A systematically conducted scoping review of the evidence and fidelity of treatments for verb deficits in aphasia: verb-in-isolation treatments.

Authors: Julie Hickin,[1] Madeline Cruice[1] and Lucy Dipper[1]

[1] Division of Language and Communication Science, City, University London, UK

Corresponding Author: Julie Hickin, Division of Language and Communication Science, City, University of London, Northampton Square, London EC1V 0HB. Email: jule.hickin.city.ac.uk. Telephone: +447496079850

Published in the American Journal of Speech Language Pathology, This article is part of the Special Issue: Select Papers From the 48th Clinical Aphasiology Conference.

Available at: https://pubs.asha.org/doi/10.1044/2019_AJSLP-CAC48-18-0234

Conflict of Interest: There are no relevant conflicts of interest.

Funding Statement: This research was supported by a doctoral studentship to the first author jointly funded by the Worshipful Company of Saddlers and the School of Health Sciences, City, University of London.

Purpose. Aphasia research demonstrates increasing interest in the treatment of verb retrieval deficits. This systematically conducted scoping review reports on the level and fidelity of the current evidence for verb treatments, on its effectiveness regarding the production of trained and untrained verbs, functional communication, sentences and discourse, and on the potential active ingredients. Recommendations to guide clinical decision-making and future research are made.

Method. The computerized database search included studies January 1980 to September 2018. The level of evidence of each study was documented, as was fidelity in terms of treatment delivery, enactment and receipt. Studies were also categorised according to the treatment methods used.

Results. Thirty seven studies were accepted into the review and all but one constituted a low level of evidence. Thirty three studies (89%) described treatment in sufficient detail to allow replication, dosage was poorly reported, and the fidelity of treatment was rarely assessed. The most commonly reported treatment techniques were phonological and semantic cueing in 25 (67.5%) and 20 (54%) of studies respectively. Retrieval of trained verbs improved for 80% of participants, and improvements generalised to untrained verbs for 15% of participants. There was not sufficient detail to evaluate the impact of treatment on sentence production, functional communication and discourse.

Conclusions. The evidence for verb treatments is predominantly of a low level. There are encouraging findings in terms of treatments being replicable, however this is tempered by poor monitoring of treatment fidelity. The quality of verb treatment research would be improved by researchers reaching consensus regarding outcome measures (including generalisation to e.g. sentences and discourse), by manualising treatment to facilitate implementation and exploring the opinions of participants. Finally, whilst treatment is largely effective in improving production of trained verbs, lack of generalisation to untrained items leads to the recommendation that personally relevant verbs are prioritised.

Introduction.

Aphasia research demonstrates increasing interest in the treatment of verb production (e.g. Webster & Whitworth, 2012). This likely reflects a greater awareness of the centrality of verbs in sentence processing (e.g. Edmonds, 2016) and thus of the potential for improved verb retrieval to impact real life discourse. This review aimed to review verb-in-isolation treatments. Studies that explore the treatment of verbs in the context of a sentence such as Verb Network Strengthening Treatment (VNeST: e.g. Edmonds, 2016; Edmonds, Obermeyer & Kernan, 2015) and mapping treatments (e.g. Marshall, Chiat & Pring, 1997; Rochon, Laird, Bose, & Scofield, 2005) are the subject of a future manuscript in preparation. The review reports on a) the level of evidence for verb-in-isolation treatments, b) the fidelity of the reviewed verb treatment studies in terms of treatment delivery, treatment enactment and treatment receipt as defined by Hinckley and Douglas (2013); c) the evidence for the effectiveness of treatment on the production of trained and untrained verbs, sentences, functional communication and discourse; and d) the potential active ingredients of verb treatments.

Existing reviews of verb treatments.

The existing reviews of verb treatments will be discussed briefly to demonstrate the additional contributions made by the study reported here. Conroy, Sage & Ralph (2006) reviewed the relationship between theory relating to verbs and verb treatments in aphasia. They present a very detailed and insightful overview of the relationship between theory and practice but do not consider the level of evidence for verb treatments, the fidelity of treatment or its overall effectiveness. All of these are addressed in this review.

Boyle (2017) conducted a review of semantic treatments for word and sentence production in aphasia i.e. her review was not exclusive to verb treatments as it included

studies investigating semantic treatments for nouns as well. The current review adds to the existing evidence because it is restricted to verb treatments only thus allowing a more detailed consideration of the impact of treatment on verbs per se. Additionally, whilst Boyle included studies that explored semantic treatments for nouns, she excluded studies which *combined* a semantic treatment with another treatment technique (such as phonological cueing). This review encompasses not only studies which explore *single* treatment techniques (e.g. semantic treatments) but also those which *combine* treatment techniques (e.g. semantic + gesture treatment), the latter of which constitute most verb treatment studies.

Webster and Whitworth (2012) conducted a review of verb treatments which did include studies that used techniques other than semantic, those which used a combination of techniques and those which treated verbs as single words in the context of a sentence. They concluded that there is insufficient evidence to establish what type of treatment is most effective, but they did note that treatment that targeted verbs *and their arguments* appeared to result in better generalisation to sentence production than treatment that targeted single verbs. They found no clear relationship between the treatment given to participants and the deficit/s underlying their verb/sentence level difficulties and thus no evidence to advocate particular treatments for specific deficits. They state that there is a need for a more systematic approach to evaluation of the outcomes of therapy for spoken verb deficits, including evaluation of sentence production before and after treatment and evaluation of the impact on connected speech/communication in real life. Our review aims to address this need and updates Webster and Whitworth's review which only included studies published up to March 2011. In an attempt to elucidate the active ingredients of treatment there is more detailed reporting of the techniques used in treatment and of the

impact of treatment (in that effect sizes are reported for individual participants whenever possible) in the current review than was reported in Webster and Whitworth. Finally, this review uniquely includes an evaluation of the level of evidence of verb treatment studies and their fidelity which has not been carried out before.

As well as the qualitative reviews discussed above, three systematic reviews of verb treatments have been conducted since 2013 (Efstratiadou, Papathanasiou, Holland, Archonti, & Hilari, 2018; Maddy, Capilouto & McComas, 2014; Rose, Raymer, Lanyon & Attard, 2013) and all of these contribute to our understanding of the effectiveness of verb treatments. However, none of these encompassed *all* researched treatments for verbs or included an evaluation of treatment fidelity. Maddy et al. (2014) and Efstratiadou et al. (2018) both conducted systematic reviews restricted to *Semantic Feature Analysis* (*SFA) treatment*, including studies that used SFA to treat nouns or discourse, as well as those which targeted verbs. The methodological quality of the reviewed studies was evaluated using the Single Case Experimental Design (SCED) Scale (Tate et al., 2008) and effect sizes were calculated (using Cohen's d and benchmarks specified by Beeson and Robey (2006)) to evaluate the effectiveness of SFA. Efstratiadou et al., and Maddy et al., rated reviewed studies highly on the SCED with an average of 9.55 and 9.3/11 respectively (range 8-11), but only small or negligible treatment effect sizes were found for the majority of participants. Rose and colleagues (2013) also conducted a systematic review this time restricted to treatments which used *gesture* either in isolation or in combination with verbal techniques. They included both group and single case studies in their review and found that whilst the quality of single case studies was high (as rated on the SCED scale), group studies were of low methodological quality (as rated on the Physiotherapy Evidence Database scale (PEDro-P scale, see Verhagen et al.,1998). They found positive effects of treatment on verb

production in over 50% of participants with indications that combined gesture and verbal treatment were more effective than verbal treatments alone. However they highlighted the need for further research in this neglected field in order to reach firmer conclusions about the effectiveness of verb treatments and hence to make clinical recommendations about the treatments of choice. The current review evaluates the evidence for *all verb-in-isolation* treatments with the aim of making such recommendations to guide clinical practice.

Finally, de Aguiar, Bastiaanse & Miceli (2016) conducted a meta-analysis of single case studies of verb treatments with the specific aim of identifying factors that predicted response to treatment including generalization of treatment to untrained verbs. Predictive factors were grouped into demographic (e.g. age, gender), clinical (e.g. severity of aphasia, size of lesion) and treatment related factors (e.g. amount, intensity, type of treatment), and Random Forests were used to assess the contribution of each factor (see de Aguiar et al., for a detailed discussion). Whilst the review provides a very valuable insight into the factors which influence response to verb treatments, it excludes case series studies (which comprise the majority of verb treatment studies). In addition, de Aguiar et al., do not evaluate the fidelity of the studies included in the meta-analysis (not an aim for their study).

In summary, the review reported here included all verb treatments that treated verbs in isolation that is, semantic treatments including but not restricted to SFA, and studies that used phonological cues, gestural cues, orthographic cues, video cues or a combination thereof. The review included studies which used group, case series or single case designs and studies which delivered verb-in-isolation treatments via computer, as these have not been included in any review of verb treatments to date. The degree to which verb treatment studies have attempted to evaluate the impact of treatment beyond the naming of treated and untreated verbs is systematically reviewed as recommended by Webster and

Whitworth (2012). Finally this review is unique in considering the level of evidence and the fidelity of verb treatment studies.

Verb retrieval deficits in aphasia: theory and therapy.

It is beyond the scope of this paper to present a detailed review of the relationship between theories of verb retrieval deficits in aphasia and their treatment (see Conroy et al., 2006 for such a review). However, this will be discussed briefly to highlight the contributions made by current theories to verb-in-isolation treatments and the challenges that remain.

Verbs have been the subject of much discussion in the aphasia research literature for a number of years. For example, Berndt, Haediges, Mitchum & Sandson (1997) investigated the ability of eleven participants with aphasia to produce words of different grammatical classes and found that whilst verbs were harder to retrieve than nouns for five participants, there was no straight forward relationship between verb retrieval and aphasia type. This finding contradicted the generally held *belief* that a double dissociation exists between noun and verb impairments in aphasia: i.e. that people with nonfluent (and particularly agrammatic) aphasia show greater impairment of verbs than nouns, and the reverse is true for fluent aphasia. In contrast, a review of the *evidence* suggests that the pattern of noun and verb impairment in people with aphasia is more complicated than this. For example, Luzzatti, Raggi, Zonca, Pistarini, Contardi & Pinna (2001) compared verb and noun retrieval in 58 participants with aphasia and found that whilst people with non-fluent aphasia showed a strong tendency to have poorer verb than noun retrieval, this was not always the case, with some people with nonfluent aphasia not exhibiting a difference and some showing the reverse pattern. In fluent aphasia the pattern is more variable with the naming of verbs and of nouns almost equally liable to selective impairment.The theoretical standpoint taken regarding the factors that underlie the differing patterns of impairment of

nouns and verbs in aphasia is split in terms of whether this reflects that nouns and verbs are stored and/or processed *separately* or, that the differing patterns of impairment are an inevitable consequence of their differing psycholinguistic properties.

The psycholinguistic properties of verbs have given rise to differences in verb treatment design, in the studies reviewed here, and so merit further discussion. Firstly, verbs are thought to be both less richly represented *semantically* than nouns (i.e. they are more abstract) and to have "looser" connections with their semantic networks than nouns. This is because whilst nouns represent relatively unchanging physical entities, verbs represent actions and therefore have a temporal component as well as attributes which can vary according to context (e.g. consider "drive" in the following sentences: "The man drives a car," "The sheep dog drove the sheep into the pen" and "The crying child drove his mother to distraction"). Thus their relationship with their semantic networks is more fluid than that of nouns. (See Black and Chiat (2003) for a detailed review.)

Verbs can also be categorized according to their *syntactic* properties, in terms of their transitivity. The fact that verbs encode syntactic information means that they have a pivotal role in sentence production. Garrett (1988) and Levelt (1989) proposed that information contained within the semantic representation of verbs was essential to sentence production because this encoded *syntactically relevant* information about predicate argument structure as well as core semantic information pertaining to a verb's meaning. Thus being unable to retrieve the semantic representations of "buy" and "sell" for example, would entail not only difficulty retrieving information about the verbs' core meanings (that they are change of possession verbs) but also difficulty retrieving information about their *argument structures,* and how to map these arguments onto the

syntax of the sentence according to their *thematic roles.* For "buy/sell", the lexicon contains information about argument structure and thematic roles such that both have two obligatory arguments and an optional one, and both have arguments carrying the thematic roles of theme, source and goal, but they differ as to the role that the optional argument takes (optional source for "buy" and optional goal for "sell"). The rules for mapping this information onto syntax are similarly distinct: "goal" is mapped onto the subject of the sentence for "buy" but onto a prepositional phrase for "sell". Several studies have indeed attested to the destructive effect of impaired semantic representations of verbs on sentence production (e.g. Jones 1986; Marshall, Chiat & Pring, 1997). That verbs encode additional information (in comparison to nouns) which is essential for sentence production makes them potentially very fruitful targets for treatment in that improving the retrieval of a verb should have a greater impact on improving sentence production than improving the retrieval of a noun (or indeed other word classes such as adjectives). In other words, verb retrieval treatments have greater potential to *generalize* beyond improving single word production to sentence production (across level generalization).

Whilst verbs store information at the lexical level which is essential to sentence production, perhaps surprisingly this can also affect their production *in isolation*. For example, Thompson, Lange, Schneider & Shapiro (1997) found that the complexity of a verb's argument structure influenced verb production in both a single word and sentence context for ten participants with agrammatic aphasia. This finding was replicated by Kim and Thompson (2000) whose seven agrammatic participants demonstrated a hierarchy of difficulty between one, two and three place verbs in a single word context for both production (naming) and comprehension (categorization) tasks.

Verbs are also more complex *morphologically* (e.g. being marked for person and time – for instance, look/looks/looked). There are also *phonological* features that could make English verbs vulnerable to impairment: verbs tend to be shorter in duration than nouns, have fewer syllables and tend to carry their stress on the second rather than the first syllable making them less salient. And finally, there are other psycholinguistic properties that distinguish verbs from nouns, including relative imageability, age-of-acquisition and frequency. The combination of these differences between nouns and verbs may account for their differential impairment in aphasia.

The differences between nouns and verbs may mean that verbs are harder to process both cognitively and linguistically than nouns (a theory borne out by verbs being harder to acquire developmentally) and this may also account for their relative neglect in the treatment literature because it may have fostered the belief that they may be both *harder to treat* and *less responsive* to treatment than nouns. (e.g. Conroy, Sage and Lambon-Ralph, 2009b and c). However despite this, the treatment of verb retrieval deficits has attracted increased attention in the aphasia treatment literature in recent years as demonstrated by a notable increase in the number of studies published (for example, all but three of the thirty-seven studies included in the current review were published after the millennium). As well as trying to establish the effectiveness of verb retrieval treatments, studies have tried to elucidate how verb treatments work – the active ingredients and mechanisms of treatment – in order to develop a "theory of (verb) therapy" (Byng, Nickels & Black, 1994). Regarding so-called active ingredients of treatment these have been defined as:

> "a behavior-influence procedure shown through experimental analysis to affect a specific behavior and that is indivisible in the sense that removing any of its components would render it inert" (Embry & Biglan, 2008, p. 75).

Whilst active ingredients may be easy to identify in pharmacological treatments (e.g. a particular antibiotic or analgesic), this is often not the case for behavioral treatments such as those applied in the treatment of aphasia. Indeed, most aphasia treatments are likely to be complex interventions as defined by the Medical Research Council in that they "contain several interacting components" (www.mrc.ac.uk/complexinterventionsguidance, p.7). Thus identifying the active ingredients of aphasia treatment is likely to be a difficult process but this has not stopped researchers attempting it. So for example, studies have attempted to adapt SFA treatment (applied initially to nouns) for verbs by adapting the features generated during treatment to be more appropriate (or active) for verbs (e.g. Wambaugh Mauszycki & Wright, 2014; Wambaugh & Ferguson, 2007). In SFA treatment, the likely active ingredient of treatment is the *generation of semantic features* for a target word (see Gravier et al., 2018). Thus SFA for nouns includes generating the semantic feature of category whereas in SFA adapted for verbs, features unique to verbs (and therefore potentially active ingredients of treatment for them) are generated (e.g. the instrument of a verb). Because it is hypothesized that SFA strengthens the semantic network of treated verbs, it is predicted that treatment should generalize to semantically related verbs (within level generalisation) but not to semantically unrelated verbs (in line with the prediction for SFA targeting nouns). However, to date most of the participants in SFA studies have not demonstrated such generalization.

Other verb treatment studies have used phonological and orthographic cues cue to try to improve verb retrieval (e.g. Conroy, Sage & Lambon-Ralph, 2009a, b & c), once again borrowing from the evidence base for anomia treatment. Because it is hypothesized that these treatments will work on a lexical basis (i.e. that they will strengthen the link between semantics and the phonological form of a verb), it is predicted that treatment will not

therefore generalize to untrained verbs, and indeed Conroy et al. found word retrieval improvements almost entirely restricted to treated verbs only in their series of studies (Conroy et al., 2009a, b and c).

Finally, some verb treatment studies have investigated techniques which are designed to target unique qualities of (action) verbs i.e. that the lexical representations of an action verb in the brain may be intimately linked with the representation of the sensory motor features which encode its actions. This leads to the prediction that gesture will be an effective treatment for spoken production of verbs and this has been investigated in a number of studies (see Rose et al. (2013) for a review). Very recently, studies have also investigated whether the observation of an action alone (or in combination with gesture) can facilitate the spoken production of the related verb based on the belief that this will also activate the sensory motor representation of a verb and hence facilitate retrieval of its spoken form. Initial results have proved to be promising, including for action observation alone (e.g. Marangolo et al., 2012). Treatments are also starting to exploit emerging research which suggests that using a dynamic depiction of a verb (i.e. a video) rather than a static (picture) representation of a verb as a stimulus may be more effective in evoking action verb production (e.g. Blankestijn-Wilmsen et al., 2017).

In summary, whilst theory relating to verb deficits in aphasia is beginning to explicitly inform the development of verb treatments there remain many gaps in our understanding of how verb treatments work, how to facilitate generalization of treatment effects beyond the production of trained verbs in isolation and what the treatment of choice should be for any one individual with a verb deficit. Indeed as Conroy et al. (2006) concluded, there is disappointingly little evidence that our understanding of how verbs and nouns differ from each other has informed treatment. Thus for example, theory predicts that verbs are more

difficult to process than nouns (because they are less imageable and less phonologically salient, but more complex syntactically) and that they are thus likely require to more processing resources during communication and *during treatment*. However, this has not been taken into account in planning or delivering treatments (e.g. by using errorless learning techniques such as those investigated in a series of studies of the treatment of anomia (e.g. Fillingham, Sage & Lambon-Ralph, 2005).

Fidelity of verb treatment studies.

The fidelity of verb treatment studies has not been reported on to date. With regard to the fidelity of aphasia treatment more generally, Hinckley and Douglas (2013) and Kaderavek and Justice (2010) found that this had been poorly monitored. Hinckley and Douglas reviewed aphasia treatment studies published in the previous ten years and found that only 14% of 149 studies reviewed assessed treatment fidelity (usually by rating a sample of videotaped treatment sessions). They recommended that three levels of treatment fidelity needed to be addressed to improve the quality of aphasia therapy research namely, *treatment delivery* (e.g. by the use of treatment manuals and training), *treatment receipt* (e.g. by the use of homework record sheets and establishing the views of recipients regarding their treatment), and *treatment enactment* (e.g. by observation of treatment delivery). Kaderavek and Justice (2010) argued that the neglect of treatment fidelity has undermined the implementation of evidence-based practice because, for example, treatments are not described in enough detail to enable replication in clinical practice. The fidelity of the studies included in this review is reported according to the three levels recommended by Hinckley and Douglas (2013).

Review Methodology.

A scoping review paradigm was used for this study. Scoping reviews are used to categorise existing literature in terms of its nature, features, and volume; and are best used when a body of literature exhibits a large, complex, or heterogeneous nature not amenable to a more precise systematic review (Dijkers, 2015; Peters, Godfrey, Khalil, McInerney, Parker, & Soares, 2015). Verb treatments are complex in that the method and content varies considerably, as do the outcome measures and methods used to analyse the effectiveness of treatment. The review followed the five steps recommended by Dijkers (ibid). These were firstly to identify the research questions, and secondly to search for relevant papers via established methods (a database search). The third step was to select papers pertinent to the research questions and the fourth to chart the collected data. The final, fifth step was to collate, summarize and report the results of the scoping review. The review was conducted using systematic procedures to ensure that these were rigorous, explicit and replicable.

The research questions that the review set out to answer were:

1. What is the level of evidence for verb-in-isolation treatments?
2. What is the fidelity of the research in terms of treatment delivery, receipt, and enactment (Hinckley & Douglas, 2013)?
3. What is the evidence of positive gains for treatment in terms of improved production of a) trained verbs in isolation and in sentences and b) untrained verbs in isolation and in sentences (within and across level generalization)?
4. What is the evidence of positive gains for verb production treatments in terms of a) improved functional communication and b) improved production of discourse?
5. What are the potential active ingredients of verb treatments?

The evaluation and charting of the reviewed studies in terms of a) the level of evidence of a study and b) treatment fidelity was guided by established hierarchies of evidence for healthcare research (e.g. the highly influential Cochrane hierarchy (http://consumers.cochrane.org/levels-evidence)), and by the small treatment fidelity literature pertaining to aphasia treatment research respectively. With regard to level of evidence, hierarchies of evidence used to guide evidence-based healthcare generally place Systematic Reviews (SRs) of Randomised Controlled Trials (RCTs) at the top of the hierarchy, with single case studies placed at the bottom. Intermediate levels of evidence are placed in the following, descending, order: RCTs, pseudoRCTs, nonrandomized group studies which include a control group alongside the treated group, followed by nonrandomized group studies where treatment effects are determined by comparison with previous performance/historical data (e.g. case series).

Method.

Cinahl Complete and Medline Complete databases were searched using the following terms:

Verb-in-isolation treatment Searches:

Face-to-face delivery of verb treatments:

verb and *aphasia* and *treatment* or *therapy*

Computer delivery of verb treatments:

verb and *aphasia* and *technology* or *computer*

Sentence treatment Searches:

Face-to-face delivery of sentence treatments:

sentence or *sentence production* and *aphasia* and *treatment* or *therapy*

Computer delivery of sentence treatments:

sentence and *aphasia* and *technology* or *computer*

Studies which were original research and which were published in peer-reviewed journals, in English from 1980 up to September 2018 were considered for inclusion. The

titles of all papers identified in the database search were read. If the purpose of the study was not clear from the title, the abstract was read to determine whether the study should be included. All papers accepted into the review were then read in full and the contents charted including the aim of the study; type of design; the number of participants; the type, amount, intensity and duration of treatment; the results of treatment; and any limitations of the study. The reference lists of all papers were scrutinized to identify any additional relevant studies. Where possible higher-level evidence was included, but the literature identified almost exclusively represented lower levels of evidence, that is single cases and case series designs.

As there is no pre-established protocol for evaluating fidelity, a tailored approach was developed to evaluate the studies in this review. This was based on the elements of aphasia treatment needed for a 'theory of therapy' (e.g. Byng et al., 1994), to determine the optimal dose of treatment (e.g. Baker, 2012a; 2012b), and the active ingredients of treatment or 'kernels of knowledge' (Embry & Biglan, 2008). The following data were extracted from the reviewed studies, and used to judge fidelity: 1) the amount of treatment given (at least the number of sessions); 2) the number of verbs treated; 3) the type of stimulus (e.g. picture vs video); 4) the treatment hierarchy; 5) the modality of cues; and 6) the contingency under which cues were given and how many times they were given. The impact of treatments on trained and untrained verb production in isolation and in sentences was charted, as was the impact of treatment on functional communication and discourse. The results of treatment for individual participants are described whenever these are reported. When a study reported results for the group only, this is reported separately i.e. the participants in these studies were not added to the totals for individual participants

because it was not clear how many participants in the group had (or had not) benefitted from treatment.

Results.

The results of the four seaeches carried out for the systematically conducted scoping review are summarized in Figures X - X1, and are reported according to PRISMA guidelines (Moher, Liberati, Tetzlaff & Altman, 2009). The combined searches resulted in 331 papers, with another nine papers identified from additional sources. Once duplications were removed, the titles and abstracts of 178 papers were screened. As a result of this screening, 141 papers were excluded (see Figure 1). Studies that were excluded investigated other types of aphasia treatment (e.g. anomia treatment, dysgraphia treatment, conversation training) or investigated spoken verb production treatment but were:

a) published in a language other than English,

b) had participants with another form of aphasia (e.g. progressive aphasia),

c) were without at least one pre and post treatment measure of verb retrieval (e.g. when the main aim of the study was to investigate the neurophysiological response to treatment),

d) used techniques which focused on verbs in sentences (i.e. verbs with their arguments) rather than verb-in-isolation treatments,

e) targeted correct production of verb morphology rather than correct production of the verb itself,

f) studies whose primary focus was assessment of aphasia, and

g) studies which were reviews or meta-analyses rather than original research.

Following the exclusion of 141 papers the full text of 37 papers was screened and all papers were included in the subsequent review. The details of the included papers are summarized in Table 1. Thirty two of the reviewed papers reported verbs treatment delivered face-to-face, whilst five reported treatment delivered via computer. Results are reported according to research questions.

Insert Figure 1 about here

Insert Table 1 about here

1. What is the level of evidence for verb production treatments?

The majority of studies comprised lower levels of evidence because they were case series or single case studies. Specifically, verb treatment studies comprised 22 case series, seven case series which also reported group results, six single-case studies, one group study (Marangolo et al., 2012) and one pilot RCT (Palmer et al., 2012). The maximum number of participants in a study was 15 (Carragher et al., 2013; Conroy et al., 2009c; Palmer et al., 2012; Raymer et al., 2006), and the total number of participants in the reviewed studies was 182 (see Table 1, column 2).

2. What is the fidelity of the research in terms of treatment delivery, receipt, and enactment?

In terms of *treatment delivery,* none of the studies reported the existence of a published manual for the treatment investigated[1]. However, most studies (33 or 89%) reported the treatment *procedure* in sufficient detail to enable replication (excluding Fink, Schwartz, Sobel & Myers, 1997; McNeil et al., 1998; Palmer et al., 2012; Raymer et al., 2005). In particular, some studies gave a detailed description of the treatment protocol in an appendix (e.g. Boo & Rose, 2011; Wambaugh, Doyle, Martinez, & Kalinyak-Fliszar, 2002).

Regarding *dose,* only eight studies (22%) reported the exact amount of treatment given (see Table 1, column 4): 30 hours over 4 months (Adrian, Gonzalez, Buiza & Sage, 2011), 20 hours over 5-6 weeks (Kristensson, Behrns & Saldert, 2015), 24 hours over 14 weeks (Marshall, Pring, & Chiat, 1998), 25 hours over 5 months (Palmer et al., 2012), 10 hours 2 - 4 times a week (Raymer et al., 2006), and 7 hours twice weekly (Rochon & Reichman, 2003). Kurland et al. (2014) and Mortley et al. (2004) were able to report the amount of treatment given in their studies in precise detail because this was monitored by the computer programs used in their self-delivered treatments (26-67 total hours in Kurland et al.'s study and 46–93 hours in Mortley et al.). Edwards and Tucker (2006) state the exact amount for two of three participants (17 and 19 hours). Carragher, Sage and Conroy (2013) gave their participants 8 hours of treatment plus an unspecified amount of homework. Three other studies reported an approximate amount: between 13.5 hours and 22 hours (Knoph, Lind & Simonsen, 2015; Rose & Sussmilch, 2008; Routhier, Bier & Macoir, 2015). Six studies reported the number of treatment sessions given but not their length: 6 (Fink et al., 1997); 10 each (Conroy & Scowcroft, 2012; Raymer & Kohen, 2006; Raymer et al., 2005 and

[1] However the StepByStep® therapy software programme used in the studies by Mortley et al. (2004) and Palmer et al. (2012) is commercially available and does have an instruction manual.

2007); and 12 (Schneider & Thompson, 2003). Ten studies reported the minimum and maximum length of sessions as well as the number of sessions which allowed the estimation of the minimum and maximum total amount of treatment (Bonifazi et al.,; 2013; Boo & Rose, 2011; Conroy, Sage & Lambon-Ralph, 2009 a, b and c; Marangolo et al., 2010; McCann & Doleman, 2011; McNeil et al., 1998; Rodriguez, Raymer & Gonzalez-Rothi, 2006; Routhier, Bier & Macoir, 2016). Wambaugh and colleagues reported a series of studies where the amount of treatment was dependent on reaching a predetermined performance criterion, meaning amount of treatment given differed (Wambaugh et al., 2014; Wambaugh & Ferguson, 2007; Wambaugh, Cameron, Kalinyak-Fliszar, Nessler & Wright, 2004; Wambaugh, Doyle, Martinez, & Kalinyak-Fliszar, 2002). Takizawa, Nishida, Ikemoto, & Kurauchi (2015) reported huge variation in the duration and intensity of treatment given in their clinical study: one to five 40-minute sessions per week over a span of 2–8 months. Three studies did not report the amount of treatment given (Faroqi-Shah & Graham, 2011; Marangolo et al., 2012; Raymer & Ellsworth, 2002).

Moving on to *treatment receipt*, the only studies that reported the *views* of participants about their treatment were those investigating self-delivered treatments via the computer program StepByStep. Thus Mortley and colleagues report the views of the participants in their 2004 study in a companion paper (Wade, Mortley and Enderby, 2003), and Palmer and colleagues (2012) report their participants' views in Palmer, Enderby and Paterson (2013). Both studies used structured interviews and thematic framework analysis to investigate the views of the participants with aphasia and their carers, with the focus being on the *acceptability* of self-delivered computer treatment and any perceived *advantages* and *disadvantages* of treatment delivered in this way rather than on the verb

treatment itself. Thus there is a significant gap in the evidence base regarding verb treatment receipt. Two studies attempted to monitor the amount of *home practice* carried out in addition to face-to-face treatment: Carragher and colleagues (2013) reported a range in homework hours completed during their study from 0.5 to 23.75 hours, and Conroy and Scowcroft (2012) mentioned home practice, but did not report it because it was not recorded accurately by their participants.

In terms of *treatment enactment*, only studies delivering face-to-face treatments (n=32) were considered. (Treatment fidelity of self-delivered computer treatments is an emerging area of research: for example Ball, de Riesthal & Steele (2018) investigated the degree to which participants complied with recommended treatment procedures during self-administered (computer) anomia treatment, and whether adherence influenced accuracy of performance.) In terms of face-to-face treatment, five studies (16%) reported that the fidelity of treatment enactment was assessed (by rating how closely the treatment protocol was followed using a sample of videotaped treatment sessions). These studies were: Carragher et al., (2013), Faroqi-Shah and Graham (2011), Rose & Sussmilch (2008), Wambaugh & Ferguson (2007) and Schneider & Thompson (2003). Whilst technically not treatment enactment, it was encouraging to see that twelve studies (37.5%) addressed *assessment* fidelity, most commonly by establishing inter-rater reliability using another assessor who was sometimes independent of the study and/or blind to the pre/post-treatment status of the assessment.

3. What is the evidence of positive gains for treatment in terms of improved production of a) trained verbs in isolation and in sentences and b) untrained verbs in isolation and in sentences (within and across level generalization)?

The evidence for treatment effectiveness on the production of i) treated and ii) untreated verbs, and on sentences using iii) treated and iv) untreated verbs is summarized in Table 2 Columns 4 - 7. Significance levels and effect sizes are given for each individual participant when available.

Insert Table 2 about here.

Encouragingly all but one of the 37 reviewed studies reported significant improvement for *at least one individual* or for the group of participants (Table 2, column 4). (The remaining study Kristenssen et al., (2015) did not assess treated verbs). This represents improvement in trained verb retrieval for 104 of the 130 (80%) participants for whom individual results are reported. Thirty one studies (84%) reported improvement in treated verbs via inferential statistics (reporting either probability levels ($p<.05$) and/or effect sizes), and four (11%) reported improvements according to pre-established performance criteria, with three also using visual inspection of graphs of performance before, during and after treatment. In terms of effect sizes for trained verbs, when reported these were predominantly small in size (reported on 17 occasions across all studies – see Table 2 column 4), with medium effect sizes reported on 11 occasions and large effect sizes on 9. Whilst there is debate about the best way to analyse improvement in single case studies and case series (see e.g. Howard, Best & Nickels, 2015), the finding that 84% of studies used

statistical analysis to assess the effectiveness of treatment adds to the rigour of the evidence for verb treatments.

In terms of (ii) untrained verbs (Table 2, column 5), 31 studies investigated this, with 11 (30%) of these studies reporting significant improvement in untrained verbs following verb treatment. This represented significant improvement for 18 out of 119 participants (15%) for whom individual results are reported.

Regarding the impact of verb treatment on *sentence production* involving either (iii) trained verbs or (iv) untrained verbs, interpretation of the evidence is hampered by inconsistent assessment (Table 2, columns 6 and 7). For trained verbs, only five studies (13%) assessed sentence production. Four of these studies (with a total of five particiants) report *individual* results, finding significant improvement for two of the five participants (Marshall et al., 1998; Raymer & Ellsworth, 2002), and no significant improvement for the remaining three participants (Raymer & Kohen, 2006; Rochon & Reichman, 2003). Schneider & Thompson (2003) report significant improvement in sentence production using trained verbs for the *group* (n=7). For untrained verbs, nine studies (24%) assessed sentence production: five studies reported significant improvement for six of the 23 participants (26%) for whom individual results are reported; Schneider & Thompson (2003) again report significant improvement for their group study (n=7); and three studies found no significant improvement.

4. What is the evidence of positive gains for verb production treatments in terms of a) improved functional communication and b) improved production of discourse?

The evidence concerning the impact of verb treatment on functional communication and discourse is also limited by lack of assessment (see Table 2, columns 8 and 9). Only four studies (11%) investigated the impact of verb treatment on functional communication. Three of these reported effectiveness as measured by the LaTrobe Communication Questionnaire (LCQ: Douglas, O'Flaherty & Snow, 2000) or the Communicative Effectiveness Index (CETI: Lomas et al., 1989). These studies were Boo and Rose (2011), Raymer et al., (2007), and Rose and Sussmilch (2008). This represented self-reported improvement for four of the five participants for whom individual results are reported, with spouses also reporting improvements for two of these participants. Raymer et al. report improvements in CETI ratings for their group as a whole (n=8) as rated by participants' carers. Kristenssen et al. (2015) found no effect of SFA treatment on functional communication for their three participants, as measured by the Communication Outcome after Stroke scale (COAST: Long, Hesketh, Paszek, Booth, & Bowen, 2008) given to both participants and carers.

Twelve studies (32%) investigated the impact of treatment on discourse. The outcome measures used to assess discourse were varied ranging from complex picture description to analysis of narrative production and conversation (see Table 2, column 9). Seven studies reported a significant effect of treatment. This represented improvement for 15 of the 48 (31%) participants for whom individual results were reported.

5. What are the potential active ingredients of verb treatments?

The treatment techniques (i.e. potential active ingredients) reported in the reviewed studies are summarized in column 6 of Table 1. The most commonly reported technique was the use of *phonological cues* (including repetition, initial phoneme and rhyme cues)

which was reported in 25 (67.5%) of studies.[2] The next most common technique reported was *semantic cueing* reported in 20 studies (54%) including those using SFA. The use of *orthographic cues* was reported in 13 (35%) of studies as were *comprehension tasks*. *Gesture cues* were reported in 10 (31%) of studies and *sentence closure* in 8 (22%). The use of *video (verb) stimuli* was reported in 5 (13.5%) of studies. The least commonly reported techniques (reported in 3 (8%) of studies respectively) were 1) *construction of a sentence* using a treated verb, 2) *action observation* and 3) video cues (of the articulation) of a target verb.

The frequency with which a treatment technique is reported does not *per se* indicate its potency as an active ingredient, that is, we cannot assume that the most frequently reported techniques are necessarily the most effective. Indeed, approximately half of the reviewed studies (17 or 46%) attempted to establish the active status of treatment techniques by using them in isolation and *comparing* their effectiveness (*comparison* studies) whilst 20 studies (54%) assessed the effectiveness of one treatment *(treatment evaluation* studies).

Given the lack of certainty regarding the active ingredients of verb treatment, it is perhaps unsurprising that most of the studies in the review investigated treatment which used a *combination of cues*: 21 of the 37 studies (57%) investigated treatment which involved a combination of cues (these are highlighted by italics in Table 2). For example, Edwards and Tucker (2006) and McCann & Doleman (2011) used sentence completion, naming to definition, semantic and progressive phonemic cues in their clinical studies.

[2] When a study used more than one type of cue (i.e. a combination treatment) it is counted as a study for each type of cue used (e.g. Raymer et al. (2007) would be counted as a study using semantic cues and as a study using phonological cues).

Carragher et al., (2013) used repetition, SFA and gesture cues, whilst Marshall et al., (1998) used comprehension tasks paired with reading aloud followed by a verb generation task in their single case study. All of these studies reported significant benefits to the participants. It is interesting to note that of the eighteen participants for whom generalization of treatment to untrained verb production is reported, sixteen of these received a combination treatment. (Wambaugh et al., (2002) report partial generalization for one of their 3 participants after (single) semantic cueing treatment and Marangolo et al. (2012) report generalization for one of their seven participants following action observation treatment).

All five of the studies evaluating verb treatments *delivered by computer* also used a combination of cues with the finding that 23 of the 28 participants (82%) (for whom individual results were reported) demonstrating improved retrieval of treated verbs (Adrian et al., 2011; Kurland et al., 2014; Mortley et al., 2004; Routhier et al., 2016). Additionally, significant improvement was reported for the group as a whole (n=15) in the pilot RCT conducted by Palmer and colleagues (Palmer et al., 2012). These findings are particularly impressive given that in all but one of these five studies (Adrian et al., 2011), treatment was self-administered.

Comparison studies of treatments utilising a single type of cue compared the use of: semantic cues with phonological cues (e.g. Raymer & Ellsworth, 2002; Wambaugh et al., 2004; Wambaugh et al., 2002;), semantic, phonological and gestural cues (e.g. Boo & Rose, 2011; Rodriguez et al., 2006); a word versus a sentence cue (Conroy et al., 2009a; Raymer & Kohen, 2006) and increasing versus decreasing cues (Conroy et al., 2009b and c). An emerging area of research is the comparison of action observation alone versus observation accompanied by execution of the action as a treatment technique, with initial findings being promising (Bonifazi et al., 2013; Marangolo et al., 2012; Marangolo et al., 2010).

None of these studies was able to reach a clear conclusion as to whether any type of cue is more effective than another. However, there are indications that participants whose verb retrieval deficit is semantic (rather than phonological) in nature may be less responsive to treatment (Bonifiazi et al., 2013; Marangolo et al., 2010; Rodriguez et al., 2006; Wambaugh et al., 2004), and participants with a more severe deficit may also be less responsive (Conroy et al., 2009c; Palmer et al., 2012).

In summary, the review found that the evidence for verb treatments is currently at a low level. In terms of the fidelity of verb treatment studies, whilst the vast majority of studies reported treatment in sufficient detail to enable replication, the fidelity of studies was poor. In particular, the dose of treatment was not accurately reported and treatment receipt was almost entirely neglected in terms of the views of participants about their treatment. Regarding the effectiveness of verb treatments, they resulted in improvements in the production of treated verbs in isolation for 80% of participants. However, generalization of treatment effects to untrained verbs occurred for only 15% of participants. The impact of verb treatments on sentence production, functional communication and discourse could not be evaluated because these were not consistently assessed in the reviewed studies. Regarding the active ingredients of verb treatments, it was not possible to identify these, although the review highlighted potentially active ingredients which merit further investigation such as action observation.

Discussion.

The existing body of evidence for verbs in isolation treatments is almost entirely of a low level, with the exception of one pilot RCT and one group study. This represents a challenge to researchers to conduct studies using designs which constitute higher levels of

evidence such as well-designed, larger scale RCTs (see, for example, Leff & Howard, 2012). Before progressing with larger scale studies however, researchers need to be cognizant of the current phase of verb treatment research as defined by Robey and Schultz (1998). In terms of Robey and Schultz's model, verb treatment research is very predominantly at Phase I and II in that it is still seeking to establish therapeutic effects, refine treatment protocols and establish optimal dosages for example. Researchers must therefore be careful not to "put the cart before the horse" and ensure that they are trialing treatments where the effective ingredients have been clearly established (see discussion below in relation to research question 5).

Regarding the *fidelity of verb treatment* research, there are also challenges for researchers particularly in terms of the reporting of treatment dose, the gathering of data on treatment receipt and the monitoring of treatment enactment. Taking *treatment delivery* first, there were encouraging findings in that treatment protocols were almost always described in sufficient detail to enable replication. However, because this detail was most commonly contained within the body of the paper (e.g. within the Method section) rather than in an appendix or supplement, it was not easy to extract. This may be an issue for busy clinicians who might wish to implement a treatment in their practice and for researchers conducting replication studies. It is therefore recommended that treatment protocols are given in the appendices of papers (e.g. Boo & Rose, 20111; Wambaugh et al., 2002) and that this is clearly flagged to facilitate easy extraction of treatment protocols and thus potentially their implementation into clinical practice. This would also represent a step towards the development of treatment manuals which state explicitly how treatment should be delivered and which can therefore enhance the fidelity of treatment delivery.

Reporting the *dose* of treatment is vital to ensure that treatment is delivered to the recommended amount and yet only eight studies reported the exact treatment schedule. To address this, it is recommended that the minimum detail studies should report is dose, dose form, dose frequency, session duration, and total intervention duration (as recommended by Baker, 2012a & 2012b). Treatments delivered by computer have the potential to report these data in detail but only two of the five studies which reported treatment delivered by computer had the technological capacity to do this (Mortley et al., 2004; Palmer et al., 2012). Reporting of this detail would facilitate accurate replication of dose and also enable progress towards identifying the optimal dose of treatment. Dosages reported in the reviewed studies varied hugely (from a total of 7 to nearly 93 hours of treatment, delivered over an estimated duration of a minimum of 2 weeks to a maximum of 8 months). Studies delivering a relatively small amount of treatment non intensively (e.g. Carragher et al., 2013: 8 hours over 8 weeks) reported significant benefit to participants as well as those delivering large amounts of treatment (e.g. Mortley et al., 2004: c46 hours - c93 hours over 6 months). Mortley et al.'s treatment was self-delivered by computer and treatments delivered in this way clearly have the potential to increase the dosage delivered significantly. However the amount of treatment received in the four studies using this delivery mode still varied hugely between participants (from 20 hours delivered over five weeks in Routhier et al. (2016), 25 and 26 hours delivered over 5 and 6 months in Kurland et al. (2014) and Palmer et al. (2012) respectively, to c46 – c93 hours in Mortley et al. (2004) as described above). Thus it is currently not possible to evaluate the potential efficacy of intensively delivered verb treatments or indeed to identify the optimal dose of treatment. This has implications not only in terms of the efficacy and efficiency with which treatment is delivered but also for the compliance of participants. Brady et al., (2016) conducted a Cochrane review of aphasia

treatment and found that whilst treatment delivered to a high dose at a high intensity could be beneficial, this was confounded by a higher drop-out rate of participants in comparison to treatment of a lower dose and intensity. Determining the optimal dose for verb treatments is vital at many levels.

Turning now to *treatment receipt*, only two studies of verb treatment investigated the views of the participants on their treatment (Wade et al., 2003; Palmer et al., 2013) and this largely focused on the participants' views of the *mode* of treatment delivery (self-delivered via computer) rather than the treatment itself. This is a very significant gap in the evidence base but the pioneering studies by Wade et al. and Palmer et al. represent a potential way forward here in their use of structured interviews and thematic framework analysis to investigate the views of both people with aphasia and their carers on treatment.

Finally in terms of the fidelity of treatment*, treatment enactment* was monitored in only 16% of studies, and the way in which this was done varied considerably. For example, Carragher et al., (2103) state that they "regularly" reviewed an unspecified number of videoed taped sessions (p. 858), whilst adherence to the treatment protocol was rated for approximately 50% of sessions in Faroqi-Shah and Graham's (2011) study versus 17% for Wambaugh and Ferguson (2007) albeit with approximately 100% adherence found in both studies. As well as routinely monitoring treatment enactment, it is therefore recommended that a standard protocol to do this be developed. Apropos of this, Kaderavek & Justice (2010) regard direct observation as the gold standard for assessing the fidelity of treatment enactment and give an example of a Fidelity Coding Catalog for use during direct observation (Kaderavek & Justice, 2010, Appendix A, p.377), as well as making a number of

recommendations to assess and enhance treatment fidelity in clinical practice (ibid, Table 1, p.375).

Turning now to treatment efficacy, this review found that the production of treated verbs in isolation improved for 80% of participants (for whom individual results were reported). This encouraging finding is slightly in excess of that of the meta-analysis carried out by de Aguiar et al. (2016) who found that treated verbs improved for 76.1% of participants. This may partly be accounted for by the inclusion of studies of self-delivered treatment via computer (n=4). These studies reported improvement in trained verb production for 100% of participants, which bodes well for the development of more treatments delivered in this way. Optimism must however be tempered by the as yet small number of participants for whom (individual) data are available (n=13).

The efficacy of verb treatments is also tempered by lack of generalization of treatment to untrained verbs which only occurred for 15% of participants (similar to 14.5% in de Aguiar et al. (ibid)). The lack of generalization of verbs treatments is perplexing in that verbs are proposed to have looser semantic networks than nouns. Verbs tend not to be tied to particular actors or objects but instead constrain their semantic network through their argument structure, which is expressed in looser semantic terms (e.g. the looser notions of 'agent' and 'theme' rather than specific concepts such as 'teacher' and 'pen'). This should facilitate generalization at least to semantically related verbs, such as all verbs that have an 'agent' and a 'theme' (see e.g. Boyle, 2017). Given that 54% of studies incorporated semantic cues in treatment the lack of generalization is again hard to explain. There are several possibilities. Boyle (ibid) speculates that SFA treatment may have resulted in disappointing generalization because "the current lack of agreement about the semantic

representation of verbs means that we have not yet identified the features that might be most potent in promoting generalization of improved verb retrieval" (p.58). So, for example, generating the instrument of the action denoted by a verb, during SFA, might be more potent than generating the agent, if the former is stored as a feature of the verb but the latter is not. Another possibility is that generalisation of treatment to untrained verbs may require verbs to be treated in the context of a sentence. This explanation is supported by the suggestion that sentence treatments result in generalization more frequently than verb treatments (see the review by Webster & Whitworth, 2012). The reportedly more frequent generalization seen in sentence treatments may occur because sentence treatments require the production not just of the targeted verb during treatment but also its arguments. As noted above, most verbs are unconcerned about the particular phrase that appears in an argument slot, so long as the general semantic requirement of 'agent' or 'theme' is met. As a consequence, the production of a verb with its arguments may enable activation to spread more easily through the network. Thus sentence treatments may exploit verbs' looser semantic networks better than verb treatments. Alternatively, sentence treatments may result in greater generalisation because they are activating the syntactic information that verbs encode, resulting in a "syntactic bootstrapping" effect which verb treatments fail to capitalize upon.

Another possible explanation is drawn from anomia treatment research. The lack of generalisation of verb treatments mirrors that for anomia treatments in that they also predominantly improve only treated nouns (see e.g. Best et al., 2013). This is unsurprising in that verb treatments unashamedly use the same techniques as anomia treatments. It may indeed be the case that, as Howard (2000) has argued in relation to anomia treatments, all treatments whether they claim to be semantic, phonological (or, in the case of verbs,

gestural) actually function as “mapping” treatments, here in the sense of strengthening mapping between semantic and phonological stages of lexical retrieval. If this is the case, then treatment effects will be item specific (i.e verb or noun specific) because the mechanism of treatment is lexically based. Finally it could be the case that verb treatments do not result in generalisation simply because they are not powerful enough and/or not enough treament was given to enable generalization.

Whatever the reason, the lack of generalization of verb treatments strongly suggests that treatments should target verbs which are *functionally useful* to participants. However, only two reviewed studies did this: Carragher et al. (2013) included five functionally relevant verbs in a treated set of forty, and Palmer et al. included 48 personally relevant nouns/verbs in a treated set of 96, although the results for these verbs are not reported on specifically in either study. Recommendations that functionally relevant verbs should be the target of treatment have been made before (e.g. Webster & Whitworth, 2012) but as yet this recommendation has not been followed in verb treatment research. As the evidence to date indicates that treated verbs only improve in the vast majority of cases, then we owe it to our participants to work on targets that are maximally meaningful to their lives.

Turning to the impact of verb treatments on sentence production using treated and untreated verbs, interpretation of the evidence here is seriously restricted by lack of assessment as only 13% and 24% of studies assessed sentence production using trained verbs and untrained verbs respectively. Whilst these studies reported improvements in sentence production for some of their participants, no firm conclusions can be drawn due to the small numbers involved, and it is thus recommended that verb treatments routinely include an assessment of sentence production as an outcome measure.

Regarding broader impact, the impact of verb treatments on functional communication and discourse was also difficult to evaluate because it was infrequently measured. Only four studies assessed functional communication, and whilst they found some evidence of improvement, the very small number of participants (16) means again that no firm conclusions can be drawn. Similarly, it is recommended that functional communication is routinely assessed in verb treatment research, and that ideally, researchers should reach agreement on the measure used to do this.

Discourse was more frequently assessed (in twelve or 32% of studies) and improvements were reported for 31% of participants (for whom individual results were reported). Given the limited generalization of verb treatments, it is perhaps not surprising that the impact of treatment on discourse is limited. It seems plausible that treating functionally relevant verbs might produce a greater improvement in discourse as these verbs may be used more frequently in discourse in real life. This however leads to the problem of the best way to assess discourse as the reviewed studies used a variety of methods, with few analyzing samples of discourse from daily life. Indeed, the assessment of discourse is currently a topic of debate for many reasons, including the plethora of measures available (Bryant, Ferguson & Spencer, 2016), the varied psychometric quality of those measures (Pritchard, Hilari, Cocks, & Dipper, 2017), and the potential for a core set of discourse outcome measures (e.g. Dietz & Boyle, 2017). Whilst there continues to be disagreement about the best way to assess discourse (Wallace, Worrall, Rose & Le Dorze, 2017), it is difficult to reach agreement regarding a Discourse Core Outcomes Set to be used in verb treatment studies and in aphasia treatment research in general. However, this is high priority for future research.

Finally, regarding what induces change in verb treatment, the most commonly reported treatment technique was *phonological* cueing (reported in 67.5% of studies) followed by *semantic* cueing (54% of studies). Verb treatments are similar to anomia treatments here as the latter also commonly use semantic and phonemic cues (e.g. Nickels, 2002) and this is unsurprising given that most verb-in-isolation treatments are derived from anomia treatments. What is perhaps surprising is that only 22 of the reviewed studies (59%) incorporated treatment techniques which were designed to exploit the unique, action related properties of verbs (such as the use of gesture cues, action observation, video stimuli and the adaptation of SFA for verbs). All of these treatments were designed to capitalize upon the unique features of verbs, and their impact is promising. Indeed it seems plausible that treatments which deliberately target verbs' unique properties have the potential to be more powerful than those which transpose techniques directly from anomia treatments because these unique properties may be the (most) active ingredients of verb treatments. Thus, for example, the use of *video* rather than static picture stimuli in treatment may be more effective in eliciting verb production (Blankestijn-Wilmsen et al., 2017) because seeing an action performed primes production of the related verb's lexical form (either by mirror neurons or some other mechanism). This is also supported by the emerging area of research which indicates that *action observation* (either face-to-face or via video stimuli) has the potential to be an active ingredient of treatment (Bonifazi et al., 2013; Marangolo et al., 2012 and 2010). The use of both video stimuli and action observation warrants further investigation.

Also warranting further investigation is the *self-delivery* of verb treatments via *computer* as studies investigating this reported improvements in treated verbs for 100% of participants. Whilst the significance of this finding is greatly tempered by the small number

of participants (n=13), that verb treatment remained as (or possibly even more) effective when self-delivered is an alluring prospect, especially given that treating verbs has been considered more complex than treating nouns and might therefore be deemed unsuitable for self-delivery. There are many potential reasons why computer delivered verb treatments appear to work which may include the delivery of higher dosages of treatment (see Mortley et al., 2004). However there are potential disadvantages of computer-delivered verb treatments which include the need to train participants to use the treatment program, as well as the lack of supervision and feedback from a clinician (which was indeed seen as a disadvantage of self-delivered computer treatment by some participants and carers when interviewed by Palmer et al. (2013) and Wade et al. (2003)). There is thus an urgent need to explore in more detail how computer delivery affects (all) aphasia treatments whose efficacy has been established face-to-face, and indeed this is an emerging area of research. For example, Ball et al. (2018) found that whilst participants complied with the treatment protocol in only 45% – 61% of sessions of self-delivered anomia treatment, there were actually *more* successful naming attempts when participants did *not* comply with the treatment protocol than when they did. Ball et al. speculate that this may be because, through repeated interaction with the treatment program, participants identified the level of cueing which was most successful for them and adopted this. This points to the need for greater research into how participants *qualitatively* interact with computers during self-delivered computer treatment as well as how much, *quantitatively*, they receive.

The active status of verb treatment ingredients whether delivered face-to-face or by computer may depend on other factors, two of which will now be briefly discussed. The number of verbs treated in studies may influence the outcome. This review found that the

minimum number of verbs treated in the reviewed studies was five (Fink et al., 1997) and the maximum 120 (Bonifazi et al., 2013). However, studies which treated larger number of verbs tended to treat them in smaller sets, usually to compare treatment techniques. For example, Boo & Rose (2011) and Rose and Sussmilch (2008) each treated 80 verbs, in sets of 20, with either semantic cues, repetition, gesture or combined cues to compare the effectiveness of each of these treatment techniques. Indeed, treating verbs in sets of 20 was the commonest design found in the reviewed studies, with four studies treating a total set of 20 verbs and seven studies treating 80 verbs in sets of 20. However, it is unclear whether this is the optimal number for treatment. Edwards & Tucker (2006) treated a set of 50 verbs. This was replicated by McCann & Doleman (2011) who treated 100 verbs but in two sets of 50, in a crossover design. These would appear to be the largest sets of verbs treated in the studies of face-to-face treatments, with significant improvement in treated verbs for five of the six participants. The largest set of verbs treated in the studies of computer delivered treatments however was 100 (alongside a set of 162 nouns) in Mortley et al. (2004). The use of computers to deliver treatment clearly has the potential to greatly increase the number of verbs treated. However, given the self-administered nature of treatment in this study it is not guaranteed that every verb will be treated in every treatment session (see Ball et al. (2018) who found that participants adhered to the treatment protocol in only 45-61% of self-administered sessions).

Whilst it may seem logical to treat a larger number of verbs (given that treatment is likely to improve retrieval of treated verbs only), treating a large number of verbs might effectively “dilute” the dose of treatment. The number of times each verb might be treated (i.e. go through the prescribed treatment protocol) during a one-hour treatment session will be considerably less for a verb which is part of a 50-verb set compared to one from a set of

20. In this scenario, treatment of a large set of verbs might conceivably be less effective than treatment of a small set. Indeed it is possible that an active ingredient of verb treatments which aim to improve the *production* of verbs is the *number of attempts at producing* a verb, which are required during a treatment session, and this may be less for a large set of treated verbs. However, of the reviewed studies only Conroy et al., (2009a; 2009b; 2009c) specified that verb naming was attempted at least 100 times for each of the treated verbs during their treatment programs. Reporting the number of times a verb was attempted has the additional benefit of allowing the optimal dose of treatment to be more accurately determined than if just the number and length of sessions is reported (although even this is inconsistently reported as discussed earlier) and it is therefore recommended. Two recent studies underline the importance of detailed reporting not just of the amount of treatment delivered but also how much of each potential active ingredient is delivered. Quique, Evans and Dickey (2018) conducted a meta-analysis of SFA (for anomia) and found that response to treatment was positively correlated with the amount of treatment. However, Gravier et al. (2018) found that response to treatment was more accurately predicted by the *number of features generated* by a participant during SFA than either total treatment time or the average number of treatment trials in their treatment study also using SFA for anomia with 17 participants.

Thus it seems that the number of features generated by a participant in SFA may be an active ingredient of treatment. Indeed Gravier et al. (ibid) go on to hypothesize that it may be the generation of features *specific* to a noun which predict improvement of *treated* items (e.g. *fuzzy* in relation to the treated noun peach), whereas generation of features *shared* within a category (such as *fruit, round* and *stone*) may be predict generalisation to *untreated* items (such as plum and nectarine) and this is the subject of future research for

the authors. Whilst it is unclear how these findings might translate to SFA for verbs, It is such fine-grained research as this which has the potential to tease apart the active ingredients of aphasia treatments (see for example the series of studies by Marangolo and colleagues which are teasing apart the active ingredients of action observation treatment.)

A second factor which may influence the effectiveness of verb treatment ingredients is the *nature* and *severity* of the *deficit* underlying a verb retrieval difficulty. So for example, we might speculate that verb retrieval difficulties caused by a semantic difficulty would respond better to semantic treatment and likewise that phonologically based retrieval difficulties would respond better to phonological treatments. That is, semantic ingredients would be active during treatment of semantic deficits and phonological ingredients inactive, and vice versa for phonologically based deficits. We might also speculate that more severe verb retrieval deficits may be less responsive to treatment. The findings of this review only allow very tentative conclusion to be drawn here, with the suggestion that a variety of treatments (including semantic) were less effective in treating semantically based verb retrieval difficulties (Bonifiazi et al., 2013, Marangolo et al., 2010 Rodriguez et al., 2006; Wambaugh et al., 2004), and that participants with a more severe verb deficit may also be less responsive (Conroy et al., 2009c; Palmer et al., 2012). However, far more research is needed to establish what are the most effective (or active) ingredients of treatment for different types of verb retrieval deficits as this review also found no clear relationship between treatment given to participants and the nature of their underlying verb retrieval difficulties in line with Webster and Whitworth (2012).

Finally, whilst nearly half of the studies included in this review (16) compared treatment techniques to try to establish their status as active ingredients, no clear winners

emerged from these studies. Indeed, what was clear from this review is that verb treatments generally constitute complex interventions (as defined by the Medical Research Council) and it is thus likely to be difficult to tease apart what are the individual, active ingredients of treatment. It is indeed entirely possible that it is the *combination* of ingredients in verb treatments that is actually the "active" ingredient. The finding that generalisation to untreated verbs was almost entirely restricted to treatment that combined techniques adds weight to this suggestion.

Before final conclusions are drawn limitations of this review will be briefly discussed. The review is not a systematic review. Thus studies were not blind reviewed by two or more reviewers but solely by the first author. This is because the review was completed as part of a doctoral research study. Studies were also not evaluated with a published, standard tool as this is lacking for case series which was the design used for 29 (78%) of the studies reviewed. Whilst the systematic reviews of SFA (Efstratiadou et al., 2018; Maddy et al., 2014) and of gestural treatment for verbs (Rose et al., 2013) used the Single-Case Experimental Design (SCED) Scale (Tate et al., 2008) to review both single cases and case series, this does not address those aspects of case series which are beyond the scope of single case design, such as whether it is appropriate to report group results with the small n of case series (as was the case for six studies in this review).

The review also does not cover treatments which target verbs *together with their arguments* (as opposed to *in isolation.)* Although SFA adapted for verbs does include arguments in the array of features (e.g. via the question 'Who usually does this?') the excluded treatments target a verb's arguments *within a sentence structure* (such as Verb Network Strengthening Treatment (VNeST: e.g. Edmonds, 2016, Edmonds, Obermeyer & Kiernan, 2015) and mapping treatments (e.g. Byng, Nickels & Black, 1994; Marshall, Chiat &

Pring, 1997). This is an important limitation because the current evidence suggests that these sentence level treatments are more likely to results in improvements in *sentence production* (using treated verbs) than verb-in-isolation treatments (e.g. Webster & Whitworth, 2012). Sentence level treatments are therefore reviewed in a companion paper (Hickin, Cruice & Dipper, in preparation)

In conclusion, whilst this review identified many challenges for verb treatment research, researchers should not be despondent. The predominant finding that treatment improves the retrieval of trained verbs for 80% of participants is very encouraging. The challenge is to demonstrate generalisation of treatment to untrained verbs, and this remains a challenge for anomia treatment too. Functional communication and discourse need to be routinely measured to establish if verb treatments affect these. As Carragher, Sage and Conroy (2015) vividly express it, demonstrating that treatment improves communication in daily life remains the "holy grail" for all aphasia treatments. We encourage verb treatment researchers to don their hats, crack their whips and, Indiana Jones like, pursue their quest.

Acknowledgements.

This research was supported by a doctoral studentship to the first author jointly funded by the Worshipful Company of Saddlers and the School of Health Sciences, City, University of London.

References.

Adrian J.A., Gonzalez. M., Buiza, J.J. & Sage, K. (2011). Extending the use of spanish

computer-assisted anomia rehabilitation program (CARP-2) in people with aphasia. *Journal of Communication Disorders, 44*, 666–677.

Auerbach, C., Schnall, D., Heft LaPorte, H., & SingWin. (1995). In Bloom, M., Fischer, J. & Orme J.G, (editors). *Evaluating practice: guidelines for the accountable professional*. (2nd ed.) Boston (MA): Allyn & Bacon

Baker, E. (2012a). Optimal intervention intensity. *International Journal of Speech-Language Pathology, 14*, 401-409. https://doi.org/10.3109/17549507.2012.700323

Baker, E. (2012b). Optimal intervention intensity in speech-language pathology: Discoveries, challenges, and unchartered territories. *International Journal of Speech-Language Pathology,* 14, 478-485. https://doi.org/10.3109/17549507.2012.717967

Ball, A.L., de Riesthal, M. & Steele, R.D. (2018). Exploring treatment fidelity in persons with aphasia autonomously practicing with computerized therapy materials. *American Journal of Speech-Language Pathology, 27*, 454–463.

Beeson, P., & Robey, R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychological Review*, 16, 161–169.

Berndt, R.S.,Haediges, A.N., Mitchum,C.C. and Sandson, J. (1997). Verb Retrieval in aphasia. 1. Characterizing single word impairments. *Brain and Language,* 56, 68–106.

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J. & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex, 49*, 2345-2357. https://doi.org/10.1016/j.cortex.2013.01.005

Black, M & Chiat, S. (2003). Noun - verb dissociations: a multi-faceted phenomenon. *Journal of Neurolinguistics*, 16, 231-250.

Blankestijn-Wilmsen, J., Damen, I., Voorbraak-Timmerman, V., Hurkmans, J., Koning, B. D.,

Pross, A., & Jonkers, R. (2017). The effect of static versus dynamic depictions of actions in verb and sentence production in aphasia. *Aphasiology,* 31, 1166-1182. https://doi.org/10.1080/02687038.2016.1258537

Bonifazi, S., Tomaiuolo, F., Altoè, G., Ceravolo, M. G., Provinciali, L. & Marangolo, P. (2013). *European Journal of Physical and Rehabilitation Medicine*, 49, 473-481.

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*, 154-175. https://doi.org/10.1080/02687031003743789

Boyle, M. (2017). Semantic treatments for word and sentence production deficits in aphasia. *Seminars in Speech & Language, 38*, 52-61. https://doi.org/10.1055/s-0036-1597256

Brady, M.C., Kelly, H., Godwin, J., Enderby, P. & Campbell, P. (2016). 'Speech and language therapy for aphasia following stroke', *Cochrane Database of Systematic Reviews,* 6(Art. No.: CD000425). https://doi.org/10.1002/14651858.CD000425.pub4.

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, *30*, 489-518.

Byng, S., Nickels, L. & Black, M. (1994). Replicating therapy for mapping deficits in agrammatism: Remapping the deficit?, *Aphasiology, 8*, 315-341, https://doi.org/10.1080/02687039408248663

Carragher, M., Sage, K. & Conroy, P. (2015). Outcomes of treatment targeting syntax production in people with Broca's-type aphasia: evidence from psycholinguistic assessment tasks and everyday conversation. *International Journal of Language & Communication Disorders*, *50*, 322-336.

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*, 846-887. https://doi.org/10.1080/09602011.2013.832335

Cochrane Consumer Network (undated) Available at: http://consumers.cochrane.org/levels-evidence. [Accessed: 22 November 2017].

Conroy, P. & Scowcroft, J. (2012). Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study. *Neuropsychological Rehabilitation*, *22*, 295-318.

Conroy, P., Sage, K., & Lambon-Ralph, M. (2009a). A comparison of word versus sentence cues as therapy for verb naming in aphasia. *Aphasiology, 23*, 462 – 482.

Conroy, P., Sage, K., & Lambon-Ralph, M. (2009b). The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia, *Aphasiology*, 23, 707–730.

Conroy, P., Sage, K. & Ralph, M.A.L. (2009c). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology, 23*, 1311-1337.

Conroy, P., Sage, K. & Ralph, M.A.L. (2006). Towards theory-driven therapies for aphasic verb impairments: A review of current theory and practice. *Aphasiology,* 20, 1159-1185. https://doi.org/10.1080/02687030600792009

de Aguiar, V., Bastiaanse, R. & Micel, G. (2016). Improving production of treated and untreated verbs in aphasia: a meta-analysis. *Frontiers in Human Neuroscience, 10,* 468. https://doi.org/10.3389/fnhum.2016.00468

Dietz, & Boyle, M. (2017). Discourse measurement in aphasia research: have we reached the tipping point? *Aphasiology*, https://doi.org/10.1080/02687038.2017.1398803

Dijkers, M. (2015). What is a scoping review. *KT Update, 4. An e-newsletter from the Center on Knowledge Translation for Disability and Rehabilitation Research*.

[http://ktdrr.org/products/update/v4n1]

Douglas, J. M., O'Flaherty, C. A., & Snow, P. C. (2000). Measuring perception of communicative ability: The development and evaluation of the La Trobe communication questionnaire. *Aphasiology, 14*, 251–268.

Edmonds, L. A. (2016). A review of verb network strengthening treatment. *Topics in Language Disorders, 36*(2), 123-135. https://doi.org/10.1097/TLD.0000000000000088.

Edmonds, L. A., Obermeyer, J., & Kernan, B. (2015). Investigation of pretreatment sentence production impairments in individuals with aphasia: Towards understanding the linguistic variables that impact generalisation in verb network strengthening treatment. *Aphasiology, 29*, 1312-1344. https://doi.org/10.1080/02687038.2014.975180

Edwards, S. & Tucker, K. (2006). Verb retrieval in fluent aphasia: A clinical study. *Aphasiology*, *20*, 644-675. https://doi.org/10.1080/02687030600631827

Efstratiadou, E.A., Papathanasiou, I., Holland, R., Archonti, A. & Hilari, K. (2018). A Systematic Review of Semantic Feature Analysis Therapy Studies for Aphasia. *Journal of Speech, Language, and Hearing Research*, *61*, 1261–1278.

Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review, 11*, 1573 – 2827.

Faroqi-Shah. Y & Graham, L.E. (2011). Treatment of semantic verb classes in aphasia: acquisition and generalization effects. *Clinical Linguistics & Phonetics, 25*, 399–418.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005a). Further explorations and an overview of errorless and errorful therapy for aphasic word finding difficulties: The number of naming attempts during therapy affects outcome. *Aphasiology, 19*(7),

597–614.

Fink, R.B., Schwartz, M.F., Sobel, P.R. & Myers, J.L. (1997). Effects of multilevel training on verb retrieval: is more always better? *Brain & Language, 60*, 1, 41-45.

Garrett, M. (1988). Processing in language production. In F. Newmeyer (Ed.), *Linguistics: The Cambridge Survey, Vol 3.* Cambridge: Cambridge University Press.

Gravier, M.L., Dickey, M.W., Hula, W.D, Evans, W.S., Owens, R.L., Winans-Mitrik, R.L. & Doyle, P.J. (2018) What Matters in Semantic Feature Analysis: Practice-Related Predictors of Treatment Response in Aphasia. American Journal of Speech-Language Pathology, 27, 438–453. https://doi.org/10.1044/2017_AJSLP-16-0196

Hinckley, J. J. & Douglas, N. F. (2013). Treatment fidelity: Its importance and reported frequency in aphasia treatment studies. *American Journal of Speech-Language Pathology*, 22, 279–284.

Howard, D., Best, W. & Nickels, L. (2015) Optimising the design of intervention studies: critiques and ways forward, *Aphasiology*, *29*, 526-562, https://doi.org/10.1080/02687038.2014.985884

Howard D. (2000). Cognitive neuropsychology and aphasia therapy: the case of word retrieval. In: Papathanasiou, I, ed. *Acquired Neurogenic Communication Disorders: A Clinical Perspective*. London: Whurr.

Jones, E. V. (1986). Building the foundations for sentence production in a non-fluent aphasic. *International Journal of Language & Communication Disorders, 21*, 63-82. https://doi.org/10.3109/13682828609018544

Kaderavek, J.N. and Justice L.M. (2010). Fidelity: An Essential Component of Evidence-Based Practice in Speech-Language Pathology. *American Journal of Speech-Language Pathology, 19*, 369–379.

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: implications for lexical organization. *Brain and Language, 74,* 1-25.

Knoph, M. I. N., Lind, M., & Simonsen, H. G. (2015). Semantic feature analysis targeting verbs in a quadrilingual speaker with aphasia. *Aphasiology, 29*, 1473-1496. https://doi.org/10.1080/02687038.2015.1049583

Kristensson, J. Behrns, I. & Saldert, C. (2015). Effects on communication from intensive treatment with semantic feature analysis in aphasia, *Aphasiology, 29*, 466-487, https://doi.org/10.1080/02687038.2014.973359

Kurland, J., Wilkins, A. R., & Stokes, P. (2014). iPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatment. *Seminars in Speech & Language, 35*, 51-63. https://doi.org/10.1055/s-0033-1362991.

Leff, A.P. & Howard, D. (2012). Stroke: Has speech and language therapy been shown not to work? Nature Reviews Neurology. https://doi.org/10.1038/nrneurol.2012.211

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge MA: MIT Press.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A. & Zoghaib, C. (1989). The Communicative Effectiveness Index: development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*, 113-124.

Long, A. F., Hesketh, A., Paszek, G., Booth, M., & Bowen, A. (2008). Development of a reliable self-report outcome measure for pragmatic trials of communication therapy following stroke: The Communication Outcome after Stroke (COAST) scale. *Clinical Rehabilitation*, *22*, 1083–1094. https://doi.org/10.1177/0269215508090091

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. (2001). On the nature of the selective impairment of verb and noun retrieval. *Cortex, 37*, 724–726.

Maddy, K.M., Capilouto, J.G. & McComas, ,K.L. (2014). The effectiveness of semantic feature analysis: An evidence-based systematic review. *Annals of Physical and Rehabilitation Medicine, 57*, 254–267. http://dx.doi.org/10.1016/j.rehab.2014.03.002

Marangolo P, Cipollari S, Fiori V, Razzano C. & Caltagirone C. (2012). Walking but Not Barking Improves Verb Recovery: Implications for Action Observation Treatment in Aphasia Rehabilitation. *PLoS ONE 7(*6): e38610. https://doi.org/10.1371/journal.pone.0038610

Marangolo P, Bonifazi S, Tomaiuolo F, Craighero L, Coccia M, et al., (2010). Improving Language without words: first evidence from aphasia. *Neuropsychologia, 48*, 3824–33.

Marshall, J., Pring, T. & Chiat, S. (1998) Verb Retrieval and Sentence Production in Aphasia. *Brain and Language, 63*, 159–183.

Marshall, J., Chiat, S. & Pring, T. (1997). An impairment in processing verbs' thematic roles: A therapy study. *Aphasiology, 11*, 855 — 876. http://dx.doi.org/10.1080/02687039708250461

McCann C. & Doleman, J. (2011). Verb retrieval in nonfluent aphasia: A replication of Edwards & Tucker, 2006. *Journal of Neurolinguistics, 24*, 237–248.

McNeil , M.R., Doyle, P.J. , Spencer, K. Jackson Goda, V.,Flores, D. & Small, S.L. (1998) Effects of training multiple form classes on acquisition,generalization and maintenance of word retrieval in a single subject, *Aphasiology, 12*, 575-585. https://doi.org/10.1080/02687039808249559

Medical Research Council (2006) Developing and evaluating complex interventions: new

guidance. Available at: www.mrc.ac.uk/complexinterventionsguidance. [Accessed 7th February 2018].

Moher D., Liberati A., Tetzlaff J. & Altman DG. The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(6): e1000097. https://doi.org/10.1371/journal.pmed1000097

Mortley, J., Wade, J. & Enderby, P. (2004). Superhighway to promoting a client-therapist partnership? Using the Internet to deliver word-retrieval computer therapy, monitored remotely with minimal speech and language therapy input. *Aphasiology*, *18*, 193-211.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology, 16*, 935–979.

Palmer, R., Enderby, P. & Paterson, G. (2013). Using computers to enable self-management of aphasia therapy exercises for word finding: the patient and carer perspective. *International Journal of Language & Communication Disorders, 48*, 508-521.

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., . . . Dixon, S. (2012). Computer therapy compared with usual care for people with long-standing aphasia poststroke: A pilot randomized controlled trial. *Stroke (00392499), 43*, 1904-1911. https://doi.org/10.1161/STROKEAHA.112.650671

Peters, M.D.J., Godfrey, C.M., Khalil, H., McInerney, P., Parker, D. & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare,* 13 3: 141-146. https://doi.org/10.1097/XEB.0000000000000050

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. International Journal of Language & Communication Disorders, 52, 689-732.

Quique, Y.M., Evans, W.S. & Dickey, M.W. (2018). Acquisition and Generalization Responses in Aphasia Naming Treatment: A Meta-Analysis of Semantic Feature Analysis Outcomes. *American Journal of Speech-Language Pathology, 28*, 1-17. https://doi.org/10.1044/2018_AJSLP-17-0155

Raymer, A.M. , Ciampitti , M., Holliway, B., Singletary, F., Blonder, L.X., Ketterson, T., Anderson , S., Lehnen, J., Heilman, K.M. & Gonzalez Rothi, L.J. (2007). Semantic-phonologic treatment for noun and verb retrieval impairments in aphasia, *Neuropsychological Rehabilitation, 17*, 244-270, https://doi.org/10.1080/09602010600814661

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*, 867–882. https://doi.org/10.10170S1355617706061042

Raymer, A., & Kohen, F. (2006). Word-retrieval treatment in aphasia: Effects of sentence context. *Journal of Rehabilitation Research & Development, 43*, 367-377.

Raymer, A., Rodriguez, A., Ciampitti, M., Singletary, F. Fuller, R., Heilman, K. & Gonzalez Rothi, L. (2005). Effects of Pantomime Training for Noun and Verb Retrieval in Aphasia. *Neurorehabilitation and Neural Repair 19*, 378.

Raymer, A.M. & Ellsworth, T.A. (2002). Response to contrasting verb retrieval treatments: A case study. *Aphasiology*, *16*, 1031-1045. https://doi.org/10.1080/026870401430000609

Robey, R. R., & Schultz, M. C. (1998). A model for conducting clinical outcome research: An adaptation of the standard protocol for use in aphasiology. *Aphasiology, 12*, 787–810.

Rochon, E., Laird, L., Bose, A. & Scofield, J. (2005). Mapping therapy for sentence production impairments in nonfluent aphasia. *Neuropsychological Rehabilitation, 15*, 1-36. https://doi.org/10.1080/09602010343000327

Rochon, E. & Reichman, S. (2003). A Modular Treatment for Sentence Processing Impairments in Aphasia: Sentence Production. *Journal of Speech-Language Pathology and Audiology,* 27, 202 -210.

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*, 286-297. https://doi.org/10.1080/02687030500474898

Rose, M.L., Raymer, A.M., Lanyon, L.E. & Attard, M.C. (2013). A systematic review of gesture treatments for post-stroke aphasia, *Aphasiology, 27*, 1090-1127. https://doi.org/10.1080/02687038.2013.805726

Rose, M., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology, 22*, 691-706. https://doi.org/10.1080/02687030701800800.

Routhier, S., Bier, N. & Macoir, J. (2016) Smart tablet for smart self-administered treatment of verb anomia: two single-case studies in aphasia. *Aphasiology, 30*, 269-289, https://doi.org/10.1080/02687038.2014.973361

Routhier, S., Bier, N. Macoir, J. (2015). The contrast between cueing and/or observation in therapy for verb retrieval in post-stroke aphasia. *Journal of Communication Disorders 54*, 43–55. http://dx.doi.org/10.1016/j.jcomdis.2015.01.003

Schneider, S. & Thompson, C. (2003). Verb production in agrammatic aphasia: The influence of semantic class and argument structure properties on generalization. *Aphasiology, 17*, 213-241. https://doi.org/10.1080/729255456

Takizawa, T., Nishida, N., Ikemoto, A., & Kurauchi, N. (2015). Comparison of single-word therapy versus sentence therapy for verb retrieval and sentence production in Broca's aphasia. *Aphasiology, 29*, 169-194. https://doi.org/10.1080/02687038.2014.961121

Tate, R.L., McDonald, S., Perdices, M., Togher, L., Schultz, R. & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation, 18*, 385–401. https://doi.org/10.1080/09602010802009201

Thompson, C.K., Lange, K.L, Schneider, S.L. & Shapiro, L. P. (1997). Agrammatic and non-brain-damaged subjects' verb and verb argument structure production. Aphasiology, 11, 473-490, https://doi.org/10.1080/02687039708248485

Verhagen, A., de Vet, H., de Bie, R., Kessels, A., Boers, M., Bouter, L., & Knipschild, P. (1998). The Delphi List: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi Consensus. *Journal of Clinical Epidemiology*, *51*, 1235–1241.

Wade, J., Mortley, J. & Enderby, P. (2003). Talk about IT: Views of people with aphasia and their partners on receiving remotely monitored computer-based word finding therapy. *Aphasiology, 17*, 1031-1056. https://doi.org/10.1080/02687030344000373

Wallace, S.J., Worrall, L.E. Rose, T. & Le Dorze, G.(2017). Discourse measurement in aphasia research: have we reached the tipping point? A core outcome set … or greater standardisation of discourse measures? *Aphasiology,*

https://doi.org/10.1080/02687038.2017.1398811

Wambaugh, J. L., Mauszycki, S., & Wright, S. (2014). Semantic feature analysis: Application to confrontation naming of actions in aphasia. *Aphasiology, 28*, 1-24. doi:10.1080/02687038.2013.845739

Wambaugh, J.l. & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in Aphasia. *Journal of Rehabilitation Research and Development*, *44*, 381-394.

Wambaugh, J., Cameron, R., Kalinyak-Fliszar, M., Nessler, C., & Wright, S. (2004). Retrieval of action names in aphasia: Effects of two cueing treatments. *Aphasiology, 18*, 979-1004. https://doi.org/10.1080/02687030444000471

Wambaugh, J.L., Doyle, P.J., Martinez, A.L., & Kalinyak-Fliszar, M. (2002). Effects of two lexical retrieval cueing treatments on action naming in aphasia. *Journal of Rehabilitation Research and Development*, *39*, 455-466.

Webster, J. and Whitworth, A. (2012). Treating verbs in aphasia: exploring the impact of therapy at the single word and sentence levels. *International Journal of Language & Communication Disorders, 47*, 619–636. doi:10.1111/j.1460-6984.2012.001

Figure 1. Results of the systematically conducted scoping review.