Music and musical elements in the treatment of childhood speech sound disorders

van Tellingen, Mirjam; Hurkmans, Joost; Terband, Hayo; Jonkers, Roel; Maassen, Ben

Published in:
International Journal of Speech-Language Pathology

DOI:
10.1080/17549507.2022.2097310

Publication date:
2023

International Journal of Speech-Language Pathology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iasl20

Music and musical elements in the treatment of childhood speech sound disorders: A systematic review of the literature

Mirjam van Tellingen, Joost Hurkmans, Hayo Terband, Roel Jonkers & Ben Maassen

To cite this article: Mirjam van Tellingen, Joost Hurkmans, Hayo Terband, Roel Jonkers & Ben Maassen (2023) Music and musical elements in the treatment of childhood speech sound disorders: A systematic review of the literature, International Journal of Speech-Language Pathology, 25:4, 549-565, DOI: 10.1080/17549507.2022.2097310

To link to this article: https://doi.org/10.1080/17549507.2022.2097310

International Journal of Speech-Language Pathology, 2023; 25: 549–565

Music and musical elements in the treatment of childhood speech sound disorders: A systematic review of the literature

MIRJAM VAN TELLINGEN[1], JOOST HURKMANS[1], HAYO TERBAND[2,3], ROEL JONKERS[4] AND BEN MAASSEN[4]

[1]Rehabilitation Centre "Revalidatie Friesland", Beetsterzwaag, The Netherlands, [2]Utrecht Institute of Linguistics-OTS, Utrecht University, Utrecht, The Netherlands, [3]Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA, and [4]Centre for Language and Cognition Groningen, University of Groningen, Groningen, The Netherlands

Abstract

Purpose: Music-based interventions are used in the treatment of childhood speech sound disorders (SSDs). Hypotheses on working mechanisms are being developed, focussing on shared neural processes. However, evidence of the effect of treatment with musical elements in SSDs in children is lacking. This study reviews the literature regarding the use of music-based interventions in the treatment of childhood SSDs.
Method: A systematic search in six databases was conducted, yielding 199 articles, eight of which met the inclusion criteria. Included articles were reviewed on study characteristics, patient characteristics, interventions, outcomes and methodological quality.
Result: This review included four case studies, three single-subject design studies and one cohort study. Seven studies reported positive outcomes on speech production, but outcome measures in the four studies with experimental design were not all aimed at the level of speech (motor) processes. Methodological quality was sufficient in one study.
Conclusion: Seven out of eight studies in this review report positive outcomes of music-based interventions in the treatment of SSDs. However, these outcomes are not sufficiently supported by evidence due to insufficient methodological quality. Suggestions for improving methodological quality in future research are presented.

Keywords: *speech-language pathology; music therapy; speech sound disorder; childhood apraxia of speech; systematic review*

Introduction

Humans communicate and express themselves through speech. It is an essential part of full participation in society (Ruben, 2000). Most children acquire speech effortlessly during the first years of their lives, in a gradual process of acquiring speech sounds, organising speech sounds into speech patterns and fine tuning of phonological knowledge. From producing their first words around their first birthday, children show progress in intelligibility, articulatory development and phonological development, and reach adult-like target production at the age of five (Dodd, 2011). During development children use strategies, such as substitutions, omissions and distortions to approximate adult-like speech. In typical development, these strategies disappear with increasing speech production skills. Persistence of these strategies in the speech production of children is a sign of a speech delay or speech sound disorder (SSD; Dodd, 2011).

SSD is defined as a range of difficulties producing speech sounds in children, due to a variety of limitations in perceptual, speech motor or linguistic processes or a combination of these limitations (McLeod & Baker, 2017; Namasivayam, Coleman, O'Dwyer, & van Lieshout, 2019; Shriberg, 2010). Overlap in symptoms due to a combination of limitations is to be expected because the various processes are interdependent. Therefore, a limitation in one process affects development in adjacent processes (Terband, Maassen, & Maas, 2019).

Children with SSD require (sometimes intensive) treatment to improve their speech production and reduce the risk of acquiring associated developmental and participation problems (Terband, Maassen,

Correspondence: Mirjam van Tellingen, Revalidatie Friesland, Hoofdstraat 3, 9244 CL Beetsterzwaag, The Netherlands. Phone: +31 885801787. E-mail: m.van.tellingen@revalidatie-friesland.nl

ISSN 1754-9507 print/ISSN 1754-9515 online
Published by Taylor & Francis
DOI: 10.1080/17549507.2022.2097310

et al., 2019). Various interventions have been developed, aiming at different processes that might be limited, such as the approach developed by Hodson and Paden (1991) for phonological disorders, Rapid Syllable Transition Treatment (ReST; Ballard, Robin, McCabe, & McDonald, 2010) and Dynamic Temporal and Tactile Cueing (DTTC; Strand, 2020) that focus on the planning and programming of speech and PROMPT (Hayden, Eigen, Walker, & Olsen, 2010), that focusses on motor-speech processes.

Also, treatment including musical elements has been described as an intervention for SSD, for example in Childhood Apraxia of Speech (CAS; Rosenbek, Hansen, Baughman, & Lemme, 1974). A specific method using musical elements is Melodic Intonation Therapy (MIT; Sparks & Holland, 1976), originally developed for adults with aphasia and adapted by Helfrich-Miller (1984) for the treatment of children with CAS. In MIT, the musical elements melody, tempo and rhythm are used to slow the speech rate and exaggerate rhythm and stress. Simplifying melody, using two tones, leads to what is described as "intoned utterances". In the original application of MIT, the rhythmic pattern is supported by tapping. Helfrich-Miller (1984) substituted tapping for signed English to support recollection of the target words. In three levels children move towards increased length of unit, increased phonemic difficulty, diminished dependency on the therapist and diminished reliance on the intonation as induced by MIT (Helfrich-Miller, 1984).

Another method using musical elements and originally developed for adults with aphasia and Apraxia of Speech (AoS) is Speech-Music Therapy for Aphasia (SMTA; De Bruijn, Zielman, & Hurkmans, 2005; Hurkmans, De Bruijn, Reitsma, & Koek, 2018). SMTA is a combination of speech-language pathology and music therapy, in which both therapists work with the patient at the same time. In SMTA, the musical elements melody, tempo, rhythm, metre and dynamics are used by a Music Therapist to compose new melodies that closely follow the prosodic features of speech production. In exercises, the dependency on musical support is diminished by moving from singing to rhythmic chanting and speaking (De Bruijn et al., 2005; Hurkmans et al., 2018). SMTA enhances accuracy, consistency and fluency of articulation in adults with AoS and aphasia (Hurkmans et al., 2015). This method is being used in clinical practice in the treatment of children with CAS (Van Tellingen et al., 2022), since both AoS and CAS are described as a disorder in the planning and programming of speech movements (American Speech-Language-Hearing Association, 2007; Hurkmans, 2016) and share various characteristics, such as inconsistent errors in the realisation of phonemes, segmentation, vowel distortions, groping and effect of articulatory complexity (Iuzzini-Seigel & Murray, 2017; Ziegler, 2008).

While working mechanisms of SMTA, MIT and other interventions using musical elements are still unclear, hypotheses on the underlying mechanisms have been proposed. Recent hypotheses focus on shared neural processes for processing speech and music. With the OPERA-hypothesis, Patel (2014) proposes a framework in which treatment with music can lead to better speech processing, assuming shared sensory and cognitive mechanisms in the brain. This framework is expanded by Fujii and Wan (2014) to hypothesise about the role of rhythm in the rehabilitation of speech production. In their framework, rhythm supports speech production for patients with speech disorders through synchronisation and entrainment to a pulse. While both hypotheses focus on shared neural processes, the OPERA-hypothesis is additionally based on other features of music. These include eliciting emotions, repetition and increasing attention. These factors, which relate to motivation and mood, are also described by Merrett, Peretz, and Wilson (2014) as part of the working mechanism of MIT.

In typical development, engagement with music plays a major role in developing perceptual processing systems which facilitate the encoding and identification of speech sounds and patterns (Hallam, 2010). Consistent with the OPERA-hypothesis, the hypothesis on the working mechanism of this effect is based around shared processing systems (Hallam, 2010). While the aforementioned studies by Patel (2014) and Hallam (2010) both hypothesise on the benefits of music on speech processing through shared processing systems, the evidence supporting these hypotheses is focussed on speech perception and rehabilitation in stuttering and autism, rather than speech production and rehabilitation in childhood SSDs.

Studies on the use of musical elements in the treatment of neurological language and speech disorders in adults have been reviewed by Hurkmans et al. (2012). All studies reported improvement after treatment including melody and rhythm, but methodological quality was rated low. Therefore, no conclusions could be drawn on the effect of musical elements in the treatment of neurological language and speech disorders.

The use of musical elements in the treatment of SSD in children has been described in the literature (e.g. Helfrich-Miller, 1984; Rosenbek et al., 1974). However, evidence of the effectiveness of these interventions is lacking. Additionally, hypotheses have been proposed on the working mechanisms of music in the treatment of SSDs, but evidence to support these hypotheses is also lacking. This study aims to identify these gaps by reviewing existing literature on the use of music or musical elements in the treatment

of SSDs and provide recommendations for future research, as we plan to evaluate SMTA in the treatment of children with CAS.

Method

A systematic review of the literature was conducted, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA; Liberati et al., 2009).

The list of search terms below was compiled with the aim to find literature on the use of musical elements in the treatment of a wide variety of childhood speech and language disorders.

(1) "speech disorder", "dyspraxi*", "phonological disorder", "dysarthria", "communication disorder", "speech sound disorder", "speech impairment", "apraxia", "language disorder", "oral".
(2) "music", "melodic", "rhythm", "singing".
(3) "intervention*", "treatment", "therap*"
(4) "child*", "infant*.

This compiled list was expanded with the following terms based on the underlying terms from the MeSH descriptor of PsychINFO.

"Speech Disorders", "Articulation Disorders", "Dysphonia", "Echolalia", "Mutism", "Stuttering", "Aphasia", "Dysphasia", "Articulation Disorders", "Dysarthria".

These terms were added under (1). This expanded list was used to search all below mentioned databases. Studies including some of these terms, such as "aphasia", would be excluded based on the exclusion criteria for this review (see below). However, all terms were added to ensure inclusion of all studies on musical elements in the treatment of SSD. All search terms were linked using combinations of (1), (2), (3) and (4), so that articles including at least one term from each category would be included. This led to the search strings included in Appendix A.

The search was conducted in the databases Pubmed, PsychINFO, AMED, ERIC, CINAHL and Web of Science on 1 July 2020, by an expert from Research Institute SHARE at University Medical Centre Groningen. Endnote was used to remove duplicates.

Remaining articles were included or excluded by the first and third author independently. Criteria for inclusion were formulated in terms of PICOS, i.e. patients, intervention, comparison, outcome and study design (Liberati et al., 2009). Criteria were (P) children <18 years, diagnosed with SSD (or similar term), (I) therapy using musical elements, and (COS) all study designs. Articles in all languages were included. Articles were first screened, based on title and abstract. Articles not matching the criteria for inclusion were excluded. When a decision could not be made based on the abstract, the full text was read and judged according to the criteria. Selections were compared and 12 articles were further discussed, leading to the inclusion of two articles in addition to the six articles that were included based on agreement of individual selections.

The search for articles yielded a substantial number of studies that used music in the treatment of childhood speech and language disorders, including hearing disorders, stuttering, speech-language disorders in children with autism and dyslexia. These studies did not include children with SSDs or did not provide treatment focussed on SSDs, and were therefore excluded.

Variables used to describe the studies were characteristics of the study, participants, intervention and outcomes. Several indicators were used to describe each variable. Information on these indicators was based on information provided in the article. Any missing information was indicated as "not reported".

The characteristics of the study were summarised through four indicators: the study design, the linguistic level of the main outcome measures, the intervention and the target group of each study. We chose not to note the designs as described in the articles, but instead we used the terminology from the American Speech-Language-Hearing Association (2020) to determine study designs and allow for easy comparison of studies.

Indicators for the characteristics of the participants were the number of participants, age, sex, handedness, language(s) and primary language in cases of multilingualism. The speech disorder of participants was summarised, including criteria used for the diagnosis and the severity of the disorder. Possible comorbidities and specific medical diagnoses were also noted. The final indicators were previous interventions and musical background of participants.

Interventions were summarised through the indicators treatment method, therapist (Music Therapist or speech-language pathologist [SLP]), condition (individual or group), dose (session time, number of sessions per week and total intervention duration), linguistic levels included in the intervention, the level of items used (e.g. syllables and words), musical elements (e.g. melody and rhythm) and musical forms (e.g. singing and tapping). The International Classification of Functioning (ICF; World Health Organization (Red.), 2007) was used to determine whether interventions targeted body functions and structures, activities and/or participation. Finally, hypothesised working mechanisms of the interventions, as described in the original articles, were summarised.

Outcomes were summarised by describing outcome measurements and their validity and reliability. The level of ICF targeted by the outcome measurements was determined. Data on the measured improvement and the statistical analyses were extracted.

For the appraisal of the quality of the studies, multiple tools were needed because various study designs were included. We were unable to find existing tools to fit our needs and therefore compiled our own, using the appraisal tools by Boles (2015), supplemented with some items from the RoBiNT-scale (Tate et al., 2013), and the Quality Assessment Tool for Quantitative studies by the Effective Public Health Practice Project (EPHPP; Thomas, Ciliska, Dobbins, & Micucci, 2004). The appraisal scales that were used are available in Appendix B. We appraised the different study designs with varying combinations of the scales.

All study designs were rated for descriptive quality indicators, as described by Boles (2015). Experimental design (i.e. single subject and group) studies were rated for additional phase and assessment descriptive quality indicators, with a tool based on Boles (2015) and items from the RoBiNT scale (Tate et al., 2013). Single subject designs were rated for basic quality indicators, with a tool including indicators based on Boles (2015) and items from the RoBiNT-scale (Tate et al., 2013). Group studies were rated for basic quality indicators using the Quality Assessment Tool for Quantitative studies from the EPHPP (Thomas et al., 2004). Single subject design studies that would meet design standards or meet design standards with reservations would additionally be analysed for evidence of effect quality indicators (Boles, 2015).

All tools use a three-point scale, with the levels described as sufficient, minimal or insufficient.

The first author summarised and appraised all articles. The second and third author each summarised and appraised half of the articles. All summaries and appraisals were conducted independent of each other and compared afterwards. Disagreements were first resolved through discussion between the two authors that appraised the same articles. Any remaining disagreements were resolved through discussion between the first three authors.

Result

The search yielded 263 publications in total. After duplicates were removed, 199 publications remained. A total of 165 publications were excluded because they did not include participants with a primary diagnosis of SSD (but dyslexia, autism or hearing disorders) based on screening of title and abstract. An additional 26 publications were excluded after screening abstracts and full text. Six studies did not include any participants with a primary diagnosis of SSD (two of which included participants that stutter), and 10 publications provided descriptions or rationale for treatment without patient data. In one case, the abstract and full text of the publication could not be obtained. Nine publications were excluded for other reasons, such as including participants with anatomical deviations or neurological disorders. The remaining eight publications were included for further analysis, comprising seven articles from peer-reviewed journals and one book chapter. Figure 1 shows the PRISMA flowchart and summary of excluded articles.

Study characteristics

An overview of the eight studies and corresponding study characteristics is available in Table I. Four articles described case studies. Three articles described single-subject design studies, all with multiple phases. In the study by Krauss and Galloway (1982) each child served as their own control for two months of traditional speech therapy followed by two months of traditional speech therapy combined with MIT. Lagasse (2012) used a design with alternating treatments, speech-language pathology and MIT, in nine phases. A design with five phases was used by Martikainen and Korpilahti (2011), which included baseline, MIT, no treatment, Touch Cue Method (TCM), which is a method without musical elements (Bashir, Grahamjones, & Bostwick, 1984) and follow up. Gross, Linden, and Ostermann (2010) conducted a cohort study in a reversal design with four phases, alternating creative music therapy based on the Nordoff-Robbins approach (Mahoney, 2010) and no treatment.

Objectives of the case studies were described as reporting on the development of speech and language in relation to treatment, such as Music Therapy or MIT. The single-subject design studies all reported to aim at assessing the use of a specific method in the treatment of CAS. The prospective cohort study by Gross et al. (2010) aimed at examining the effect of Music Therapy on what is described by the authors as verbal reasoning.

Patient characteristics

Table II provides an overview of the patient characteristics for the eight studies. Three articles on case studies described one participant, Helfrich-Miller (1994) included three case studies. Two single-subject design studies included two participants, and Martikainen and Korpilahti (2011) included one. Gross et al. (2010) included 18 participants in their cohort study. All studies reported the gender of participants, including a total of 20 males and nine females. Ages ranged from 2.9 to 8 years, the age of participants was not reported in one study (Krauss & Galloway, 1982). Two studies reported the primary language of the participant. Beathard and Krout (2008) reported multilingualism and languages spoken by the participant. However, information on the primary language was not reported. In the five remaining studies, there was no information reported on the language(s) of participants. Handedness was only reported by Krauss and Galloway (1982).

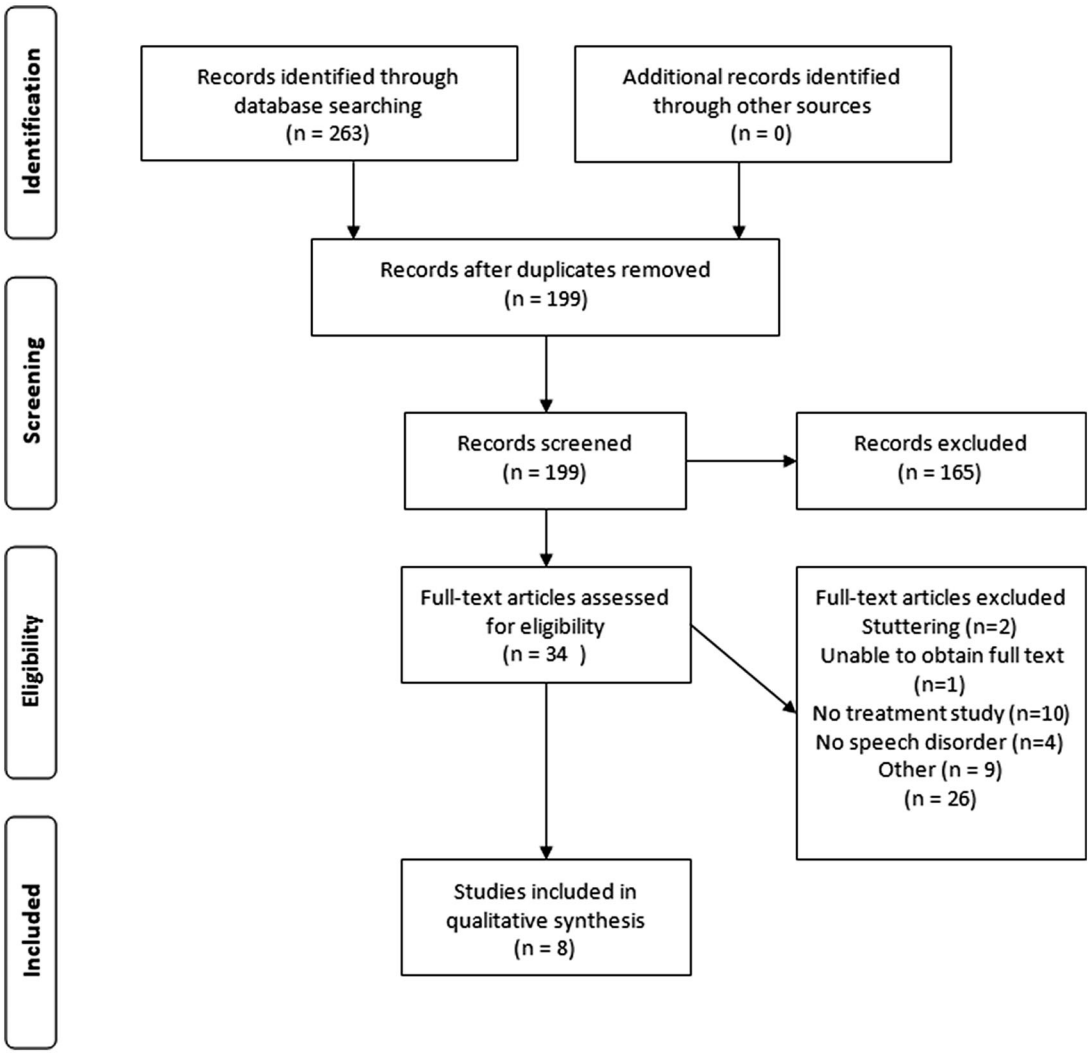

Figure 1. PRISMA 2009 Flow Diagram (Moher, Liberati, Tetzlaff, Altman, & The PRISMA GROUP, 2009).

Table I. Study characteristics.

Study	Study design	Outcome	Intervention	Target group
Beathard and Krout (2008)	Case study	Speech	MT	CAS
Catrini and Lier-DeVitto (2019)	Case study	Speech	SLT; focussing on prosody through rhythm and singing	CAS
Gross et al. (2010)	Prospective cohort study; ABAB reversal design	Language production and comprehension	Creative MT based on Nordoff-Robbins approach	Developmental speech disorder
Helfrich-Miller (1994)	Case studies	Speech	MIT	Oral apraxia
Krauss and Galloway (1982)	Single subject design; AB design	Language production and comprehension	MIT	Language delay with apraxia
Lagasse (2012)	Single subject design; ABABABABA design	Speech	MIT	CAS
Martikainen and Korpilahti (2011)	Single subject design; 5 phases design: baseline, MIT, no treatment, TCM, follow-up	Speech	combination MIT/TCM	CAS
Sutton (1993)	Case study	Speech and language development	MT	Speech and language impairment

MIT: Melodic Intonation Therapy; MT: Music Therapy; SLT: speech and language therapy; TCM: Touch Cue Method; CAS: Childhood Apraxia of Speech

Study design as defined by the authors of this review based on ASHA guidelines (American Speech-Language-Hearing Association, 2020). Level of outcome measurements and target group as defined by the authors of this review.

All articles reported speech and language diagnoses, but not for all participants. Reported diagnosis was CAS in five studies, the remaining studies reported diagnoses of specific developmental speech disorder, oral apraxia and lexical syntactic deficit syndrome. Diagnostic criteria were only reported by Martikainen and Korpilahti (2011). Severity and severity criteria were not reported in any of the

Table II. Participant characteristics.

Study	Nr.	Sex	Age	Language	Handed-ness	Speech-language diagnosis	Diagnostic criteria	Severity	Comorbidity	Medical diagnosis	Musical background	Previous interventions
Beathard and Krout (2008)	1	F	3	American Sign Language, English (NR)	NR	Childhood Apraxia of Speech	NR	NR	Hypotonia, gross motor difficulties, sensory issues.	Hypotonia	Engaged in music through play and signing (ASL).	Early Childhood Intervention Programme with occupational, speech, play therapy.
Catrini and Lier-DeVitto (2019)	1	F	5;0	NR (Brazilian Portugese)	NR	Speech apraxia	Phonemic inconsistency	NR	Gross motor difficulties	No organic issues	NR	None
Gross et al. (2010)	18	6F, 12M	3;5 – 6	German	NR	Specific developmen-tal speech disorder	NR	NR	None	No	No previous experience with music therapy	SLT and early intervention programs.
Helfrich-Miller (1994)	3	M	1: 2;9 2: 2;10 3: 8	NR (English)	NR	1: Oral apraxia 2: NR 3: NR	1: Programming deficit 2: NR 3: NR	1: NR 2: NR 3: NR	1: poor voluntary tongue control 2: NR 3: NR	1: History of otitis media; mild right facial asymmetry 2: Attentional problems 3: None	1: NR 2: NR 3: NR	1: SLT 2: SLT 3: SLT
Krauss and Galloway (1982)	2	M	NR	NR	R	Verbal apraxia, expressive language below age level	Motor sequencing and programming problems	NR	NR	NR	NR	SLT
Lagasse (2012)	2	M	1: 5 2: 6	English	NR	Developmen-tal Apraxia of Speech	NR	NR	No known cognitive deficits	NR	NR	SLT
Martikainen and Korpilahti (2011)	1	F	4;7	Finnish	NR	Childhood Apraxia of Speech	Met 10 of the 11 characteristics for CAS (Thoonen et al., 1997; Davis et al., 1998)	NR	None	None	NR	SLT
Sutton (1993)	1	M	5	NR (English)	NR	Lexical Syntactic Deficit Syndrome	Syntactic and lexical access problems	NR	Developmental delay	Global developmental delay.	NR	SLT

Nr.: number; NR: not reported; Sex: F: female; M: male; Handedness: R: right; ASL: American Sign Language; SLT: speech and language therapy.

articles, but five studies provided a subjective description relating to severity.

All articles reported existing or absence of comorbidities, such as speech-language, motor function or cognitive problems, except for Krauss and Galloway (1982). Absence or presence of a medical diagnosis was reported in six of the articles. Musical background was described by Beathard and Krout (2008) and Gross et al. (2010). All articles described previous interventions, in six of the articles this was reported as speech-language pathology.

Intervention characteristics

The intervention characteristics for the eight studies are summarised in Table III. Four of the studies reported MIT as the treatment program, with adjustments to the method in three studies. In three remaining studies, the treatment program was Music Therapy, one of which was not further defined. Catrini and Lier-DeVitto (2019) reported the use of musical elements in general speech-language pathology.

Three studies reported a treatment goal specific to the participant, such as increased speech generation or teaching the child developmentally appropriate phonemes.

Five articles mentioned the (linguistic) level(s) of the intervention. Articulation was the most reported, in four studies. Vocalisation and phonology were reported in two studies. Other reported levels include prosody, social interaction and musical expression. The linguistic level(s) of the intervention were not reported in three articles.

The type of items that was used in the interventions consisted of words and sentences in four studies. Beathard and Krout (2008) and Martikainen and Korpilahti (2011) additionally used phonemes and syllables. Two studies did not report which items were used in the treatment.

Musical elements were reported in all articles, including rhythm and melody in all studies. Tempo was used in seven studies, metre in five. These elements were carried out through singing in five studies, intoning and so-called "sprechgesang" in two studies, tapping in one study and playing instruments in three studies.

Participants received individual therapy in all studies. Regarding dose, the session time was reported in four studies and varied from 25 to 40 min. The number of sessions per week varied from 1 to 3 times in the five studies that reported frequency of sessions. Seven studies reported on the total intervention duration, which varied from 4 weeks to 2 years and 9 months.

Three articles reported that a Music Therapist conducted the intervention. In one study both an SLP and a Music Therapist each conducted different interventions. The remaining four articles did not report the type of therapist conducting the intervention. The ICF level of the intervention was not specifically reported in any studies. From subjective judgement of the eight studies, it is concluded that all interventions focussed on levels of body functions and structures and/or activities.

Working mechanisms

None of the studies included means to examine the working mechanism of the interventions in their methodology. Five articles proposed a hypothesis on the working mechanism of the intervention as interpretations of their results. Catrini and Lier-DeVitto (2019) speculated on the beneficial effect of music on fluency in movement and speech, through the flow of music as they believe it reaffirms the flow of early vocalisations. Gross et al. (2010) hypothesise that improvements occur because music therapy addresses listening, perception, processing and memorising sounds and musical structures.

Three studies hypothesised about the beneficial effects of slowing speech rate and supporting rhythm and stress through MIT on speech production, albeit different in their approach. Helfrich-Miller (1984) combines research on the effectiveness of slowing speech rate in the treatment of apraxia (Rosenthal, Williams, & Ingham, 1981) with research on the effect of music in the teaching of children (Kallan, 1972) to hypothesise on possible beneficial effects of MIT in the treatment of CAS. Krauss and Galloway (1982) hypothesise that through directing the child's attention to slow tempo, precise rhythm and distinct stress, the child is better able to process the structural aspect of intoned verbal utterances, as described in the original study on MIT by Sparks, Helm, and Albert (1974). Martikainen and Korpilahti (2011) focus on heightening sensory feedback during articulation and providing necessary time for motor planning and programming through lengthening speech gestures (Wambaugh & Martinez, 2000) as rationale for the use of MIT in the treatment of CAS.

Outcome characteristics

An overview of the outcome characteristics for the eight studies is presented in Table IV. Four of the studies were case studies, describing course of treatment without outcome measurements. The four studies with an experimental design used language-specific versions of articulation and language tests. One study also included a non-verbal intelligence test (Gross et al., 2010). Two studies reported on the reliability and validity of these instruments.

Gross et al. (2010) reported statistically significant improvement over the whole study period for subtests measuring phonological memory and understanding sentences, as well as three subtests from a non-verbal intelligence test. In their ABAB reversal design, they reported greater improvement during treatment periods. Krauss and Galloway (1982) reported mixed

Table III. Intervention characteristics.

Study	Treatment program	Treatment goal	Condition	Dose Session time	Freq.	Duration	Therapist	(Linguistic) levels of intervention	Trained levels	Musical elements	Musical forms	ICF
Beathard and Krout (2008)	Music therapy using a data-based model for clinical music therapy (Hanser, 1999)	(1) Increased speech recognition and generation, (2) improved social integration and communication skills	I	30 min	1	9 months, including a break during two months	MT intern supervised by MT.	Vocalisation, articulation, social interaction	Sounds, phonemes, words	Rhythm, melody, tempo, metre	Singing, playing instruments	FA
Catrini and Lier-DeVitto (2019)	SLT using nursery rhymes, paced reading	To invite her to speak	I	NR	NR	NR	NR	Articulation, prosody (rhythm), syntax, phonology, voice	Words and sentences	Tempo, rhythm, metre	Singing	FA
Gross et al., 2010	Creative music therapy based on the Nordoff-Robbins approach (Mahoney, 2010)	NR	I	25 min	NR	32 weeks, 2x eight weeks therapy, 2x eight weeks waiting time.	MT	Musical and vocal expression	NR	Rhythm, melody, tempo	Singing, dancing, being sung to, play instruments	A
Helfrich-Miller (1994)	MIT	1: Teach developmentally appropriate phonemes; Focus on combining and sequencing skills; stimulate carryover. 2: Sequencing of consonant sounds and sequencing words into sentences. 3: NR.	I	1: 10 sentences 2: 10 sentences 3: 20 sentences	1: 2 2: 1 3: 3	1: 1 year 2: 2;9 years 3: 28 weeks	1: NR 2: NR 3: NR	Articulation, phonology, sequencing	Words and sentences	Rhythm, tempo, melody, metre	Intoning, sprech-gesang	F
Krauss and Galloway (1982)	Traditional SLT, and traditional SLT with MIT	NR	I	NR	2	4 months (2 months per phase)	NR (SLP)	NR (articulation)	Phrases and sentences	Tempo, rhythm, metre (stress)	NR	F
Lagasse (2012)	SLT and MIT	NR	I	40 min	SLT: 1 MIT: 1	4 weeks	SLT by SLP MIT by MT	NR (articulation)	Words and phrases	Rhythm, melody, tempo	Tapping, singing	F
Martikainen and Korpilahti (2011)	MIT and TCM	NR	I	30 min	3	6 weeks MIT and 6 weeks TCM	NR (SLP)	Articulation	Sentences (MIT); syllables, words (TCM)	Tempo, melody, rhythm, metre (stress)	Intoning	FA
Sutton (1993)	Music therapy	NR	I	NR	NR	2 years	MT	NR	NR	Rhythm, melody	Singing, playing instruments	F

NR: not reported; SLT: Speech and Language Therapy, condition: I: individual, freq: frequency sessions per week, TCM: Touch Cue Method; MT: music therapist; SLP: speech and language pathologist; ICF: International Classification of Functioning: F: functions; A: activities.

Table IV. Outcome characteristics.

Study	Outcome measurements	Validity and reliability of measurements	Measured improvement	Conclusion original authors	ICF
Beathard and Krout (2008)	NA	NA	NA	Positive effect of MT on (1) verbal communication, (2) socialisation, (3) cognitive/ emotional and (4) motor skills/ movement.	FAP
Catrini and Lier-DeVitto (2019)	NA	NA	NA	Increase in speech production and intonation contours	AP
Gross et al. (2010)	SETK 3-5, SON-R, Nordoff-Robbins assessment scales (Mahoney, 2010)	SETK 3-5: sufficient, SON-R: sufficient, Nordoff-Robbins: reliability is sufficient	Significant improvement on Phonological Memory and Understanding sentences and SON-R parameters cognitive structures, action patterns and IQ.	Possible positive effect of MT on speech development	FA
Helfrich-Miller (1994)	1: NA 2: NA 3: NA	1: NA 2: NA 3: NA	1: NA 2: NA 3: NA	1: Steady acquisition of consonant sounds; normalisation of conversational speech 2: Speech improved 3: Speech improved	FA
Krauss and Galloway (1982)	Verbal and auditory subtests of the PICAC and Language Sampling, Analysis, and Training (Gottsleben & Tyack, 1974).	NR	Significant effects in PICAC tasks Names Objects and Imitates Words and Phrases, Language Sampling measure MLU and Intelligibility.	Positive effect of MIT on verbal naming, phrase length, verbal imitation and articulation.	F
Lagasse (2012)	GFTA2 , KLPA2, Speech Production Test (SPT), modified from data collection sheets created by Dauer, Irwin, and Schippits (1996).	The GFTA2 and KLPA2: sufficient. SPT: NR	No significant differences.	MIT was no more successful than SLT for improving speech production.	F
Martikainen and Korpilahti (2011)	PVC, PCC, PMLU, PWP and PWC in Picture-naming task, modified from the Finnish articulation test (Remes and Ojanen, 1996).	NR	Significant improvement on PMLU, PWC.	Positive effect of the combination of two motor-based treatments on the speech of a child with CAS.	F
Sutton (1993)	NA	NA	NA	Positive effect of MT on sound vocabulary and communication	P

NA: not applicable; NR: not reported; SET-K 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen (Grimm et al., 2001); SON-R: Snijders-Oomen Non-Verbal Intelligence Test (Tellegen et al. 2002); PICAC: Porch Index of Communicative Ability in Children (Porch, 1974); GFTA2: Goldman-Fristoe Test of Articulation, 2nd ed. (Goldman & Fristoe, 2000); KLPA2: Khan-Lewis Phonological Analysis, 2nd ed. (Khan & Lewis, 2002); PVC: percentage of vowels correct; PCC: percentage of consonants correct; PMLU: phonological mean length of utterance; PWP: proportion of whole-word proximity; PWC: proportion of whole-word correctness; MT: Music Therapy; MIT: Melodic Intonation Therapy; SLT: Speech and Language Therapy; ICF: International Classification of Functioning: F: functions; A: activities; P: participation.

results in their AB design study, with significant improvement on subtests measuring object naming, imitation of words and sentences, but no significant improvement on auditory tasks and subtests measuring description of functions and sentence completion. A significant increase of Mean Length of Utterance in a language sampling task was reported for both children in this study. Additionally, the authors reported an increase in intelligibility due to improved articulation on an imitation task. Martikainen and Korpilahti (2011) reported significant improvement for Percentage of Vowels Correct directly after MIT treatment, and at the follow-up six weeks later. Percentage of Consonants Correct was reported to decrease significantly after MIT, but increased significantly in the following six weeks of no treatment. Significant improvement six weeks after MIT was reported for Phonological Mean Length of Utterance

(PMLU), which was measured by the authors as a mean value of the entire sample by scoring the produced segments and correctly produced phonemes and dividing the total by the total number of words in the sample. Vowels were included, in reflection of the high proportion of vowels in Finnish. Directly after MIT, no change in PMLU was reported. For all three measures additional significant improvement over the entire study period was reported, including treatment with TCM and follow up. Lagasse (2012) reported no significant differences in their reversal design study.

No study specifically mentioned the ICF level of the outcome measurements. After assessment of the measurements, it is concluded that six studies reported on the outcomes on the ICF-level of body functions and structures. Two of these studies also reported outcomes at the level of activities, while one also reported at the level of participation. The two remaining studies reported outcomes at the level of participation (Sutton, 1993) or activities and participation (Catrini & Lier-DeVitto, 2019).

Quality appraisal

A summary of the quality appraisal is presented in Tables V–VII. The scoring guidelines are available in Appendix B. All indicators are scored on a three-point scale, described as sufficient, minimal or insufficient.

Agreement on the quality appraisal between the first and second author was 50%, and 58% between the first and third author. The first three authors then discussed the appraisal guidelines to determine what accounted for this low rate of agreement. It was concluded that interpretations of the appraisal guidelines differed. After discussing the guidelines and interpreting them in a similar way, agreement was 100%.

An overview of the descriptive quality indicators is presented in Table V. Overall, two studies were minimally described; all remaining studies had insufficient description.

Participants' characteristics were minimally or sufficiently described in most studies, except for two studies, one of which did not report on the age of participants (Krauss & Galloway, 1982), and one not reporting on diagnosis in two of the case studies (Helfrich-Miller, 1994).

Description of the setting varied across studies, with three studies describing the setting sufficiently, two giving a minimal description and in three studies settings described insufficiently. Four studies gave a description of the therapist or interventionist conducting the intervention. Procedures in baseline and intervention phases were sufficiently described by Gross et al. (2010), Lagasse (2012) and Martikainen and Korpilahti (2011). None of the four case studies had a sufficient description of the baseline phase, but Beathard and Krout (2008) and Helfrich-Miller (1994) described the intervention phase in sufficient

Table V. Descriptive quality indicators.

Study	Participant	Setting	Therapist	Baseline/intervention	Dependent variable	Overall
Beathard and Krout (2008)	1	2	2	1	0	0
Catrini and Lier-DeVitto (2019)	1	0	0	0	0	0
Gross et al. (2010)	2	2	2	2	1	1
Helfrich-Miller (1994)	0	0	0	1	0	0
Krauss and Galloway (1982)	0	1	0	1	1	0
Lagasse (2012)	1	2	2	2	1	1
Martikainen and Korpilahti (2011)	2	0	0	2	1	0
Sutton (1993)	1	1	1	0	0	0

0 = insufficient description, 1 = minimal description, 2 = sufficient description.

Table VI. Additional phase and assessment descriptive quality indicators.

Study	Maintenance	Generalisation	Raw data	Fidelity	Social validity	Overall
Gross et al. (2010)	0	2	1	2	1	0
Krauss and Galloway (1982)	0	0	0	0	0	0
Lagasse (2012)	0	0	0	0	0	0
Martikainen and Korpilahti (2011)	2	0	2	0	0	0

0: insufficient measure; 1: minimal measure; 2: sufficient measure.

Table VII. Experimental design basic quality indicators.

Study	IV/intervention	Inter-rater reliability	Effect	Data points	Blinding	Data analysis	Overall
Krauss and Galloway (1982)	2	0	0	0	0	2	0
Lagasse (2012)	2	0	0	0	1	1	0
Martikainen and Korpilahti (2011)	2	1	0	0	0	2	0

Study	Selection bias	Study design	Confounders	Blinding	Data collection	Withdrawal and drop-outs	Intervention integrity	Analysis	Overall
Gross et al. (2010)	2	1	2	2	2	2	2	2	2

0: does not meet design standards; 1: meets design standards with reservations; 2: meets design standards.

detail. All the experimental design studies described the dependent variable, but none operationally defined the dependent variable, described the data collection on these target behaviours AND gave a reason for targeting these behaviours.

Experimental design studies were appraised for additional phase and assessment descriptive quality indicators, as presented in Table VI. Overall, all four studies provided insufficient measures. Gross et al. (2010) provided minimal or sufficient information for most indicators, with the exception of maintenance. Martikainen and Korpilahti (2011) provided sufficient information on maintenance and raw data. Krauss and Galloway (1982) and Lagasse (2012) met or reported on none of the indicators that were assessed.

The single-subject design studies were additionally appraised on basic design standards quality indicators. An overview is presented in Table VII. Overall, the three single-subject design studies did not meet basic design standards. They had a limited amount of data points and attempts to present an effect. The studies by Krauss and Galloway (1982) and Martikainen and Korpilahti (2011) were not blinded, and inter-rater reliability was not reported by Krauss and Galloway (1982) and Lagasse (2012).

The cohort study by Gross et al. (2010) does meet design standards, as appraised with the EPHPP-tool described in the methods section, scoring strong on all indicators except for study design, where the score is moderate, considering this is not a Randomised Controlled Trial or Controlled Clinical Trial.

Discussion

The purpose of this study was to review existing literature on the effect of music or musical elements in the treatment of childhood SSDs, and the working mechanisms explaining potential effects. A broad search yielded four studies with an experimental design and four case studies. Seven out of eight articles reported positive outcomes after interventions including music or musical elements in the treatment of SSDs. However, the evidence supporting these claims is limited, since only one of the experimental studies had a strong methodological quality. Additionally, the significant effects in this study by Gross et al. (2010) were on measurements regarding language and cognition rather than speech. This review shows that evidence of effectiveness on the use of music in the treatment of childhood SSDs is limited.

Defining speech and language

The articles included in this review describe speech and language disorders in various ways. Based on inclusion criteria this review intended to only include children with SSDs. During the review process it became clear that "speech" was defined differently across studies regarding participant inclusion, intervention choice, measurements and outcome descriptions. All four studies from a speech-language pathology perspective included children with SSDs, used an intervention aimed at speech and included outcome measurements or provided outcome descriptions relating to speech.

However, three of the studies by Music Therapists had a broader focus in speech and language disorders, intervention and outcomes. Yet, all these studies came to conclusions on "speech" or "speech development".

The study by Lagasse (2012) is an exception among the articles by Music Therapists regarding the specificity of the description of "speech" in participants, interventions and outcome measurements. In this study, there was collaboration between a Music Therapist and a SLP, with an intervention (MIT) and outcome measurements both at the level of articulation. This is an example of how collaboration across both fields can help to improve the specificity of the speech and language disorders addressed and the measurements chosen to operationalise "speech" in studies that use Music Therapy as an intervention in the treatment of SSDs.

Participant description

All studies included speech- and language disorders of the participants, but only one described the criteria that were used to reach this diagnosis. Several others provided a description of symptoms to support the diagnosis. In the case of CAS, the diagnosis in five of the studies in this review, stating just the diagnosis is not sufficient, since there is debate on the diagnostic criteria (Iuzzini-Seigel & Murray, 2017). In the case of CAS, but also in other diagnoses, clear description of diagnostic criteria in future studies could enable comparison of participants across studies and allow clinicians to assess the relevance of the study for their patients.

Since there are no methods available to objectively determine the severity of SSDs, this was reported in none of the studies. Five studies gave a subjective description of the severity of the speech and language disorder but did not include this factor in their discussion of treatment outcomes. However, severity can play a major role in the outcome of interventions. Variation in severity of childhood speech disorders is one of the factors that contribute to contradictory research findings (Crosbie, Holm, & Dodd, 2005). Future inclusions of severity in the description of SSDs could possibly clarify these contradictions. Indicators of severity that can be used for both research and clinical practice are speech accuracy, intelligibility in different contexts, and children's perception of their own speech (Van Doornik, Gerrits, Terband, & McLeod, 2020).

Intervention description

The studies in this review were difficult to compare, due to variation in applications of methods and treatment dose. MIT was the most prevalent method used in the studies described, even though all studies using MIT made alterations to the method. While all these studies could be summarised as using MIT, this would be inaccurate. These different applications of MIT not only lead to difficulties comparing studies, but also make it difficult for clinicians to assess the applicability of MIT for their patients. Ideally, studies on specific methods should apply the same treatment protocol, to allow for adequate comparison and application in clinical practice. Where use of an original protocol is not possible, deviations should be clearly stated and conclusions on the efficacy of the method should be drawn with caution.

Reports on dose were incomplete, with dose (teaching episodes per session; Warren, Fey, & Yoder, 2007) not being reported in any of the studies and therefore replaced with session time in the summary. The reported total intervention durations varied widely. Dose could be an important feature in the effect of intervention in specific SSDs, such as CAS (Namasivayam et al., 2015). Clear descriptions on dose may contribute to clarifying differences in treatment outcomes and provide directions for clinical application of an intervention.

Selection of outcome measurements

The ICF level of interventions and outcome measurements was reported in none of the studies. From judgement, it was clear that all interventions focus on levels of body function and structures, and/or activities. All outcome measurements described speech (or language) at the level of functioning. It is unclear if measured improvement has led to improved communication in daily life and improved participation in social situations and/or school. Case studies reported treatment outcomes at the level of participation, but these statements are not supported by data. While outcome measurements at the level of functioning provide insight in the effect of an intervention, the ultimate goal in clinical practice is to improve children's participation in daily life. Measurements at this level should therefore be included in effect studies on speech and language interventions.

While there is measured improvement in the single-subject design studies in this review, the methodological quality of these studies is insufficient. Studies did not include enough data points, providing insufficient support for reported positive results. In the study of Martikainen and Korpilahti (2011) for example, most significant improvements seem to occur during no-treatment phases and positive trends directly after MIT treatment cannot be clearly separated from baseline trends. Inclusion of sufficient data points in single-subject design studies is necessary to provide clear evidence for any outcome.

The methodological quality of the cohort study by Gross et al. (2010), however, is adequate. With measurements focussing, in our opinion, on indicators of language processing and psychological development, it is unclear whether the treatment with music therapy influenced (motor) speech abilities. However, the measures of language processing and psychological development yielded measurable improvement, leading to conclusions by the authors of improvement in speech development. In a review of the outcomes of music-based interventions in adult patients with aphasia, Zumbansen and Tremblay (2019) reported that these interventions were most likely to yield positive outcomes for patients with motor speech disorders. Positive outcomes were also found on language measures, but mostly in patients with motor speech disorder. As Terband, Maassen, et al. (2019) reported, speech processes are interdependent and a limitation in one process affects development in adjacent processes. It is likely that improvement in one process will lead to improvement in adjacent processes. While language improvement through music-based intervention is thus likely to co-occur with speech improvement, studies in children that include both measures for (motor) speech abilities and language abilities are needed to see if music-based interventions mostly influence motor speech disorders in children with SSDs as reported in adults with acquired speech and language disorders.

Identification of working mechanisms

Working mechanisms were not examined in the studies included in this review. Five of the studies interpreted their findings, with three studies mostly focussing on beneficial effects on speech production of slowing speech rate and supporting rhythm and stress through MIT. Two studies hypothesised mainly about the beneficial effects of music and music therapy on (joint) attention, interaction and perception of sounds and musical structures. The focus of the hypotheses in these studies appears to be influenced by the researcher's background. Theories regarding speech rate and rhythm were proposed by SLPs and theories regarding the beneficial effects of music were suggested by Music Therapists. As Merrett et al. (2014) concluded in their review of MIT, these theories should not be seen as opposing explanations regarding the working mechanism, but rather as contributing mechanisms at different levels. While it is to be expected that this also applies to the other interventions using music, further research is needed to understand the working mechanisms of music in the treatment of SSDs.

Six studies in this review included participants with CAS, suggesting that music-based interventions are mostly applied in the treatment of speech disorders at the level of planning and programming. This

use is supported by evidence that the music-based intervention SMTA improved accuracy, consistency and fluency of articulation in adults with AoS (Hurkmans et al., 2015), since CAS and AoS share characteristics such as inconsistency and inappropriate prosody. Speech and music are highly associated in prosody, e.g. stress is expressed by a combination of pitch, loudness and duration (Terband, Namasivayam, et al., 2019) that relate to the musical elements melody, dynamics (volume) and rhythm, respectively (Hurkmans et al., 2015). While there is debate on how these associations support speech rehabilitation, there is growing evidence that these associations contribute to the effect of music-based interventions (Merrett et al., 2014).

There is growing evidence of an overlap in neural processing of speech and language (e.g. Brown, Martinez, & Parsons, 2006). The OPERA-hypothesis is based around the idea that through this overlap, musical training can improve speech processing (Patel, 2014). Future research may allow us to identify the working mechanism responsible for the effect of music on speech production. Implementing and testing (combinations of) hypotheses on music in the treatment of SSD could contribute to this goal and will additionally contribute to improving efficacy of treatment for children with SSD.

Conclusion

While most studies in this review on musical elements in the treatment of childhood SSDs included children with CAS, the evidence for the effectiveness of this kind of interventions with this group is insufficient. Studies with a high-quality study design, comparing different kinds of interventions and outcome measures specific to (motor) speech abilities are necessary to provide more insight in the effect of music in the treatment of SSD and CAS specifically. Implementing various hypotheses on working mechanisms of music in the treatment of SSD in such studies could contribute to the identification of these mechanisms.

Acknowledgements

The authors thank Truus van Ittersum (Research Institute Share, University Medical Centre Groningen) for the literature search. The authors thank Marja Meinsma (Rehabilitation Centre "Revalidatie Friesland") for comments on the manuscript.

Declaration of interest

The authors have no potential competing interests to report.

Supplemental material

Supplemental data for this article can be accessed at https://doi.org/10.1080/17549507.2022.2097310

ORCID

Mirjam van Tellingen http://orcid.org/0000-0001-7559-7209
Joost Hurkmans http://orcid.org/0000-0003-4155-9882
Hayo Terband http://orcid.org/0000-0001-7265-3711
Roel Jonkers http://orcid.org/0000-0002-0223-1596
Ben Maassen http://orcid.org/0000-0002-3330-1462

References

American Speech-Language-Hearing Association. (2020). *Understanding research designs and external scientific evidence*. ASHA.org. https://www.asha.org/research/ebp/research-designs/

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech [Technical report]*. American Speech-Language-Hearing Association. www.asha.org/policy

Ballard, K.J., Robin, D.A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 53, 1227–1245. doi:10.1044/1092-4388(2010/09-0130)

Bashir, A., Grahamjones, F., & Bostwick, R. (1984). A touch-cue method of therapy for sevelopmental verbal apraxia. *Seminars in Speech and Language*, 5, 127–137. doi:10.1055/s-0028-1082519

Beathard, B., & Krout, R.E. (2008). A music therapy clinical case study of a girl with childhood apraxia of speech: Finding Lily's voice. *The Arts in Psychotherapy*, 35, 107–116. doi:10.1016/j.aip.2008.01.004

Boles, M.B. (2015). *Effectiveness and quality of employment skill interventions for individuals with developmental disabilities: A meta-analysis, quality review, and single-case analysis*. College Station, TX: Texas A&M University.

Brown, S., Martinez, M.J., & Parsons, L.M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, 23, 2791–2803. doi:10.1111/j.1460-9568.2006.04785.x

Catrini, M., & Lier-DeVitto, M.F. (2019). Apraxia of speech and language delay: The complexity of diagnosis and treatment of symptomatic children. *Codas*, 31, e20180121. doi:10.1590/2317-1782/20192018121

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language & Communication Disorders*, 40, 467–491. doi:10.1080/13682820500126049

Dauer, K. E., Irwin, S. S., & Schippits, S. R. (1996). *Becoming verbal and intelligible: A functional motor programming approach of children with Developmental Verbal Apraxia*. Austin: Pro-Ed.

Davis, B.L., Jakielski, K.J., & Marquardt, T.P. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics and Phonetics*, 12, 25–45. doi:10.3109/02699209808985211

De Bruijn, M., Zielman, T., & Hurkmans, J.J.S. (2005). *Speech-music therapy for aphasia (SMTA)*. Beetsterzwaag, Netherlands: Revalidatie Friesland.

Dodd, B. (2011). Differentiating speech delay from disorder: Does it matter? *Topics in Language Disorders*, 31, 96–111. doi:10.1097/TLD.0b013e318217b66a

Fujii, S., & Wan, C.Y. (2014). The role of rhythm in speech and language rehabilitation: The SEP hypothesis. *Frontiers in Human Neuroscience*, 8, 777. doi:10.3389/fnhum.2014.00777

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe test of articulation* (2nd ed.). Circle Pines: American Guidance Service.

Gottsleben, R., & Tyack, D. (1974). *Language sampling, analysis and training*. Palo Alto: Consulting Psychologists Press.

Grimm, H., Aktas, M., & Frevert, S. (2001). *SETK 3-5 Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose*

von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen. Gottingen: Hogrefe-Verlag.

Gross, W., Linden, U., & Ostermann, T. (2010). Effects of music therapy in the treatment of children with delayed speech development—results of a pilot study. *BMC Complement Altern Med*, 10, 39. doi:10.1186/1472-6882-10-39

Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. *International Journal of Music Education*, 28, 269–289. doi:10.1177/0255761410370658

Hanser, S.B. (1999). *The new music therapist's handbook*. Thomann: Berklee Press.

Hayden, D., Eigen, J., Walker, A., & Olsen, L. (2010). PROMPT: A tactually grounded model for the treatment of childhood speech production disorders. In L. Williams, S. McLeod, & R. McCauley (Eds). *Treatment for speech sound disorders in children* (pp. 453–474). Baltimore, MD: Brookes.

Helfrich-Miller, K.R. (1984). Melodic intonation therapy with developmentally apraxic children. *Seminars in Speech and Language*, 5, 119–126. doi:10.1055/s-0028-1082518

Helfrich-Miller, K.R. (1994). A clinical perspective: Melodic intonation therapy for developmental apraxia. *Clin Commun Disord*, 4, 175–182. PMID: 7994292.

Hodson, B., & Paden, E. (1991). *Targeting intelligible speech: A phonological approach to remediation* (2nd ed.). Austin, TX: ProEd.

Hurkmans, J., de Bruijn, M., Boonstra, A.M., Jonkers, R., Bastiaanse, R., Arendzen, H., & Reinders-Messelink, H.A. (2012). Music in the treatment of neurological language and speech disorders: A systematic review. *Aphasiology*, 26, 1–19. doi:10.1080/02687038.2011.602514

Hurkmans, D., De Bruijn, M., Reitsma, T., & Koek, P. (2018). *Speech-music therapy for aphasia (e-learning)*. Beetsterzwaag, Netherlands: Revalidatie Friesland.

Hurkmans, J.J.S. (2016). *The treatment of apraxia of speech*. Groningen, Netherlands: Rijksuniversiteit Groningen.

Hurkmans, J., Jonkers, R., de Bruijn, M., Boonstra, A.M., Hartman, P.P., Arendzen, H., & Reinders-Messelink, H.A. (2015). The effectiveness of Speech–Music Therapy for Aphasia (SMTA) in five speakers with Apraxia of Speech and aphasia. *Aphasiology*, 29, 939–964. doi:10.1080/02687038.2015.1006565

Iuzzini-Seigel, J., & Murray, E. (2017). Speech assessment in children with childhood apraxia of speech. *Perspectives of the ASHA Special Interest Groups*, 2, 47–60. doi:10.1044/persp2.SIG2.47

Kallan, C.A. (1972). Rhythm and sequencing in an intersensory approach to learning disability. *Journal of Learning Disabilities*, 5, 68–74. doi:10.1177/002221947200500202

Khan, L., & Lewis, N. (2002). *Khan-Lewis phonological analysis* (2nd ed). Circle Pines: American Guidance Service.

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. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T.R., Hitchcock, J.H., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M., & Shadish, W.R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34, 26–38. doi:10.1177/07419325

Krauss, T., & Galloway, H. (1982). Melodic intonation therapy with language delayed apraxic children. *Journal of Music Therapy*, 19, 102–113. doi:10.1093/jmt/19.2.102

Lagasse, B. (2012). Evaluation of melodic intonation therapy for developmental apraxia of speech. *Music Therapy Perspectives*, 30, 49–55. doi:10.1093/mtp/30.1.49

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, 62, e1–e34. doi:10.1016/j.jclinepi.2009.06.006

Mahoney, J. (2010). Interrater agreement on the Nordoff-Robbins evaluation scale 1: Client-therapist relationship in musical activity. *Music and Medicine*, 2, 23–28. doi:http://dx.doi.org/10.47513/mmd.v2i1.238

Martikainen, A. L., & Korpilahti, P. (2011). Intervention for childhood apraxia of speech: A single-case study. *Child Language Teaching and Therapy*, 27, 9–20. doi:10.1177/0265659010369985

McLeod, S., & Baker, E. (2017). *Children's speech: An evidence-based approach to assessment and intervention*. London: Pearson.

Merrett, D.L., Peretz, I., & Wilson, S.J. (2014). Neurobiological, cognitive, and emotional mechanisms in melodic intonation therapy. *Frontiers in Human Neuroscience*, 8, 401. doi:10.3389/fnhum.2014.00401

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G, & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6, e1000097. doi:10.1371/journal.pmed.1000097

Namasivayam, A.K., Coleman, D., O'Dwyer, A., & van Lieshout, P. (2019). Speech sound disorders in children: An articulatory phonology perspective. *Frontiers in Psychology*, 10, 2998. doi:10.3389/fpsyg.2019.02998

Namasivayam, A.K., Pukonen, M., Goshulak, D., Hard, J., Rudzicz, F., Rietveld, T., … van Lieshout, P. (2015). Treatment intensity and childhood apraxia of speech: Treatment intensity and childhood apraxia of speech. *International Journal of Language & Communication Disorders*, 50, 529–546. doi:10.1111/1460-6984.12154

Patel, A.D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108. doi:10.1016/j.heares.2013.08.011

Porch, B. (1974). *Porch index of communicative ability in children*. Palo Alto: Consulting Psychologists Press.

Remes, K., & Ojanen, A. (1996). *Artikulaatiotesti: Äänteenmukainen sanakuvatesti [Test of articulation: Picture naming articulation test]*. Helsinki: Early Learning Oy.

Rosenbek, J., Hansen, R., Baughman, C.H., & Lemme, M. (1974). Treatment of developmental apraxia of speech: A case study. *Language, Speech, and Hearing Services in Schools*, 5, 13–22. doi:10.1044/0161-1461.0501.13

Rosenthal, J., Williams, R., & Ingham, R.J. (1981). *An experimental therapy programme for developmental articulatory dyspraxia*. Launceston Tasmania: Annual Convention of the Australian Association of Speech and Hearing.

Ruben, R.J. (2000). Redefining the survival of the fittest: Communication disorders in the 21st century: Redefining the survival of the fittest: Communication disorders in the 21st century. *The Laryngoscope*, 110, 241–241. doi:10.1097/00005537-200002010-00010

Shriberg, L.D. (2010). Childhood speech sound disorders: From post-behaviorism to the post-genomic era. In R. Paul, & P. Flipsen (Eds.), *Speech sound disorders in children* (pp. 1–34). San Diego, CA: Plural Publishing.

Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, 10, 303–316. doi:10.1016/S0010-9452(74)80024-9

Sparks, R.W., & Holland, A.L. (1976). Method: Melodic intonation therapy for aphasia. *Journal of Speech, Language, and Hearing Research*, 41, 287–297. doi:10.1044/jshd.4103.287

Strand, E.A. (2020). Dynamic temporal and tactile cueing: A treatment strategy for childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 29, 30–48. doi:10.1044/2019_AJSLP-19-0005

Sutton, J. (1993). The guitar doesn't know this song: An investigation of parallel development in speech/language and music therapy. In M. H. Heal & T. Wigram (Eds.), *Music therapy in*

health and education (pp. 264–272). London: Jessica Kingsley Publishers; APA PsycInfo. http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=1995-98464-021&site=ehost-live&scope=site

Tate, R.L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item risk of bias in N-of-1 trials (RoBiNT) scale. *Neuropsychological Rehabilitation*, 23, 619–638. doi:10.1080/09602011.2013.824383

Tellegen, P., Laros, J., & Petermann, F. (2002). *SON-R 2 1/2 - 7. Nonverbaler Intelligenztest*. Hogrefe-Verlag.

Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica et Logopaedica*, 71, 216–227. doi:10.1159/000499426

Terband, H., Namasivayam, A., Maas, E., van Brenk, F., Mailend, M.-L., Diepeveen, S., … Maassen, B. (2019). Assessment of childhood apraxia of speech: A review/tutorial of objective measurement techniques. *Journal of Speech, Language, and Hearing Research*, 62, 2999–3032. doi:10.1044/2019_JSLHR-S-CSMC7-19-0214

Thomas, B.H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, 1, 176–184. doi:10.1111/j.1524-475X.2004.04006.x

Thoonen, G., Maassen, B., Gabreëls, F., Schreuder, R., & de Swart, B. (1997). Towards a standardised assessment procedure for developmental apraxia of speech. *International Journal of Language & Communication Disorders*, 32, 37–60. doi:10.3109/13682829709021455

Van Doornik, A., Gerrits, E., Terband, H., & McLeod, S. (2020). *Severity of speech sound disorders (SSD)*. [Poster] Conference on Motor Speech 2020, Santa Barbara, CA.

Van Tellingen, M., Hurkmans, J., Van der Zande, A.M., Terband, H., Meinsma, M., Maassen, B., & Jonkers, R. (2022). *Speech and Music Therapy in the treatment of Childhood Apraxia of Speech: A case study.* [Poster] European Academy of Childhood Disability, Barcelona.

Wambaugh, J.L., & Martinez, A.L. (2000). Effects of rate and rhythm control treatment on consonant production accuracy in apraxia of speech. *Aphasiology*, 14, 851–871. doi:10.1080/026870300412232

Warren, S.F., Fey, M.E., & Yoder, P.J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, 13, 70–77. doi:10.1002/mrdd.20139

World Health Organization (Red.). (2007). *International classification of functioning, disability and health: Children & youth version; ICF-CY.* Geneva, Switzerland: World Health Organization.

Ziegler, W. (2008). Apraxia of speech. In G. Goldenberg, & B. Miller (Eds.), *Handbook of clinical neurology.* (pp. 269–285). Amsterdam, Netherlands: Elsevier.

Zumbansen, A., & Tremblay, P. (2019). Music-based interventions for aphasia could act through a motor-speech mechanism: A systematic review and case–control analysis of published individual participant data. *Aphasiology*, 33, 466–497. doi:10.1080/02687038.2018.1506089

Appendix A: Search strings

PubMed

((“Speech Disorders”[Mesh] OR “Apraxias”[Mesh] OR (speech[tiab] AND (problem*[tiab] OR disorder[tiab] OR impair*[tiab])) OR aprax*[tiab] OR mutism[tiab] OR aphasia[tiab] OR articulation[tiab] OR stuttering[tiab] OR dysphasia[tiab] OR dysphonia[tiab] OR echolalia[tiab] OR mutism[tiab]) AND (“Music Therapy”[Mesh] OR ((music[tiab] OR melodic[tiab] OR singing[tiab] OR rhythm[tiab]) AND (intervention*[tiab] OR therap*[tiab] OR treatment[tiab])))) AND (“Child”[Mesh] OR “Infant”[Mesh] OR child*[tiab] OR infant*[tiab])

The search strings for the databases PsycINFO, AMED, ERIC, Cinahl and Web of Science are available in the Supplementary Materials.

Appendix B: Quality appraisal guidelines

Descriptive quality indicators

Participant description	Setting description	Interventionist description	Baseline and intervention procedure description	Dependent variable description	Overall score
0 = Does NOT include all 3 of the main participant characteristics: 1. Age 2. Gender 3. Diagnosis.	0 = Includes NO description or only 1 detail of the setting (i.e. only materials involved, only location or only presence of other individuals in the setting).	0 = No description of the interventionist is provided.	0 = Both baseline and intervention procedures are described in too little detail to create any accurate replication **OR** description is not included for either baseline or intervention phase.	0 = The target behaviours are not operationally defined and the procedures for data collection are not described thoroughly enough for replication.	0 = Insufficient description (scores 0 on one or more of the indicators).
1 = Includes the age (may include age range), gender, and the primary diagnosis of each participant.	1 = Includes the location of the setting (i.e. classroom, home) as well as other individuals present (i.e. teacher, family members) **OR** the the materials involved (**A total of 2 setting details**).	1 = includes either the inventionist’s occupation/ relationship to the participant (i.e. teacher, parent, researcher, therapist); **OR** the interventionist’s expertise (i.e. training, experience).	1 = Describes most of the elements of the procedures in sufficient detail (e.g. replicable detail for materials used and assessed behaviours, but the session time limit is not given) **OR** only one phase (either baseline or intervention) is described with sufficient detail but the other phase does not allow for accurate replication.	1 = Either the target behaviours are not operationally defined, the data collection procedures are not described thorough enough for replication, **OR** the reason for targeting certain behaviours is not given.	1 = Minimal description (scores a 1 on one or more of the indicators and no 0).

(*Continued*)

(*Continued*).

Participant description	Setting description	Interventionist description	Baseline and intervention procedure description	Dependent variable description	Overall score
2 = Includes the age (specific age per participant), gender, primary and other diagnoses if applicable, inclusion criteria for participants, and other relevant characteristics (i.e. IQ, skill deficits, previous therapy/ training).	2 = Includes the materials involved in the setting, presence of other individuals and location.	2 = Includes both interventionist's occupation/ relationship to participant and the interventionist's expertise.	2 = Includes thorough descriptions of both the baseline and intervention procedures (i.e. materials used, session time limit, steps for implementation, behaviours of the interventionist, and data collection) to allow for accurate replication.	2 = Each target behaviour is operationally defined, the procedures for taking data on these target behaviours are described thoroughly and there is a reason given for targeting these behaviours.	2 = Sufficient description (scores a 2 on all indicators).

Single case experimental design basic quality design standards

IV/Intervention	Inter-rater reliability	Effect	Data points	Blinding	Data analysis	Overall score
0 = There is no purposeful manipulation of the independent variable/ intervention.	0 = IRR is not reported or IRR is less than 70%.	0 = There are less than 3 attempts to present an effect at 3 separate points in time or less than 6 minimum phases for MBD or 4 phases for reversal designs.	0 = contains less than 3 data points per phase.	0 = Assessor is not independent to therapist and intervention phase.	0 = no visual or statistical analysis conducted.	0 = does not meet design standards (scores 0 in one or more design indicators).
	1 = IRR is 70–79%.		1 = Contains 3–4 data points per phase.	1 = Assessor is independent to therapist but not to intervention phase.	1 = Systematic/aided visual analysis is incomplete/not conducted for every phase change OR no rationale is provided for statistical analysis.	1 = Meets design standards with reservations (scores a 1 on at least one design indicator).
2 = There is purposeful manipulation of the independent variable/ intervention.	2 = IRR is sampled in at least 20% of data and reaches at least 80% agreement.	2 = There are 3 or more attempts to present an effect at 3 separate points in time AND at least 6 minimum phases for MBD or 4 phases for reversal design.	2 = Contains 5 or more data points per phase.	2 = Assessor is independent to therapist and blinded to intervention phase.	2 = Visual analysis according to steps by Kratochwill et al. (2010; 2013), OR visual analysis aided by quasi statistical techniques OR statistical methods with rationale for their suitability.	2 = Meets design standards (scores a 2 on all design indicators).

Additional Phase and Assessment Descriptive quality indicators

Maintenance	Generalisation	Raw data	Fidelity	Social validity	Overall score
0 = Maintenance phase or maintenance data is NOT reported.	0 = Generalisation phase or generalisation data is NOT reported.	0 = Ad hoc selection of data.	0 = Fidelity is recorded for less than 20% of overall data (or is not reported) and/or overall fidelity scores are less than 80%; or fidelity measures are NOT reported.	0 = Social validity is NOT reported or only includes one out of the 5[a] components of social validity.	0 = Insufficient Measure (scores a 0 in two or more of the indicators).
1 = Maintenance data is collected after the intervention is implemented **AND** there are less than 3 data points in at least one maintenance phase **OR** all maintenance probes are recorded within one month or less from the conclusion of the intervention phase.	1 = Generalisation data is collected only after the intervention is implemented (i.e. no generalisation data is taken during baseline or intervention phases) OR there are less than 3 total data points in at least one generalisation phase.	1 = Incomplete data such as multiple probe designs, aggregated data.	1 = Procedural/ Treatment fidelity is recorded for at least 20% of overall data recorded; overall fidelity scores are 80% or above; and fidelity is only recorded in either the baseline phase or intervention phase.	1 = Social validity measure includes at least 2 out of the 5[a] components of social validity.	1 = Minimal Measure (scores a 0 in only one of the indicators AND scores a 1 or higher for the other indicators).
2 = Maintenance data is collected after the intervention is implemented, there are 3 or more data points in each maintenance phase, and all maintenance probes are recorded for more than 1 month from the conclusion of the intervention phase.	2 = Generalisation data is recorded in baseline and intervention phases (in addition, there can be a specified generalisation phase at the end of the intervention) and there is a total of 3 or more data points per generalisation phase.	2 = Complete record of raw data at a session-by-session level.	2 = Procedural/ Treatment fidelity is recorded for at least 20% of data in each condition with overall scores of 80% or above; and fidelity is recorded for procedures in both baseline and intervention phases.	2 = Social validity measure includes at least 4 out of the 5[a] components of social validity.	2 = Sufficient Measure (scores a 1 in one of the indicators AND scores a 2 in the other indicators).

[a]Social validity components: (1) social significance of the dependent variables (i.e. the target behaviours are beneficial to the participant and relevant to the context), (2) the intervention was efficient and cost effective, (3) the change in behaviour or intervention effects were significant according to the criterion or goals set for individual studies, (4) all individuals involved are satisfied with the procedures and outcomes and (5) the intervention contains a natural component (i.e. the interventionist is an individual that is present in the participant's natural setting, or the intervention is implemented in the natural setting).

Quality assessment tool for quantitative studies (EPHPP)

Selection bias	Study design	Confounders	Blinding	Data collection methods	Withdrawal/ drop-outs	Intervention integrity	Analyses	Overall score
0 = The selected individuals are not likely to be representative of the target population (Q1 is 3); or there is less than 60% participation (Q2 is 3) or selection is not described (Q1 is 4); and the level of participation is not described (Q2 is 5).	0 = will be assigned to those that used any other method or did not state the method used.	0 = will be assigned when less than 60% of relevant confounders were controlled (Q1 is 1) and (Q2 is 3) or control of confounders was not described (Q1 is 3) and (Q2 is 4).	0 = The outcome assessor is aware of the intervention status of participants (Q1 is 1); and the study participants are aware of the research question (Q2 is 1); or blinding is not described (Q1 is 3 and Q2 is 3).	0 = The data collection tools have not been shown to be valid (Q1 is 2) or both reliability and validity are not described (Q1 is 3 and Q2 is 3).	0 = will be assigned when a follow-up rate is less than 60% (Q2 is 3) or if the withdrawals and drop-outs were not described (Q1 is No or Q2 is 4).	0 = Very likely that subjects received an unintended intervention (contamination or co-intervention; Q3 is yes) **OR** Less than 60% of participants received the allocated intervention (Q1 = 3 or 4).	0 = Analysis are not appropriate (Q3 is no) **AND** analyses are not performed with intention to treat.	0 = weak (two or more weak ratings)
1 = The selected individuals are at least somewhat likely to be representative of the target population (Q1 is 1 or 2); and there is 60–79% participation (Q2 is 2). "Moderate" may also be assigned if Q1 is 1 or 2 and Q2 is 5 (cannot tell).	1 = will be assigned to a cohort analytic study, a case control study, a cohort design, or an interrupted time series.	1 = will be given to those studies that controlled for 60–79% of relevant confounders (Q1 is 1) and (Q2 is 2).	1 = The outcome assessor is not aware of the intervention status of participants (Q1 is 2); or the study participants are not aware of the research question (Q2 is 2).	1 = Data collection tools have been shown to be valid (Q1 is 1); data collection tools have not been shown to be reliable (Q2 is 2) or reliability is not described (Q2 is 3).	1 = will be assigned when the follow-up rate is 60–79% (Q2 is 2) OR Q1 is 4 or Q2 is 5.	1 = Unclear if subjects received an unintended intervention (q3 = cannot tell), **OR** consistency of intervention was not measured (Q2 is no) **AND** Q1 is 1 or 2.	1 = Statistical analyses are appropriate (Q3 is yes), intention to treat is NOT applied or unclear.	1 = moderate (one weak rating)
2 = The selected individuals are very likely to be representative of the target population (Q1 is 1) and there is greater than 80% participation (Q2 is 1).	2 = will be assigned to those articles that described RCTs and CCTs.	2 = will be assigned to those articles that controlled for at least 80% of relevant confounders (Q1 is 2); or (Q2 is 1).	2 = The outcome assessor is not aware of the intervention status of participants (Q1 is 2); and the study participants are not aware of the research question (Q2 is 2).	The data collection tools have been shown to be valid (Q1 is 1); and the data collection tools have been shown to be reliable (Q2 is 1).	2 = will be assigned when the follow-up rate is 80% or greater (Q1 is 1 and Q2 is 1).	2 = More than 80% of subjects received allocated inter-vention (Q1 is 1) **AND** consistency of intervention is measured **AND** subjects receiving unintended inter-vention is unlikely.	2 = Statistical analyses are appropriate **AND** performed with intention to treat.	2 = strong (no weak rating)