

## ***Case Study 1: An Evidence-Based Practice Review Report***

### ***Are behavioural treatments for selective mutism effective at increasing verbal behaviour for older children and adolescents?***

#### **Summary**

According to the American Psychiatric Association (APA) selective mutism is a rare anxiety disorder, characterised by failure to speak in some contexts but not others (APA, 2013). Behavioural treatments, such as modelling and stimulus fading, are recommended for younger children, but there is no advice known to the reviewer for how to treat older children and adolescents.

This systematic literature review aims to determine the efficacy of behavioural interventions for older children and adolescents with selective mutism, by examining whether they lead to an increase in verbal behaviour. The behavioural interventions reviewed included reinforcement, contingency management, stimulus fading, gradual exposure and video self-modelling. The review finds that, while some participants made promising progress, the overall lack of convincing evidence is due to variation in measurement, insufficient methodological rigour in some cases, and failure to generalise results from clinic-based interventions to school.

## **Introduction**

### **Selective Mutism**

Selective mutism describes when children and young people who have learned to speak in the typical way persistently do not speak in some contexts but can still speak in others. Commonly, children who are selectively mute speak comfortably at home but remain silent in school (Frederickson & Cline, 2015). Selective mutism tends to develop when children begin preschool or primary school, and has been reported to be “strongly resistant to traditional therapeutic strategies and classroom behaviour management strategies” (Frederickson & Cline, 2015, p. 279). The prevalence is rare, affecting less than one percent of children, and even fewer adolescents and adults (Sharp, Sherman & Gross, 2007).

To meet the diagnostic criteria for selective mutism, the child or young person should show a failure to speak in one setting, which significantly impacts their functioning, and persists for at least one month (APA, 2013). There is “substantial evidence” of an association between selective mutism and anxiety-based conditions (Viana, Beidel & Rabian, 2009, p.62) and it is now classified as an anxiety disorder (APA, 2013). There is mixed evidence for oppositional or externalising behaviour in children with selective mutism (Viana et al., 2009), although Diliberto and Kearney (2015) identified two distinct profiles of children with selective mutism: oppositional and anxious. Viana et al.’s (2009) review also reported comorbidity with developmental delays and enuresis.

Psychological explanations for selective mutism include psychodynamic (unresolved conflict), behavioural (maladaptive reinforcement patterns) and family systems theories (Viana et al., 2009). Selective mutism has also been connected to traumatic experiences and a genetic vulnerability to anxiety (Viana et al., 2009).

### **The Need for Selective Mutism Intervention in Schools**

Facon, Sahiri and Riviere (2008) emphasise that selective mutism does not tend to improve without targeted intervention. As school tends to be the setting in which children with selective mutism do not speak, this is a natural context for intervention. Not only does this provide the greatest opportunity for generalisation of the child or young person's new skills to school contexts, it also provides opportunities to bring about changes in school staff's interactions with the child or young person.

### **The Need for Selective Mutism Intervention for Older Children and Adolescents**

A best evidence synthesis (Stone, Kratochwill, Sladeczek & Serlin, 2002) found that the duration of selectively mute behaviour was one of the largest predictors of effect size in behavioural treatments for selective mutism; this means that the longer it has persisted, the harder it is to treat. While selective mutism is more common in younger children, it does sometimes persist into adolescence. By this time, the young person has had several years of reinforcement and the behaviour is often entrenched. For example, teachers tend to stop asking questions of pupils who do not talk. Christon et al. (2012) adds that when a selectively mute pupil avoids an uncomfortable situation by

not talking, the subsequent decrease in anxiety acts as further reinforcement for this behaviour.

### **Behavioural Treatments for Selective Mutism**

Behavioural therapies stem from a Behavioural Psychology approach. These include a range of techniques which are based on the principles of conditioning and reinforcement. Cohan, Chavira and Stein (2006) explain that behavioural models understand selective mutism as a learned behaviour, either in order to avoid anxiety or gain “attention” (p.1087).

In Stone et al.’s (2002) best evidence synthesis, some of the elements of effective treatments were positive reinforcement (including social reinforcement), contingency management, stimulus fading, shaping (with tangible reinforcements), and modelling. No differences in effectiveness were found between different behavioural approaches.

*Reinforcement* refers to any stimulus which increases the response; this can be positive (i.e. adding something desirable) or negative (i.e. removing something undesirable).

*Contingency management* and *shaping* both involve giving the young person rewards for successive approximations towards a desired behaviour (Kehle, Bray, Byer-Alcorace, Theodore & Kovac, 2012). Bunnell, Procci, Beidel and Bowers (2015) describe that the difference between contingency management and shaping is simply that with contingency management a new behaviour is not being learned (the young person already knows *how* to speak).

*Stimulus fading* involves starting in a context where the pupil is comfortable speaking with reinforcement (e.g. a small resource room) and gradually

changing the environment (e.g. to a larger room and then the classroom) while reducing the number of prompts allowed for reinforcement of speaking (Kehle et al., 2012).

*Gradual Exposure* also begins in a context where the pupil is comfortable speaking, such as in a small room with a parent or sibling, and involves allowing the child to acclimatise to anxious feelings by gradually adding new people in a non-threatening way (Kehle et al., 2012). *Intensive exposure* (or flooding) involves exposing the pupil to the feared stimulus (e.g. crowds of people) until their anxiety reduces (Bunnell & Beidel, 2013).

*Video self-modelling* involves showing the pupil edited footage of themselves that shows only the desired behaviour. For selective mutism, this might involve videoing the pupil talking to a parent in the classroom and editing it so that it looks like they are talking to their teacher (Kehle et al., 2012).

### **Evidence for Behavioural Treatments with Older Children and Adolescents**

Although a number of reviews support the use of behavioural treatments for selective mutism (e.g. Cohan et al., 2006; Stone et al., 2002; Viana et al., 2009), these have all focused on treatments for children. Behavioural approaches have been effective with adolescents for other conditions, for example anxiety (Jones, Dickter, Beard, Perales & Bunge, 2016).

### **Relevance to Educational Psychology Practice and Rationale for Review**

Imich (1998) calculated that on average, educational psychologists in the United Kingdom will work with a selective mutism case every five years (as cited in Cleave, 2009). Due to the rarity of selective mutism, schools often do

not understand the function of the behaviour (Kehle et al., 2012) and school staff can find themselves feeling hopeless or frustrated (Cleave, 2009). Educational psychologists are well placed to explore the multiple factors influencing a pupil's selective mutism behaviour and to recommend individualised support in collaboration with parents and teachers (Cleave, 2009).

### **Review Question**

Are behavioural treatments for selective mutism effective at increasing verbal behaviour for older children and adolescents?

### **Critical Review of the Evidence**

#### **Literature Search and Screening Strategy**

In January 2018, searches for peer-reviewed journals were carried out in the electronic databases PsycINFO, BEI (British Educational Index) and ERIC, chosen to cover the literature in educational and psychological research.

The following search criteria were used:

(behav\*)

and (intervention or program\* or treatment or therapy)

and ("selective mutism" or "elective mutism"))

Fields: title, abstract, heading word, table of contents, key concepts, original title, tests & measures.

Limits: since 2005, peer reviewed journals

As shown in Figure 1, this search yielded 99 studies: 78 from PsycINFO; 20 from ERIC and 1 from BEI. The abstracts and titles were screened together to eliminate studies which did not meet the inclusion criteria (see Table 1). Ten articles were screened by reading the full text. Excluded studies are listed in Appendix A. Six articles were included in the review (see Table 2). Each paper details a single-subject study with only one participant.

For narrative simplicity, the papers may be referred to by their participants' pseudonyms.

Table 1:

Criteria for Inclusion in Systematic Literature Review

<b>Criteria</b>	<b>Inclusion</b>	<b>Exclusion</b>	<b>Rationale</b>
<b>Intervention</b>	Intervention includes behavioural therapies	Intervention does not include a behavioural component	This review aims to explore the effect of behavioural therapies
<b>Outcome Variables</b>	Study includes quantitative data on verbal behaviour	Study does not include quantitative data on verbal behaviour	This review aims to focus on selective mutism
<b>Population</b>	9 – 17 year olds	Study includes adults or children younger than 9	Selective mutism is less prevalent but more difficult to treat in this age group
<b>Type of publication</b>	Peer reviewed journal	Not published in a peer reviewed journal	To ensure methodological quality
<b>Study design</b>	Empirical with at least single subject research design	Designs that are not empirical and/or not quantitative.	To facilitate comparison across studies
<b>Date</b>	From 2005 until January 2018	Prior to 2005 or after January 2018	To ensure studies are recent. Search completed in January 2018.
<b>Language</b>	Study published in English	Study not published in English	To ensure reviewer understanding.
<b>Database</b>	Indexed in PsycINFO, ERIC or BEI	Not indexed in PsycINFO, ERIC or BEI	These are relevant databases to education and psychology.

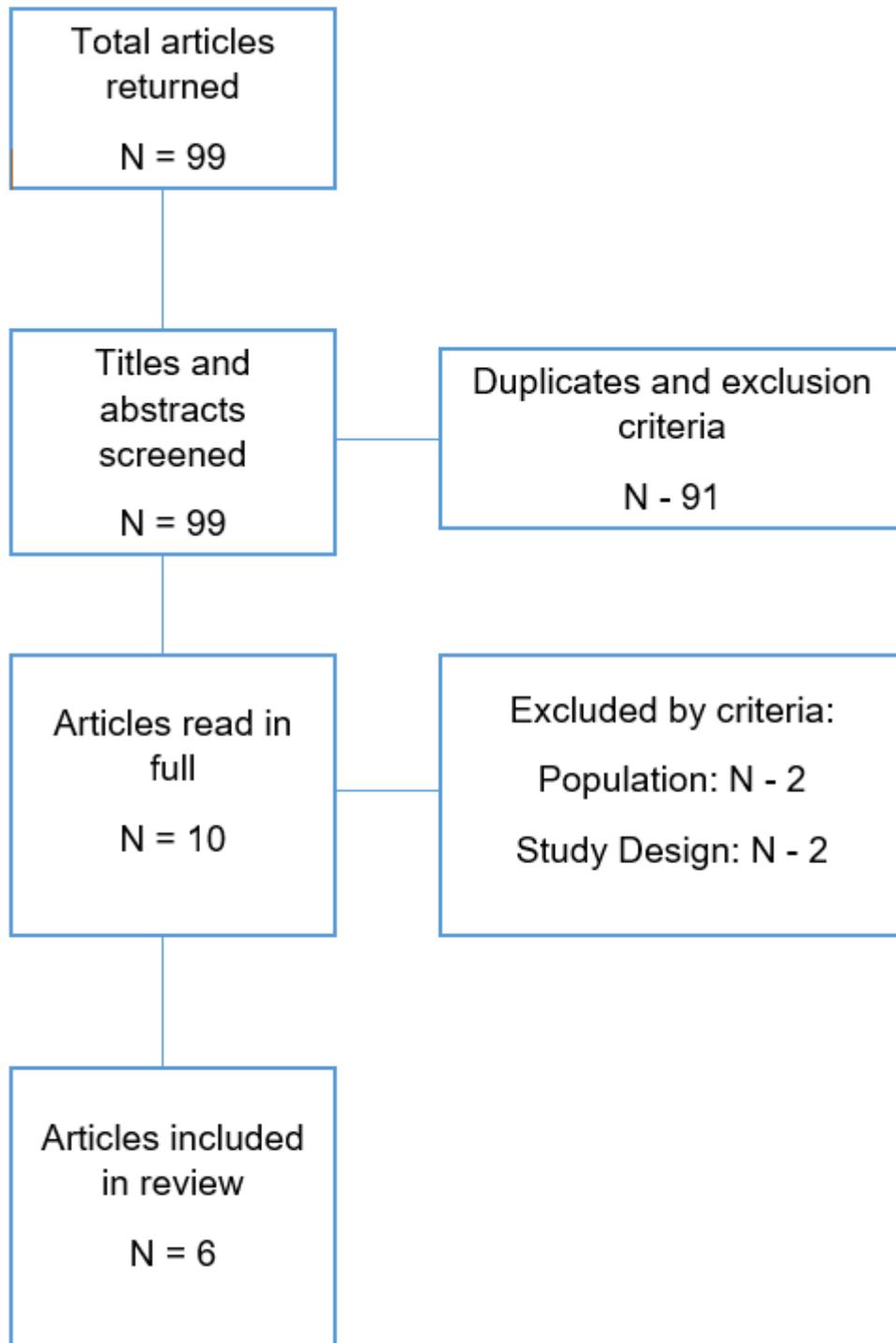


Figure 1: Literature Search Process

Table 2:

Details of Included Studies

Participant pseudonym (ppt.)	Full reference of study
Luke	Beare, P., Torgerson, C., & Creviston, C. (2008). Increasing verbal behavior of a student who is selectively mute. <i>Journal of Emotional and Behavioral Disorders</i> , 16(4), 248-255.
Nina	Bunnell, B. E., & Beidel, D. C. (2013). Incorporating technology into the treatment of a 17-year-old female with selective mutism. <i>Clinical Case Studies</i> , 12(4), 291-306.
Ava	Christon, L. M., Robinson, E. M., Arnold, C. C., Lund, H. G., Vrana, S. R., & Southam-Gerow, M. A. (2012). Modular cognitive-behavioral treatment of an adolescent female with selective mutism and social phobia: A case study. <i>Clinical Case Studies</i> , 11(6), 474-491.
Sami	Facon, B., Sahiri, S., & Rivière, V. (2008). A controlled single-case treatment of severe long-term selective mutism in a child with mental retardation. <i>Behavior Therapy</i> , 39(4), 313-321.
LM	Fisak Jr, B. J., Oliveros, A., & Ehrenreich, J. T. (2006). Assessment and behavioral treatment of selective mutism. <i>Clinical Case Studies</i> , 5(5), 382-402.
Leslie	Lang, R., Regester, A., Mulloy, A., Rispoli, M., & Botout, A. (2011). Behavioral intervention to treat selective mutism across multiple social situations and community settings. <i>Journal of Applied Behavior Analysis</i> , 44(3), 623-628.

### Weight of Evidence

Using Gough’s (2007) Weight of Evidence Framework (WOE) each of the six selected papers was evaluated in three areas: methodological quality (WOE A); methodological relevance (WOE B); and topic relevance (WOE C). For each paper, the three scores were averaged to determine Overall Weight of Evidence (WOE D). Table 3 outlines the WOE scores for each paper.

The papers were evaluated for methodological quality (WOE A) using a coding protocol devised from Horner et al.'s (2005) guidelines for evaluating single-subject research (see Appendix C).

Criteria for methodological relevance and topic relevance (WOE B and C) were devised to assess how well the design and topic of each paper can answer the review question (see Appendix C).

Table 3:  
Weight of Evidence Ratings for Included Studies

Authors	WOE A Methodological Quality	WOE B Methodological Relevance	WOE C Relevance to Review Question	WOE D Overall Weight of Evidence
Facon et al.	2.57	2.00	2.60	2.39 (medium)
Beare et al.	2.43	2.00	2.60	2.34 (medium)
Christon et al.	1.43	1.00	2.30	1.58 (medium)
Bunnell & Beidel	2.00	2.00	2.70	2.23 (medium)
Lang et al.	2.43	2.00	2.40	2.28 (medium)
Fisak et al.	1.00	1.00	1.70	1.23 (low)

Where: Low 1.4 or less, Medium 1.5-2.4, High 2.5 or above

### Participants

Each reviewed paper described its single participant in detail, earning high WOE A ratings for this criteria. As only six participants are covered in this

review, there is scope to describe them in more detail. Table 4 outlines some of their demographic characteristics. As shown, three boys and three girls between 9 and 17 years were included (mean age 12.5 years). All but one of the studies described significant additional areas of concern; notably four participants had comorbid anxiety (Sami, Ava, Nina and LM) and another displayed “emotional and behavioural difficulties” (Luke, p. 249).

*Table 4:*

*Participant Characteristics*

<b>Ppt.</b>	<b>Gender</b>	<b>Age</b>	<b>Additional Diagnosis/Difficulty</b>	<b>Languages</b>	<b>Location</b>
Sami	Boy	12	"Mental retardation" and anxiety (p. 313)	Arabic and French	France
Luke	Boy	12	Emotional and Behavioural Difficulties	English (presumed)	USA
Ava	Girl	15	Social phobia and panic attacks. Separation anxiety. Depression. School refusal 2/week.	English	USA
Nina	Girl	17	Depression and anxiety disorder	English (presumed)	USA
Leslie	Girl	9	None described	English (presumed)	USA
LM	Boy	10	Anxiety	Spanish and English	USA

All but one of the studies was conducted in the United States of America, which has a similar school system to the United Kingdom. Table 4 shows that two of the participants were bilingual, speaking another language at home with their family (Sami and LM). Although neither paper discusses this, it could be that these participants’ selective mutism is restricted to their additional language. It

is common for children learning an additional language to experience a silent period until they feel comfortable talking in their new language (Toppelberg et al., 2005). The diagnostic criteria for selective mutism insists that it cannot be diagnosed if a child is unfamiliar or uncomfortable with the language (APA, 2013). Sami's learning difficulty or LM's anxiety may have impacted their acquisition or comfort with speaking their additional language, and as such selective mutism may not be an appropriate diagnosis for them. Toppelberg et al. (2005) argue that the risks of inappropriate diagnosis in bilingual children, including pharmaceutical intervention, necessitate caution on clinicians' part when diagnosing selective mutism in bilingual children. Further information on their proficiency in the additional language or whether they show selective mutism across languages should be provided.

Due to the heterogeneity across young people who show selective mutism, Table 5 outlines some characteristics of each participants' selective mutism in order to qualitatively explore them further. It shows that most of the participants developed selective mutism either on entering Kindergarten or changing to a new school or new area (and country, in Sami's case). In Ava's case, selective mutism became apparent after her brother was hospitalised for a traumatic car accident and worsened after the birth of her younger sister. During the time her brother was hospitalised, she stayed with friends; afterwards her brother had an intellectual disability and her parents devoted much of their time to caring for him and the younger sister. Unusually, her symptoms were also apparent at home.

For the two oldest participants, Ava and Nina, previous treatment attempts were unsuccessful. This is consistent with literature describing selective

mutism’s resistance to treatment (Frederickson & Cline, 2015). Interestingly, pharmaceutical intervention was prescribed to both of these participants at around the same age (15 for Ava and 16 for Nina). Previous treatment was not described for the other participants, suggesting that within this group, there may have been no intervention attempted before age nine, despite symptoms commonly being apparent by around age four or five.

Table 5:

Features of Participants' Selective Mutism

<b>Ppt.</b>	<b>Onset of Selective Mutism</b>	<b>Speaks well ...</b>	<b>Difficulty speaking ...</b>	<b>Previous Treatment</b>
Sami	After moving to France from Morocco aged 8 and placed in integration class	Mother and family (in Arabic)	Peers and staff at school (daycare centre)	Not described
Luke	Kindergarten	Information not provided	School	Multiple setting changes with “aversive intervention” (p.250)
Ava	Since she was 4 years old. Following hospitalisation of brother for 3 months after car injury. Worsened after birth of sister.	Minimally at home: never about feelings or thoughts	Any other contexts including school. Rarely interacted with siblings.	"talk therapy" (p.488) in 5th Grade. Began taking SSRIs during intervention.
Nina	Since age 4 on entering Kindergarten.	Immediate family only. Initially whispered at school.	Everyone except immediate family.	Play therapy (ages 9 and 11). Fluoxetine (age 16)

Leslie	Not described (at least one year previous)	Clearly and frequently at home	Community settings, summer camp, swimming lessons, restaurants, meeting new adults or playing with a new peer.	Not described
LM	Kindergarten: became more apparent age 9 when moved	At home (in "baby talk" to mother, p. 396) (possibly in Spanish)	School, peers, neighbours	Not described

### Study Design

This review includes four single subject designs (Sami, Luke, Leslie and Nina) and two case studies (LM and Ava). Single subject designs are a valid methodology to evaluate the effectiveness of interventions (Horner, 2005; Kratochwill et al., 2010), particularly when the phenomenon studied is rare. Ava and LM’s interventions were described with case studies, which provide rich contextual information but are less appropriate for evaluation, and as such were weighted lower for methodological relevance (WOE B).

### Intervention Content

The behavioural components of each intervention are listed in Table 6. The interventions are described in more detail in Appendix B. Each of the interventions showed some degree of individualisation.

Table 6:

*Behavioural Techniques Used in Included Studies*

<b>Ppt.</b>	<b>Behavioural Techniques</b>
Sami	Shaping and stimulus fading
Luke	Positive reinforcement and stimulus fading
Ava	Exposure-based therapy and contingency management
Nina	Contingency management, Intensive exposure, graduated exposure, social skills
Leslie	Role play, video self modelling, reinforcement
LM	Social skills training, exposure, modelling, shaping, contingency management

Ava and Nina also received psychoeducation as part of their treatment. These were the oldest two participants (15 and 17). Viana et al.'s (2009) review cautions that there is limited evidence for cognitive behavioural approaches on selective mutism, however this is likely because it is more commonly found in much younger children who would find it harder to access cognitive therapy. Nina and Ava were also the only two taking medication for their symptoms. This is likely because their symptoms were so persistent that the risks of pharmaceutical intervention were outweighed.

Interestingly, while the loudness of Sami's verbalisations was measured with a decibel meter, he was never shown the values. Nina's intervention, on the other hand, included the use of an iPad Decibel Meter app as a visual reinforcement to shape her to speak louder. The intervention included other apps that respond to sound, for example a candle that "blows out" when the iPad registers a sound. The researchers claim that the use of apps in Nina's intervention gamified the tasks, retained her engagement, and paired her vocalisations with positive emotions. In a follow-up chapter in a book, they also assert that distracting from the pressure to speak supports the therapeutic

relationship (Bunnell et al., 2015). As many schools in the UK already have iPads, this could be an inexpensive and engaging way to apply behavioural techniques with young people who have selective mutism.

### **Intervention Duration and Intensity**

Information about the duration and intensity of the interventions was inconsistent. Luke and Nina had multiple sessions per week, while LM had sessions weekly and Ava had about 3 sessions per month.

The average number of sessions was 41, but this figure includes Sami and Leslie's shorter sessions: Leslie had 13 sessions between 2 and 45 minutes (community sessions were shorter), though it was not described how frequent these were; Sami had four 15 minute sessions per week for about 5 months.

The older participants' interventions were longer; Ava's intervention was delivered over 21 months and Nina's for approximately 8 months.

### **Intervention Setting**

Each of the papers reviewed included more than one intervention setting. Sami and Luke both began sessions in a small classroom and gradually moved to their normal classroom. Leslie's sessions began at home and were assessed in community settings such as restaurants. The other participants' interventions began in clinic settings and moved to community settings, including LM whose class teacher had limited involvement via email. Borger, Bartley, Armstrong, Kaatz and Benson (2007) emphasise the importance of strong collaboration between parents, classroom teachers, school administration, speech and language therapists and other professionals in addressing selective mutism. The features of the clinical setting where LM and

Nina's interventions were held was not described, so they were scored lower for WOE A.

### **Baseline and Experimental Control**

Most of the studies reviewed included a multiple baseline stage, as recommended by Horner et al. (2005) and continued measurements throughout the intervention stage. The papers detailing Ava and LM's interventions did not include a baseline stage, and instead began measurements at the same time as the intervention. Neither of these two studies described with adequate detail whether systematic manipulations of the independent variable (the intervention) corresponded with changes to the dependent variable (verbal behaviour). As a result, they were unable to demonstrate sufficient experimental control (Horner et al., 2005) and were scored lower for WOE A in this domain. Additionally, as Ava began taking Selective Serotonin Reuptake Inhibitors during the course of her behavioural treatment, it is impossible to attribute improvements to behavioural treatments alone. As a result of these factors, Ava's case study was weighted lower for WOE A.

Maturation is unlikely to have affected the participants as their selective mutism had persisted for between one and 13 years without improvement.

### **Measures**

Each of the reviewed studies described their dependent variables with replicable precision, and each of these was quantifiable, earning higher WOE A scores in this domain as a result. Although a number of the studies also included measures of anxiety and depression, the current review is limited to

verbal behaviour. For Sami, this was achieved through measuring the loudness of this verbalisations with a decibel meter (decibels measure the intensity of sound on a logarithmic scale). The use of an objective measure and device eliminates the need for inter-rater reliability (presuming the device was working reliably). An adapted rating scale for elective mutism was used before, but not after Sami's intervention.

Luke, Ava and LM's progress was measured by counting how often they spoke in sessions: for Luke this was the number of words spoken in session and the rate of words spoken per minute in session; for Ava this was the number of words spoken in session; and for LM it was the number of verbalisations in session (where a verbalisation was counted as any utterance including at least one word). Leslie's progress was determined by counting the number of responses, initiations and breakdowns per session in the community (e.g. ordering food in a restaurant).

A standardised Selective Mutism Questionnaire was also used with Ava, but it was not standardised for her age group and the validity was questionable due to items designed for younger children (e.g. talking to babysitters).

Nina's vocal behaviour was self-reported; she kept a diary of the frequency of her speech in school and outside of school. This is a subjective measure, vulnerable to bias or inaccuracy. Showing faith in Nina's capacity to self-report as a 17 year old preparing for university may have been an appropriate therapeutic choice, but a more objective measure of her verbal behaviour would be more appropriate to evaluate the effectiveness of her intervention on verbal behaviour (though the researchers did include pre-post standardised

measures of depression and anxiety). As a result, measurement of Nina’s dependent variable was not rated as valid or replicable for WOE A.

Overall, measurement was unreliable. Interrater reliability is crucial for researcher-devised measures and observations (Horner et al., 2005) though Ava, Nina and LM did not achieve this, and as such were rated lower for WOE A in this domain

**Outcomes**

Table 7 shows the Percentage Non-Overlapping Data effect sizes for each relevant outcome. This refers to the percentage of data points after the baseline that exceed the highest baseline measurement. It is calculated by totalling the number of measurements that meet this criteria and dividing this figure by the total number of post-baseline measurements. This method is appropriate for interpreting visual data provided by single-case experimental designs.

*Table 7:  
Effect Sizes for Verbal Behaviour*

<b>Ppt.</b>	<b>Variable</b>	<b>Baseline</b>	<b>Outcome</b>	<b>Effect Size</b>
Sami	Loudness in dB	42 dB	72-72 dB	98.84% PND
Luke	Number of words spoken in resource room	0 words per session (wps)	25-50 wps	100% PND
	Number of words spoken in study room	0 – 3 wps	20-40 wps	100% PND
	Number of words spoken in classroom	0-3 wps	30-40 wps	100% PND
	Rate of words spoken (per minute) in resource room	0 words per minute (wpm)	3-4 wpm	100% PND

	Rate of words spoken (per minute) in study room	0 wpm	3-4.5 wpm	100% PND
	Rate of words spoken (per minute) in classroom	0-1 wpm	3-4.5 wpm	100% PND
Ava	Selective Mutism Questionnaire	No baseline phase First score = 13	35	Could not be calculated
	Number of words spoken in sessions.	No baseline phase First score = 0	160	Could not be calculated
Nina	Self-reported frequency of speech at school	0	0-5	8% PND
	Self-reported frequency of speech outside home	0-1	0-20	34.08% PND
Leslie	Responses when ordering in a restaurant	0	3-7	100% PND
	Initiations when ordering in a restaurant	0	0-1	40% PND
	Breakdowns when ordering in a restaurant	2	0-1	100% PND
	Responses when meeting new adults	0	4-8	100% PND
	Initiations when meeting new adults	0	0-5	80% PND
	Breakdowns when meeting new adults	2	0-1	100% PND
	Responses when playing with peers	0	15-18	100% PND
	Initiations when playing with peers	0	4-8	100% PND
	Breakdowns when playing with peers	2	0-3	25% PND
LM	Verbalisations during sessions	No baseline phase First score = approx.. 5 wps	Above 40 wps	Could not be calculated

Note:

*dB = decibels*

*wps = words per session*

*wpm = words per minute*

*PND = Percentage Non-overlapping Data*

Comparison of participants' progress is difficult due to the variation in measurements used. While all participants showed some improvement (see Table 7), it is difficult to determine how meaningful these improvements are.

Sami, Luke and Leslie appeared to show substantial improvements in their verbal behaviour. These three studies also had the highest Weight Of Evidence ratings. Sami's speech increased markedly in volume and this was reported to generalise to everyday life in the form of increased interaction with peers and adults. Given that Sami enrolled in a new school before the follow up, these improvements are impressive.

Luke's verbalisations increased considerably across three settings, including the classroom, but this was only measured in response to immediate tangible reward. Unfortunately, by the end of that school year Luke had not yet generalised this to spontaneous speech, but his teachers did report that he benefited from the reinforcements.

Leslie responded to more interactions after her intervention, however her improvements in initiating conversations with adults were less impressive. Importantly, no information is provided on Leslie's behaviour in school.

Ava's score on the Selective Mutism Questionnaire improved but not to the score of a non-clinical population; however since this questionnaire was inappropriate for her age group this may not be meaningful. She also appeared to gradually speak more in sessions throughout the intervention but as there was no baseline phase this cannot be attributed to the intervention. The researchers claim that following treatment she no longer met the diagnostic criteria for selective mutism, however due to lack of experimental control this

cannot be attributed to the intervention alone. Again, it is not described if Ava's improvements generalised to school.

LM's intervention included no baseline, but his speech increased across sessions to 40 verbalisations per session; despite this, his teacher reported that he had made limited progress. This may be due to the limited involvement of the teacher in his clinic-based intervention.

Nina's self-report of her verbal behaviour showed minimal improvement in school but better progress outside of school; however in both contexts there were several days that she recorded zero utterances. Due to the clinical setting of her intervention, and her longstanding victimisation in school, it is understandable that she found it hard to increase her speech there. Follow up sessions and measurements in her new university would have been ideal. Table 7 shows low Percentage of Non Overlapping Data scores for Nina's treatment, but it is important to note that this includes the intensive exposure phase (which was ineffective) for two reasons: one, because this review covers all behavioural approaches; and two because it is impossible to determine if the phases had a cumulative effect on Nina's behaviour.

### **External Validity**

All studies demonstrated low external validity. Single subject designs can have high external validity when experimental effects can be demonstrated across participants, settings and measures (Horner et al., 2005). The three strongest studies (Sami, Luke and Leslie's) were assessed in three contexts (as recommended by Horner et al., 2005), although for Luke and Leslie this was

with researcher-devised measures. It would be feasible to also use multiple validated measures to increase external validity.

### **Social Validity**

Only two of the reviewed studies were judged to be practical and cost effective (Sami and Luke). These interventions were both implemented in the young person's school over a relatively shorter period of time. Luke's intervention was also delivered by staff from his school, further reducing implementation cost. Although Luke's paper scored highly on WOE A, the dependent variable was not judged to be socially important as it only measured Luke's verbal behaviour *in response to immediate reward*.

### **Conclusions and Recommendations**

In conclusion, evidence for the effectiveness of behavioural therapies in increasing verbal behaviour of older children and adolescents remains limited. While some participants made promising progress, the overall lack of convincing evidence is due to variation in measurement, insufficient methodological rigour in some cases, and failure to generalise results from clinic-based interventions to school. In addition, this review may be limited by publication bias as clinicians are unlikely to publish results of interventions which have not been successful (Cohan et al., 2006).

### **Recommendations for Researchers**

Future researchers should be aware that single-subject designs have the potential to answer questions about efficacy, and could use a framework such as Horner et al.'s (2005) guidelines to ensure methodological rigour. Future

researchers should work closely with schools and teachers to ensure effects can be generalised to the classroom, as the only interventions which were able to confirm generalisation of effects to the classroom were school-based. An effective way of doing this appears to be through using three different school contexts.

As the most convincing measures were objective measures of the young person's speech (either through interrater reliability or using a decibel meter), future researchers should strive to include these. Follow-up observations of the young person in their regular classrooms are also needed.

Where participants are bilingual, efforts to establish their proficiency in the additional languages as well as whether their selective mutism is evident in both languages should be described.

### **Recommendations for Educational Psychologists**

While the evidence is limited, this review points to some promising approaches for Educational Psychologists to employ when working with young people with selective mutism. Behavioural interventions, especially when delivered at school, are less intrusive than pharmaceutical intervention, and should be tried as early as possible. Given the austere economic climate, educational psychologists may need to work creatively with the resources that schools already have, such as employing freely available apps for use with existing equipment. The most promising intervention (Sami's), was delivered in school in fifteen minute sessions by school staff. Educational psychologists are well-placed to train support teams in secondary schools to do this, and could extend

this casework to rigorous single-subject design e.g. by establishing multiple baselines and acting as a blind data reviewer for a trained staff member.

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Beare, P., Torgerson, C., & Creviston, C. (2008). Increasing verbal behavior of a student who is selectively mute. *Journal of Emotional and Behavioral Disorders*, 16(4), 248-255.
- Borger, G. W., Bartley, D. L., Armstrong, N., Kaatz, D., & Benson, D. (2007). The importance of a team approach in working effectively with selective mutism: A case study. *TEACHING Exceptional Children Plus*, 4(2), 2-11.
- Bunnell, B. E., & Beidel, D. C. (2013). Incorporating technology into the treatment of a 17-year-old female with selective mutism. *Clinical Case Studies*, 12(4), 291-306.
- Bunnell, B. E., Procci, K., Beidel, D. C., & Bowers, C. A. (2015). Gamification of Therapy: Treating Selective Mutism. In D. Novak, B. Tulu and H. Brendryen (Eds.), *Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice* (pp. 390-410). Pennsylvania: IGI Global.
- Christon, L. M., Robinson, E. M., Arnold, C. C., Lund, H. G., Vrana, S. R., & Southam-Gerow, M. A. (2012). Modular cognitive-behavioral treatment of an adolescent female with selective mutism and social phobia: A case study. *Clinical Case Studies*, 11(6), 474-491.
- Cleave, H. (2009). Too anxious to speak? The implications of current research into Selective Mutism for educational psychology practice. *Educational Psychology in Practice*, 25(3), 233-246.
- 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.
- Diliberto, R. A., & Kearney, C. A. (2016). Anxiety and oppositional behavior profiles among youth with selective mutism. *Journal of Communication Disorders*, 59, 16-23.
- Facon, B., Sahiri, S., & Rivière, V. (2008). A controlled single-case treatment of severe long-term selective mutism in a child with mental retardation. *Behavior Therapy*, 39(4), 313-321.
- Fisak Jr, B. J., Oliveros, A., & Ehrenreich, J. T. (2006). Assessment and behavioral treatment of selective mutism. *Clinical Case Studies*, 5(5), 382-402.
- Frederickson, N., & Cline, T. (2015). *Special Educational Needs, Inclusion and Diversity* (3<sup>rd</sup> ed.). Barkshire: Open University Press.
- Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 37–41.

- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional children*, 71(2), 165-179.
- Jones, M. K., Dickter, B., Beard, C., Perales, R., & Bunge, E. L. (2016). Meta-analysis on cognitive behavioral treatment and behavioral intervention technologies for anxious youth: More than a BIT effective. *Contemporary Behavioral Health Care*, 1, 1-9.
- Kehle, T. J., Bray, M. A., Byer-Alcorace, G. F., Theodore, L. A., & Kovac, L. M. (2012). Augmented self-modeling as an intervention for selective mutism. *Psychology in the Schools*, 49(1), 93-103.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What works clearinghouse*.
- Lang, R., Regeister, A., Mulloy, A., Rispoli, M., & Botout, A. (2011). Behavioral intervention to treat selective mutism across multiple social situations and community settings. *Journal of Applied Behavior Analysis*, 44(3), 623-628.
- Sharp, W. G., Sherman, C., & Gross, A. M. (2007). Selective mutism and anxiety: A review of the current conceptualization of the disorder. *Journal of anxiety disorders*, 21(4), 568-579.
- Stone, B., Kratochwill, T. R., Sladeczek, I., & Serlin, R. C. (2002). Treatment of selective mutism: A best-evidence synthesis. *School Psychology Quarterly*, 17(2), 168-190.
- Toppelberg, C. O., Tabors, P., Coggins, A., Lum, K., Burger, C., & Jellinek, M. S. (2005). Differential diagnosis of selective mutism in bilingual children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(6), 592-595.
- 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.

**Appendix A: Excluded Studies**

Table 8:

References of Excluded Studies and Reason for Exclusion

<b>Study</b>	<b>Reason for Exclusion</b>
Mitchell, A. D., & Kratochwill, T. R. (2013). Treatment of selective mutism: Applications in the clinic and school through conjoint consultation. <i>Journal of Educational and Psychological Consultation</i> , 23(1), 36-62.	Population (children under 9 years)
Turkiewicz, G., Castro, L. L., Morikawa, M., Costa, C. Z. G. D., & Asbahr, F. R. (2008). Selective mutism and the anxiety spectrum: a long-term case report. <i>Revista Brasileira de Psiquiatria</i> , 30(2), 172-173.	Study design (not quantitative)
Vecchio, J., & Kearney, C. A. (2007). Assessment and treatment of a Hispanic youth with selective mutism. <i>Clinical Case Studies</i> , 6(1), 34-43.	Study design (not quantitative)
Vecchio, J., & Kearney, C. A. (2009). Treating youths with selective mutism with an alternating design of exposure-based practice and contingency management. <i>Behavior Therapy</i> , 40(4), 380-392.	Population (children under 9 years)

**Appendix B: Summary of Included Studies**

Table 9: Summary of Included Studies

<b>Authors (Pt.)</b>	<b>Context</b>	<b>Aim</b>	<b>Setting</b>	<b>Details</b>	<b>Intensity</b>	<b>Duration</b>	<b>No. of sessions</b>
Beare et al., 2008 (Luke)	Given “emotional and behavioural difficulties” (p. 249) label. Mainstream class with aid and 30m/day resource support with EBD teacher. Within average range on WISC. Communicated through smiling and gestures.	Increase verbal comm.	3 rooms in school (resource room, study room, classroom)	A-B-B’ multiple baseline. Delivered by EBD teacher. Stimulus fading via changing the setting towards the classroom and fading prompts to obtain speech (from full voice to whisper). Repeated sentences for incentives.	Not reported	5 sessions in resource room, 14 sessions in study room, 27 days in mainstream classroom	46

<p>Bunnell &amp; Beidel, 2013 (Nina)</p>	<p>Bullied by peers for SM – called her “stupid” (p. 2), stole her belongings. Nina communicated via writing. Strong academic achievement. Athletic. Enuresis until 7. Articulation disorder for th, ch, s and r sounds.</p>	<p>Decrease social anxiety and speak during social situations</p>	<p>Clinic and community , one session in high school</p>	<p>Psychoeducation, contingency management, intensive exposure (9 sessions), graduated exposure (6 sessions), social skills (12 sessions). Used iPad app with decibel meter as visual reinforcement.</p>	<p>Average 3 or 4 per week</p>	<p>Approx 8 months( 238 days</p>	<p>27</p>
<p>Christon et al., 2012 (Ava)</p>	<p>Parents spend a lot of time caring for her siblings (one with LD due to traumatic car accident).</p>	<p>Increase audible speech via anxiety reduction</p>	<p>Clinic, then community</p>	<p>Psychoeducation, cognitive and relaxation strategies, exposure to feared situations, social skills, parent training.</p>	<p>About 3 a month</p>	<p>21 months (60 sessions in total)</p>	<p>60</p>
<p>Facon et al., 2008 (Sami)</p>	<p>Bilingual learner (Arabic and French). Born in Morocco, neonatal problems and developmental delay. Attending day-care centre for children with mental retardation</p>	<p>Increase loudness of speech and allow him to talk to peers and staff</p>	<p>School (small room gradually increasing to classroom)</p>	<p>Asked to give verbal answers to questions about interests or pictures. Reinforced at increasing pre-determined decibel levels and sentence lengths. Other persons added.</p>	<p>Four 15-min per week</p>	<p>20 weeks</p>	<p>80</p>

Fisak et al., 2006 (LM)	High anxiety. Parental disagreement/argument about severity of SM. Mother very protective. Father attempting behavioural control.	Increase "social skills" (p. 389)	Clinic and a university classroom (with cooperation via email from teacher)	Adapted SET C social skills training, exposure, modelling, shaping, contingency management and parent training	Weekly	22 sessions	22
Lang et al., 2011 (Leslie)	School context not discussed by researchers.	Speak in social and community settings	Sessions at home and assessments at community settings	Role play and video self modelling of three situations: ordering at a restaurant, meeting a new adult, and playing with a peer.	RP+ VSM 30-45 min. Others 2-10 min.	13 sessions (time period not provided)	13

---

**Appendix C: Weight of Evidence Ratings**

**WOE A Methodological Quality**

Coding Protocol: Quality Indicators within Single Subject Research (adapted from Horner et al., 2005). Coding for each study is outlined in Table 10.

Table 10:

*Completed Methodological Quality Coding Protocol for Each Included Study*

---

	Bunnell & Beidel, 2013 (Nina)
<hr/>	
<b>Description of Participants and Setting</b>	
Participants are described with sufficient detail to allow others to select individuals with similar characteristics; (e.g., age, gender, disability, diagnosis).	✓
The process for selecting participants is described with replicable precision.	✓
Critical features of the physical setting are described with sufficient precision to allow replication.	x
<hr/>	
Total	2
<hr/>	
<b>Dependent Variable</b>	
Dependent variables are described with operational precision.	✓
Each dependent variable is measured with a procedure that generates a quantifiable index.	✓
Measurement of the dependent variable is valid and described with replicable precision.	x
Dependent variables are measured repeatedly over time.	✓

Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA = 80%; Kappa = 60%).

x

*All criteria fulfilled = 3; three or four criteria fulfilled = 2; one or two criteria fulfilled = 1*

---

Total	2
-------	---

---

**Independent Variable**

Independent variable is described with replicable precision. ✓

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

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

---

Total	2
-------	---

---

**Baseline**

The study includes a baseline phase that provides repeated measurement of a dependent variable ✓

The baseline establishes a pattern of responding that can be used to predict the pattern of future performance if introduction or manipulation of the independent variable did not occur. ✓

Baseline conditions are described with replicable precision. x

---

Total	2
-------	---

---

**Experimental Control/internal Validity**

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

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

The results document a pattern that demonstrates experimental control. ✓

---

Total	3
-------	---

---

**External Validity**

Experimental effects are replicated across participants x

Experimental effects are replicated across settings ✓

Experimental effects are replicated across materials	x
Total	1
<b>Social Validity</b>	
The dependent variable is socially important.	✓
The magnitude of change in the dependent variable resulting from the intervention is socially important.	✓
Implementation of the independent variable is practical and cost effective.	x
Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.	x
<i>All criteria fulfilled = 3; two or three criteria fulfilled = 2; one of the criteria fulfilled = 1</i>	
Total	2
<b>Overall Total</b>	
	<b>14</b>

Table 11:

Weight of Evidence A Calculations for each Included Study

<b>Author (Ppt.)</b>	<b>Overall Total</b>	<b>WOE A</b> <i>(average across 7 criteria)</i>
Beare et al., 2008 (Luke)	17	2.43 (medium)
Bunnell & Beidel, 2013 (Nina)	14	2.00 (medium)
Christon et al., 2012 (Ava)	10	1.43 (medium)
Facon et al., 2008 (Sami)	18	2.57 (high)
Fisak et al., 2006 (LM)	7	1.00 (low)
Lang et al., 2011 (Leslie)	17	2.43 (medium)

Where: Low 1.4 or less, Medium 1.5-2.4, High 2.5 or above

**Weight of Evidence B (Methodological Relevance)**

Table 12:

Coding Protocol for Methodological Relevance

	<b>3 points</b>	<b>2 points</b>	<b>1 point</b>
<b>Study Design</b>	Randomised Controlled Trials or Meta Analyses	Experimental or quasi-experimental designs, including Single Case Experimental Designs	Case studies
<b>Rationale</b>	RCTs show a high degree of experimental control and are therefore suitable for evaluation of the effectiveness of an intervention.	Single subject designs are a valid methodology to evaluate the effectiveness of interventions (Horner, 2005; Kratochwill et al., 2010), particularly when the phenomenon studied is rare.	Case studies provide rich contextual information but lack experimental control and are therefore less suited to evaluations of effectiveness.

Table 13:

Weight of Evidence B ratings for each included study

<b>Study (ppt.)</b>	<b>Study Design</b>	<b>WOE B</b>
Beare et al., 2008 (Luke)	SCED	2 (medium)
Bunnell & Beidel, 2013 (Nina)	SCED	2 (medium)
Christon et al., 2012 (Ava)	Case study	1 (low)
Facon et al., 2008 (Sami)	SCED	2 (medium)
Fisak et al., 2006 (LM)	Case study	1 (low)
Lang et al., 2011 (Leslie)	SCED	2 (medium)

Where: Low = 1, Medium = 2, High =3 or above

## Weight of Evidence C (Topic Relevance)

Table 14:

Coding Protocol for Topic Relevance

	<b>3</b>	<b>2</b>	<b>1</b>	<b>Rationale</b>
<b>Country</b>	English-speaking OECD countries	Non-English speaking OECD countries	Non OECD countries	English-speaking countries are the most relevant when considering treatment for pupils who are selectively mute in English
<b>Age of Ppt.</b>	16-25	11-15	9-10	Older participants are more relevant when considering selective mutism treatments for older pupils
<b>Intervention setting</b>	School	Community	Clinic	School-based interventions are most replicable in school settings
<b>Outcome variable</b>	Speech in a natural setting	Speech in sessions	Speech in response to reward	Outcome variables that are more natural are more generalizable to wider contexts
<b>Intervention</b>	Behavioural only	Behav. and cognitive	Behav. and other	Interventions using only behavioural approaches are most relevant when considering the efficacy of behavioural approaches.

Table 15:

Weight of Evidence C calculations for each included study

	<b>Beare et al., 2008 (Luke)</b>	<b>Bunnell &amp; Beidel, 2013 (Nina)</b>	<b>Christon et al., 2012 (Ava)</b>	<b>Facon et al., 2008 (Sami)</b>	<b>Fisak et al., 2006 (LM)</b>	<b>Lang et al., 2011 (Leslie)</b>
Country	3	3	3	2	3	3
Age	2	3	2	2	1	1
Setting	3	2	1.5	3	1.5	2
Outcome	2	3	3	3	2	3
Intervention	3	2.5	2	3	1	3
<b>Total</b>	13	13.5	11.5	13	8.5	12
<b>WOE C (Average)</b>	2.6 (high)	2.7 (high)	2.3 (medium)	2.6 (high)	1.7 (medium)	2.4 (medium)

Where: Low 1.4 or less, Medium 1.5-2.4, High 2.5 or above