Constraint and Multimodal approaches to therapy for chronic aphasia: A systematic review and meta-analysis

John E. Pierce[a,b], Maya Menahemi-Falkov[a], Robyn O'Halloran[a], Leanne Togher[c] and Miranda L. Rose[a]*

[a]School of Allied Health, La Trobe University, Melbourne, Australia; [b]Speech Pathology, Cabrini Health, Melbourne, Australia; [c]Speech Pathology, Faculty of Health Sciences, The University of Sydney, Lidcombe, Australia

John E Pierce. 494 Glen Huntly Road, Elsternwick 3185, Australia. +61430 130 893. ORCID identifier: 0000-0001-5164-5106 Twitter:@johnpierce85 pierce.john.e@gmail.com

Maya Menahemi-Falkov. Health Sciences 1, La Trobe University, Plenty Road, Bundoora 3083, Australia. mayam10@gmail.com

Robyn O'Halloran. Health Sciences 1, La Trobe University, Plenty Road, Bundoora 3083, Australia. ORCID identifier: 0000-0002-2772-2164 R.OHalloran@latrobe.edu.au

Leanne Togher. Faculty of Health Sciences, The University of Sydney, 75 East St, Lidcombe 2141, Australia. ORCID identifier: 0000-0002-4518-6748 Leanne.togher@sydney.edu.au

*Miranda L. Rose. Health Sciences 1, La Trobe University, Plenty Road, Bundoora 3083, Australia. ORCID identifier: 0000-0002-8892-0965 M.rose@latrobe.edu.au

Disclosure of interest: The authors report no conflicts of interest

Constraint and Multimodal approaches to therapy for chronic aphasia: A systematic review and meta-analysis

Aphasia is a significant cause of disability and reduced quality of life. Two speech pathology treatment approaches appear efficacious; Multimodal and Constraint Induced aphasia therapies. In Constraint Induced therapies, nonverbal actions (e.g. gesture, drawing) are believed to interfere with treatment and patients are therefore constrained to speech. In contrast, Multimodal therapies employ non-verbal modalities to *cue* word retrieval. Given the clinical and theoretical implications, a comparison of these two divergent treatments was sought. This systematic review investigated both approaches in chronic aphasia at the levels of impairment, participation and quality of life. After a systematic search, the level of evidence and methodological quality were rated. Meta-analysis was conducted on 14 single case experimental designs using Tau-U, while heterogeneity in the four group designs precluded meta-analysis. Results showed that high-quality research was limited; however, findings were broadly positive for both approaches with neither being judged as clearly superior. Most studies examined impairment-based outcomes without considering participation or quality of life. The application and definition of constraint varied significantly between studies. Both constraint and multimodal therapies are promising for chronic post-stroke aphasia, but there is a need for larger, more rigorously conducted studies. The interpretation of "constraint" also requires clearer reporting.

Keywords: systematic review; aphasia; constraint; multimodal; therapy

Word count: 8050 word (inc. citations)

Introduction

The presence of aphasia after stroke results in significantly poorer quality of life than stroke alone (Hilari, 2011). Aphasia is perceived as more detrimental for quality of life than any other illness including cancer, Alzheimer's Disease and quadriplegia (Lam & Wodchis, 2010). There are also significant financial and carer burdens (Flowers, Silver, Fang, Rochon, & Martino, 2013; Patrício, Jesus, & Cruice, 2013) and thus, effective treatments for aphasia are highly sought after.

Constraint and multimodal are two treatment approaches in aphasia. Multimodal treatments have a long history of use in aphasia research and treatment, while constraint principles were first introduced in 2001. However, these two approaches have theoretically distinct rationales. Authors of constraint therapies such as Constraint Induced Aphasia Therapy (CIAT) posit that the use of other communication modalities distract from, and therefore weaken, recovery of verbal output, while authors of multimodal therapies suggest that these additional modalities can facilitate word retrieval and learning. This results in a key difference between treatments regarding the cueing of patient responses. These two therapy approaches and their rationales will be described before their evidence is compared.

Multimodal

The concept of multimodal cueing in aphasia rehabilitation dates back at least to the 1940s, when Luria put forward the principle of *Intersystemic Reorganisation* (Luria, 1970). This principle proposes that a defective system can be supported and supplemented by another, less damaged, system. Luria gave the example of recruiting the visual system to compensate for impaired proprioception or balance during walking. In language, the impaired system (spoken word retrieval) might be assisted by gesture production, for example. It is important to note that in multimodality therapies, other modalities are used to *facilitate* spoken output and not to replace it, though compensation may be a secondary goal if improvement of spoken output is not successful. The most common communicative modalities used to promote word retrieval and their proposed mechanisms are described below.

Reading.

Orthographic cues, typically manifested as first letter cues, are widely used in aphasia speech pathology practice (Lorenz & Nickels, 2007). Research has found orthographic cues to be generally successful for spoken naming (Nickels, 2002). The orthographic cue is thought to provide complete or partial activation of the phonological form of the target word using grapheme-phoneme conversion (Lorenz & Nickels, 2007); in other words, the letter/s provide an internal phonemic cue to facilitate word production.

Writing.

As with reading, writing takes advantage of the deep links between graphemes and phonemes. When a patient writes all or part of a target word, this might provide an alternative route to the phonological output lexicon (a store of all phonological forms of known words) via the reading route(s) described above (Nickels, 1992). Alternatively, the connection between the orthographic and phonological output lexicons may be bidirectional, meaning that phonological information is active whenever written word forms are accessed, as well as vice versa (Kiran, 2005). This connection is proposed based on the fact that written naming alone can improve spoken naming (DeDe, Parris, & Waters, 2003; Wright, Marshall, Wilson, & Page, 2008).

Gesture.

The *lexical retrieval hypothesis* holds that, even in non-aphasic speakers, gestures used in conversation are more for the speaker's benefit than the listener's in that they aid in word retrieval (Beattie & Shovelton, 2006). Evidence for this hypothesis comes from findings that a) more gestures are produced during word finding difficulties in the speech of both normal and aphasic individuals, and b) restricting gestures during speech increases the frequency of dysfluencies (Rose, 2006). There is "intense theorising" about exactly how gestures assist with word finding with language (Hadar & Rumiati, 2006, p. 141), but two possible mechanisms are described here.

Language is traditionally viewed as a discrete system in the brain. *Embodied language*, a subset of *embodied cognition*, is the theory that language is connected to action and sensory systems and there is experimental support for this theory (Fischer & Zwaan, 2008). For example, one study showed that reading of "action" words was immediately followed by activation of a relevant motor cortex area, such as the word "kick" activating the leg motor areas (Pulvermüller & Berthier, 2008). It follows from such close motor-language connections that where a word cannot be produced, its corresponding gesture might aid retrieval. The other explanation is that gesture aids the speaker in preverbal message planning. That is, gesture helps the speaker mentally arrange the spatial and visual thoughts behind a message, and this stimulates processing of semantic features which assist word retrieval (Feyereisen, 2006).

Drawing.

In aphasia, drawing has typically been used to augment or compensate for speech loss (Sacchett, 2002). However, there are case reports of individuals using drawing to self-cue verbal word retrieval, or individuals whose naming has improved after treatment targeting drawing accuracy (Cubelli, 1995). Farias, Davis and Harrington (2006) explored drawing as a facilitator of word finding in 22 people with aphasia. While participants were drawing the target, confrontation naming improved compared to both baseline and attempts at written naming. Interestingly, this effect was not influenced by drawing quality (as measured by recognisability ratings), suggesting that it could be a suitable treatment even for those with limb apraxia or significant hemiplegia.

How would drawings assist word retrieval? Drawings, like gesture, have the advantage of being free from linguistic symbolism. Drawing quality in aphasia correlates with the integrity of the semantic system (Farias et al., 2006), which suggests that the process of drawing requires access to semantic features. For naming, drawing is thought to stimulate the semantic aspects of objects by placing attention on visual features (Farias et al., 2006) and may do so for longer than naming attempts alone (Makuuchi, Kaminaga, & Sugishita, 2003). For example, in drawing a truck, a person needs to focus more deeply on features such as its large size, square shape and additional wheels, and perhaps its function, than when merely naming it from sight. This process might suppress competing concepts whose features do not match the target while activating sufficient semantic features to assist word retrieval.

Music.

Melodic Intonation Therapy (MIT; Sparks, Helm, & Albert, 1974) is probably the most widely used and recognised music-related therapy for aphasia. MIT utilises “intoned speech”, a song-like prosody, for phrases and utterances. Notes can be either high or low depending on syllable stress, and syllables are rhythmicised. Simultaneous left hand tapping is also employed during word or phrase production (Zumbansen, Peretz, & Hébert, 2014a). There are other variations of music and rhythm treatments, including Modified MIT, Singen Intonation Prosodie Atmung Rhytmusübungen Improvationen (SIPARI) and Speech-Music

Therapy in Aphasia (Hurkmans et al., 2012). As with other multimodal treatments, all these treatments aim to encourage verbal output. Generalisation to untreated phrases in conversational speech is the ultimate goal and patients are not expected to sing in everyday life (Zumbansen, Peretz, & Hébert, 2014a).

There are several suggested mechanisms for melodic therapies. Sparks et al. (1974) were initially uncertain as to how to explain the positive results of MIT, but did not believe that the right hemisphere was learning to take over language production. Instead, they proposed that the right hemisphere was assisting the left hemisphere. Tapping in MIT is left handed for this reason — to encourage activation of the right hemisphere. Neuroimaging evidence is mixed for the theory of increased right hemisphere activation in MIT, with some showing increased perilesional left hemisphere activation and others, increased right hemisphere activation (Zumbansen, Peretz, & Hébert, 2014a). Even if present, right hemisphere activity might only increase due to the high intensity of "singing" that occurs during such treatments, with the improvements in speech due to the repetition of phrases. That is, right hemisphere activity and verbal improvement could be independent events (Stahl, Kotz, Henseler, Turner, & Geyer, 2011).

A more recent hypothesis is that the rhythmic component of these treatments, particularly the tapping of the left hand, is the true underlying mechanism (Stahl et al 2011). This has some experimental support (Zumbansen, Peretz, & Hébert, 2014a), while other studies found superiority of combined rhythm and pitch (Zumbansen, Peretz, & Hébert, 2014b).

Combined Multimodal

Multiple non-verbal modalities may be combined within a treatment to maximise cueing of verbal output. M-MAT (Multi-modality Aphasia Therapy) is a high intensity combined multimodal treatment (Attard, Rose, & Lanyon, 2013) which utilises a structured cueing hierarchy of gesture, drawing and writing to cue word retrieval. A structured and detailed protocol delineates that each time a patient is unable to produce a target, they are asked to gesture, draw and write or copy the word while repeating it verbally. M-MAT has shown improvements across receptive and expressive language measures, at both impairment and activity/participation levels (Rose, Attard, Mok, Lanyon, & Foster, 2013).

Constraint

At the heart of constraint therapy lies the concept of *learned nonuse*. Taub and colleagues developed the term based on observation of monkeys with deafferented upper limbs (Taub, 1976). They proposed that the deficit from an injury is not wholly due to the physiological impairment, but also a subconscious preference not to use the affected body part (Taub, Uswatte, Mark, & Morris, 2006). They hypothesised that nonuse is learned through (a) punishment when trying to use the affected limb (e.g. issues with incoordination or dropping), and (b) positive reinforcement when using the alternative limb. A "vicious spiral" (Taub et al., 2006, p. 245) of nonuse then commences as reduced use leads to shrinkage of the relevant cortical area, which results in less use of the limb, and so on. However, by constraining function to the damaged limb, through either restraining the unaffected limb or presenting tasks which necessitate the use of both limbs, improvements in the function of the deafferented upper limbs were seen.

Constraint Induced Movement Therapy (CIMT) was subsequently developed to address learned nonuse in adult stroke patients. The key elements of CIMT are intensive training, use of functional transfer tasks, and constraint applied to the affected limb (Taub et al., 2006). The efficacy of CIMT has been demonstrated extensively in upper and lower limbs in stroke as well as other conditions (Smania, 2006; Taub et al., 2006).

CIAT was developed through combining the principles of CIMT with an existing aphasia treatment, *Communicative Aphasia Therapy* (Pulvermüller & Berthier, 2008), which employs language games that rely on the correct verbal response of patients. First reported in 2001 (Pulvermüller et al., 2001), CIAT has three primary principles: 1. Massed practice over a short period, 2. Action-embedded, relevant language, and 3. Constraint to possible, but avoided, verbal output. It is the third principle that addresses learned nonuse. In the context of language, learned nonuse is proposed to occur where patients avoid problematic words or phrases, reduce their attempts at verbal communication overall, or use alternative modalities to compensate, such as gesture or writing. In CIAT, constraint is applied through progressive difficulty of stimuli, gradual shaping of responses into more complex utterances, and prohibiting nonverbal modes of communication.

More recently, the potentially detrimental effects of inhibiting multimodal self-cueing have been recognised in CIAT and its use has now been permitted (Difrancesco, Pulvermüller, & Mohr, 2012). Even so, any multimodal cues used within CIAT remain incidental and patient-generated, in contrast to the systematic, clinician-prompted cueing of multimodal treatments. A number of variations on the original CIAT protocol have been described, including CIAT II (Johnson et al., 2014) and CIAT Plus (Attard et al., 2013; Meinzer, Djundja, Barthel, Elbert, & Rockstroh, 2005). CIAT Plus builds upon CIAT by assigning home tasks in order to improve carryover of language skills into real life and using written stimuli as well as photographs (Meinzer et al., 2005). CIAT II uses a wider variety of language activities than CIAT, including a role-playing task, picture description and repetition drills (Johnson et al., 2014).

Summary

As outlined, constraint aphasia therapies seek to avoid learned nonuse of language partly through constraining participants to the verbal modality, whereas multimodal treatments seek to leverage intact modalities to aid verbal output.

The most recent Cochrane review of aphasia therapy called for further data comparing different therapies in order to identify the most effective treatments (Brady, Kelly, Godwin, Enderby, & Campbell, 2016). A comparison of constraint and multimodal approaches will improve outcomes for people with chronic aphasia if one is found to be more effective (Rose et al., 2013). If patients clearly respond better to being constrained to the verbal modality, then use of multimodal cues should be re-examined. Conversely, if patients improve more with multimodal cues, the contribution of verbal constraint in the CIAT protocol may be questioned.

CIAT and M-MAT are two treatments that are operationally similar despite being constraint and multimodal treatments, respectively. Both use group language games which provide a social imperative for successful communication, have a high intensity of treatment (30 hours over two weeks), and use shaping of responses to gradually increase the complexity of utterances. To date, CIAT and M-MAT have been directly compared in two studies. In a pilot study (Attard et al., 2013), a single case crossover design was used to provide CIAT Plus and M-MAT to two participants. Confrontation naming of treated items was marginally superior for M-MAT compared to CIAT Plus. In a phase 1 trial (Rose et al., 2013), 11 participants underwent both treatments in a multiple baseline crossover design. Both produced strong positive effects for the primary outcome measure (confrontation naming), with comparable mean effect sizes (M-MAT = 8.00, CIAT Plus = 8.58) according to Busk and Serlin's *d* (Busk & Serlin, 1992). These similar results in direct comparisons are puzzling given the contrasting nature of cues provided.

In a narrative review of constraint and multimodal treatments, Rose (2013) examined the theoretical explanations for constraint and multimodal treatments in detail as well as critically examining the literature to date. The review concluded that there was limited theoretical support for constraining people with aphasia to the verbal modality and that doing so was potentially counterproductive for word retrieval. The research examined within the review did not favour either approach. However, as a narrative review, the literature reviewed was not exhaustive and was not subjected to meta-analysis.Therefore, the purpose of this review was to systematically examine and compare the efficacy of constraint and multimodal therapies more broadly than CIAT Plus and M-MAT. There were three key questions. Two examined outcomes based on the World Health Organisation's *International Classification of Functioning, Disability and Health* (ICF). Quality of life was also considered, frequently described as a missing component of the ICF and under consideration for inclusion in future versions of the ICF (Ravenek, Skarakis-Doyle, Spaulding, Jenkins, & Doyle, 2015). Finally, the outcome of carer burden was included due to the significant negative impacts of aphasia on carers (Patrício et al., 2013).

> For stroke-induced chronic aphasia, what is the influence of constraint and multimodal treatments on measures of *(1) language impairment*, *(2) communication activity/participation*, and *(3) quality of life and carer burden*?

Definitions

Multimodal therapy or *multimodal training* has many variations in speech pathology and even within the field of aphasiology, so it is important to describe our definition for this systematic review. In this paper we use *multimodal therapy* to mean treatments which expect the patient to produce output in different modalities alongside speech production in order to cue speech. Our own operationalised definition of *multimodal therapy* also formed the basis for our inclusion criteria:

(1) The patient carries out a communication task (particularly writing, gesture, singing, or drawing) at the same time as or immediately before an attempt at speech.
(2) Speech targets are practised repeatedly in a therapy task with the intention of improving spontaneous speech in the long term rather than as a short-term facilitation effect.

Based on this definition, studies looking at a simultaneous action that is not communicative are not eligible. There is a series of research looking at the effects of "intention gesture", for example, where participants make a nonsymbolic, circular gesture with their left hand while repeating the target word. These are not eligible as the nonsymbolic gesture is designed to activate the right hemisphere and is not related to the target word in a linguistic sense. Outside of this restriction we included any gesture, as classified in McNeill's (1992) model which differentiates gesticulation, pantomime, emblems and sign languages.

MIT and other melodic treatments were included as per the above definition as they combine the modalities of singing and speech in a song-like prosody. Studies investigating the use of rhythm on language without a melodic component do not meet our definition as rhythm is not a communicative act in isolation.

Studies including orthographic cueing (i.e., reading the target word or part of it) as their only multimodal cue were not included as they are so commonly used as to include the majority of the aphasia literature. We also limited the writing studies to those focussing on writing or typing the whole target word and not those that taught phoneme-grapheme correspondences as an explicit strategy.

Table 1. Search strategy example – Medline (OVID, 1946-present)

Keywords:	aphasia, dysphasia, anomia, NOT "primary progressive"	**AND**	Therapy, Intervention, Treatment	**AND**	CIAT*, CILT*, "Constraint induced language", "constraint induced aphasia", "Constraint language", "Constraint aphasia", "Forced language treatment", ILAT, "intensive language action"	**OR**	M-MAT, MMAT, "Multi-modality aphasia therapy", multimodal*, cross modal*, Draw*, Amerind, Amer-ind, Pantomim*, Sign, signs, signing, Makaton, Gestur*, gestic*, iconic, Writ*, graphem*, orthograph*, Read*, music*, melod*, sing, singing, rhythm
	OR						
Subject headings:	*Aphasia*, *Anomia*, NOT *Aphasia, primary progressive*						

Note. **All subject headings were exploded, and all subheadings were included.**

Method

Search terms relating to aphasia, constraint and multimodal therapies (including drawing, writing, gesture and music) were used. An example of the search strategy employed is displayed in Table 1. Three major databases were searched in September 2015: Medline (OVID, 1946-present), CINAHL (Ebscohost) and Psycinfo (OVID, 1987-present). No limits for date of publication were applied as we wanted to find as many relevant articles as possible.

Duplicate results were excluded using citation software and manual checking. Resulting titles and abstracts were then screened by each of the first two authors as per the following inclusion criteria:

- Original data
- Experimental design
- Article from peer reviewed journal, text in English
- Adults (18 years +) with stroke-induced, chronic aphasia (≥6 months)
- Uses constraint to verbal output *or* uses multimodal therapy (see earlier definition)
- Purposefully investigates effects on verbal output

If a study included participants who did not meet the demographic criteria, it was excluded unless results for eligible participants could be separated.

After the screening process, the first two authors met to achieve consensus on discrepancies regarding inclusion of studies. Full texts of the included articles were obtained and checked again according to the inclusion criteria. Any discrepancies were discussed until consensus was reached. The screening process is visualised in Figure 1. This selection process resulted in 60 papers.

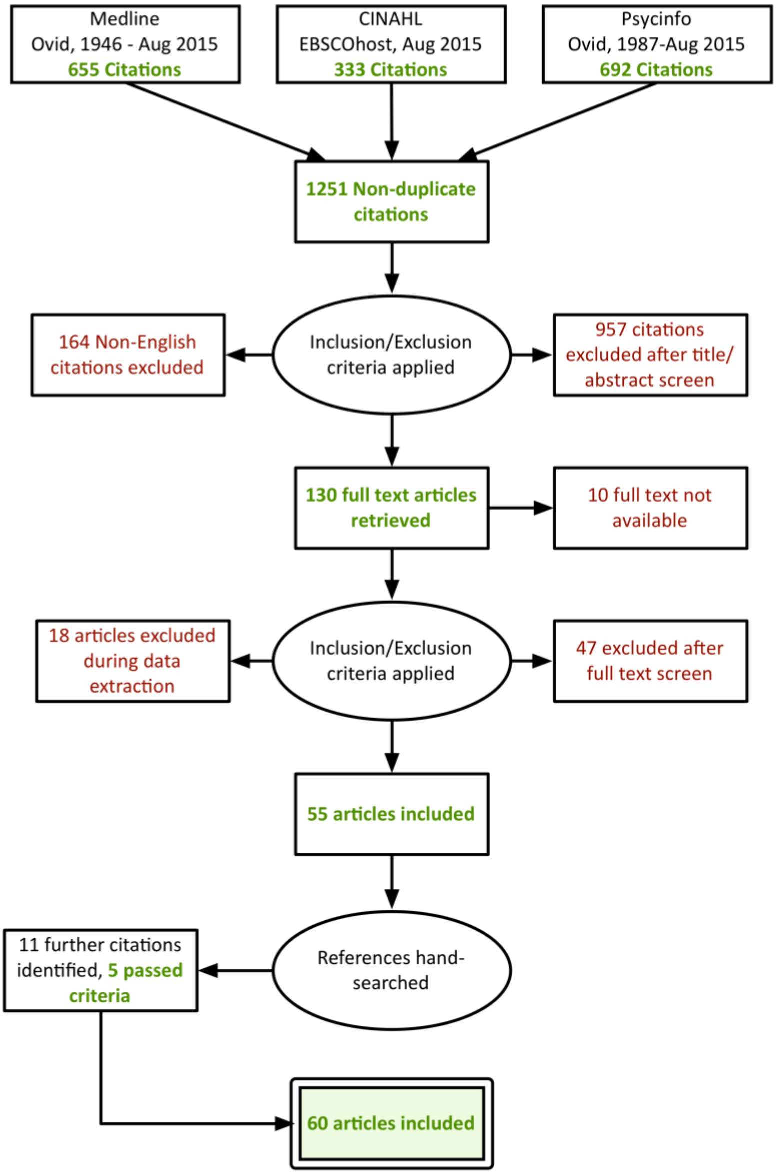


Figure 1. Flow diagram of search results

The final articles were categorised according to study type (Oxford Centre for Evidence Based Medicine - Levels of Evidence; 2011), treatment type (constraint or modality type) and outcome (Impairment, Activity/Participation, Quality of life, Carer burden). Author ROH resolved discrepancies during this process. The OCEBM levels do not have a manual that clearly describes categories and the Level "Case series" presumably includes pre/post group studies, case reports and Single Case Experimental Designs (SCEDs) that are not the traditional, medical N-of-1 design. However, there is a vast difference in rigour between SCEDs, which can provide evidence of cause and effect of a treatment, and other single participant studies such as case reports (Tate et al., 2013) or pre/post group designs. Therefore, we have included an additional level of evidence for SCEDs, below randomised trials and above non-randomised trials and have used this method in our categorisation. This more accurately represents the levels of rigour of study designs. Using the guidelines in Tate et al. (2013), this category included any multiple baseline, alternating treatments, withdrawal/reversal or changing criterion designs. It also included quasi-experimental bi-phasic (AB) designs.

Outcomes were classified as Activity/Participation where they rated or measured performance in either real or simulated everyday life activities. Quality of life outcomes were those using subjective ratings of life satisfaction, while carer burden outcomes were any that proposed to measure the carer's wellbeing or distress of any sort. An extraction template was created and included data fields for sample size, treatment, details of cueing and constraint used, and outcome measures. The completed data extraction for all 60 articles is included in the appendix.

Randomised controlled trials (RCTs) and nonrandomised controlled trials (nonRCTs) were assessed for methodological quality using the PEDro-P, an adapted version of the PEDro scale (Sherrington, Herbert, Maher, & Moseley, 2000) used by the PsycBITE and speechBITE teams (see Fitzpatrick, 2008). Ratings by the first author were compared against those found on SpeechBITE and PsycBITE and any discrepancies were resolved with an independent rater from the SpeechBITE team.

The RoBiNT scale (Tate et al., 2015) was used to assess methodological quality for SCEDs. Non-experimental single case designs, including one phase designs, pre/post designs and case descriptions, were not rated. Pre/post group designs were also not rated as these have no experimental control and form a low level of evidence. Each study was rated by two authors and discrepancies were discussed until agreement was reached.

For RCTs and nonRCTs, scores five and above on PEDro-P were considered moderate to high quality based on commonly accepted consensus (Centre for Evidence-Based Physiotherapy, 2016). For the RoBiNT, benchmarks have not yet been formally established, but in a paper examining the reliability of the scale on a small number of papers (Tate et al., 2013), the mean score was 12 with the highest score being 18 (Tate et al., 2015). In lieu of existing benchmarks, SCEDs that scored 12 or higher were therefore considered moderate to high quality. Studies that did not reach these quality cutoff scores (≥5 and ≥12) are included in the Appendix but they were not considered further in the results.

Effect size calculation

Group Designs

We planned to calculate effect sizes for relevant group studies of Level 2 or 3 evidence (RCTs and nonRCTs); however, studies were too heterogeneous in terms of outcome measures and treatments (see results).

Single Case Experimental Designs

Calculation of effect sizes for SCEDs is a domain that continues to develop but it has a number of promising methods (Parker, Vannest, Davis, & Sauber, 2011). Tau-U is a recent effect size measure which is resistant to effects of autocorrelation, considers baseline trend, deals well with only few data points and has the highest power of the non-overlap indices (Brossart, Vannest, Davis, & Patience, 2014). Tau-U can also produce confidence intervals.

In this review, to be eligible for SCED effect size calculation, studies needed to be an experimental design (viz., we excluded one phase designs, pre/post designs and case descriptions), have raw data presented on case charts and have at least 2 data points in each of the baseline and intervention phases. Only the outcome measure *confrontation naming of treated items* was used for effect sizes, which included 94% of SCEDs with a RoBiNT score of 12 or greater. Data was extracted by the first author and calculated using the online Tau-U calculator at www.singlecaseresearch.org. A weighted average Tau-U was calculated across all relevant participants and word sets for each paper. Positive baseline trend was corrected where baselines had a Tau score of greater than 0.4 (Parker et al., 2011) and an increasing trend apparent on visual inspection. Negative baseline was not corrected, nor was trend in treatment phases, as this would have boosted effect sizes. These steps are consistent with recommendations to correct for trend conservatively and to check statistical results against visual analysis (Brossart et al., 2014; Parker et al., 2011). Maintenance and follow up phases were not included in calculations.

Results

As outlined in Figure 1, of 1680 original results (1251 non-duplicates), 60 papers met the criteria for this systematic review. Figure 2 summarises the number of studies for each treatment approach.

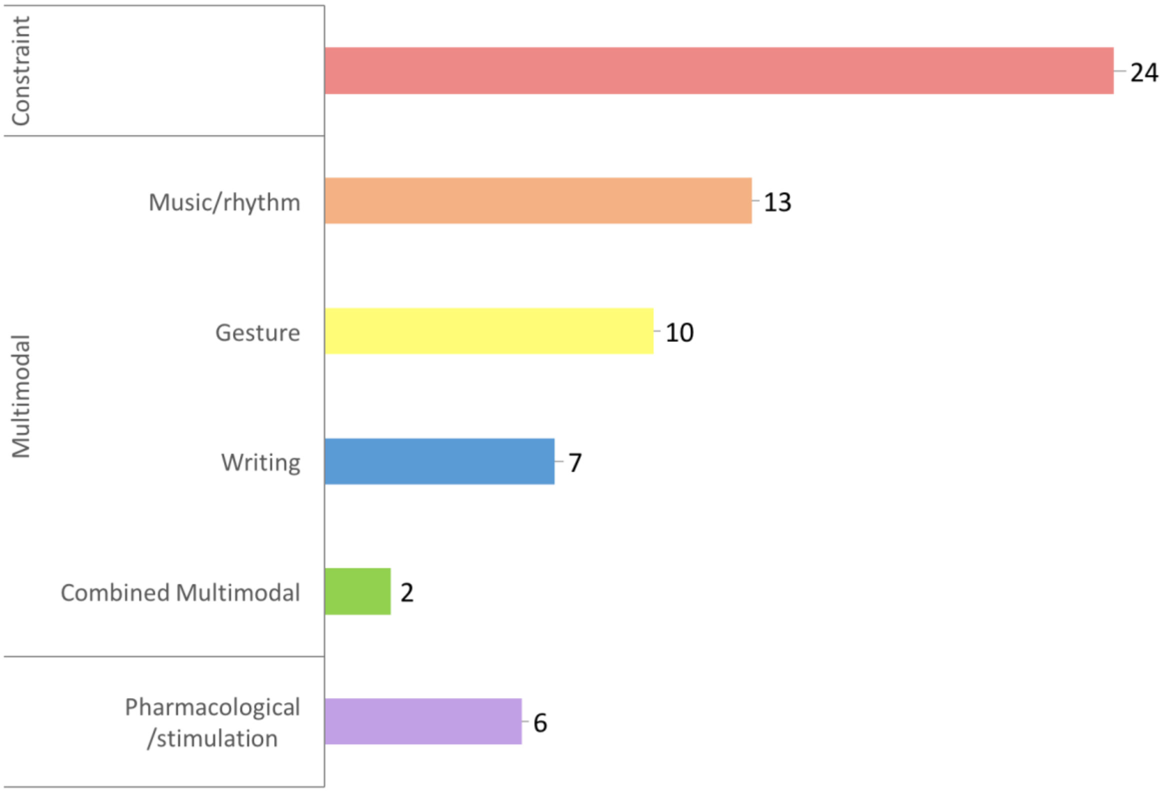


Figure 2. Number of results by treatment approach

Descriptive characteristics of data

Overall, there were more multimodal treatment studies than constraint (36 vs. 24) but more constraint studies than studies on any single modality. The first constraint paper was published in 2001 (Pulvermüller et al.), while the earliest multimodal paper was published in 1975 (Sparks et al.), demonstrating the rapid growth in constraint therapy research.

No studies were identified that used drawing alone as a means of cueing verbal output. Six studies utilised neurological stimulation (e.g. rTMS) or pharmacological treatments (e.g., memantine) in combination with constraint or multimodal therapies. Two of these were RCTs (Barbancho et al., 2015; Berthier et al., 2009), one a SCED (Al-Janabi et al., 2014) and three were pre/post designs (Abo et al., 2012; Martin et al., 2014; Vines, Norton, & Schlaug, 2011). While results of these mixed treatments were broadly positive, the contribution of the constraint and multimodal therapies could not be differentiated from the pharmacological and stimulation treatment aspects. These studies were therefore not considered further in this review.

The majority of the twenty-four constraint studies reported use of the original CIAT approach. Other variations included CIAT Plus (Attard et al., 2013; Meinzer et al., 2005), CIAT II (Johnson et al., 2014), lower intensity CIAT (Goral & Kempler, 2009; Kempler & Goral, 2011; Maul, Conner, Kempler, Radvanski, & Goral, 2014), or modifications such as inclusion of grammatical shaping (Faroqi-Shah & Virion, 2009) or drill tasks (Kempler & Goral, 2011).

The way "constraint" was applied varied considerably and was not well described in many studies. While all studies constrained communication between participants to the verbal modality only, three studies explicitly prevented participants from using gestures to self-cue (Breier, Maher, Novak, & Papanicolaou, 2006; L. M. Maher et al., 2006; Martin et al., 2014) whereas five studies allowed such gesture (Attard et al., 2013; MacGregor, Difrancesco, Pulvermüller, Shtyrov, & Mohr, 2015; Meinzer, Streiftau, & Rockstroh, 2007; Mohr, Difrancesco, Harrington, Evans, & Pulvermüller, 2014). The remaining 19 did not specify whether they allowed self-cueing or not, including the original CIAT paper (Pulvermüller et al., 2001). Similarly, while the majority of studies (14) allowed therapists to provide cueing when necessary, one reported not providing cues (Maul et al., 2014) and the remaining 11 provided no description of whether therapists provided cues.

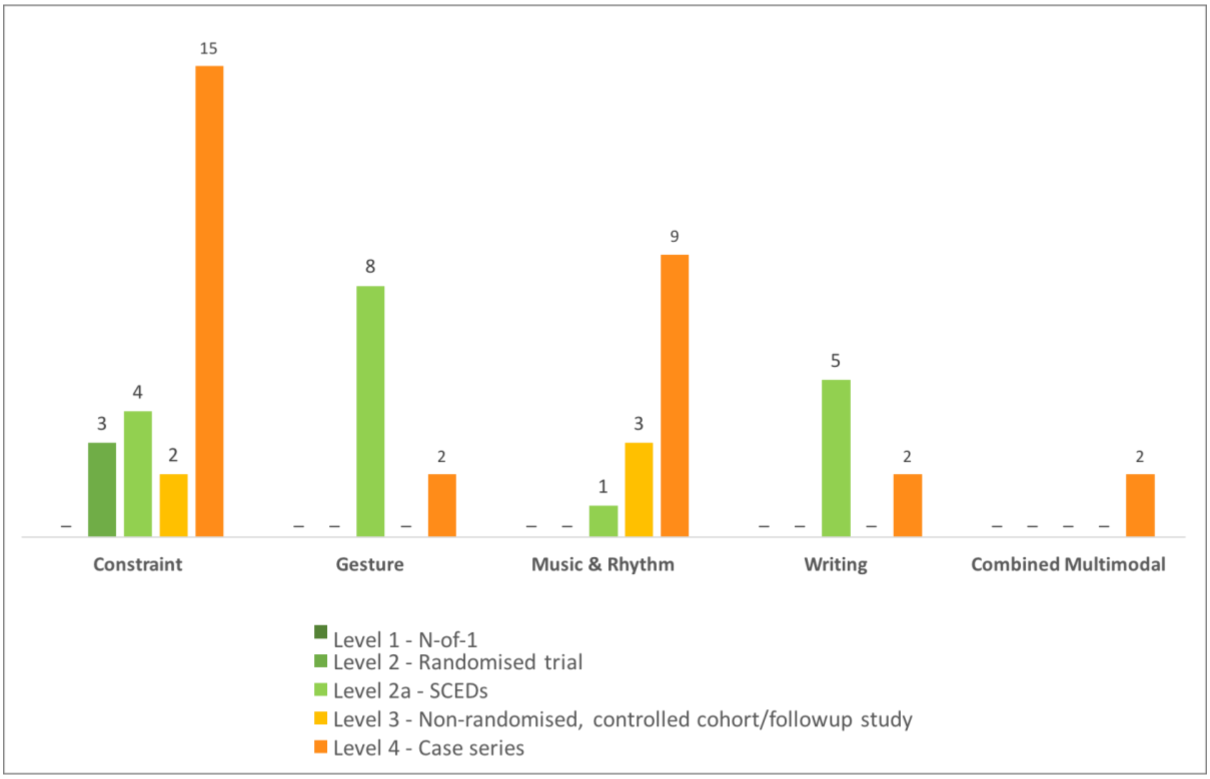


Figure 3. Number of papers per level (OCEBM levels of evidence)

Note. Level 4 studies (Case Series) were not eligible for quality evaluation and meta-analysis. In addition to N-of-1 designs, Level 1 ordinarily includes Systematic Reviews but these were not eligible as they are not original data.

Figure 3 shows the levels of evidence found for each therapy type according to OCEBM levels. There were a very limited number of RCTs and nonRCTS – five for constraint and three for music – however, there were a number of SCEDs (19), particularly in gesture studies. The remaining papers were low quality designs.

Figure 4 presents the methodological scores for papers for controlled trials (PEDro-P, 4b) and SCEDs (RoBiNT, 4b) and the red line shows the cutoff scores for inclusion in further analysis. As visualised in Figure 4a, the methodological quality scores ranged from 3-7 on PEDro-P (/10), with half the eight controlled trials (Level 2-3) below the cutoff score of 5. Quality scores on the RoBiNT ranged from 9-21 of a possible 30, with three of the 18 SCED studies (Level 2a) below the cutoff score of 12 (Figure 4b).

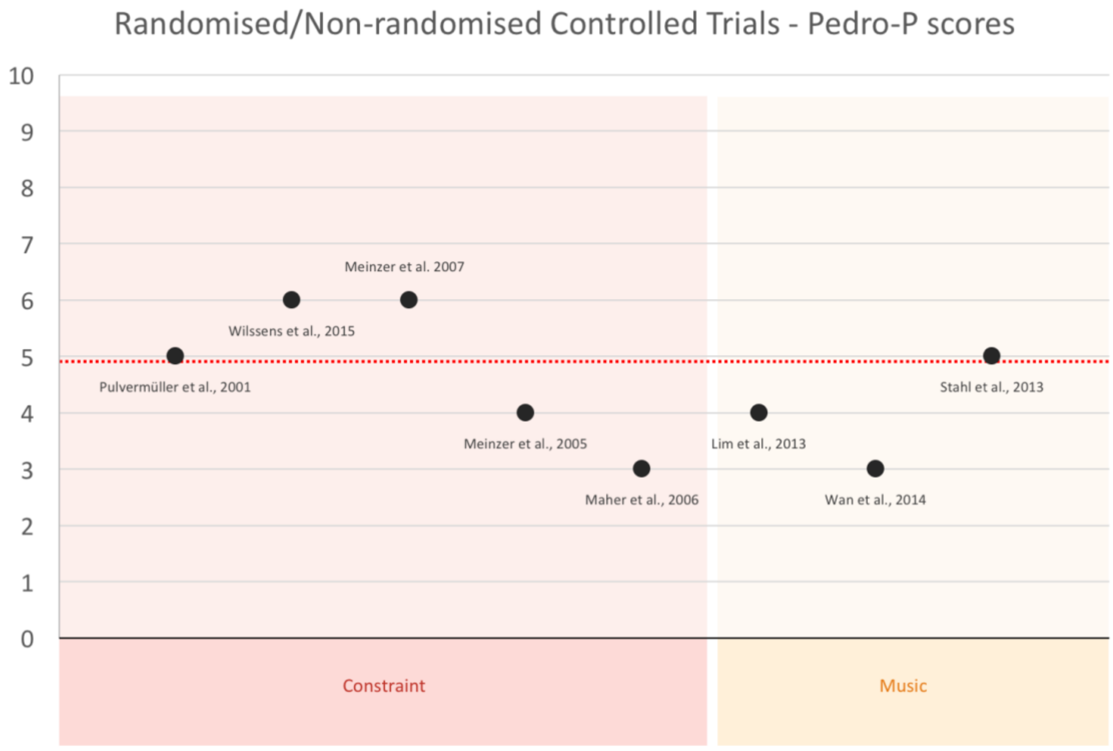


Figure 4a.

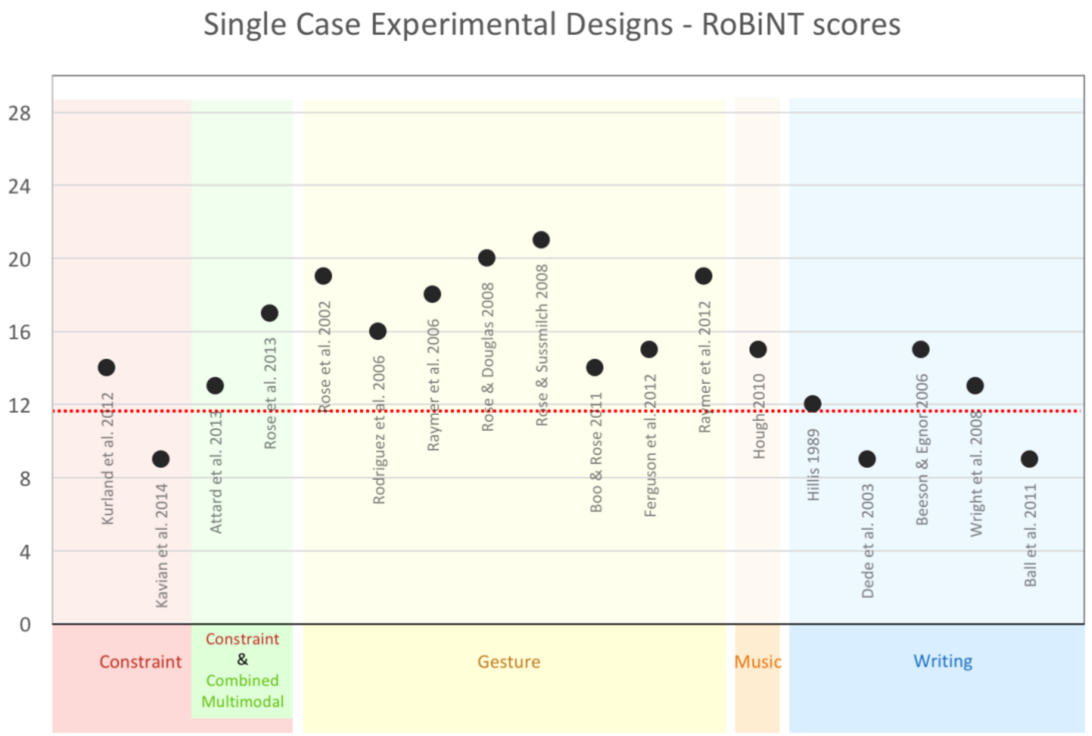


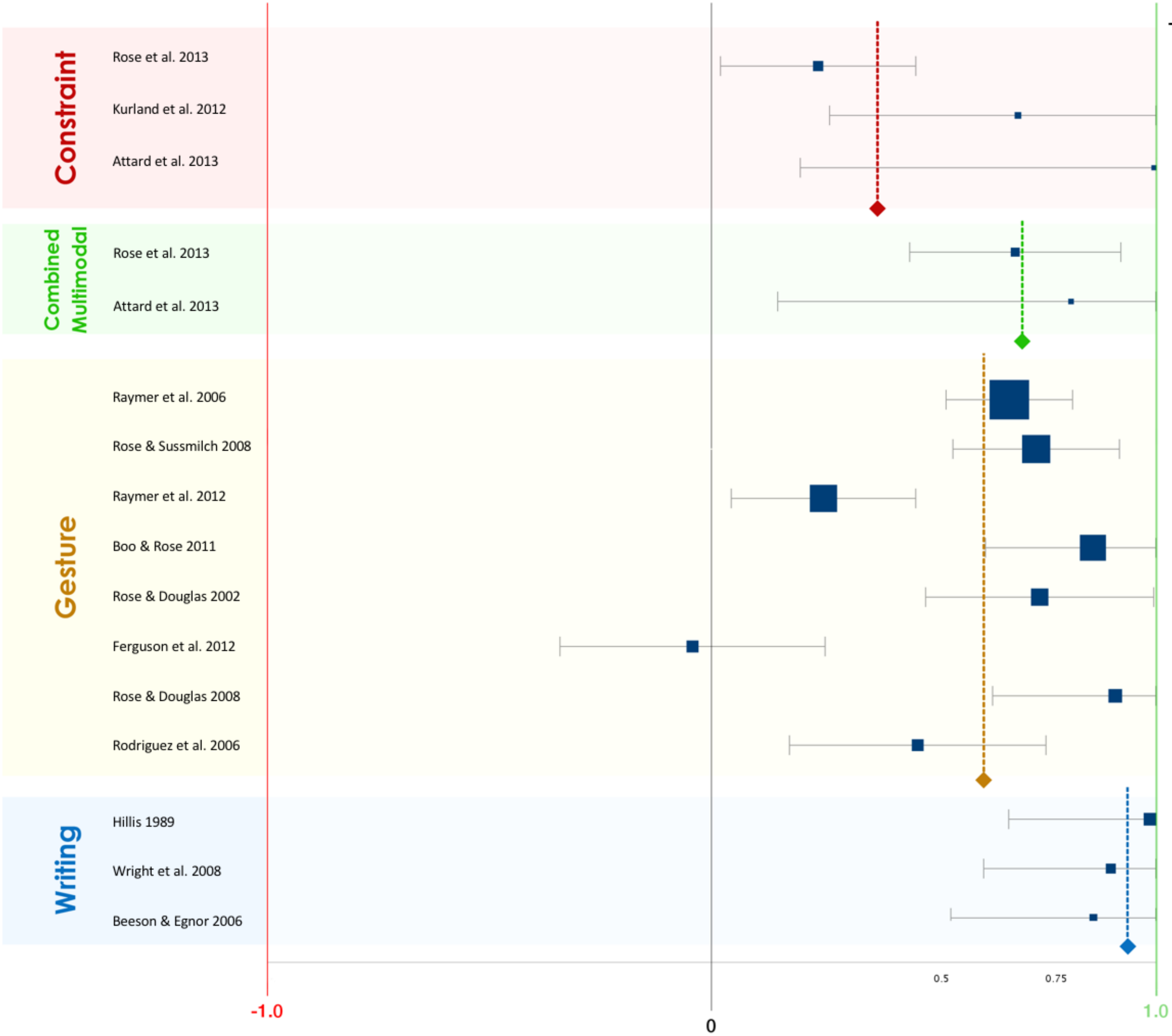
Figure 4b.

Figure 4. Pedro-P and RoBiNT scores

Note. Red line indicates cutoff score for methodological quality.

The Tau-U effect sizes for confrontation naming of treated items across 14 eligible studies are shown in Figure 5. Many lacked regular intervals between probes and/or had significant gaps in probing, and these are noted. Interpretation of Tau-U effect sizes is difficult without benchmarks, which are not available for Tau-U at present. However, Tau-U scores are an indication of the percentage of data in the treatment phase that has improved over time compared to baseline (Parker et al., 2011). The limits are -1.0 and 1.0, which would indicate 100% reduction and improvement in scores respectively, while 0.0 would indicate no change in scores. A total effect size weighted by study was calculated for each modality. Studies with more data points, whether due to more participants or more probes during phases, have a greater weighting (Parker et al., 2011).

Results are described below in relation to the key questions of this review: by outcome level (Impairment, Activity/Participation, Quality of life, Carer burden) and within outcome level by treatment type (constraint, multimodal). Only studies Level 3 or higher which met minimum quality cutoff scores are discussed.



Tau-U	95% C.I.	Weight	Sets calculated	Notes
0.240	0.020 - 0.460	3.3	CIAT-plus	Irregular probing frequency
0.692	0.266 - 1.000	1.2	Constraint items only	Seven day intermission between baseline and treatment phase with reduced probe frequency
1.000	0.200 - 1.000	0.2	CIAT-plus	Irregular probing frequency
0.684	0.446 - 0.921	2.5	M-MAT	Irregular probing frequency
0.813	0.149 - 1.000	0.4	M-MAT	Irregular probing frequency
0.670	0.528 - 0.812	21.0	Trained nouns & verbs	One month break beween baseline and treatment phases for some participants
0.731	0.543 - 0.918	14.1	Gest, Sem + Gesture	Delayed treatment of control set included
0.252	0.045 - 0.460	13.6	Gest. facilitation	Duration of intermission period unclear
0.857	0.616 - 1.000	12.8	Gest, sem + gesture	
0.738	0.482 - 0.995	7.6	Gest, Gest + Verbal	Delayed treatment of control set included
-0.043	-0.341 - 0.256	4.3	Pantomime set only	Seven day intermission between baseline and treatment phase for some participants
0.913	0.633 - 1.000	5.3	Gest, Gest + Verbal	
0.464	0.175 - 0.753	4.2	Verbal naming scores, Verbal + Gesture	One month break between baseline and treatment phase for Participant 2
0.997	0.669 - 1.000	4.7	P2, verb & noun verbal naming (treatment 2)	
0.896	0.612 - 1.000	3.1	All treated sets	Significant, irregular gaps in baseline probing frequency
0.864	0.539 - 1.000	1.8	Verbal naming (CART + Repetition)	

Graph Generated by DistillerSR

Figure 5. Tau-U meta-analysis

Note. Solid lines -1.0 and 1.0 indicate the possible limits of Tau-U scores – negative values demonstrate reduced naming performance and positive clues demonstrate improvement. Dotted lines with diamonds indicate the weighted average for each area or modality. Larger squares indicate larger relative weighting.
CIAT = Constraint Induced Aphasia Therapy, M-MMAT = Multi-Modal Aphasia Therapy, Gest = Gesture, Sem = Semantic

Impairment Outcomes

Constraint Treatment

Impairment-based results are detailed in Table 2. Constraint had the highest number of top tier studies according to the OCEBM levels; however, only three met quality criteria on the PEDro-P and all had small sample sizes (range 9 - 27). In addition, two of the three did not compare constraint to an equivalent non-constraint control. Meinzer, Streiftau and Rockstroh (2007) compared CIAT to CIAT run by family, and Pulvermüller et al. (2001) compared CIAT to conventional therapy of equal total hours but lower intensity. Wilssens et al. (2015) did compare CIAT to equal intensity conventional therapy and included impairment based assessments as secondary outcome measures.

Of the three constraint SCEDs that met quality criteria, two compared CIAT Plus to M-MAT (Attard et al., 2013; Rose et al., 2013) while the other (Kurland, Pulvermüller, Silva, Burke, & Andrianopoulos, 2012) compared CIAT to PACE (Promoting Aphasic Communicative Effectiveness, Davis, 2005). Figure 5 shows Tau-U effect sizes with constraint studies at the top followed by multimodal studies. In Attard et al. (2013), the Tau-U effect size for CIAT Plus (1.0) was greater than M-MAT (0.81), though M-MAT was within the 95% confidence interval of CIAT Plus. In Rose et al. (2013) the CIAT Plus effect size was markedly lower than M-MAT (0.24 and 0.68) and the upper and lower confidence interval bounds just overlapped.

The weighted Tau-U effect size for all three high-quality constraint SCEDs was 0.374, which means that approximately 37% of treatment phase naming scores were higher than baseline scores. This was the lowest effect size relative to other modalities and was reduced by Rose et al. (2013) which, due to its larger number of participants and probes, held a greater weighting. However, Rose et al. (2013) was an outlier and the other three constraint SCEDs were more in line with the effect sizes of other modalities (see Figure 5), though they had lower weighting and RoBiNT scores.

As displayed in Table 2, group studies showed improvements on secondary, pre/post impairment measures such as the Aachen Aphasia Test (AAT; Huber, Poeck, & Willmes, 1984). AAT results in Wilssens et al. (2015) appeared to favour CIAT over traditional therapy but subtest results were reported without any overall profile scores. There were indications of positive effects on blinded clinician ratings on the Communicative Activity Log (CAL; Pulvermüller et al., 2001), while Boston Naming Test changes were variable between studies (BNT; Kaplan, Goodglass, & Weintraub, 1983). Changes in pre/post measures within SCEDs were obscured due to the crossover designs.

Multimodal Treatment

Combined Multimodal. There were two studies on combined multimodal treatment that investigated M-MAT. Both used multiple impairment-based outcomes. As reported above, the Tau-U effect size for M-MAT in Attard et al. (2013) was 0.81, lower than CIAT Plus but suggesting a strong effect of the treatment compared to baseline nonetheless. The Tau-U effect size was 0.68 for Rose et al. (2013), notably higher than CIAT Plus. The combined, weighted Tau-U for combined multimodal was 0.70, the second highest of all total effect sizes calculated; however, an effect size based on only two studies is far from conclusive.

Aside from confrontation naming of treated items, other impairment-based measures were obscured by the difficulty of calculating pre/post changes in crossover designs.

Gesture. All studies using gesture to facilitate verbal output that met the quality and inclusion criteria for gesture were SCEDs. Most compared the efficacy of gesture to another treatment (see appendix), including repetition, semantic treatment and intention gestures. All used impairment outcomes, with most investigating confrontation naming and/or the Western Aphasia Battery (WAB, Kertesz, 2007).

As all eight studies used confrontation naming of treated items for probes, Tau-U was calculated for each one. The narrower confidence intervals in some of the Tau-U effect sizes and high relative weightings reflect the high number of data points, due to more participants or more probes in each phase (Figure 5). The overall weighted effect size for the eight SCED studies on gesture was 0.62, indicating that approximately 62% of treatment phase naming scores were higher than baseline scores. This is promising for the effects of gesture on confrontation naming of verbs and nouns. However, this included one negative (Ferguson, Evans, & Raymer, 2012), low (Raymer et al., 2012) and moderate effect size (Rodriguez, Raymer, & Gonzalez Rothi, 2006). Gesture therefore had the widest range of Tau-U effect sizes. These lower effect sizes are not attributable to poor quality methodology, as Raymer et al. (2012) scored 19 on the RoBiNT.

The small number of secondary, impairment-based outcome measures taken before and after gesture treatment alone showed variable changes. Mean changes were small for naming batteries. The mean change in those using the WAB Aphasia Quotient (WAB AQ) reached the critical difference of five for one study (Raymer et al., 2012) but not the other (Raymer et al., 2006).

Music. Despite 13 articles for music meeting the initial inclusion criteria — the second highest yielding category — nine were pre/post designs, along with two low quality nonRCTs. Only two studies with impairment-based outcomes reached adequate methodological quality (Hough, 2010; Stahl, Henseler, Turner, Geyer, & Kotz, 2013).

While Stahl et al. (2013) aimed to disentangle the contributions of rhythm and singing in melodic therapy, the results of the rhythm arm are not reported for this review. The singing group rehearsed common phrases whilst the conventional therapy group received treatment on other stimuli. After treatment, patients were assessed in their ability to sing and speak in unison with recordings of both trained phrases and untrained phrases. Written prompts were also provided. Unsurprisingly, the music group who had practised the trained phrases showed greater improvement than the conventional therapy group. The conventional therapy group demonstrated gains on the untrained phrases, however, while the singing group did not improve.

Hough (2010) used successful *repetition* of phrases as the primary outcome for their single participant. Thus, while the data presented in their case chart appears to show improvement, it is not comparable to confrontation naming tasks, and a Tau-U effect size was not calculated. Regardless, improvement of repetition performance was highly significant for both common and personalised phrases ($p < 0.0001$), and the WAB AQ and CQ (cortical quotient) showed large improvements (13 and 13.6 respectively).

Neither repetition (Hough, 2010) nor speaking in unison and with prompts (Stahl et al., 2013) is a test of generative verbal output. These two papers show only that singing specific phrases improves participants' ability to produce those phrases with maximum modelling and prompting.

Writing. There were no group studies utilising writing as a prompt for verbal output. All three high quality SCEDs used the same primary outcome measure of confrontation naming of treated items. These SCEDs had a high overall Tau-U effect size of 0.94, greater than all other modalities and constraint. This is close to the ceiling of 1.0 and suggests that the majority of treatment probes were improved over baseline.

In regards to pre/post impairment-based outcomes, the two participants in the multiple baseline design of Wright et al. (2008) demonstrated improvement in pre/post WAB AQ (3.1 and 9.7) and contradictory improvement in pre/post BNT (-3 and 5).

Activity/Participation Outcomes

Constraint Treatment

Four high quality constraint studies included activity/participation outcomes: two RCTs (Pulvermüller et al., 2001; Wilssens et al., 2015) and two SCEDs (Attard et al., 2013; Rose et al., 2013). As displayed in Table 3, outcomes included the Communication Effectiveness Index (CETI; Lomas, Pickard, Bester, & Elbard, 1989), the Amsterdam-Nijmegen Everyday Language Test (ANELT, Blomert, Kean, Koster, & Schokker, 1994), the Scenario Test, and one used a customised measure, the CAL.

The improvements on the CAL self-ratings in Pulvermüller et al. (2001) were significant for constraint and not the control group, but no between-group comparisons were made and the control group received lower intensity therapy.

Wilssens et al. (2015) was the only controlled trial to compare constraint to a non-constraint treatment of the same intensity, a Dutch drill-based lexical-semantic therapy program, BOX . They did not find a statistically significant *between* group change on the CETI (p=.332). However, within group changes were significant for the BOX group and not the CIAT group. The mean CETI change for the BOX group was also above the clinically significant improvement level of 12 while it was less for the CIAT group (Lomas et al., 1989). There were no significant between group differences on the ANELT.

There was no clinically significant improvement in CETI scores for CIAT in Attard, Rose & Lanyon (2013) or in mean changes in Rose et al. (2013). Scenario Test scores improved in both participants for CIAT in Attard, Rose & Lanyon but the mean change was negligible for Rose et al.

Multimodal Treatment

Combined Multimodal. The CETI score changes post M-MAT in Attard, Rose and Lanyon (2013) did not reach the minimum clinically significant change of 12. In Rose et al. (2013) the mean change in CETI scores was 8.5 (range -2 to 33), again lower than the clinically significant change. Scenario Test changes in both studies were minimal.

Gesture. There was limited use of activity/participation measures in gesture studies, but the three studies that did use these were moderate to high quality SCEDs (RoBiNT 14-21). However, two of these were comparisons of multiple treatments (Boo & Rose, 2011; Rose & Sussmilch, 2008), and therefore pre/post measures of activity/participation cannot be considered as they represent changes from all treatments. The third, Raymer et al. (2012), used family member ratings on two measures — the CETI and the Functional Outcomes Questionnaire for Aphasia (FOQ-A; Glueckauf et al., 2003). Changes on the CETI were inconsistent. Two participants had a negative change — one clinically significant — and two had a positive change — one clinically significant. The mean change was 5.20 (range -15 to 33). On the FOQ-A changes were positive but small, with only one score changing greater than one standard deviation from the original FOQ-A paper (Glueckauf et al., 2003). The mean change was 0.44 (-0.19 to 1.21).

Music. A single high quality study investigated activity/participation outcomes in a multiple baseline design (Hough, 2010). The CETI was completed by the participant's caregiver and improved by 28.2, which is well above the clinically significant change of 12. The ASHA FACS (American Speech-Language Hearing Association Functional Assessment of Communication Skills; Frattali, Thompson, Holland, & Wohl, 1995) improvement was 2.05/7 (to our knowledge there are no established benchmarks for clinically significant change for the ASHA FACS).

Writing. No high quality studies on writing used activity/participation outcomes.

Quality of Life/Carer Burden Outcomes

Carer Burden

No studies were retrieved that used assessments of carer burden.

Constraint Treatment

In two constraint studies (Attard et al., 2013; Rose et al., 2013), quality of life was measured using the Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39; Hilari, Byng, Lamping, & Smith, 2003), but no others investigated this domain. Both administered SAQOL-39 before and after their crossover design comparing CIAT Plus and M-MAT. As they were not measured between crossover of treatments, it is not possible to attribute changes to one of M-MAT or CIAT Plus.

Multimodal Treatment

Combined Multimodal. As reported above, Rose et al. (2013) and Attard, Rose and Lanyon (2013) did not measure the SAQOL-39 between CIAT and M-MAT, therefore individual contributions of the treatments cannot be determined.

Gesture. No gesture studies were found that investigated quality of life.

Music. One study meeting quality criteria investigated quality of life (Hough, 2010). The single participant demonstrated a 25 point increase on the ASHA QCL (Paul et al., 2003) — a 64% improvement from baseline score.

Writing. No writing studies investigated quality of life.

Discussion

The aim of this systematic review was to examine and compare evidence for constraint and multimodal therapies in chronic aphasia. Such a review is important for the treatment of people living with aphasia long term and for the evidence-based application of learned nonuse and multimodal cueing. The effectiveness of these approaches also inform our theoretical understanding of language processing. This review shows that there is limited high quality evidence to support the use of either constraint or multimodal approaches for impairment, activity/participation and quality of life outcomes. The amount and strength of evidence varied between communication modalities.

Impairment outcomes

There were few high-quality studies comparing constraint therapies to equivalent intensity controls, and for those that did, results did not favour constraint. Indeed, Tau-U scores for one study favoured multimodal over constraint. The existing research therefore fails to demonstrate superiority of CIAT over any non-constraint treatment in chronic aphasia. Comparisons aside, overall evidence does suggest positive effects for impairment outcomes in constraint, but data is far from conclusive and further research is required.

There is scant evidence for combined multimodal treatment with only two SCED studies. While this preliminary evidence has positive results in impairment measures with an effect size of 0.702 for naming, the effect size could change with further research. Changes in noun production during semi-structured conversation also hinted at positive outcomes, but more research is indicated before any conclusions can be drawn.

Gesture evidence came from high quality SCED papers, yet effect sizes varied widely for impairment outcomes. Further investigation is needed to explain this variability. In addition to improvements in confrontation naming, there were signs of positive effects on other impairment-based assessments. However, these data are only preliminary as results were inconsistent and obscured at times by crossover designs. Group designs with control groups would address the limitations of crossover designs and allow direct comparison of changes on multiple outcome measures.

It is also worth noting that all eight SCEDs for gesture were conducted by Rose or Raymer. Though a variety of participants and study designs were used, the evidence would be enhanced with research from other authors. Group designs such as RCTs are also needed to confirm current findings.

Despite MIT being a well-known treatment for aphasia, in chronic aphasia the evidence is of low quality in terms of both study design and methodology. Future studies into music-based treatment of verbal output in aphasia need to employ more rigorous research designs. Rather than investigating speech production in unison or repetition, research should investigate the presumed end goal of such treatments — the independent production of trained words or phrases — and probe for generalisation to discourse and conversation.

Results for impairment outcomes from writing are very positive thus far but inconclusive due to low replication. It is possible that research specifically examining the effects of writing on verbal output are limited because pairing writing with speech is already widely accepted and frequently embedded in cueing hierarchies. Nonetheless, further research is necessary to confirm the impact of writing on verbal output, especially group studies.

No studies were found investigating drawing. Research on this promising modality is needed.

In summary, there is encouraging but insufficient evidence for constraint or multimodal treatments on impairment-based measures, in terms of both quality and quantity.

Activity & Participation

In group studies on constraint therapy, there was insufficient evidence for effects on Activity/Participation following treatment, both in terms of pre/post improvement and between-group comparisons. CIAT Plus and M-MAT changes in the CETI, though not clearly attributable to either treatment due to the crossover designs, suggested positive outcomes, again with high variability amongst participants. In gesture studies, the limited results for Activity/Participation measures were contradictory and inconclusive. Improvement and deterioration of scores effectively cancelled each other out, with disparity even between self and carer improvement ratings at times. In music studies, there was only quality data for a single participant, though this was clinically and statistically significant.

Quality of Life and Carer Burden

Quality of life outcomes were rarely investigated. A subjective improvement in a person's life experience should be a goal of aphasia therapy, yet in this review, only five studies employed quality of life outcomes, and only three were high quality.

Similarly, the effect of treatment on carer burden remains unexplored. Carers of people with aphasia experience changes such as loneliness, anxiety, increased responsibilities and need for support (Patrício et al., 2013), and supporting them should also be a crucial goal of rehabilitation (Rombough, Howse, & Bartfay, 2006). Yet the impact of treatment on carers and family was not investigated in any study.

Future research should include both quality of life and carer burden as outcome measures.

Systematic review design

In this systematic review, we chose to include high quality SCEDs. The inclusion of SCEDs allowed closer inspection of research areas that would have returned no results if a traditional evidence hierarchy was used. We made this decision based on increasing recognition that high quality SCEDs can have equal or even superior rigour to RCTs (e.g. Medical N-of-1 designs) (Tate et al., 2016).

Use of the RoBiNT scale allowed us to rank SCEDs according to specific features that contribute to internal and external validity. This approach is more fine-grained than simply classifying SCEDs by design subtype or as experimental/non-experimental. As RoBiNT is a relatively new scale there is not yet an established cutoff for what constitutes a quality study. Our use of 12 as the cutoff was based on early uses of this new scale. While this formed a limited empirical basis, the widely accepted cutoff of 5 for the PEDro/PEDro-P is also based on common scores (Teasell et al., 2007). Neither approach considers relative weighting of individual items. Nevertheless, higher scoring studies will have stronger methodological quality and internal validity and in our results there were no borderline papers on the RoBiNT that were excluded.

As far as we are aware, this is the first time that the Tau-U effect size calculations have been applied within a systematic review of aphasia treatments. The Tau-U effect size provided ranking of SCEDs based on improved data points in the treatment phases, after correction for baseline trend. These effect sizes allowed comparisons *between* papers in this study but we know of no external benchmarks for Tau-U effect sizes. In addition, many SCEDs analysed with Tau-U had gaps in treatment probe intervals (e.g. washout periods between crossover trials) or irregular probing, which may have influenced the validity of resulting effect sizes. Another challenge in utilising Tau-U is that there is no agreed protocol yet; for example, the threshold for applying baseline trend correction varies between studies. The first author therefore combined visual inspection with the baseline Tau to eliminate overcorrection (see methods). While this did introduce a subjective element to calculation, the overall process remains a defensible calculation of effect sizes in a family of designs previously resistive to meta-analysis.

Finally, the exclusion of non-English articles is an unfortunate, but in our case, unavoidable, criterion. It is possible that relevant research published in languages other than English was missed by this systematic review.

Treatment reporting

Interventions should be reported thoroughly to allow future replication and synthesis of results (Hoffmann, Glasziou, Boutron, & Milne, 2014). A problem noted across retrieved constraint studies, regardless of quality or design, was the disparity in what was considered to constitute constraint. While the original authors have recently clarified that self-cueing with actions (e.g., gesture) is permitted as long as it is not communicative (Difrancesco et al., 2012), prior to this publication some protocols banned gesture and the majority made no comment on this aspect, including the original CIAT article (Pulvermüller et al., 2001). Likewise, whether or not therapists provided cues to participants was often not described in methods and is not addressed in Difrancesco, Pulvermüller and Mohr (2012). The term "constraint" therefore currently represents therapies with significant procedural differences. Without clearer reporting of methods, there is a risk of continued bleeding of the term "constraint" to an increasingly diverse range of game-based language treatments. All future constraint-based research should provide comprehensive description of methods and state explicitly in which ways they depart from the outline in Difrancesco, Pulvermüller and Mohr (2012). A template such as TIDieR (Hoffmann et al., 2014) also provides a framework for thorough reporting. Without this detail, it will be difficult to determine the aspects of CIAT that contribute to effectiveness.

Outcome measures

The majority of studies in this review focused on improvements at the Impairment level. However, outcomes in aphasia research should be those that are important to people with aphasia and their families (Wallace, 2016). Recent work has shown that these include a range of outcomes across the ICF as well as quality of life and patient satisfaction with treatment (Wallace et al., 2016). While these constructs can be difficult to measure directly, especially in people with aphasia (e.g.; Szaflarski et al., 2015), without the inclusion of such outcomes in future, the benefit of these treatments to people living with aphasia will remain unknown.

There was also a high number of different outcomes within this review. Heterogeneity of outcome measures reduces research efficiency by limiting synthesis and meta-analysis of results, which is an important way to overcome the small sample sizes that are common in aphasia research (Brady et al., 2016). The development of a core outcome for aphasia research set is currently underway which will recommend preferred outcomes measures for the constructs identified as important to those with aphasia and other stakeholders (Wallace, Worrall, Rose, & Le Dorze, n.d.). The core outcome set should be adhered to in future research wherever possible.

Conclusion

Overall, this review has found a limited evidence base for constraint therapy in chronic aphasia, especially in proportion to its prominence in research and clinical practice. While studies indicated positive outcomes, there is a need for rigorous high level studies comparing CIAT and its derivatives to non-constraint therapies or controls. We also found a very limited evidence base for multimodal therapies. Studies on some modalities had limited research of any quality (drawing, writing, combined multimodal) while others had more research but little of adequate quality (music, gesture).

Accordingly, there is insufficient data to suggest superiority of either constraint or multimodal approaches in chronic aphasia. There were not enough comparable high-quality group studies to perform meta-analysis. Meta-analysis of SCEDs favoured multimodal treatments but this is not yet conclusive.

In addition, there was insufficient examination of “real world” endpoints. Aphasia research needs to expand beyond the use of basic impairment outcomes such as confrontation naming and toward consistent outcome measures based on the wishes of people with aphasia and their families.

Examination of constraint and multimodalities against control treatments of equal intensity and duration are needed, as well as direct, rigorous comparison between constraint and multimodal treatments. It is further recommended that future research provides a comprehensive description of treatment methods and readily accessible treatment materials arising from clinical trials to enable translation into practice.

Clinicians should not adopt either treatment approach exclusively until further research is published demonstrating the superiority of one treatment, or, more likely, the suitability of each to particular patient characteristics.

Acknowledgements

This work was supported by an Australian Government Research Training Program Scholarship.

References

Abo, M., Kakuda, W., Watanabe, M., Morooka, A., Kawakami, K., & Senoo, A. (2012). Effectiveness of low-frequency rTMS and intensive speech therapy in poststroke patients with aphasia: A pilot study based on evaluation by fMRI in relation to type of aphasia. *European Neurology*, *68*(4), 199–208. https://doi.org/10.1159/000338773

Al-Janabi, S., Nickels, L. A., Sowman, P. F., Burianová, H., Merrett, D. L., & Thompson, W. F. (2014). Augmenting melodic intonation therapy with non-invasive brain stimulation to treat impaired left-hemisphere function: Two case studies. *Frontiers in Psychology*, *5*(37), 1–12. https://doi.org/10.3389/fpsyg.2014.00037

Attard, M. C., Rose, M. L., & Lanyon, L. E. (2013). The comparative effects of Multi-Modality Aphasia Therapy and Constraint-Induced Aphasia Therapy-Plus for severe chronic Broca's aphasia: An in-depth pilot study. *Aphasiology*, *27*(1), 80–111. https://doi.org/10.1080/02687038.2012.725242

Ball, A. L., de Riesthal, M., Breeding, V. E., & Mendoza, D. E. (2011). Modified ACT and CART in severe aphasia. *Aphasiology*, *25*(6-7), 836–848. https://doi.org/10.1080/02687038.2010.544320

Barbancho, M. A., Berthier, M. L., Navas-Sanchez, P., Dàvila, G., Green-Heredia, C., Garcia-Alberca, J. M., et al. (2015). Bilateral brain reorganization with memantine and constraint-induced aphasia therapy in chronic post-stroke aphasia: An ERP study. *Brain and Language*, *145-146*, 1–10. https://doi.org/10.1016/j.bandl.2015.04.003

Beattie, G., & Shovelton, H. (2006). A critical appraisal of the relationship between speech and gesture and its implications for the treatment of aphasia. *International Journal of Speech-Language Pathology*, *8*(2), 134–139. https://doi.org/10.1080/14417040600667392

Beeson, P. M., & Egnor, H. (2006). Combining treatment for written and spoken naming. *Journal of the International …*, *12*, 816–827. https://doi.org/10.10170S1355617706061005

Berthier, M. L., Green, C., Lara, J. P., Higueras, C., Barbancho, M. A., Dàvila, G., & Pulvermüller, F. (2009). Memantine and constraint-induced aphasia therapy in chronic poststroke aphasia. *Annals of Neurology*, *65*(5), 577–585. https://doi.org/10.1002/ana.21597

Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam—Nijmegen everyday language test: construction, reliability and validity. *Aphasiology*, *8*(4), 381–407. https://doi.org/10.1080/02687039408248666

Bonakdarpour, B., Eftekharzadeh, A., & Ashayeri, H. (2003). Melodic Intonation Therapy in Persian aphasic patients. *Aphasiology*, *17*(1), 75–95. https://doi.org/10.1080/02687030244000464

Boo, M., & Rose, M. L. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology*, *25*(2), 154–175. https://doi.org/10.1080/02687031003743789

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *The Cochrane Database of Systematic Reviews*, *6*(6), CD000425–CD000425. https://doi.org/10.1002/14651858.CD000425.pub4

Breier, J. I., Juranek, J., & Papanicolaou, A. C. (2011). Changes in maps of language function and the integrity of the arcuate fasciculus after therapy for chronic aphasia. *Neurocase*, *17*(6), 506–517. https://doi.org/10.1080/13554794.2010.547505

Breier, J. I., Juranek, J., Maher, L. M., Schmadeke, S., Men, D., & Papanicolaou, A. C. (2009). Behavioral and neurophysiologic response to therapy for chronic aphasia.

Archives of Physical Medicine and Rehabilitation, *90*(12), 2026–2033. https://doi.org/10.1016/j.apmr.2009.08.144
Breier, J. I., Maher, L. M., Novak, B., & Papanicolaou, A. C. (2006). Functional imaging before and after constraint-induced language therapy for aphasia using magnetoencephalography. *Neurocase*, *12*(6), 322–331. https://doi.org/10.1080/13554790601126054
Breier, J. I., Maher, L. M., Schmadeke, S., Hasan, K. M., & Papanicolaou, A. C. (2007). Changes in language-specific brain activation after therapy for aphasia using magnetoencephalography: A case study. *Neurocase*, *13*(3), 169–177. https://doi.org/10.1080/13554790701448200
Breier, J. I., Randle, S., Maher, L. M., & Papanicolaou, A. C. (2010). Changes in maps of language activity activation following melodic intonation therapy using magnetoencephalography: two case studies. *Journal of Clinical and Experimental Neuropsychology*, *32*(3), 309–314. https://doi.org/10.1080/13803390903029293
Brossart, D. F., Vannest, K. J., Davis, J. L., & Patience, M. A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation*, *24*(3-4), 464–491. https://doi.org/10.1080/09602011.2013.868361
Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Research Design and Analysis: New Directions for Psychology and Education* (pp. 187–212). Hillsdale, NJ: Lawrence Erlbaum Associates.
Carragher, M., Sage, K., & Conroy, P. (2013). The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation. *Neuropsychological Rehabilitation*, *23*(6), 846–887. https://doi.org/10.1080/09602011.2013.832335
Centre for Evidence-Based Physiotherapy. (2016). PEDro Statistics. Retrieved June 14, 2016, from http://www.pedro.org.au/english/downloads/pedro-statistics/
Cubelli, R. (1995). More on drawing in aphasia therapy. *Aphasiology*, *9*(1), 78–83. https://doi.org/10.1080/02687039508248693
Davis, G. A. (2005). PACE revisited. *Aphasiology*, *19*(1), 21–38. https://doi.org/10.1080/02687030444000598
DeDe, G., Parris, D., & Waters, G. (2003). Teaching self-cues: A treatment approach for verbal naming. *Aphasiology*, *17*(5), 465–480. https://doi.org/10.1080/02687030344000094
Difrancesco, S., Pulvermüller, F., & Mohr, B. (2012). Intensive language-action therapy (ILAT): The methods. *Aphasiology*, *26*(11), 1317–1351. https://doi.org/10.1080/02687038.2012.705815
Farias, D., Davis, C., & Harrington, G. (2006). Drawing: Its contribution to naming in aphasia. *Brain and Language*, *97*(1), 53–63. https://doi.org/10.1016/j.bandl.2005.07.074
Faroqi-Shah, Y., & Virion, C. R. (2009). Constraint-induced language therapy for agrammatism: Role of grammaticality constraints. *Aphasiology*, *23*(7-8), 977–988. https://doi.org/10.1080/02687030802642036
Ferguson, N. F., Evans, K., & Raymer, A. M. (2012). A comparison of intention and pantomime gesture treatment for noun retrieval in people with aphasia. *American Journal of Speech-Language Pathology*, *21*(2), s126–s139.
Feyereisen, P. (2006). How could gesture facilitate lexical access? *International Journal of Speech-Language Pathology*, *8*(2), 128–133. https://doi.org/10.1080/14417040600667293
Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the

motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Fitzpatrick, R. B. (2008). PsycBITE™: Psychological Database for Brain Impairment Treatment Efficacy. *Journal of Electronic Resources in Medical Libraries*, *5*(2), 171–178. https://doi.org/10.1080/15424060802064428

Flowers, H. L., Silver, F. L., Fang, J., Rochon, E., & Martino, R. (2013). The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke. *Journal of Communication Disorders*, *46*(3), 238–248. https://doi.org/10.1016/j.jcomdis.2013.04.001

Frattali, C. M., Thompson, C. M., Holland, A. L., & Wohl, C. B. (1995). ASHA functional assessment of communication skills (FACS). Rockville, MD: American Speech-Language Hearing Association.

Glueckauf, R. L., Blonder, L. X., Ecklund-Johnson, E., Maher, L. M., Crosson, B., & Gonzalez-Rothi, L. (2003). Functional Outcome Questionnaire for Aphasia: overview and preliminary psychometric evaluation. *NeuroRehabilitation*, *18*(4), 281–290.

Goldfarb, R., & Bader, E. (1979). Espousing melodic intonation therapy in aphasia rehabilitation: A case study. *International Journal of Rehabilitation Research*, *2*(3), 333–342.

Goral, M., & Kempler, D. (2009). Training verb production in communicative context: Evidence from a person with chronic non-fluent aphasia. *Aphasiology*, *23*(12), 1383–1397. https://doi.org/10.1080/02687030802235203

Hadar, U., & Rumiati, R. I. (2006). I gesture therefore I speak: Cartesian cures. *International Journal of Speech-Language Pathology*, *8*(2), 140–142. https://doi.org/10.1080/14417040600667301

Hilari, K. (2011). The impact of stroke: are people with aphasia different to those without? *Disability and Rehabilitation*, *33*(3), 211–218. https://doi.org/10.3109/09638288.2010.508829

Hilari, K., Byng, S., Lamping, D. L., & Smith, S. C. (2003). Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): Evaluation of Acceptability, Reliability, and Validity. *Stroke*, *34*(8), 1944–1950. https://doi.org/10.1161/01.STR.0000081987.46660.ED

Hillis, A. E. (1989). Efficacy and generalization of treatment for aphasic naming errors. *Archives of Physical Medicine & Rehabilitation*, *70*(8), 632–636.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., & Milne, R. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *Bmj*.

Hough, M. S. (2010). Melodic Intonation Therapy and aphasia: Another variation on a theme. *Aphasiology*, *24*(6-8), 775–786. https://doi.org/10.1080/02687030903501941

Huber, W., Poeck, K., & Willmes, K. (1984). The Aachen Aphasia Test. *Advances in Neurology*, *42*, 291–303.

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), 1–19. https://doi.org/10.1080/02687038.2011.602514

Johnson, M. L., Taub, E., Harper, L. H., Wade, J. T., Bowman, M. H., Bishop-McKay, S., et al. (2014). An enhanced protocol for constraint-induced aphasia therapy II: A case series. *American Journal of Speech-Language Pathology*, *23*(1), 60–72. https://doi.org/10.1044/1058-0360(2013/12-0168)

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). The Boston naming test. Philadelphia, PA: Lea & Febiger.

Kavian, S., Khatoonabadi, A. R., Ansari, N. N., Saadati, M., & Shaygannejad, V. (2014). A

single-subject study to examine the effects of constrained-induced aphasia therapy on naming deficit. *International Journal of Preventive Medicine*, *5*(6), 782–786.
Kempler, D., & Goral, M. (2011). A comparison of drill- and communication-based treatment for aphasia. *Aphasiology*, *25*(11), 1327–1346. https://doi.org/10.1080/02687038.2011.599364
Kertesz, A. (2007). Western Aphasia Battery (Revised). San Antonio, TX: PsychCorp.
Kiran, S. (2005). Training phoneme to grapheme conversion for patients with written and oral production deficits: A model-based approach. *Aphasiology*, *19*(1), 53–76. https://doi.org/10.1080/02687030444000633
Kurland, J., Baldwin, K., & Tauer, C. (2010). Treatment-induced neuroplasticity following intensive naming therapy in a case of chronic Wernicke's aphasia. *Aphasiology*, *24*(6-8), 737–751. https://doi.org/10.1080/02687030903524711
Kurland, J., Pulvermüller, F., Silva, N., Burke, K., & Andrianopoulos, M. (2012). Constrained versus unconstrained intensive language therapy in two individuals with chronic, moderate-to-severe aphasia and apraxia of speech: Behavioral and fMRI outcomes. *American Journal of Speech-Language Pathology*, *21*(2), S65–S87. https://doi.org/10.1044/1058-0360(2012/11-0113)
Lam, J. M. C., & Wodchis, W. P. (2010). The Relationship of 60 Disease Diagnoses and 15 Conditions to Preference-Based Health-Related Quality of Life in Ontario Hospital-Based Long-Term Care Residents. *Medical Care*, *48*(4), 380–387. https://doi.org/10.1097/MLR.0b013e3181ca2647
Lim, K.-B., Kim, Y.-K., Lee, H.-J., Yoo, J., Hwang, J. Y., Kim, J.-A., & Kim, S.-K. (2013). The therapeutic effect of neurologic music therapy and speech language therapy in post-stroke aphasic patients. *Annals of Rehabilitation Medicine*, *37*(4), 556–562. https://doi.org/10.5535/arm.2013.37.4.556
Lomas, J., Pickard, L., Bester, S., & Elbard, H. (1989). The Communicative Effectiveness Index: Development and Psychometric Evaluation of a Functional Communication Measure for Adult Aphasia. *Journal of Speech and …*, *54*, 113–124.
Lorenz, A., & Nickels, L. A. (2007). Orthographic cueing in anomic aphasia: How does it work? *Aphasiology*, *21*(6-8), 670–686. https://doi.org/10.1080/02687030701192182
Luria, A. R. (1970). Traumatic aphasia: Its syndromes, psychology and treatment. (D. Bowden, Trans.). The Hague, Netherlands: Mouton & Co.
MacGregor, L. J., Difrancesco, S., Pulvermüller, F., Shtyrov, Y., & Mohr, B. (2015). Ultra-rapid access to words in chronic aphasia: The effects of intensive language action therapy (ILAT). *Brain Topography*, *28*(2), 279–291. https://doi.org/10.1007/s10548-014-0398-y
Maher, L. M., Kendall, D. L., Swearengin, J. A., Rodriguez, A., Leon, S. A., Pingel, K., et al. (2006). A pilot study of use-dependent learning in the context of Constraint Induced Language Therapy., *12*(6), 843–852. https://doi.org/10.1017/S1355617706061029
Makuuchi, M., Kaminaga, T., & Sugishita, M. (2003). Both parietal lobes are involved in drawing: a functional MRI study and implications for constructional apraxia. *Cognitive Brain Research*, *16*(3), 338–347. https://doi.org/10.1016/S0926-6410(02)00302-6
Marangolo, P., Bonifazi, S., Tomaiuolo, F., Craighero, L., Coccia, M., Altoè, G., et al. (2010). Improving language without words: First evidence from aphasia. *Neuropsychologia*, *48*(13), 3824–3833. https://doi.org/10.1016/j.neuropsychologia.2010.09.025
Martin, P. I., Treglia, E., Naeser, M. A., Ho, M. D., Baker, E. H., Martin, E. G., et al. (2014). Language improvements after TMS plus modified CILT: Pilot, open-protocol study with two, chronic nonfluent aphasia cases. *Restorative Neurology and Neuroscience*, *32*(4), 483–505. https://doi.org/10.3233/RNN-130365
Maul, K. K., Conner, P. S., Kempler, D., Radvanski, C., & Goral, M. (2014). Using

informative verbal exchanges to promote verb retrieval in nonfluent aphasia. *American Journal of Speech-Language Pathology*, *23*(3), 407–420. https://doi.org/10.1044/2014_AJSLP-13-0004

McNeill, D. (1992). Hand and Mind. Chicago, IL: University of Chicago Press.

Meinzer, M., Djundja, D., Barthel, G., Elbert, T., & Rockstroh, B. (2005). Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke*, *36*(7), 1462–1466. https://doi.org/10.1161/01.STR.0000169941.29831.2a

Meinzer, M., Streiftau, S., & Rockstroh, B. (2007). Intensive language training in the rehabilitation of chronic aphasia: Efficient training by laypersons., *13*(5), 846–853. https://doi.org/10.1017/S1355617707071111

Mohr, B., Difrancesco, S., Harrington, K., Evans, S., & Pulvermüller, F. (2014). Changes of right-hemispheric activation after constraint-induced, intensive language action therapy in chronic aphasia: fMRI evidence from auditory semantic processing. *Frontiers in Human Neuroscience*, *8*(309), 919. https://doi.org/10.3389/fnhum.2014.00919

Morrow-Odom, K. L., & Swann, A. B. (2013). Effectiveness of melodic intonation therapy in a case of aphasia following right hemisphere stroke. *Aphasiology*, *27*(11), 1322–1338. https://doi.org/10.1080/02687038.2013.817522

Nickels, L. A. (1992). The autocue? self-generated phonemic cues in the treatment of a disorder of reading and naming. *Cognitive Neuropsychology*, *9*(2), 155–182. https://doi.org/10.1080/02643299208252057

Nickels, L. A. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10-11), 935–979. https://doi.org/10.1080/02687030244000563

OCEBM Levels of Evidence Working Group. (2011). *The Oxford Levels of Evidence 2*. Oxford Centre for Evidence-Based Medicine. Retrieved from http://www.cebm.net/index.aspx?o=5653

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, *42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Patrício, B., Jesus, L., & Cruice, M. (2013). *Quality of life of the caregivers of people with aphasia. A systematic review.*

Paul, D. R., Frattali, C. M., Holland, A. L., Thompson, C. K., Caperton, C. J., & Slater, S. C. (2003). ASHA Quality of Communication Life Scale (QCL). Rockville, MD: American Speech-Language-Hearing Association.

Pulvermüller, F., & Berthier, M. L. (2008). Aphasia therapy on a neuroscience basis. *Aphasiology*, *22*(6), 563–599. https://doi.org/10.1080/02687030701612213

Pulvermüller, F., Hauk, O., Zohsel, K., Neininger, B., & Mohr, B. (2005). Therapy-related reorganization of language in both hemispheres of patients with chronic aphasia. *NeuroImage*, *28*(2), 481–489. https://doi.org/10.1016/j.neuroimage.2005.06.038

Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke*, *32*(7), 1621–1626. https://doi.org/10.1161/01.STR.32.7.1621

Ravenek, M. J., Skarakis-Doyle, E., Spaulding, S. J., Jenkins, M. E., & Doyle, P. C. (2015). Enhancing the conceptual clarity and utility of the international classification of functioning, disability & health: the potential of a new graphic representation. *Disability and Rehabilitation*, *35*(12), 1015–1025. https://doi.org/10.3109/09638288.2012.717582

Raymer, A. M., McHose, B., Smith, K. G., Iman, L., Ambrose, A., & Casselton, C. (2012). Contrasting effects of errorless naming treatment and gestural facilitation for word retrieval in aphasia. *Neuropsychological Rehabilitation*, *22*(2), 235–266. https://doi.org/10.1080/09602011.2011.618306

Raymer, A. M., Singletary, F., Rodriguez, A., Ciampitti, M., Heilman, K. M., & Gonzalez Rothi, L. J. (2006). Effects of gesture+verbal treatment for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society*, *12*(6), 867–882. https://doi.org/10.10170S1355617706061042
Richter, M., Miltner, W. H. R., & Straube, T. (2008). Association between therapy outcome and right-hemispheric activation in chronic aphasia. *Brain*, *131*(Pt 5), 1391–1401. https://doi.org/10.1093/brain/awn043
Rodriguez, A. D., Raymer, A. M., & Gonzalez Rothi, L. J. (2006). Effects of gesture+verbal and semantic-phonologic treatments for verb retrieval in aphasia. *Aphasiology*, *20*(2-4), 286–297. https://doi.org/10.1080/02687030500474898
Rombough, R. E., Howse, E. L., & Bartfay, W. J. (2006). Caregiver strain and caregiver burden of primary caregivers of stroke survivors with and without aphasia. *Rehabilitation Nursing : the Official Journal of the Association of Rehabilitation Nurses*, *31*(5), 199–209.
Rose, M. L. (2006). The utility of arm and hand gestures in the treatment of aphasia. *Advances in Speech Language Pathology*, *8*(2), 92–109. https://doi.org/10.1080/14417040600657948
Rose, M. L. (2013). Releasing the constraints on aphasia therapy: the positive impact of gesture and multimodality treatments. *American Journal of Speech-Language Pathology*, *22*(2), S227–39. https://doi.org/10.1044/1058-0360(2012/12-0091)
Rose, M. L., & Douglas, J. (2008). Treating a semantic word production deficit in aphasia with verbal and gesture methods. *Aphasiology*, *22*(1), 20–41. https://doi.org/10.1080/02687030600742020
Rose, M. L., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology*, *22*(7-8), 691–706. https://doi.org/10.1080/02687030701800800
Rose, M. L., Attard, M. C., Mok, Z., Lanyon, L. E., & Foster, A. M. (2013). Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study. *Aphasiology*, *27*(8), 938–971. https://doi.org/10.1080/02687038.2013.810329
Rose, M. L., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology*, *16*(10-11), 1001–1030. https://doi.org/10.1080/02687030143000825
Sacchett, C. (2002). Drawing in aphasia: moving towards the interactive. *International Journal of Human-Computer Studies*, *57*(4), 263–277. https://doi.org/10.1006/ijhc.2002.1018
Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: why singing may lead to recovery of expressive language function in patients with broca's aphasia. *Music Perception*, *25*(4), 315–323. https://doi.org/10.1525/MP.2008.25.4.315
Sherrington, C., Herbert, R. D., Maher, C. G., & Moseley, A. M. (2000). PEDro. A database of randomized trials and systematic reviews in physiotherapy. *Manual Therapy*, *5*(4), 223–226. https://doi.org/10.1054/math.2000.0372
Smania, N. (2006). Constraint-induced movement therapy: an original concept in rehabilitation.
Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, *10*(4), 303–316.
Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: Evidence for two paths of speech recovery. *Frontiers in Human Neuroscience*, *7*(Article 35), 1–12. https://doi.org/10.3389/fnhum.2013.00035

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*, *134*(Pt 10), 3083–3093. https://doi.org/10.1093/brain/awr240

Sugishita, M., Seki, K., Kabe, S., & Yunoki, K. (1993). A material-control single-case study of the efficacy of treatment for written and oral naming difficulties. *Neuropsychologia*, *31*(6), 559–569.

Szaflarski, J. P., Ball, A. L., Vannest, J., Dietz, A. R., Allendorfer, J. B., Martin, A. N., et al. (2015). Constraint-Induced Aphasia Therapy for Treatment of Chronic Post-Stroke Aphasia: A Randomized, Blinded, Controlled Pilot Trial. *Medical Science Monitor*, *21*, 2861–2869. https://doi.org/10.12659/MSM.894291

Szaflarski, J. P., Ball, A., Grether, S., Al-fwaress, F., Griffith, N. M., Neils-Strunjas, J., et al. (2008). Constraint-induced aphasia therapy stimulates language recovery in patients with chronic aphasia after ischemic stroke. *Medical Science Monitor*, *14*(5), CR243–CR250.

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., et al. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. *Aphasiology*, *30*(7), 862–876. https://doi.org/10.1080/02687038.2016.1178022

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*(5), 619–638. https://doi.org/10.1080/09602011.2013.824383

Tate, R. L., Rosenkoetter, U., Wakim, D., Sigmundsdottir, L., Doubleday, J., Togher, L., et al. (2015). The Risk of Bias in N-of-1 Trials (RoBiNT) Scale: An expanded manual for the critical appraisal of single-case reports. Sydney, Australia: Author.

Taub, E. (1976). Movement in nonhuman primates deprived of somatosensory feedback. *Exercise and Sport Sciences Reviews*, *4*, 335–374.

Taub, E., Uswatte, G., Mark, V. W., & Morris, D. M. (2006). The learned nonuse phenomenon: implications for rehabilitation. *Europa Medicophysica*, *42*(3), 241–255.

Teasell, R., Bayona, N., Marshall, S., Cullen, N., Bayley, M., Chundamala, J., et al. (2007). A systematic review of the rehabilitation of moderate to severe acquired brain injuries. *Brain Injury*, *21*(2), 107–112. https://doi.org/10.1080/02699050701201524

van der Meulen, I., van de Sandt-Koenderman, M. E., & Ribbers, G. M. (2012). Melodic Intonation Therapy: Present controversies and future opportunities. *Archives of Physical Medicine and Rehabilitation*, *93*(S1), S46–S52. https://doi.org/10.1016/j.apmr.2011.05.029

Vines, B. W., Norton, A. C., & Schlaug, G. (2011). Non-invasive brain stimulation enhances the effects of melodic intonation therapy. *Frontiers in Psychology*, *2*(Article 230), 230. https://doi.org/10.3389/fpsyg.2011.00230

Wallace, S. (2016, December 18). *Improving Research Outcome Measurement in Aphasia (ROMA): Development of a Core Outcome Set*. Retrieved from http://espace.library.uq.edu.au/view/UQ:415571

Wallace, S. J., Worrall, L. E., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., et al. (2016). Which outcomes are most important to people with aphasia and their families? an international nominal group technique study framed within the ICF. *Disability and Rehabilitation*, *91*, 1–16. https://doi.org/10.1080/09638288.2016.1194899

Wallace, S., Worrall, L. E., Rose, T., & Le Dorze, G. (n.d.). Improving Research Outcome Measurement in Aphasia (ROMA): Development of a Core Outcome Set. Retrieved March 13, 2017, from http://www.cometinitiative.org/studies/details/287

Wan, C. Y., Zheng, X., Marchina, S., Norton, A., & Schlaug, G. (2014). Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca's aphasia. *Brain and Language*, *136*, 1–7. https://doi.org/10.1016/j.bandl.2014.03.011

Weill-Chounlamountry, A., Capelle, N., Tessier, C., & Pradat-Diehl, P. (2013). Multimodal therapy of word retrieval disorder due to phonological encoding dysfunction. *Brain Injury*, *27*(5), 620–631. https://doi.org/10.3109/02699052.2013.767936

Wilson, S. J., Parsons, K., & Reutens, D. C. (2006). Preserved singing in aphasia: A case study of the efficacy of Melodic Intonation Therapy. *Music Perception*, *24*(1), 23–36.

Wilssens, I., Vandenborre, D., van Dun, K., Verhoeven, J., Visch-Brink, E., & Marien, P. (2015). Constraint-induced aphasia therapy versus intensive semantic treatment in fluent aphasia. *American Journal of Speech-Language Pathology*, *24*(2), 281–294. https://doi.org/10.1044/2015_AJSLP-14-0018

Wright, H. H., Marshall, R. C., Wilson, K. B., & Page, J. L. (2008). Using a written cueing hierarchy to improve verbal naming in aphasia. *Aphasiology*, *22*(5), 522–536. https://doi.org/10.1080/02687030701487905

Zumbansen, A., Peretz, I., & Hébert, S. (2014a). Melodic intonation therapy: back to basics for future research. *Frontiers in Neurology*, *5*(7), 1–11. https://doi.org/10.3389/fneur.2014.00007/abstract

Zumbansen, A., Peretz, I., & Hébert, S. (2014b). The combination of rhythm and pitch can account for the beneficial effect of melodic intonation therapy on connected speech improvements in Broca's aphasia. *Frontiers in Human Neuroscience*, *8*(Article 592). https://doi.org/10.3389/fnhum.2014.00592/abstract

Appendix

Summary table of all eligible papers

Treatment	Oxford level	Publication	Title	Treatment(s)	n	Outcome type
Constraint	Level 2 - Randomised trial	(Pulvermüller et al., 2001)	Constraint-Induced Therapy of Chronic Aphasia After Stroke	CIAT *vs.* Standard therapy, less intensive	17 (10 constraint)	Impairment Activity/Participation
Constraint	Level 2 - Randomised trial	(Meinzer et al., 2007)	Intensive language training in the rehabilitation of chronic aphasia: efficient training by laypersons.	CIAT by therapists *vs.* CIAT by laypersons	20	Impairment
Constraint	Level 2 - Randomised trial	(Wilssens et al., 2015)	Constraint-Induced Aphasia Therapy Versus Intensive Semantic Treatment in Fluent Aphasia	CIAT *vs.* Standard Therapy	9 (5 constraint)	Impairment Activity/Participation
Constraint	Level 3 - Non-randomised, controlled cohort/followup study	(L. M. Maher et al., 2006)	A pilot study of use-dependent learning in the context of Constraint Induced Language Therapy.	CIAT *vs.* PACE	9 (4 constraint)	Impairment
Constraint	Level 3 - Non-randomised, controlled cohort/followup study	(Meinzer et al., 2005)	Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy	CIAT *vs.* CIAT Plus	27	Impairment Activity/Participation
Constraint	Level 4 - Case series	(Johnson et al., 2014)	An enhanced protocol for constraint-induced aphasia therapy II: a case series.	CIAT II	4	Impairment Activity/Participation
Constraint	Level 4 - Case series	(Faroqi-Shah & Virion, 2009)	Constraint-induced language therapy for agrammatism: Role of grammaticality constraints	CIAT *vs.* CIAT-G	4	Impairment
Constraint	Level 4 - Case series	(Kempler & Goral, 2011)	A comparison of drill- and communication-based treatment for aphasia	Generative CIAT *vs.* Drill CIAT	2	Impairment
Constraint	Level 4 - Case series	(Goral & Kempler, 2009)	Training verb production in communicative context: evidence from a person with chronic non-fluent aphasia	Modified CIAT (reduced intensity)	1	Impairment
Constraint	Level 4 - Case series	(Maul et al., 2014)	Using informative verbal exchanges to promote verb retrieval in nonfluent aphasia	Modified CIAT (reduced intensity, modelling of target sentences)	4	Impairment
Constraint	Level 4 - Case series	(Breier et al., 2009)	Behavioral and neurophysiologic response to therapy for chronic aphasia	CIAT	23	Impairment
Constraint	Level 4 - Case series	(Breier, Juranek, & Papanicolaou, 2011)	Changes in maps of language function and the integrity of the arcuate fasciculus after therapy for chronic aphasia	CIAT	1	Impairment
Constraint	Level 4 - Case series	(Breier et al., 2006)	Functional imaging before and after constraint-induced language therapy for aphasia using magnetoencephalography.	CIAT	6	Impairment
Constraint	Level 4 - Case series	(Breier, Maher, Schmadeke, Hasan, & Papanicolaou, 2007)	Changes in language-specific brain activation after therapy for aphasia using magnetoencephalography: A case study	CIAT	1	Impairment
Constraint	Level 4 - Case series	(MacGregor et al., 2015)	Ultra-rapid access to words in chronic aphasia: The effects of intensive language action therapy (ILAT)	CIAT	12	Impairment

Treatment	Oxford level	Publication	Title	Treatment(s)	n	Outcome type
Constraint	Level 4 - Case series	(Mohr et al., 2014)	Changes of right-hemispheric activation after constraint-induced, intensive language action therapy in chronic aphasia: fMRI evidence from auditory semantic processing.	CIAT	12	Impairment
Constraint	Level 4 - Case series	(Pulvermüller, Hauk, Zohsel, Neininger, & Mohr, 2005)	Therapy-related reorganization of language in both hemispheres of patients with chronic aphasia	CIAT	10	Impairment
Constraint	Level 4 - Case series	(Richter, Miltner, & Straube, 2008)	Association between therapy outcome and right-hemispheric activation in chronic aphasia	CIAT	24 (16 constraint)	Impairment
Constraint	Level 4 - Case series	(Szaflarski et al., 2008)	Constraint-induced aphasia therapy stimulates language recovery in patients with chronic aphasia after ischemic stroke	CIAT (individualised goals and stimuli)	3	Impairment Activity/Participation
Constraint	Level 4 - Case series	(Kurland, Baldwin, & Tauer, 2010)	Treatment-induced neuroplasticity following intensive naming therapy in a case of chronic wernicke's aphasia	CIAT *vs.* PACE	1	Impairment
Constraint	Level 4 - Single case experimental design	(Kavian, Khatoonabadi, Ansari, Saadati, & Shaygannejad, 2014)	A Single-subject Study to Examine the Effects of Constrained-induced Aphasia Therapy on Naming Deficit.	CIAT	2	Impairment
Constraint	Level 4 - Single case experimental design	(Kurland et al., 2012)	Constrained Versus Unconstrained Intensive Language Therapy in Two Individuals With Chronic, Moderate-to-Severe Aphasia and Apraxia of Speech: Behavioral and fMRI Outcomes.	CIAT *vs.* PACE	2	Impairment
Combined Multimodal & Constraint	Level 4 - Single case experimental design	(Attard et al., 2013)	The comparative effects of Multi-Modality Aphasia Therapy and Constraint-Induced Aphasia Therapy-Plus for severe chronic Broca's aphasia: An in-depth pilot study.	CIAT Plus *vs.* MMAT	2	Impairment, Activity/Participation Quality of Life
Combined Multimodal & Constraint	Level 4 - Single case experimental design	(Rose et al., 2013)	Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study	CIAT Plus *vs.* MMAT	11	Impairment, Activity/Participation Quality of Life
Gesture	Level 4 - Case series	(Carragher, Sage, & Conroy, 2013)	The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation	Semantic Feature Analysis + Gesture + phonemic cueing	9	Impairment
Gesture	Level 4 - Case series	(Marangolo et al., 2010)	Improving language without words: first evidence from aphasia	Action observation *vs.* Action observation and execution *vs.* Action observation and meaningless movement	6 (5 with stroke)	Impairment
Gesture	Level 4 - Single case experimental design	(Boo & Rose, 2011)	The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia	Repetition *vs.* semantic *vs.* gesture *vs.* semantic + gesture	2	Impairment Activity/Participation
Gesture	Level 4 - Single case experimental design	(Ferguson et al., 2012)	A comparison of intention and pantomime gesture treatment for noun retrieval in people with aphasia	Intention gesture *vs.* pantomime gesture	4	Impairment
Gesture	Level 4 - Single case experimental design	(Raymer et al., 2012)	Contrasting effects of errorless naming treatment and gestural facilitation for word retrieval in aphasia	Errorless naming *vs.* gesture	8	Impairment Activity/Participation
Gesture	Level 4 - Single case experimental design	(Raymer et al., 2006)	Effects of gesture+verbal treatment for noun and verb retrieval in aphasia	Gesture + verbal	9	Impairment
Gesture	Level 4 - Single case experimental design	(Rodriguez et al., 2006)	Effects of gesture+verbal and semantic-phonologic treatments for verb retrieval in aphasia	Gesture + verbal *vs.* Semantic + phonologic	4	Impairment

Treatment	Oxford level	Publication	Title	Treatment(s)	n	Outcome type
Gesture	Level 4 - Single case experimental design	(Rose & Douglas, 2008)	Treating a semantic word production deficit in aphasia with verbal and gesture methods	Verbal *vs.* Gesture *vs.* Verbal + gesture	1	Impairment
Gesture	Level 4 - Single case experimental design	(Rose, Douglas, & Matyas, 2002)	The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment	Verbal *vs.* Gesture *vs.* Verbal + gesture	1	Impairment
Gesture	Level 4 - Single case experimental design	(Rose & Sussmilch, 2008)	The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia	Semantic *vs.* Gesture *vs.* Semantic + Gesture *vs.* Repetition	3	Impairment Activity/Participation
Music	Level 3 - Non-randomised, controlled cohort/followup study	(Wan, Zheng, Marchina, Norton, & Schlaug, 2014)	Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca's aphasia	MIT *vs.* no therapy	20 (11 MIT)	Impairment
Music	Level 3 - Non-randomised, controlled cohort/followup study	(Lim et al., 2013)	The therapeutic effect of neurologic music therapy and speech language therapy in post-stroke aphasic patients	MIT *vs.* Standard Therapy	21 (11 chronic, 6 chronic and music)	Impairment
Music	Level 4 - Case series	(Bonakdarpour, Eftekharzadeh, & Ashayeri, 2003)	Melodic intonation therapy in Persian aphasic patients	MIT (adapted to Farsi)	7	Impairment
Music	Level 4 - Case series	(Breier, Randle, Maher, & Papanicolaou, 2010)	Changes in maps of language activity activation following melodic intonation therapy using magnetoencephalography: two case studies.	MIT	2	Impairment
Music	Level 4 - Case series	(Goldfarb & Bader, 1979)	Espousing melodic intonation therapy in aphasia rehabilitation: a case study	MIT	1	Impairment
Music	Level 4 - Case series	(Morrow-Odom & Swann, 2013)	Effectiveness of melodic intonation therapy in a case of aphasia following right hemisphere stroke	MIT	1	Impairment Activity/Participation Quality of Life
Music	Level 4 - Case series	(Schlaug, Marchina, & Norton, 2008)	From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca's aphasia	MIT *vs.* Speech Repetition Therapy	2	Impairment
Music	Level 4 - Case series	(Sparks et al., 1974)	Aphasia rehabilitation resulting from melodic intonation therapy	MIT	9	Impairment
Music	Level 4 - Case series	(van der Meulen, van de Sandt-Koenderman, & Ribbers, 2012)	Melodic Intonation Therapy: Present Controversies and Future Opportunities	MIT	2	Impairment
Music	Level 4 - Case series	(Wilson, Parsons, & Reutens, 2006)	Preserved Singing in Aphasia: A Case Study of the Efficacy of Melodic Intonation Therapy	MIT *vs.* Rhythmic therapy	1	Impairment
Music	Level 4 - Single case experimental design	(Hough, 2010)	Melodic intonation therapy and aphasia: Another variation on a theme	Modified MIT (no tapping)	1	Impairment Activity/Participation Quality of Life
Music	Level 3 - Non-randomised, controlled cohort/followup study	(Stahl et al., 2013)	How to engage the right brain hemisphere in aphasics without even singing: Evidence for two paths of speech recovery	Therapy *vs.* Repetition Therapy *vs.* Rhythmic Therapy	15 (5 music, 5 rhythm)	Impairment Activity/Participation
Music	Level 4 - Case series	(Zumbansen, Peretz, & Hébert, 2014b)	The combination of rhythm and pitch can account for the beneficial effect of melodic intonation therapy on connected speech improvements in Broca's aphasia	Melodic therapy *vs.* Rhythmic therapy *vs.* Standard Therapy	3	Impairment Quality of Life

Treatment	Oxford level	Publication	Title	Treatment(s)	n	Outcome type
Pharma/stimulation	Level 2 - Randomised trial & Level 4 - Case series	(Barbancho et al., 2015)	Bilateral brain reorganization with memantine and constraint-induced aphasia therapy in chronic post-stroke aphasia: An ERP study.	CIAT + Memantine	28	Impairment
Pharma/stimulation	Level 2 - Randomised trial & Level 4 - Case series	(Berthier et al., 2009)	Memantine and Constraint-Induced Aphasia Therapy in Chronic Poststroke Aphasia	CIAT + Memantine	27	Impairment Activity/Participation
Pharma/stimulation	Level 4 - Case series	(Abo et al., 2012)	Effectiveness of Low-Frequency rTMS and Intensive Speech Therapy in Poststroke Patients with Aphasia: A Pilot Study Based on Evaluation by fMRI in Relation to Type of Aphasia	Constraint therapy + rTMS	24	Impairment
Pharma/stimulation	Level 4 - Case series	(Martin et al., 2014)	Language improvements after TMS plus modified CILT: Pilot, open-protocol study with two, chronic nonfluent aphasia cases.	Modified CILT + TMS	2	Impairment
Pharma/stimulation	Level 4 - Case series	(Vines et al., 2011)	Non-invasive brain stimulation enhances the effects of melodic intonation therapy	MIT + tDCS	6	Impairment
Pharma/stimulation	Level 4 - Single case experimental design	(Al-Janabi et al., 2014)	Augmenting melodic intonation therapy with non-invasive brain stimulation to treat impaired left-hemisphere function: two case studies	MIT + rTMS	2	Impairment
Writing	Level 4 - Case series	(Sugishita, Seki, Kabe, & Yunoki, 1993)	A material-control single-case study of the efficacy of treatment for written and oral naming difficulties	Naming with written cueing hierarchy	22 (3 chronic)	Impairment
Writing	Level 4 - Case series	(Weill-Chounlamountry, Capelle, Tessier, & Pradat-Diehl, 2013)	Multimodal therapy of word retrieval disorder due to phonological encoding dysfunction	Therapy software “Au fil des mots” (anagrams, copying, writing)	1	Impairment Activity/Participation
Writing	Level 4 - Single case experimental design	(Ball, de Riesthal, Breeding, & Mendoza, 2011)	Modified ACT and CART in severe aphasia	ACT + CART	3	Impairment
Writing	Level 4 - Single case experimental design	(Beeson & Egnor, 2006)	Combining treatment for written and spoken naming	CART + repetition	2	Impairment
Writing	Level 4 - Single case experimental design	(DeDe et al., 2003)	Teaching self-cues: A treatment approach for verbal naming	Written naming + tactile cueing + verbal naming	1	Impairment
Writing	Level 4 - Single case experimental design	(Hillis, 1989)	Efficacy and generalization of treatment for aphasic naming errors	Written naming + verbal naming	2 (1 within chronic criteria)	Impairment
Writing	Level 4 - Single case experimental design	(Wright et al., 2008)	Using a written cueing hierarchy to improve verbal naming in aphasia	Written cueing hierarchy based on CART	2	Impairment

Table 2 – Impairment-based outcomes of high-quality studies

Treatment	Authors	Outcomes	Results
Constraint	(Pulvermüller et al., 2001)	AAT (comprehension, repetition, naming, token test)	Group x Time effect in favour of CIAT (F[1,15]=5.0; P<0.04)
		CAL (blinded clinician ratings)	*7/10 CIAT participants improved. Significant but statistical testing only conducted on seven (F[1,6] 10.5, P<0.01). Control group CAL scores not reported.*
Constraint	(Meinzer et al., 2007)	AAT	Time effect significant for both groups Clinicians: t(9) = 7.05, p<.0001 Laypersons: t(9) = 5.65, p<.002 Group x Time effect non-significant (F [1,18] = 1.26; p>.2) 19/20 improved as per critical difference in manual
Constraint	(Wilssens et al., 2015)	*AAT*	*All participants improved on at least one subtest, no overall profile scores reported.* *CIAT group: Statistically significant improvement on 4/5 subtests* *BOX group: Statistically significant improvement on 1/5 subtests*
		BNT (/60)	*Pre/post CIAT group (p = .004), Pre/post BOX group (p = .094), no between group testing.*
		SAT	
		PALPA 51, 49, 8, 5, 6)	*No significant change in 8/9, no between group testing.* *Pre/post semantic scores favoured BOX, Pre/post phonological scores favoured CIAT*
Constraint	(Attard et al., 2013)	Confrontation naming – treated items	Tau-U 1.00
		WAB AQ (0-100)	*No change pre/post CIAT Plus*
		BNT (/60)	*Non-significant changes pre/post CIAT Plus*
		Cinderella Narrative Retell	*No statistically sig. improvement*
		Semi-structured conversation	*No statistically sig. improvement*
Constraint	(Rose et al., 2013)	Confrontation naming – treated items	Tau-U 0.24

Treatment	Authors	Outcomes	Results
		WAB-R AQ (0-100) [a] *BNT (/60)* [a] *Semi-structured conversation* [a]	*Mean change 2.47 (range -3.00 to 7.70)* *Mean change 1.9 (range -6.0 to 15.0)* *Substantive nouns mean change 4.67 (range -64.0 to 76.0), substantive verbs unchanged overall*
Constraint	(Kurland et al., 2012)	Confrontation naming – treated items *BNT (/60)* *BDAE-3*	Tau-U 0.69 *P1: 25 to 35 (though pre-treatment was 32); P2: 23 to 33* *P1: Unchanged; P2: Responsive naming 6 to 11, others largely unchanged*
Combined Multimodal	(Attard et al., 2013)	Confrontation naming – treated items *WAB AQ (0-100)* *BNT (/60)* *Cinderella Narrative Retell* *Semi-structured conversation*	Tau-U 0.81 *Non-significant changes pre/post M-MAT* *Non-significant changes pre/post M-MAT* *No statistically sig. improvement* *No statistically sig. improvement*
Combined Multimodal	(Rose et al., 2013)	Confrontation naming – treated items *WAB-R AQ (0-100)* [a] *BNT (/60)* [a] *Semi-structured conversation* [a]	Tau-U 0.68 *Mean change 1.40 (range -4.2 to 9)* *Mean change 2.2 (range -9.0 to 15.0).* *Substantive nouns mean change 20.4 (range -24.0 to 87.0), substantive verbs largely unchanged*
Gesture	(Boo & Rose, 2011) [b]	Confrontation naming – treated items	Tau-U 0.85 (all gesture sets)
Gesture	(Ferguson et al., 2012)	Confrontation naming – treated items	Tau-U -0.04 (pantomime)

Treatment	**Authors**	**Outcomes**	**Results**
Gesture	(Rodriguez et al., 2006) [b]	Confrontation naming – treated items	Tau-U 0.46 (all gesture sets)
Gesture	(Raymer et al., 2006)	Confrontation naming – treated items *WAB AQ (0-100)* *BNT (/60)* *ANT (/60)*	Tau-U 0.67, one participant's results not published due to poor response *Mean change 4.8 (p = 0.15)* *Mean change 1.2 (range -3 to 9)* *Mean change -1.4 (range -11.6 to 7.4)*
Gesture	(Rose et al., 2002)	Confrontation naming – treated items	Tau-U 0.74
Gesture	(Raymer et al., 2012)	Confrontation naming – treated items *WAB AQ (0-100) (pre/post gesture)* *BNT (/60) (pre/post gesture)*	Tau-U 0.25 *Mean change 5.5 (range 0.2 to 16.1)* *Mean change 3 (range -15 to 13)*
Gesture	(Rose & Sussmilch, 2008) [b]	Confrontation naming – treated items	Tau-U 0.73
Gesture	(Rose & Douglas, 2008) [b]	Confrontation naming – treated items	Tau-U 0.91
Music	(Stahl et al., 2013)	All in unison with recordings & written prompts: Singing & speaking trained phrases (% correct) Singing & speaking untrained phrases (% correct)	 Singing therapy: Mean change = 36.47, 95% CI [28.24, 44.70]; Standard therapy: Mean change = 4.98, 95% CI [−3.25, 13.21] Singing therapy: Mean change = −0.36, 95% CI [−2.62, 1.90]; Standard therapy: Mean change = 6.21, 95% CI [3.96, 8.47]

Treatment	Authors	Outcomes	Results
Music	(Hough, 2010)	Repetition of treated phrases *WAB AQ (0-100)* *WAB CQ (0-100)*	Pre/post change ($p < 0.0001$) *Improved 13.0* *Improved 13.6*
Writing	(Wright et al., 2008)	Confrontation naming – treated items *WAB AQ (0-100)* *BNT (/60)*	Tau-U 0.90 *Pre/post improvement – P1: 3.1, P2: 9.7* *Pre/post improvement – P1: -3; P2: 5*
Writing	(Hillis, 1989)	Confrontation naming – treated items	Tau-U 0.99
Writing	(Beeson & Egnor, 2006)	Confrontation naming – treated items *PALPA 53 (/40)*	Tau-U 0.86 *Unchanged*

Note. AAT = Aachen Aphasia Test (Profile score), CAL = Communicative Activity Log, BNT = Boston Naming Test, SAT = Verbal Semantic Association Test, PALPA = Psycholinguistic Assessment of Language Processing in Aphasia, WAB-R AQ/CQ = Western Aphasia Battery [Revised] (Aphasia Quotient/Cortical Quotient), BDAE = Boston Diagnostic Aphasia Examination, ANT = Action Naming Test, VAST = Verb And Sentence Test, OANB = Object Action Naming Battery.

[a]The crossover design makes it difficult to isolate the contributions of pre/post outcome measures, but taking a conservative approach, where the treatment of interest was administered second, the post treatment measure of the first treatment was taken as the baseline.

[b]Pre/post measures represent changes from multiple treatments and so are not listed.

Table 3 – Activity/Participation outcomes of high-quality studies

Treatment	Authors	Outcomes	Results
Constraint	(Pulvermüller et al., 2001)	*CAL (self ratings)*	*Pre/post CIAT group (F[1,7]=25.0, P<0.001); Pre/post control group (F<1); no between group comparison.*
Constraint	(Wilssens et al., 2015)	CETI (/100) (self and family ratings)	No between-group difference in improvements (t(6) = 1.01, p = .332); however: Pre/post CIAT group (t(4) = 1.47, p = .216), < clinically significant difference Pre/post BOX group (t(2) = 7.40, p = .019), > clinically significant difference
		ANELT	Statistically significant improvement for both groups; no significant difference between groups t(7) = –0.85, p = .426
Constraint	(Attard et al., 2013)	*CETI (/100)*	*Both less than clinically significant difference (12). P1: 3 point increase; P2: 3 point increase*
		Scenario test (/54)	*Improved by 3 and 8 points*
Constraint	(Rose et al., 2013)	*CETI (/100)* [a]	*Mean change 4 points (range -3 to 13). 2/11 participants > clinically significant difference*
		Scenario test (/54) [a]	*Mean change -0.1/54 (range -10.0 to 7.70)*
Combined Multimodal	(Attard et al., 2013)	*CETI (/100)* [a]	*Both less than clinically significant difference (12). P1: 8 point increase; P2: 9 point increase*
		Scenario test (/54)	*P1: +1 point; P2: -3 points*
Combined Multimodal	(Rose et al., 2013)	*CETI (/100)* [a]	*Mean change 8.5 points (range -2 to 33). 3/11 participants > clinically significant difference*
		Scenario test (/54) [a]	*Mean change 0.80/54 (range -3.0 to 9.0)*
Gesture	(Raymer et al., 2012)	*CETI (pre/post gesture) (/100)* [a]	*Mean change 5.20 points (range -15 to 33). One participant > clinically significant difference, one participant negative change > clinically significant difference.*
		FOQ-A (pre/post gesture) (/5) [a]	*Mean change 0.44 (-0.19 to 1.21)*

Treatment	Authors	Outcomes	Results
Gesture	(Rose & Sussmilch, 2008)	*LCQ (/90)*[b]	*n/a*
Gesture	(Boo & Rose, 2011)	*LCQ (/90)*[b]	*n/a*
Music	(Hough, 2010)	*CETI (caregiver rating) (/100)* *ASHA FACS (/7)*	*28.2 increase (> clinically significant difference)* *2.05 increase*

Note. ANELT = Amsterdam-Nijmegen Everyday Language Test, CAL = Communicative Activity Log, CETI = Communication Effectiveness Index, LCQ = La Trobe Communication Questionnaire, FOQ-A = Functional Outcomes Questionnaire for Aphasia, ASHA FACS = American Speech-Language Hearing Association Functional Assessment of Communication Skills.

[a]The crossover design makes it difficult to isolate the contributions of pre/post outcome measures, but taking a conservative approach, where the treatment of interest was administered second, the post treatment measure of the first treatment was taken as the baseline.

[b]Pre/post measures represent changes from multiple treatments and so are not listed.

Table 4 – Quality of life outcomes of high-quality studies

Treatment	Authors	Outcomes	Results
Constraint	(Attard et al., 2013)	*SAQOL-39*[a]	*n/a*
Constraint	(Rose et al., 2013)	*SAQOL-39*[a]	*n/a*
Combined Multimodal	(Attard et al., 2013)	*SAQOL-39*[a]	*n/a*
Combined Multimodal	(Rose et al., 2013)	*SAQOL-39*[a]	*n/a*
Music	(Hough, 2010)	*ASHA-QCL (/90)*	*25 point increase*

Note. SAQOL-39 = Stroke and Aphasia Quality of Life Scale-39, ASHA-QCL = American Speech-Language Hearing Association Quality of Communication Life Scale.

[a]Pre/post measures represent changes from multiple treatments and so are not listed