, 1992.