Speech and language therapy for aphasia: Parameters and outcomes

Lucie Thomas[1], Louise Lander[2], Nicholas Cox[2] and Cristina Romani[1]

1. School of Life and Health Sciences, Aston University

2. Moor Green Outpatient Brain Injury Unit-Birmingham Community Healthcare NHS Foundation Trust

Address correspondence to:
Cristina Romani
Aston University
Aston Triangle
B4 7ET, Birmingham, UK
E.mail: C.Romani@Aston.ac.uk

Key words: review, aphasia, frequency of therapy, intensity of therapy, picture naming

Background. Speech and language therapy is effective in improving language outcomes in acquired aphasia. However, it remains unclear which therapy parameters are most important to ensure gains. Published literature reviews are limited by the heterogeneity of the protocols considered, conflation of important parameters, and/or the paucity of the studies reviewed.

Aims: We carried out two new reviews of the effects of therapy parameters on language outcomes, addressing some of the limitations of previous reviews and focusing on the effect of the number of words treated, cumulative dosage of therapy, and frequency of sessions.

Method and Procedure: In the first review (N studies =48; N participants =387), we considered only studies involving picture naming, in order to focus on a relatively homogeneous protocol. Here, we correlated therapy parameters with language outcomes. In the second review we included a broader range of protocols, in order to select studies where either dosage (N studies=8; N participants=211) or frequency of therapy (N studies=9; N participant =114) were contrasted, while other aspects of therapy were controlled for. Given the paucity of these studies, here we only presented descriptive analyses of studies that did, or did not, show significant effects of dosage or frequency on language outcomes.

Results: The therapy parameters used by the reviewed studies show use of limited resources (studies were characterised by small doses, short duration and few words treated). Percentage increase in number of words correct compared to baseline did not correlate or correlated negatively with all therapy parameters. Number of words gained, instead, correlated positively and significantly with the number of words treated and the number of words treated per hour. Controlled studies provided some evidence for the benefits of higher dosages of therapy, but no evidence in favour of either massed or distributed practice.

Conclusions: Results provided limited evidence that people with aphasia benefit from a higher dosage of therapy and no evidence at all that a massed mode of delivery is to be preferred. Instead, our results provide strong evidence of the benefits of treating larger sets of words, and more words per hour within the range of therapy durations and dosages reviewed by our study. They also suggest caution in using percentage increase as a measure of outcomes, as this will favour studies treating fewer words. Above all, our results highlight the lack of well-powered studies to assess the effects of therapy parameters on outcomes in controlled conditions.

1. INTRODUCTION

There is good evidence that speech and language therapy (SLT) is effective in ameliorating language difficulties in people with aphasia (PwA, see Albert, 2003; Greenhouse et al., 1990; Holland, Fromm, DeRuyter, & Stein, 1996; Robey, 1998; Whurr, Lorch, & Nye, 1992; see also Brady, Kelly, Godwin, Enderby, & Campbell, 2016 for a Cochrane review). However, we still have a limited understanding of the conditions under which SLT is most effective, and of the range of gains that therapists and patients should reasonably expect. Despite their importance for clinical practice, it remains unclear how therapy parameters --such as the frequency and length of sessions, the overall duration of the intervention and the number of words treated-- affect therapy outcomes (for discussions see Baker, 2012; Brady, et al., 2016; Cicerone et al., 2000).

An important clinical issue is to ascertain the ideal dosage of therapy necessary for good outcomes. This question may be answered properly only on an individual basis. However, the clinical and scientific communities would benefit from identifying, more generally, the range of parameters that produce efficacious therapy. Many studies have shown that SLT protocols delivered with high intensity are efficacious (see, for example, CILT protocols, e.g., Pulvermuller et al., 2001 or comprehensive intensive aphasia programs, e.g., Code, Torney, Gildea-Howardine, & Willmes, 2010). Conversely, studies delivering a smaller amount of therapy have shown limited benefits (e.g., Bowen et al., 2012). This gives us some indication that intensity is important. However, studies contrasting therapy schedules while keeping other aspects of therapy constant are few and, although there are a number of reviews which have examined the issue of intensity, these suffer from important limitations which reduce the strength of any conclusions that can be drawn. We summarise existing reviews which have examined relationships between therapy parameters and therapy outcomes, and discuss their limitations, before describing our own methodology.

Review of existing reviews. An early review by Whurr et al. (1992) carried out a meta-analysis of 166 studies (from 45 reviewed articles) assessing differences in outcomes between participants who did or did not receive a therapy intervention. The interventions included were very diverse, with therapies classified as: language stimulation, functional communication, cognitive retraining, automated, psychodynamic and drug intervention. They found a moderate average effect size = 0.59 in favour of participants receiving therapy. They also examined the effect sizes of therapy parameters considered categorically. They looked at duration of therapy (<4 weeks, 5-8 weeks, 9-12 weeks, >12 weeks), length of treatment sessions (<30 min, 30-60 min, >60 min), and number of treatment sessions (<60, 61-100, 101-200, >200). They also compared the mode of delivery (individual, group, or mixed). This review is an informative overview of parameters used in studies across types of therapies: the average duration of therapy was 28 weeks; the average length of session was 60 minutes; and the average number of sessions was 63. Most therapy was individually delivered. However, it provides limited evidence of the relative efficacy of different parameters, since there were no significant effects of parameters on outcomes. This, however, is not surprising given the heterogeneity of the studies compared.

More recent reviews have restricted, to an extent, the heterogeneity of the studies reviewed by concentrating on studies using impairment-based SLT. These reviews have concentrated on assessing whether delivering more intense therapy provides stronger

language benefits, but they have failed to isolate the effects of critical parameters such as frequency, duration and overall dosage of therapy. In most studies, 'intense' protocols delivered a higher dosage of therapy with increased frequency of sessions, thus confounding therapy dosage with therapy frequency.

A good example of this confound is in the often-cited review by Bhogal et al. (2003). Bhogal et al. investigated the relationship between therapy intensity and recovery in aphasia by assessing whether studies with a "positive" vs. a "negative" effect of therapy differed in intensity. They reviewed eight studies and found that those reporting a positive outcome delivered more intensive therapy. The four "positive" studies delivered on average **8.8 hours of therapy per week** over a duration of 11.2 weeks (in total, **98.4** hours of therapy) whereas the four "negative" studies delivered on average **2 hours of therapy per week** over a duration of 22.9 weeks (in total, **43.6** hours of therapy). Given these results, one may conclude that therapy delivered more frequently is to be preferred. However, this conclusion is unwarranted. Since "positive" and "negative" studies differed in both the frequency and dosage of therapy, it is unclear which variable is mostly responsible for differences in outcomes. Moreover, therapy outcomes were measured in a controversial way: by comparing the effect of therapy delivered by professional SLTs with the effects of other forms of therapy focussing on well-being and/or delivered by non-professional facilitators. In other words, studies with positive outcomes were those showing better outcomes with SLT than with other forms of therapy, and the significant effects of these studies reflected stronger gains with professionally versus non-professionally delivered SLT and not, as one would expect, stronger gains obtained with therapy versus no therapy (see also Marshall, 2008 for a discussion of the limitations of this review).

Other reviews have similarly provided evidence for the benefits of more intense therapy protocols but have not distinguished between therapy dosage and therapy frequency.

a) Basso (2005) measured benefits using differences in outcomes between treated and untreated patients. She reviewed nine studies. No statistical analyses were carried out, but the tables in the paper show that the studies reporting negative outcomes delivered fewer therapy sessions (N= 20, 40 and 48) than the studies reporting significant benefits (roughly, N = 576, 72, 60, 144, 56, 108) [1].
b) Robey (1998) carried out a meta-analysis considering 55 studies to assess efficacy of aphasia therapy and relationships with the therapy parameters. Twelve studies were assessed to evaluate effects of intensity. They were subdivided into three groups according to amount/frequency of therapy delivered: 1. Low (<1.5 hrs per week; N studies =2); 2. Moderate (2-3 hrs per week; N studies =4) and 3. High (=>5 hrs per week; N studies =6). Robey found that more intensive treatments resulted in larger effects. Moreover, he found significant, positive correlations between frequency, on one side, and therapy duration and number of treatment hours, on the other (see also Marshall et al., 1982 for a positive correlation between therapy outcomes and number of therapy sessions delivered).

[1] Therapy hours were calculated from the tables; they could not be calculated for two studies.

c) Cherney and collaborators (Cherney, 2012; see also Cherney, Patterson, & Raymer, 2011) reported the results of 11 studies which assessed language outcomes for intensive vs. less intensive treatment protocols. Five studies favoured intensive protocols, one favoured a non-intensive protocol, and three found no difference.
d) Finally, the Cochrane review by Brady et al. (2016) reviewed four trials (ORLA 2006, FUATAC, VERSE, SP-I-RIT) and four published studies (Bakheit et al., 2007; Denes, Perazzolo, Piani, & Piccione, 1996, Pulvermüller et al., 2001, and Smith, 1981) where therapy was administered with either high or low intensity. Results showed a modest advantage for intensive protocols. However, most of these studies and trials (except SP-I-RIT and Pulvermuller, 2001) confounded frequency with the overall amount of therapy delivered, not allowing an independent assessment of these parameters.

The reviews reviewed so far have embodied a clear confound between therapy frequency and therapy dosage, but they have generally supported the view that therapy of higher intensity--even within the limited range used by studies--improves language outcomes. Not all reviews, however, have found a positive effect of intensity. de Aguiar, Bastiaanse, & Miceli (2016), investigated the best predictors of efficacy for therapies treating verb production. They reviewed 30 studies, covering 166 patients, and analysed results using a forest plot methodology. They assessed outcomes in relation to a large number of variables (N=33) related both to patients' demographics (e.g., gender, age, education, months post onset), type of impairment (e.g., level of output, level of comprehension, fluency etc.), type of therapy (e.g., type of cues used), and therapy schedule (e.g., session frequency, session duration, number of treatment hours, number of days). Outcomes were measured by comparing performance pre- and post-therapy, but details are lacking. Forest plot analyses established that only three variables were significant in distinguishing between patients with significant, positive therapy outcomes and patients with non-significant outcomes: 1. level of comprehension at the beginning of therapy; 2. level of word repetition and 3. frequency of therapy (confounded with dosage). However, while the effect of the first two variables was as expected --more significant positive outcomes occurred when patients had better comprehension and repetition-- the effect of frequency was paradoxical. Better outcomes occurred with less frequent (=< 3 times per week) than with more frequent and higher-dosage treatments (=> 3 times per week). The reasons for this paradoxical result are unclear. Possible confounding factors here are the duration of therapy and the number of words treated, with more intensive therapy possibly being shorter and/or treating more words, resulting in a smaller percentage of gains from baseline (see below for a discussion). Certainly, however, the study by de Aguiar et al., (2016) highlights the need for further studies/reviews using different approaches and separating the effects of different parameters on therapy outcomes, including number of treated words, which is not commonly considered in review studies (but see Snell, Sage & Ralph, 2010).

We know of only one review by Dignam, Rodriguez and Copland (2016) where frequency of sessions (here called 'intensity') and dosage have been distinguished. Contrary to common use in the literature, Dignam et al. have reserved the term 'intensive' to describe therapy delivered with more sessions over a shorter period of time (massed delivery). They suggested that, while there is some evidence that massed practice is beneficial for motor recovery in animal studies, this evidence is not clear-cut because

benefits are not maintained long-term. Moreover, generalised stroke rehabilitation studies with human participants do not make it possible to distinguish dosage from frequency of practice, and cognitive studies which have assessed verbal memory and learning in healthy participants clearly indicate an advantage for less intensive/more distributed training (e.g., Glenberg & Lehmann, 1980; Rothkoph & Coke, 1966; for reviews see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Janiszewski, Noel, & Sawyer, 2003; Teasel, Foley, Salter et al., 2015). The picture is less clear when one focuses on rehabilitation studies with aphasic participants, but even here there is some evidence of an advantage for distributed learning, consistent with the literature on healthy participants. Dignam et al. (2006) reviewed four aphasia rehabilitation studies which controlled for patient differences by using within-subject designs and which equated total number of therapy hours administered across conditions. They found no clear evidence for either type of practice in the short-term, but some evidence in favour of distributed practice in the long-term. However, the paucity of the studies reviewed, the few participants assessed per study, and possible carry-over effects with within-subjects designs, limit the strength of any conclusion.

Finally, we know of only one review which has explicitly addressed the issue of the optimal number of words to treat in therapy. Snell et al. (2010) reviewed 21 studies and a total of 109 patients. They reported large variations in the number of words treated (5-120). Importantly, they found a *negative* correlation between outcomes in terms of percentage increase from baseline and number of words treated. However, in a cross-over study reported in the same paper, treating more words resulted in a higher raw number of words gained. This suggests that measuring outcomes as percentage increase from baseline may disadvantage studies which have treated more words, despite the fact that treating more words is crucial for functional gains.

Limitations of existing reviews. Our review of the literature highlights the following limitations of existing reviews:

First, the number of studies reviewed is small. There is some evidence that intensive therapy is more beneficial. However, reviews finding positive effects of intensity included only a limited number of studies (Basso, 2005; Bhogal et al., 2003; Cherney, 2012; Robey, 1998; reviewed respectively = 9, 8, 11 and 12 studies) or reported only small effect sizes (see Cochrane review). Reviews which included a larger number of studies found either no effect (see Whurr et al., 1992; N reviewed studies = 45) or a paradoxical effect (de Aguiar, 2016; N reviewed studies =30).

Second, existing reviews have measured therapy outcomes using different designs:

a) Bhogal et al. (2006) assessed group differences between PwA given SLT vs. another form of treatment, in studies using intense vs. non-intense protocols;
b) Basso (2005) assessed group differences between treated and untreated PwA, in studies using intense vs. non-intense protocols;
c) de Aguiar et al. (2016) considered predictors of outcomes (where outcome was not more precisely defined);
d) Robey (1998) compared effect sizes for studies using intense vs. non-intense protocols;
e) Dignam et al. (2016) assessed within-participant differences before and after therapy between studies using intense vs. non-intense protocols;

f) Cherney (2012) reviewed studies contrasting (within study) an intense vs. a less intense therapy schedule.

While all of these designs provide useful information, they face the common difficulty of comparing studies which are very heterogenous in terms of the type of therapy employed, the language tasks used for assessment, and/or the individual characteristics of the participants. Different types of therapy and different patient severity may be more important variables affecting therapy efficacy than frequency and dosage of therapy. This heterogeneity will reduce the power and significance of any effect. Comparing performance before and after therapy with a within-participant design, as done by Cherney (2012) and Dignam et al. (2016), is a good strategy, but not many studies are available when parameters are individually considered (see later).

Third, existing reviews have generally not distinguished parameters and outcomes sufficiently. Frequency of sessions and the cumulative dosage of therapy have generally been confounded (but see Dignam et al.,2015 and Dignam et al., 2016) and 'intense' has been commonly taken as synonymous with a frequent mode of delivery. This is problematic because it leads to the unwarranted conclusion that a massed mode of delivery is to be preferred. In fact, the literature presents no evidence that this is the case. If anything, the review of Dignam et al. (2016) --although based on only four studies-- suggests the opposite: that a distributed schedule is to be preferred. Finally, the number of words trained is an important parameter to consider which has been generally overlooked. The number of words trained is important since it has a clear relationship to the potential for functional gains.

Our study. In our study, we reviewed the literature to see which therapy parameters are commonly used by therapy studies and how they related to therapy outcomes, whilst also addressing some of the limitations of existing reviews. We wanted to review a large enough number of studies to power our analyses (N>40), but also to reduce the heterogeneity of the protocols reviewed to limit unwanted noise. We also wanted to separate the parameters which are often confounded in the literature (such as therapy frequency and dosage) as well as to consider other parameters (such as therapy duration and number of words treated). Finally, we wanted to measure outcomes in a systematic way, using common, meaningful measures.

The choice of which measure to use to assess outcome is important. Unfortunately, functional gains have not been measured in a sufficiently systematic way by rehabilitation studies to allow effective accruing of results. Studies focusing on picture naming, however, are plentiful, and improvement in picture naming can be used as a common outcome measure to assess the impact of different therapy parameters. In addition, one should consider which specific measure of improvement is used. Percentage increase in pictures named correctly from baseline is commonly used (e.g. Carpenter & Cherney, 2016; Kirmess & Maher, 2010; Kurland, Stanek, Stokes, Li, & Andrianopoulos, 2016). However, as suggested by the review of Snell et al. (2010), this measure may have the drawback of being confounded with number of words treated. It is easier to achieve larger percentage gains when a small number of words is treated. For example, if only 20 words are treated, one could easily achieve a 60% increase from baseline because only 12 words need to be

learned (e.g., from 4 to 16 correct). Conversely, if 200 words are treated, achieving the same percentage increase from baseline would entail relearning 120 words, which is much more difficult. However, a smaller percentage increase, say a 20% increase, will still correspond to a meaningful gain of 40 words.[2] Using percentage increase may be responsible for some of the null or paradoxical results from the literature. In our study, therefore, we will also use the raw number of words gained as a measure of outcomes with more functional significance.

In an ideal world, our review would examine one therapy parameter at a time, and consider studies where this parameter was manipulated while other aspects of the protocol were kept the same. Then, one could average the effect sizes obtained for each parameter on the same language task (e.g., picture naming) and assess significance. There are a few studies of this type in the literature, where dosage and frequency have been manipulated, but another design must be used if one wants to consider the parameters used in therapy studies more broadly. An alternative is to consider all studies where parameters and outcomes are reported in a quantitative way (even if not contrasted in a controlled way) and carry out correlations between therapy parameters and language outcomes. A drawback to this method is, of course, that variability in other therapy parameters may reduce and/or obscure correlations. Given these considerations, our study will carry out two different types of reviews to fully exploit the knowledge present in the literature and advance the debate over intensity and massed learning beyond studies such as Cherney (2012) and Cochrane (2016).

One review (Study 1) will focus on studies using relatively homogeneous protocols based on picture naming. Picture naming is one of the most common, if not the most common, type of therapy reported in the literature. Focusing on picture naming will allow analyses on a sufficiently large number of studies while keeping protocols relatively uniform. This will allow us to obtain an overview of the therapy parameters commonly used in therapy studies, of their interrelations, and of the relationship between parameters and therapy outcomes measured by percentage increase from baseline and increases in raw number of words gained. We would be particularly interested in assessing whether there are significant correlations between dosage of therapy, frequency of therapy, and number of words treated on one side, and therapy outcomes on the other. We expect a positive relationship with dosage, but possibly not with frequency. We also expect that number of words gained will produce more significant correlations than percentage increase from baseline. Finally, we expect that number of words treated will be inversely related to percentage increases from baseline, and that frequency and dosage of therapy will be positively related, consistent with possible confounding between these variables.

A drawback in our first study is that the variability, both of the patients treated and of the specifics of the protocols used across studies, may reduce correlations. A second review (Study 2) will, therefore, consider a wider range of impairment-based protocols to pull together studies where some parameters of therapy were contrasted while the type of therapy and other therapy parameters were kept the same. We will review both studies

[2] Here and elsewhere we use the term “percentage increase” to indicate the number percentage points increased, as shown in the example provided.

using a between-subjects methodology and studies using a cross-over within-subjects methodology where either: a) outcomes for high and low dosage of therapy were contrasted while keeping other therapy parameters the same, or b) massed vs. distributed therapy schedules were contrasted while keeping dosage the same (see also Dignam et al, 2016). Although we will be able to include more studies than Dignam et al. (2016), these constraints mean there are too few studies to carry out anything other than descriptive analyses. We expect that the results of this review will strengthen results of the first review. We also expect that most studies will show better outcomes when a higher dosage of therapy is used. Effects of frequency may not be present, however, because the possible advantages of massed practice for brain-plasticity may be counterbalanced by the advantages of distributed practice for memory and learning.

To avoid confusion caused by the term '*intensity'*, which may refer to either frequency of sessions or amount of therapy, following Baker et al., (2012), we will use the term '*dosage'* to reflect *total number of therapy hours* delivered. This will be calculated from number and length of sessions per week and duration over weeks. We will use the term *frequency*, rather than intensity, to refer to number of hours per week. We will use the term *duration* to refer to the total number of days, or weeks, over which therapy is delivered. Like Dignam et al. (2016), we will not consider *dose* in our reviews, since number of practice trials per word is rarely reported in therapy studies.

2. STUDY 1: OUTCOMES AND THERAPY PARAMETERS FOR PICTURE NAMING THERAPIES

Word finding difficulties and, more generally, word production difficulties are pervasive among PwA. Picture naming is a well-defined, replicable, and easy to administer treatment option for SLTs and, most importantly, it taps most processing levels impaired in aphasia. There is consensus that producing the correct word for a semantic representation requires the successful completion of three processing stages broadly defined as: selecting the correct lexical item, selecting the right phonemes, and organizing an articulatory plan. All of these stages can be affected in aphasia separately or in combination (as in anomia, phonological impairments, apraxia of speech; e.g., Best et al., 2013) and picture naming engages all of them. For these reasons, most structured, impairment-based therapy protocols for aphasia have focused on picture naming (see semantic cueing hierarchies, phonological cueing hierarchies, semantic feature analyses, phonemic training etc; e.g. Best et al., 2013; Boyle, 2004; Coelho, McHugh, & Boyle, 2000; Howard & Papathanasiou, 2000; Howard et al., 1985; Nickels & Best, 1996; Wambaugh, Linebaugh, Doyle, & Martinez, 2001). Clearly, there are limits to therapies based on picture naming. For example, it is unclear to what extent gains in picture naming generalise to functional communication. Some studies have reported limited generalisation of trained words to contexts outside of the confrontational picture naming tasks used in therapy (Best et al., 2013; Jokel et al., 2016; Nickels, 2002b). Other studies, however, have found significant generalisation from picture naming to connected speech, showing increased use of the trained words in non-confrontational contexts (e.g. see Conroy, Sage, & Ralph, 2009; Hickin et al., 2001; Romani, Thomas, Olson, & Lander, 2019). This motivates a continuing focus on this type of therapy.

The popularity of picture naming therapies justifies our focus on this type of therapy to review parameters and outcomes and carry out correlations. Surprisingly, existing reviews of aphasia therapies based on picture naming have not looked specifically at the

effects of therapy parameters on outcomes. For example, Zhang et al. (2017) assessed the outcomes of seven studies which contrasted CIAT protocols with other types of confrontational picture naming. They concluded that there was no difference in outcomes but did not assess differences in therapy parameters across studies. Wisenburn and Mahoney (2009) carried out a meta-analysis on 44 studies, investigating outcomes in picture naming tasks using semantic, phonological, and mixed cueing techniques. They assessed outcomes for both trained and untrained items (related and unrelated, exposed and unexposed). They found an overall effect size =1.66, which varied from 2.7 (SD=3.2), for treated items, to 0.44 (SD=0.13) for unrelated, unexposed items. This review provided evidence for the efficacy of cued confrontational naming therapies, but it did not include studies using structured group therapy techniques such as CIAT. Crucially, it also did not assess the effects of different therapy parameters on outcomes.

We will review results for both individually- and group-delivered picture naming therapies on word retrieval before and after therapy. Our review will also include studies using protocols based on CIAT (also referred to as ILAT) since these protocols still practice picture naming, albeit in the context of more propositional speech (requesting pictures, accepting/rejecting requests), and they generally report outcomes in terms of picture naming scores.

First, we will report statistics for therapy parameters in terms of dosage, duration, frequency of therapy, number of words treated, and number of words treated per hour. We will report statistical analysis of outcomes in terms of percentage accuracy gain from baseline for treated and untreated words, raw number of treated words gained, and number of words gained per hour of therapy (as a measure of therapy efficiency). For gains in percentage accuracy, we will also report effect sizes where available. Secondly, we will carry out correlation and regression analyses between therapy parameters and between therapy parameters and language gains.

We will comment on gains that extend beyond picture naming where available, including improvements in the use of trained words in other types of language tasks (e.g., in directed conversations or narrative productions) and/or improvements on standardised language tests. These further gains, however, are too heterogenous to lead to quantitative analyses.

2.1 Study 1: Method

The Web of Science and ASHA databases were searched (January 2019) using the following combination of search terms:
Aphas* AND
picture naming OR word retrieval OR constraint OR cueing OR ILAT OR semantic feature analysis OR verb network strengthening OR Schuell AND
therap* OR intervention OR rehabil*

'*' functions as a wildcard symbol to broaden the search by finding words that start with the same letters. Abstracts, titles and author keywords were searched. The first author read

all returned titles and abstracts and excluded irrelevant articles. The search also included reference lists from relevant previous reviews and articles.

The inclusion criteria were articles which:
1) treated participants with acquired aphasia (not primary progressive);
2) reported outcomes in confrontational picture naming before and after therapy;
3) reported parameters of therapy administration (e.g., frequency, duration, number of words treated).

The exclusion criteria were articles which:
1) were reviews or meta-analyses;
2) focused on brain stimulation techniques and/or a pharmacological intervention;
3) did not measure outcomes in terms of naming of trained items;

Eligible articles were reviewed for:

1. Type of intervention (group or individual; individual, one-to-one studies all used confrontation naming; group studies all, but two, used CIAT/CILT protocols);
2. Type of aphasic patients treated;
3. Duration of the intervention in number of days;
4. Dosage (total number of therapy hours);
5. Frequency (number of hours per day/week); this was computed by dividing duration by dosage;
6. Number of words treated;
7. Number of words treated per hour;
8. Percentage of gain on treated words from baseline;
9. Percentage of gain on untreated words from baseline;
10. Raw number of words gained;
11. Improvements in standardized assessments and/or other types of assessments;
12. Efficacy of therapy in terms of number of words gained per hour of therapy; this was computed by dividing number of words gained by dosage.

The initial keyword search yielded 1570 results. A review of titles and abstracts excluded 1491 articles, with a further 12 articles excluded as duplicates. The remaining 67 articles were then subject to full text review. Thirty-nine records were further excluded because the primary intervention was not picture naming (N=6), insufficient information was presented (N=15), articles were reviews (N=2), post-treatment naming was not assessed (N=13), or participants had Primary Progressive Aphasia (N=2). Sixteen additional articles were included based on previous reviews, examination of the references of included articles, or personal communication about a just-completed study. In total, 44 articles were included in this review (see Figure 1 for a flow chart). Where articles compared two different therapeutic techniques, these were considered separately, leading to comparison of 48 different therapeutic interventions (studies). Where results were presented in graphs, digitization software (PlotDigitizer) was used to calculate actual scores, and performance across participants was averaged.

Insert Figure 1 about here

Analyses. We will analyse results to:

1. Provide descriptive statistics for the therapy parameters used by the reviewed studies (number of participants, therapy dosage, therapy frequency etc.), and for therapy gains in terms of increase in percentage correct from baseline (for treated and untreated items) and increase in raw number of pictures named correctly.
2. Assess intercorrelations among therapy parameters using bivariate Pearson r correlations. Because there were large differences in the number of participants included, studies contributing to the correlations will be weighted by participant numbers.
3. Assess correlations between therapy parameters and measures of outcomes (percentage and raw increases in number correct from baseline) with bivariate Pearson r correlations. Therapy studies contributing to the correlations will be weighted by participant numbers.
4. For the therapy parameters showing a significant correlation with outcomes, we will assess the regression coefficients.

2.2 Study 1: Results

An overview of all reviewed studies is presented in Appendix A. Participants of all aphasia types were treated. Most studies used a multiple baseline design (39/48). Outcomes after the experimental therapy were compared with: a) usual care (N=2 studies); b) a control condition with unrelated tasks (N=12 studies); c) untreated words (N=7 studies). There was no control in 17 studies which only reported gains with treated words before/after therapy. Before/after results were reported by all the reviewed studies.

Overview of Parameters. Table 1 shows a summary of the therapy parameters used by our reviewed studies. There is variability, as one would expect, but trends are clear. The total number of therapy hours delivered was small (median=12), therapy duration was short (about a month), relatively few words were treated (median=40) and therapy was delivered to few participants (median=5.5). Treated words showed positive gains (median % increased from baseline=33.5; mean number of gained words =16) with large effect sizes both when assessed with Cohen's d (median=1.3) and with Serlin's d (median=4.3). Gains for untreated words were smaller. Gains in other tasks assessing functional generalisation were assessed in only a minority of studies (19/48 interventions; see Appendix B). Three studies reported positive functional gains, nine reported some gains in limited tests, and seven reported no gains.

Insert Table 1 about here

Table 2 compares parameters and outcomes for therapy delivered individually or in a small group (these are CIAT/CILT/ILAT protocols, except for Romani et al., 2018 and Attard, Rose, & Lanyon, 2013). Group-delivered therapies provided a higher dosage of therapy, delivered therapy more frequently, and treated more words. One-to-one therapies had

higher efficacy in terms of number of words gained per hour of therapy. This is expected given that participants receive more individual attention.

Insert Table 2 about here

Correlations among parameters. Pearson correlations among therapy parameters are shown in Table 3. There was a positive correlation between cumulative dosage and frequency (r=.66). Therapies with higher a dosage were delivered more frequently. This shows that dosage and frequency have been confounded in previous studies which have assessed intensity. An intensive study is often a study where therapy is administered both more frequently and with a higher dosage, making it impossible to disentangle the contribution of these two variables. There was no correlation between cumulative dosage and duration; therapies with a higher dosage did not last longer. Instead, there was a negative correlation between duration and frequency (r= -.46). Therapies lasting longer were delivered less frequently. There was a positive correlation between number of words treated and duration (r=.48). More words were treated in therapies that lasted longer. Finally, there were negative correlations between number of words treated per hour and both dosage and frequency. Fewer words were treated per hour in therapies which involved more hours (r= -.58) and were delivered more frequently (r= -.52). Instead, more words were treated per hour when a higher overall number of words was treated (r=.38).

Insert Table 3 about here

Correlations between parameters and outcome measures. Pearson correlations between parameters and outcomes are shown in Table 4.

Insert Table 4 about here

When outcomes were assessed in terms of a percentage increase from baseline there was a significant negative correlation with duration (r=-.52) reflecting smaller increases with therapies lasting longer and a negative although non-significant correlation with dosage. These negative correlations, however, may reflect a confounding with number of words treated. Consistent with this, there was a significant, negative correlation between percentage increase and number of treated words (r=-.55). Equivalent percentage gains are more difficult to achieve when a larger number of words are treated. Therefore, it is likely that therapies lasting longer and with a higher dosage will produce a smaller percentage gain because they treat more words.

There were no significant correlations when outcomes were assessed as % gains for untreated words, except, again, a negative correlation with number of treated words.

When outcomes were assessed in terms of raw number of words gained, correlations with dosage and duration were close to zero. The correlation with frequency was negative and did not reach significance. Instead, there were positive significant correlations with number of words treated (r=.68) and number of words gained per hour of therapy (r=.51) reflecting the fact that therapies treating more words and treating more words per hour produced more benefits.

Figure 2 shows scatter plots in different panels where number of gained words is predicted from: a) number of treated words and dosage; b) number of treated words and duration and c) number hours per week (frequency) and d) number of words treated per hour. The graphs clearly depict the positive associations between number of words gained and both number of words treated and number of words treated per hour. We ran separate regressions using these measures as independent variables and number of words gained as the depend variable. Regressions showed a significant proportion of variance predicted by number of treated words (R^2=.67, F(1,46)=40.39, p<.001;regression equation: y=6.9 + 0.154x) and by treated words per hour (R^2=.26, F(1,34)=12.04, p=.001; regression equation: y=11.0 + 1.42x). They indicated that one additional word was gained for every 6.5 additional words treated and/or one word was gained for each 0.7 additional words treated per hour. The trend with duration was flat and with frequency was slightly negative. These results, therefore, provide no evidence that more frequent, massed treatment is to be preferred.

Insert Figure 2 about here

---

3. STUDY 2: CONTRASTING DOSAGE AND FREQUENCY OF THERAPY

A limitation of the correlation analyses carried out so far is that, although all reviewed studies focused on picture naming, they still varied in the specific kind of therapy administered and in types and severity of patients treated, which may confound or obscure trends. In the following section, we will review studies where dosage and frequency of therapy have been directly contrasted while controlling for other variables.

3.1 Study 2: Method

Given the small number of studies where dosage or frequency were manipulated in a controlled fashion, we considered all relevant studies using an impairment-based approach, even if there was not a specific focus on picture naming.

The Web of Science and ASHA databases were searched (January 2019) using the following combination of search terms:
Aphas*
AND
therap* OR intervention OR rehabil*
AND
intens*OR dos* OR frequen*

Abstracts, titles and author keywords were searched. The first author read all returned titles and abstracts and excluded irrelevant articles. The search also included reference lists from relevant previous review articles and papers.

Inclusion criteria were studies which:

1) treated patients with acquired aphasia;
2) reported behavioural outcomes before and after an impairment-based intervention;
3) manipulated dosage/frequency parameters in a controlled way;
4) included an appropriate control condition (e.g., contrasted therapy parameters between- or within-subjects).

Exclusion criteria were studies which:

1) were reviews or meta-analyses;
2) used brain stimulation and/or pharmacological interventions;

The search yielded a total of 1894 studies, which was then reduced to 17 studies eligible for review (See Figure 2 for flowchart of review process). Eligible studies were reviewed for the same variables examined in the picture naming review.

Insert Figure 2 about here

We reviewed **8** studies which contrasted dosage, including **211** participants. Not all of these studies controlled for the same parameters. In one study, both frequency and duration were kept the same, while length of sessions was varied to manipulate dosage. In six studies, duration was kept the same, while frequency of sessions was varied to manipulate dosage. Finally, in one study, frequency and duration were both varied to manipulate dosage. Six of these studies used a between-subjects design and two used a within-subjects design.

We reviewed **9** studies which contrasted frequency, including **109** participants. In all of these studies, total number of hours (dosage) was kept the same, while duration was changed to manipulate frequency (shorter duration = more massed practice). Five of these studies used a between-subjects design and four used a within-subjects design.

Note that we reviewed studies where either dosage or frequency was manipulated while other therapy parameters were kept the same. Therefore, in our results, *"high dosage"* and *"low dosage"* and *"high frequency"* and *"low frequency"* will be used in relative, not absolute terms. For example, in Carpenter & Cherney (2016), one hour of therapy per day for two weeks is considered low dosage, because it is contrasted with two hours of therapy per day for two weeks, even if both schedules provide more therapy than most PwA will normally receive. The same is true for the contrast '*massed*' vs. '*distributed*'.

Analyses: Given the limited number of studies, only qualitative analyses will be conducted, with no statistical analyses**.**

3.2 Study 2: Results

Dosage. Results are reported in Table 5. Out of eight studies, five showed a better outcome with a higher dosage of therapy and three found no significant difference. The studies finding no difference included more participants (N=140 vs 71), but also manipulated dosage more weakly (for the two studies where data are available: 24 vs. 54 hours of therapy) compared to the studies finding a significant effect of dosage (27 vs. 182 hours in four studies; or 138 vs. 556 hours if one includes a fifth outlier study for number of therapy hours). Therefore, in the studies finding no difference, the high-dosage condition, included only 1.3 times more hours than the low-dosage condition, while in studies finding a difference, it included 3.0 to 5.8 times more hours depending on whether the outlier study is included. These results indicate that a higher dosage of therapy is beneficial, but given patients' individual differences in severity and motivation, significant differences emerge only when a strong difference in dosage is assessed (and/or with a relatively large number of participants).

Insert Table 5 about here

Frequency. Results are reported in Table 6. Out of nine studies, five found significantly better results with massed than distributed practice, two showed a trend in the same direction and two showed significantly better results in the opposite direction. Superficially, these results suggest better outcomes with massed practice. However, of the five studies reporting a significant advantage for massed practice, one (Pulvermuller et al. 2001) compared the effects of conventional SLT, delivered with distributed practice, with those of CIAT delivered with massed practice, thus confounding type of practice with type of therapy. The remaining four studies showing a significant effect for massed practice had very few participants (N=12) in contrast with the two studies showing a significant effect of distributed practice where the number of participants was more substantial (N=42). Furthermore, studies demonstrating a trend towards better outcomes with a massed schedule showed inconsistent results across participants. In Ramsberger & Marie's (2007) study, only one of four participants demonstrated more gains following massed treatment. In Mozeiko, Coelho, and Myers' (2016) study, four out of six participants showed improvement following massed practice, whilst two showed significant improvements only after distributed practice. Taken together, therefore, results do not clearly favour either type of practice.

Insert Table 6 about here

---

4. General Discussion

We have reviewed the literature on aphasia rehabilitation with two related aims: 1. To gain an overview of the therapy parameters used by therapy studies and their inter-relationships; 2. To acquire further evidence of how the therapy parameters of dosage, frequency, duration, and number of words treated relate to therapy outcomes. For aim 1, we reviewed studies focusing on a popular intervention, namely picture naming; for aim 2 we both reviewed the literature on picture naming and carried out a second, more general, review of impairment-based studies where either frequency or dosage were contrasted while other parameters were kept the same.

Overview of therapy parameters. Our review has highlighted how studies using protocols based on picture naming have often involved few participants and used therapy parameters which limit the likelihood of achieving positive results. It was notable that, across the reviewed studies, the median number of participants assessed per study was very small (median N=5.5), therapy was administered in a small dosage (median number of hours = 12), for a limited amount of time (median number of days = 28), and therapy targeted a relatively small number of words (median number of words = 40). It was also surprising to see that many studies did not report therapy parameters and, therefore, could not be included in our review (see also Nickels, Best & Howard, 2015 for a similar observation).

Despite resource limitations, our review shows that substantial gains are achievable in terms of increases in percentage of words correct from baseline and raw numbers of words gained. This is important. However, whether these gains translate into functional gains is more difficult to judge. Many studies either do not assess functional gains beyond naming the pictures practiced in therapy (29/48) or, if they do, they do not show any gains (7/19). The heterogeneous way in which functional gains are measured makes accruing results difficult, but, overall, they appear limited. This is not surprising, given the range of parameters used. Unfortunately, the limited resources evident in the research studies mirror limited resources available in many clinical settings. Our results here agree with those of large studies which have recently assessed functional gains after the provision of an aphasia rehabilitation programme.

Palmer, Witts & Chatter (2018) analysed the SLT provision received in the UK by 278 PwA treated in 21 different speech and language departments at different times post-stroke. Across the sample, PwA received a very small dosage of therapy in the preceding three months (only 6.3 hours of therapy). Importantly, provision remained very limited, even for PwA, in the early stages of recovery. For PwA between one- and three-months post-stroke (N=63), the median provision was one one-hour session per week. This was reduced to one one-hour session every two weeks for PwA between six-months and one-year post stroke (N=33). Not surprisingly, Palmer et al., found no difference in gains between patients receiving SLT and patients receiving support for a similar amount of time by a trained assistant. Like ours, these results indicate that the amount of therapy delivered in most research studies and clinical settings is 'homeopathic', making the value of SLT difficult to evaluate (see Leff & Howard, 2012).

In a second study, Palmer et al. (2019) have assessed the efficacy of computer-delivered therapy for word-finding difficulties in a large randomized controlled trial. PwA from 21 different speech and language departments were assigned to three different

treatment conditions: 1. Usual care (N=86); 2. Usual care + a computer-based, picture-naming based therapy (N=71); or 3. Usual care + control exercises involving puzzles (sudoku, spot the difference, word searches, or colouring; N=83). Gains were assessed after six months. The experimental group received a larger (but still limited) dosage of 28 hours of therapy on average. Unfortunately, results were disappointing. The experimental group showed significant improvement in naming the pictures trained in therapy, but no functional gains, with no between-group differences in: a) using the treated words in conversation; b) naming untreated words; c) the quality of conversation (assessed by TOM); or d) perceived quality of life. These results may indicate that functional gains may be achievable only with more hours of therapy, but they may also indicate that practicing language in a social context is a crucial ingredient to achieve functional gains. Our research group has recently piloted an approach based on playing language games in teams. This approach reduced the need for professional resources, whilst also allowing the practice of language in a social context. Published and unpublished research using this approach has shown functional gains beyond naming the pictures trained in therapy (Romani et al., 2019).

Outcome measures as percentage increase vs. raw numbers of words gained. Our review highlighted potential difficulties in using a percentage increase from baseline as a measure of therapy success. In fact, correlations with dosage and duration of therapy were *negative* (but only significant with duration); the higher the dosage and duration of therapy the smaller the % increase. This paradoxical result may be explained by a confound with the number of words treated. Percentage increase from baseline was significantly *negatively* correlated with number of words treated, as it is harder to achieve a large increase when treating more words. On the other hand, studies using more words tended to use a higher dosage and duration of therapy, although only the correlation with duration was significant. A better measure of therapy outcomes could be raw number of words gained which would also be more closely related to functional gains. With this measure, we found positive correlations with number of words treated and number of words treated per hour.

Our results are consistent with the results of Snell et al. (2010) who also found a negative correlation between percentage increase from baseline and number of words treated and with those of De Aguiar at al.(2016) that also found paradoxical effect of dosage on therapy outcomes. Snell et al. (2010) pointed out a possible confounding of aphasia severity across studies. However, a more important issue may be the difficulty of achieving a higher percentage of gain when more words are treated. In fact, in a cross-over study reported in the same paper, Snell et al., (2010) found that all the 13 PwA they tested learned more words when they practised a larger vs. a smaller set of words (N=60 vs N=20) over ten sessions of therapy.

Correlations among parameters. We found positive correlations between therapy dosage and frequency of sessions and a negative correlation between dosage and number of words treated per hour. Therefore, therapies with more hours involved more frequent sessions and treated fewer words per hour. There was also a positive correlation between duration and number of words treated. Therapies lasting longer treated more words, but the increase in dosage allowed therapies with higher dosage to treat fewer words per hour. Finally, there was a significant negative correlation between therapy frequency and therapy duration. Therapies with more frequent sessions were delivered in a shorter time period.

This highlights a confound in the literature, where "intensive" refers to therapies that both deliver more hours and use a massed mode of delivery.

Correlations between parameters and outcomes. Our analyses found no significant effects of increased dosage, frequency, or duration of therapy on number of words gained (and no or negative correlations with % of gains as already discussed). Instead, we found strong positive correlations between number of words treated and number of words gained. This result is consistent with results from Snell et al. (2010) and Laganaro, Di Pietro & Schneider (2006) who also showed numerically larger gains when larger sets of words were treated with the same amount of therapy. In addition, we found that treating more words per hour resulted in larger gains, showing that this effect is not simply the result of therapy duration. These results are far from trivial. Treating more words could have produced no advantage or even a disadvantage if the words to be learned did not receive enough practice and/or interfered with one another. Instead, our results indicated benefits, at least in the range of parameters assessed by our reviewed studies (N of treated words=10-180; median duration of therapy: about a month). This result is important since it encourages clinicians and academics to use larger sets of words during therapy (see also Snell et al., 2010 and Howard, Best & Nickels, 2015). We found that an extra word is gained for every 0.7 more words trained per hour. Treating more words will be likely to enhance functional gains, although it would be important to assess this formally.

We gathered further evidence for effects of therapy dosage and therapy frequency by reviewing studies where the effect of these variable was contrasted while controlling for other aspects of therapy, using within- or between-participants designs.

Studies contrasting dosage showed better outcomes when a larger amount of therapy was delivered. This result is consistent with results from Basso (2005), Bhogal et al. (2006), Bradely et al. (2016), Robey et al. (1998), and Cherney (2012), which reviewed studies overlapping with those we reviewed. However, the number of studies reporting significantly better outcomes with more therapy was small (N=5 with a positive effect vs. N=3 with no difference) and involved relatively few participants (N=71). Studies reporting no difference involved more participants (140) but used a weaker manipulation of dosage. These considerations, together with the lack of positive correlations across studies in our first review, and a possible publication bias favouring positive results, means that the evidence supporting more gains with a higher dosage of therapy remains limited (see Baker, 2012; Brady et al., 2016 for the same conclusion). Outcomes may be more depend on the type of patients treated and the type of therapy delivered than on dosage alone. Thus, significant differences may emerge only when a relative small amount of therapy (e.g., a dosage of 24 hours, with therapy delivered twice a week, for one hour each time, for three months) is compared to a much higher dosage (e.g., a dosage of 144 hours, with therapy delivered three times a week, for two hours each time, for six months).

Studies contrasting a distributed with a massed schedule while keeping dosage the same showed no evidence in favour of either type of schedule. Excluding the Pulvermuller (2001) study, where type of practice was confounded with type of therapy, the remaining studies showed no clear evidence for either schedule once the number of participants in the studies was taken into consideration. This, together with a lack of correlation between

frequency and number of words gained, provides no evidence that a massed practice is to be preferred (for this claim see Bhogal, Teasell & Speechley, 2003; MacLellan et al., 2011; for a conclusion similar to ours see Dignam et al., 2016).

Conclusions

Our review highlights limitations in the parameters manipulated by therapy studies using picture naming to assess therapy outcomes (limited duration, number and frequency of sessions, low-moderate number of words treated) as well as the low numbers of participants tested. The limited range of parameters studied may be, in part, responsible for the limited functional gains reported by these studies. Our review also sounds a cautionary note regarding the use of percentage increase in target words correct compared to baseline as an outcome measure. This measure is negatively correlated with raw number of words treated, as it is easier to achieve a higher percentage gain with a smaller set of words. Therefore, this measure will disadvantage studies treating more word, although this is important for functional gains. Measuring outcomes in terms of raw number of words gained is a better measure of therapy benefits. Our study provides no evidence that a massed mode of delivery is preferable to a distributed more. It does, however, provide evidence that larger gains are obtained by therapies treating a larger set of words. This last finding should encourage the use of therapy protocols where more words are treated, provided that therapy is delivered with a good overall dosage and sufficient duration. Finally, our study highlights an urgent need for well-powered studies contrasting different therapy parameters while controlling for others and using either matched groups of participants or contrasting conditions within participants.

References

* is used to denote articles included in the first review
+ is used to denote articles included in the second review

* Adrian, J. A., Gonzalez, M., & Buiza, J. J. (2003). The use of computer-assisted therapy in anomia rehabilitation: A single-case report. *Aphasiology, 17*(10), 981-1002. doi:10.1080/02687030344000256

Albert, M. L. (2003). Aphasia therapy works! Editorial comment. *Stroke, 34*(4), 992-993.

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

Baker, E. (2012). Optimal intervention intensity in speech-language pathology: Discoveries, challenges, and unchartered territories. *International Journal of Speech-Language Pathology, 14*(5), 478-485. doi:10.3109/17549507.2012.717967

+ Bakheit, A. M. O., Shaw, S., Barrett, L., Wood, J., Carrington, S., Griffiths, S., . . . Koutsi, F. (2007). A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. *Clinical Rehabilitation, 21*(10), 885-894. doi:10.1177/0269215507078486

Basso, A. (2005). How intensive/prolonged should an intensive/prolonged treatment be?. *Aphasiology, 19*(10–11), 975–984.

+ Basso, A., & Caporali, A. (2001). Aphasia therapy or the importance of being earnest. *Aphasiology, 15*(4), 307-332.

Berthier, M. L., & Pulvermuller, F. (2011). Neuroscience insights improve neurorehabilitation of poststroke aphasia. *Nature Reviews Neurology, 7*(2), 86-97. doi:10.1038/nrneurol.2010.201

* Best, W., Howard,D., Bruce, C., & Gatehouse, C. (1997) Cueing the Words: A Single Case Study of Treatments for Anomia. *Neuropsychological Rehabilitation, 7*(2), 105-141, DOI: 10.1080/096020197390211

* 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*(9), 2345-2357. doi:10.1016/j.cortex.2013.01.005

* Best, W., & Nickels, L. (2000) From theory to therapy in aphasia: Where are we now and where to next?. *Neuropsychological Rehabilitation, 10*(3), 231-247, DOI: 10.1080/096020100389147

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*(4), 987-992. doi:10.1161/01.str.0000062343.64383.d0

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A., ... & Ralph, M. A. L. (2012). Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial. *Bmj, 345*, e4407.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology, 13*(3), 236-249.

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). doi:10.1002/14651858.CD000425.pub4

+ Brindley, P., Copeland, M., Demain, C., & Martyn, P. (1989). A comparison of the speech of ten chronic Broca's aphasics following intensive and non-intensive periods of therapy. *Aphasiology, 3*(8), 695-707.

*+ Carpenter, J., & Cherney, L. R. (2016). Increasing aphasia treatment intensity in an acute inpatient rehabilitation programme: a feasibility study. *Aphasiology, 30*(5), 542-565. doi:10.1080/02687038.2015.1023695

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380. doi:10.1037/0033-2909.132.3.354

Cherney, L. R. (2012). Aphasia treatment: Intensity, dose parameters, and script training. *International Journal of Speech-Language Pathology, 14*(5), 424-431.

Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of Aphasia Therapy: Evidence and Efficacy. *Current Neurology and Neuroscience Reports, 11*(6), 560-569. doi:10.1007/s11910-011-0227-6

Cicerone, K. D., Dahlberg, C., Kalmar, K., Langenbahn, D. M., Malec, J. F., Bergquist, T. F., . . . Harrington, D. E. (2000). Evidence-based cognitive rehabilitation: recommendations for clinical practice. *Archives of physical medicine and rehabilitation, 81*(12), 1596-1615.

Code, C., Torney, A., Gildea-Howardine, E. & Willmes, (2010) Outcome of a one-month therapy intensive for chronic aphasia: variable individual responses. Seminars in Speech & Language, 31, 21-33

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology, 14*(2), 133-142.

* Conroy, P., Sage, K., & Ralph, M. L. (2009). Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders, 44*(6), 1036-1062. doi:10.3109/13682820802585975

* Conroy, P., Sotiropoulou Drosopoulou, C., Humphreys, G. F., Halai, A. D., & Lambon Ralph, M. A. (2018). Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia. *Brain, 141*(6), 1815-1827.

* Davis, C. H., & Harrington, G. (2006). Intensive semantic intervention in fluent aphasia: A pilot study with fMRI. *Aphasiology, 20*(1), 59-83.

de Aguiar, V., Bastiaanse, R., & Miceli, G. (2016). Improving Production of Treated and Untreated Verbs in Aphasia: A Meta Analysis. *Frontiers in Human Neuroscience, 10*. doi:10.3389/fnhum.2016.00468

* Deloche, G., Ferrand, I., Metz-Lutz, M. N., Dordain, M., Kremin, H., Mannequin, D., ... & Cardebat, D. (1992). Confrontation naming rehabilitation in aphasics: A computerised written technique. *Neuropsychological Rehabilitation, 2*(2), 117-124.

+ Denes, G., Perazzolo, C., Piani, A., & Piccione, F. (1996). Intensive versus regular speech therapy in global aphasia: A controlled study. *Aphasiology, 10*(4), 385-394. doi:10.1080/02687039608248418

+ Dignam, J., Copland, D., McKinnon, E., Burfein, P., O'Brien, K., Farrell, A., & Rodriguez, A. D. (2015). Intensive Versus Distributed Aphasia Therapy A Nonrandomized, Parallel-Group, Dosage-Controlled Study. *Stroke, 46*(8), 2206-2211. doi:10.1161/strokeaha.115.009522

Dignam, J. K., Rodriguez, A. D., & Copland, D. A. (2016). Evidence for Intensive Aphasia Therapy: Consideration of Theories From Neuroscience and Cognitive Psychology. *Pm&R, 8*(3), 254-267. doi:10.1016/j.pmrj.2015.06.010

* Fillingham, J., Sage, K., & Lambon Ralph, M. (2005) 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. DOI: 10.1080/02687030544000272

* Fillingham, J., Sage, K., & Lambon Ralph, M. (2006) The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation, 16*(2), 129-154. DOI: 10.1080/09602010443000254

* Francis, D. R., Clark, N., & Humphreys, G. W. (2002) Circumlocution- induced naming (CIN): A treatment for effecting generalisation in anomia? *Aphasiology, 16*(3), 243-259. DOI: 10.1080/02687040143000564

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8*(6), 528-538.

* 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*(1S), 438-453.

Greenhouse, J. B., Fromm, D., Iyengar, S., Dew, M. A., Holland, A., & Kass, R. (1990). The making of a meta-analysis: A quantitative review of the aphasia treatment literature. . In M. Straf & K. Wachter (Eds.), *The future of meta-analysis.* Beverly Hills, CA: Sage.

* Greenwald, M. L., Raymer, A. M., Richardson, M. E., & Rothi, L. J. G. (1995) Contrasting treatments for severe impairments of picture naming. *Neuropsychological Rehabilitation, 5*(1-2), 17-49. DOI: 10.1080/09602019508520174

* Harnish, S. M., Morgan, J., Lundine, J. P., Bauer, A., Singletary, F., Benjamin, M. L., ... & Crosson, B. (2014). Dosing of a cued picture-naming treatment for anomia. *American Journal of Speech-Language Pathology, 23*(2), S285-S299.

+ Harnish, S. M., Neils-Strunjas, J., Lamy, M., & Eliassen, J. (2008). Use of fMRI in the study of chronic aphasia recovery after therapy: A case study. *Topics in Stroke Rehabilitation, 15*(5), 468-483.

* Herbert, R., Webster, D., & Dyson, L. (2012). Effects of syntactic cueing therapy on picture naming and connected speech in acquired aphasia. *Neuropsychological Rehabilitation, 22*(4), 609-633.

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2001). Treatment of word retrieval in aphasia: Generalisation to conversational speech. *International journal of language & communication disorders, 36*, 13-18.

* Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology, 16*(10-11), 981-999. doi:10.1080/02687030244000509

+ Hinckley, J., & Carr, T. (2005). Comparing the outcomes of intensive and non-intensive context-based aphasia treatment. *Aphasiology, 19*(10-11), 965-974.

+ Hinckley, J. J., & Craig, H. K. (1998). Influence of rate of treatment on the naming abilities of adults with chronic aphasia. *Aphasiology, 12*(11), 989-1006. doi:10.1080/02687039808249465

Holland, A. L., Fromm, D. S., DeRuyter, F., & Stein, M. (1996). Treatment efficacy: aphasia. *Journal of Speech, Language, and Hearing Research, 39*(5), S27-S36.

Howard, D., & Papathanasiou, I. (2000). Cognitive neuropsychology and aphasia therapy: The case of word retrieval. *Acquired neurogenic communication disorders: A clinical perspective*, 76-99.

Howard, D., Best, N., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. Aphasiology. Advance online publication. doi:10.1080/02687038.2014.985884

Howard, D., Patterson, K., Franklin, S., Orchardlisle, V., & Morton, J. (1985). The facilitation of picture naming in aphasia. *Cognitive Neuropsychology, 2*(1), 49-80.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of consumer research, 30*(1), 138-149.

Jokel, R., Kielar, A., Anderson, N. D., Black, S. E., Rochon, E., Graham, S., . . . Tang-Wai, D. F. (2016). Behavioural and neuroimaging changes after naming therapy for semantic variant primary progressive aphasia. *Neuropsychologia, 89*, 191-216 doi:10.1016/j.neuropsychologia.2016.06.009

* Kirmess, M., & Maher, L. M. (2010). Constraint induced language therapy in early aphasia rehabilitation. *Aphasiology, 24*(6-8), 725-736.

* Kurland, J., Stanek, E. J., Stokes, P., Li, M. M., & Andrianopoulos, M. (2016). Intensive Language Action Therapy in Chronic Aphasia: A Randomized Clinical Trial Examining Guidance by Constraint. *American Journal of Speech-Language Pathology, 25*(4), S798-S812. doi:10.1044/2016_ajslp-15-0135

Laganaro, M., Di Pietro, M., & Schnider, A. (2006). Computerised treatment of anomia in acute aphasia: treatment intensity and training size. *Neuropsychological Rehabilitation, 16*(6), 630-640.

Leff, A. P., & Howard, D. (2012). Stroke: Has speech and language therapy been shown not to work?. *Nature Reviews Neurology, 8(*11), 600.

* Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology, 22*(9), 923-947.

* Lorenz, A., & Ziegler, W. (2009). Semantic vs. word-form specific techniques in anomia treatment: A multiple single-case study. *Journal of Neurolinguistics, 22*(6), 515-537. doi:10.1016/j.jneuroling.2009.05.003

MacLellan, C. L., Keough, M. B., Granter-Button, S., Chernenko, G. A., Butt, S., & Corbett, D. (2011). A critical threshold of rehabilitation involving brain-derived neurotrophic factor is required for poststroke recovery. *Neurorehabilitation and neural repair, 25*(8), 740-748.

+ Marcotte, K., Laird, L., Bitan, T., Meltzer, J. A., Graham, S. J., Leonard, C., & Rochon, E. (2018). Therapy-Induced Neuroplasticity in Chronic Aphasia After Phonological Component Analysis: A Matter of Intensity. *Frontiers in Neurology, 9*. Retrieved from <Go to ISI>://WOS:000429475800001. doi:10.3389/fneur.2018.00225

Marshall, R. C. (2008). The impact of intensity of aphasia therapy on recovery. *Stroke, 39*(2), E48-E48. doi:10.1161/strokeaha.107.504068

* Marshall, J., Best, W., Cocks, N., Cruice, M., Pring, T., Bulcock, G., ... & Caute, A. (2012). Gesture and naming therapy for people with severe aphasia: A group study. *Journal of Speech, Language, and Hearing Research.*

Marshall, R. C., Tompkins, C. A., & Phillips, D. S. (1982). Improvement in treated aphasia: examination of selected prognostic factors. *Folia Phoniatrica et Logopaedica, 34*(6), 305-315.

+ Martins, I. P., Leal, G., Fonseca, I., Farrajota, L., Aguiar, M., Fonseca, J., . . . Ferreira, J. J. (2013). A randomized, rater-blinded, parallel trial of intensive speech therapy in sub-acute post-stroke aphasia: The SP-I-R-IT study. *International Journal of Language & Communication Disorders, 48*(4), 421-431.

* McKissock, S., & Ward, J. (2007). Do errors matter? Errorless and errorful learning in anomic picture naming. *Neuropsychological Rehabilitation, 17*(3), 355-373. doi:10.1080/09602010600892113

+ Mozeiko, J., Coelho, C. A., & Myers, E. B. (2016). The role of intensity in constraint-induced language therapy for people with chronic aphasia. *Aphasiology, 30*(4), 339-363. doi:10.1080/02687038.2015.1070949

* Neumann, Y. (2018). A case series comparison of semantically focused vs. phonologically focused cued naming treatment in aphasia. *Clinical linguistics & phonetics, 32*(1), 1-27.

* Nickels, L. (2002a). Improving word finding: Practice makes (closer to) perfect? *Aphasiology, 16*(10-11), 1047-1060. doi:10.1080/02687040143000618

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

Nickels, L., & Best, W. (1996). Therapy for naming disorders .1. Principles, puzzles and progress. *Aphasiology, 10*(1), 21-47.

* Nickels, L., & Osborne, A. (2016). Constraint Induced Aphasia Therapy: Volunteer-led, unconstrained and less intense delivery can be effective. *Neurorehabilitation, 39*(1), 97-109. doi:10.3233/nre-161341

Nickels, L., Best, W. & Howard, D. (2015) Optimizing the ingredients for evaluation of the effects of intervention, *Aphasiology, 29*(5), 619-643, DOI: 10.1080/02687038.2014.1000613

* Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A. (2016). The impact of dose on naming accuracy with persons with aphasia. *Aphasiology*, 30(9), 983-1011. doi:10.1080/02687038.2015.1100705

Palmer R., Witts H., & Chater, T. (2018). What speech and language therapy do community dwelling stroke survivors with aphasia receive in the UK? *PLoS ONE*, 13(7): e0200096. https://doi. org/10.1371/journal.pone.0200096

* Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., ... & Harrison, M. (2019). Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. *The Lancet Neurology, 18*(9), 821-833.

+ Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke, 32*(7), 1621-1626.

*+ Ramsberger, G., & Marie, B. (2007). Self-administered cued naming therapy: A single-participant investigation of a computer-based therapy program replicated in four cases. *American Journal of Speech-Language Pathology, 16*(4), 343-358. doi:10.1044/1058-0360(2007/038)

* Raymer, A., & Kohen, F. (2006). Word-retrieval treatment in aphasia: Effects of sentence context. *Journal of Rehabilitation Research and Development, 43*(3), 367-377. doi:10.1682/jrrd.2005.01.0028

+ Raymer, A. M., Kohen, F. P., & Saffell, D. (2006). Computerised training for impairments of word comprehension and retrieval in aphasia. *Aphasiology*, 20(02-04), 257-268.

* Raymer, A., Singletary, F., Rodriguez, A., Ciampitti, M., Heilman, K., & Rothi, L. (2006). Gesture training effects for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society, 12*(6), 867-882.

* Raymer, A. M., Ciampitti, M., Holliway, B., Singletary, F., Blonder, L. X., Ketterson, T., . . . Rothi, L. J. G. (2007). Semantic-phonologic treatment for noun and verb retrieval impairments in aphasia. *Neuropsychological Rehabilitation, 17*(2), 244-270. doi:10.1080/09602010600814661

* Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L. (2008). Using semantic feature analysis to improve contextual discourse in adults with aphasia. *American Journal of Speech-Language Pathology, 17*(2), 161-172.

Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research, 41*(1), 172-187.

* Robson, J. O., Marshall, J., Pring, T. I. M., & Chiat, S. (1998). Phonological naming therapy in jargon aphasia: Positive but paradoxical effects. Journal of the International *Neuropsychological Society, 4*(6), 675-686.

* Romani, C., Thomas, L., Olson, A., & Lander, L. (2019). Playing a team game improves word production in poststroke aphasia. *Aphasiology*, 33(3) 253-288.

* Rose, M., Douglas, J., & Matyas, T. (2002) The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology, 16(*10-11), 1001-1030. DOI: 10.1080/02687030143000825

* Rose, M., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology, 22*(7-8), 691-706.

Rothkopf, E. Z., & Coke, E. U. (1966). Variations in phrasing, repetition intervals, and the recall of sentence material. *Journal of Memory and Language, 5*(1), 86.

+ Sage, K., Snell, C., & Ralph, M. A. L. (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation, 21*(1), 26-41. doi:10.1080/09602011.2010.528966

Schuchard, J., & Middleton, E. L. (2018). The roles of retrieval practice versus errorless learning in strengthening lexical access in aphasia. *Journal of Speech, Language, and Hearing Research, 61*(7), 1700-1717.

* Silkes, J. P., Dierkes, K. E., & Kendall, D. L. (2013). Masked repetition priming effects on naming in aphasia: A Phase I treatment study. *Aphasiology, 27*(4), 381-397. doi:10.1080/02687038.2012.745475

Smith 1981{published and unpublished data}Duffy, F. R. (1982), Speech therapy after stroke: a randomised controlled trial - an interim report. Demonstration Centres in Rehabilitation Newsletter, Volume 28.

Smith, D. S., Goldenberg, E., Ashburn, A., Kinsella, G., Sheikh, K., Brennan, P. J., et al. (1981) Remedial therapy after stroke: a randomised controlled trial. *BMJ, 282*(6263), 517–20.

Snell, C., Sage, K., & Ralph, M.A.L. (2010). How many words should we provide in anomia therapy? A meta-analysis and a case series study. *Aphasiology*, *24*, 9, 1064-1094. doi: 10.1080/02687030903372632

+ Stahl, B., Mohr, B., Buscher, V., Dreyer, F. R., Lucchese, G., & Pulvermuller, F. (2018). Efficacy of intensive aphasia therapy in patients with chronic stroke: a randomised controlled trial. *Journal of Neurology Neurosurgery and Psychiatry, 89*(6), 586-592. doi:10.1136/jnnp-2017-315962

* van Hees, S., Angwin, A., McMahon, K., & Copland, D. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia. *Neuropsychological Rehabilitation, 23*(1), 102-132. doi:10.1080/09602011.2012.726201

* 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*(3), 381-394. doi:10.1682/jrrd.2006.05.0038

Wambaugh, J. L., Linebaugh, C. W., Doyle, P. J., & Martinez, A. L. (2001). Effects of two cueing treatments on lexical retrieval in aphasic speakers with different levels of deficit. *Aphasiology, 15*(10-11), 933-950.

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

Whurr, R., Lorch, M. P., & Nye, C. (1992). A meta-analysis of studies carried out between 1946 and 1988 concerned with the efficacy of speech and language therapy treatment for aphasic patients. *International Journal of Language & Communication Disorders, 27*(1), 1-17.

Wisenburn, B., & Mahoney, K. (2009). A meta-analysis of word-finding treatments for aphasia. *Aphasiology, 23*(11), 1338-1352. doi:10.1080/02687030902732745

* Woolf, C., Caute, A., Haigh, Z., Galliers, J., Wilson, S., Kessie, A., . . . Marshall, J. (2016). A comparison of remote therapy, face to face therapy and an attention control intervention for people with aphasia: a quasi-randomised controlled feasibility study. *Clinical Rehabilitation, 30*(4), 359-373. doi:10.1177/0269215515582074

Zhang, J. Q., Yu, J. D., Bao, Y., Xie, Q., Xu, Y., Zhang, J. M., & Wang, P. (2017). Constraint-induced aphasia therapy in post-stroke aphasia rehabilitation: A systematic review and meta-analysis of randomized controlled trials. *Plos One, 12*(8). doi:10.1371/journal.pone.0183349.

Acknowledgement:
We would like to thank Andrew Olson for helpful comments on a version of this manuscript.

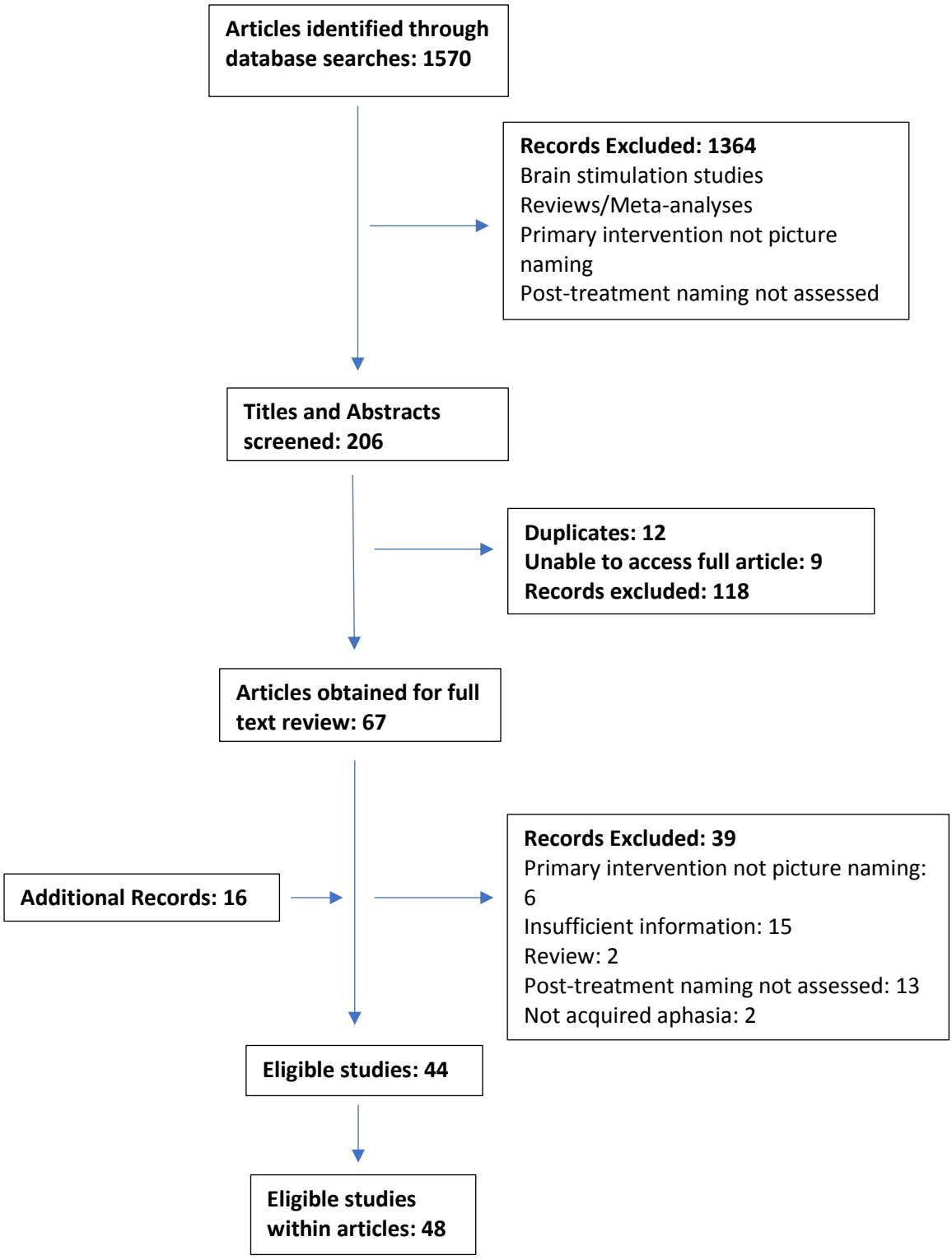

Fig.1. Flowchart for systematic review of picture naming studies

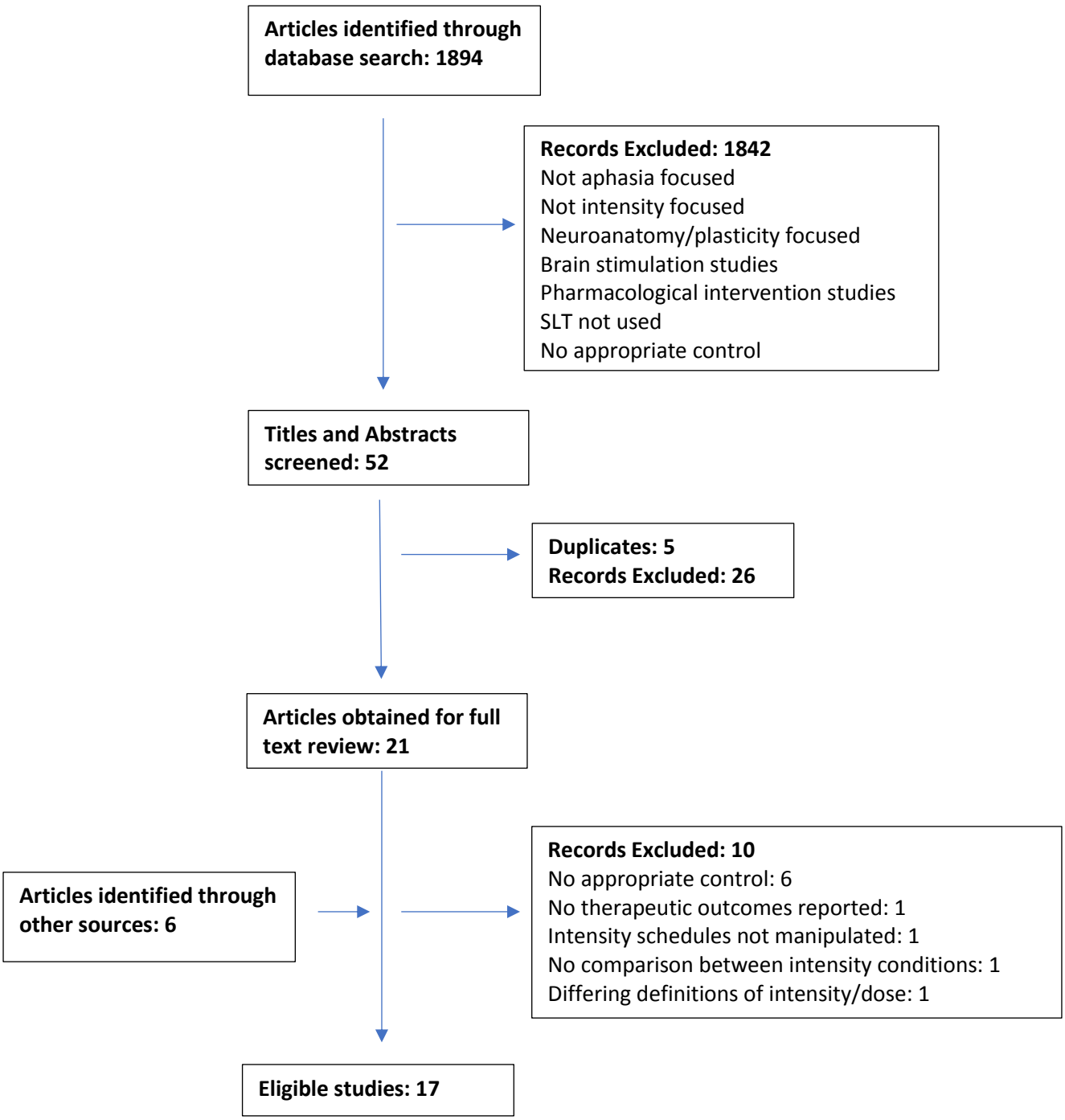

Fig.2. Flowchart for systematic review of effects of intensity and dosage in aphasia therapy.

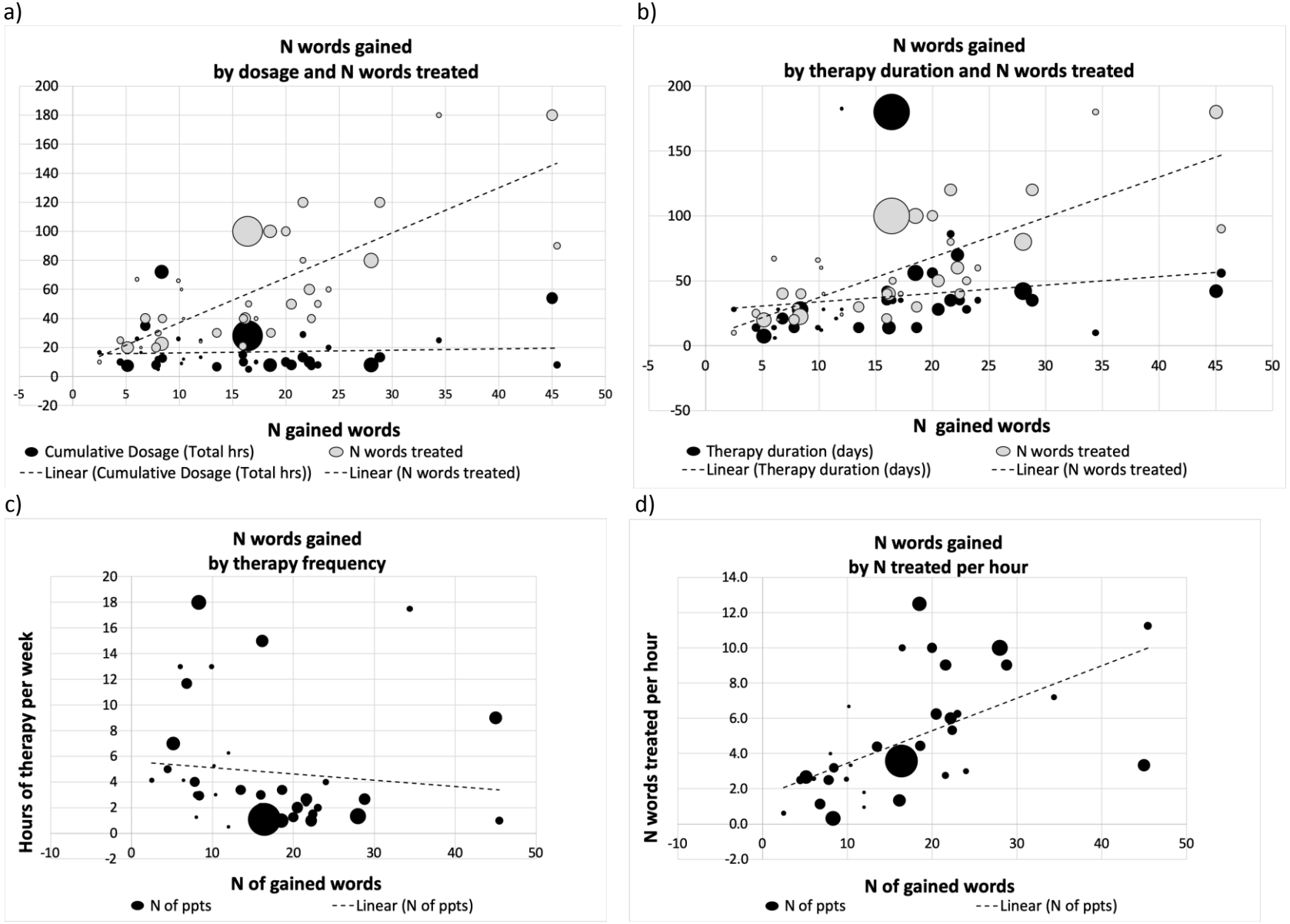

Fig.2 Regression graphs demonstrating: a) Effects of N words treated and cumulative dosage on N words gained; b) Effects of N words treated and therapy duration on N words gained; c) Effect of therapy frequency on N words gained; d)Effect of N words treated per hour on N words gained.

Table 1. Summary statistics of reviewed therapy studies focusing on picture naming; Total N comparisons =48, N of comparisons for each parameter reported in table.

**Where studies reported multiple baselines, mean baseline value was taken*

	N studies	Mean	sd	range	Median
N of participants	48	**8.1**	*12.1*	1-83	5.5
Dosage (Total hrs)	38	**16.8**	*13.8*	5-72	12
Therapy duration (days)	41	**37.8**	*36.9*	6-182.5	28
Frequency (average N hours per day)	38	**0.7**	*0.7*	0.07-2.6	0.4
Frequency (average N hours per week)	38	**4.8**	*4.8*	1-18	3
Overall N words treated	48	**55.3**	*38.4*	10-180	40
N words treated per hour	36	**4.7**	*3.3*	0.3-12.5	3.5
% gain treated words	48	**33.5**	*14.1*	9-76	33.9
Cohen's d	15	**1.8**	*1.3*	0.7-4.9	1.3
Serlin's d	18	**6**	*4.4*	1.8-20.5	4.3
% gain untreated words	37	**13.8**	*13.4*	-5-48	10
N words at baseline*	31	**20.1**	*16.3*	2-56.7	13.3
N words post-therapy	31	**37.6**	*23.7*	5-92	28
Gains N treated words	48	**16.4**	*9.7*	2.5-45.5	16

Table 2. Summary statistics of group studies and one-to-one therapy studies focusing on picture naming; Total N studies = 48, N group studies=7, N one-to-one studies=41.

	Group (N=7)			One-to-one (N=34)			T-tests		
	N	**Mean**	**sd**	N	**Mean**	**sd**	**t-value**	**df**	**p-value**
N of participants	7	**5.7**	*4.4*	41	**8.5**	*13*	-0.55	46	.58
Cumulative Dosage (total therapy hrs)	7	**26.1**	*14.5*	31	**14.7**	*13*	2.06	36	**.05**
Therapy duration (days)	7	**19.4**	*11.5*	34	**41.6**	*39.2*	-1.47	39	.15
Frequency (N hours per day)	7	**1.5**	*0.8*	31	**0.5**	*0.5*	4.43	36	**<.001****
Frequency (N hours per week)	7	**10.9**	*5.4*	31	**3.4**	*3.5*	4.54	36	**<.001****
Overall N words treated	7	**84**	*67.5*	41	**50.5**	*29.6*	2.23	46	**.03***
N words treated (per hour)	7	**3.1**	*1.9*	29	**5.0**	*3.5*	-1.4	34	.17
% gain treated words	7	**21.9**	*10.1*	41	**35.5**	*13.8*	-2.5	46	**.02***
Mean Cohen's d	3	**1.2**	*0.5*	12	**2.0**	*1.4*	-1.0	13	.34
Mean Serlin's d	3	**6.1**	*3.6*	15	**6.0**	*4.7*	.05	16	.96
% gain untreated words	6	**10.3**	*11.7*	31	**14.5**	*13.8*	-0.69	35	.49
N words at baseline* **	4	**30.5**	*25*	27	**18.6**	*14.7*	1.38	29	.18
N words post-therapy	4	**56.4**	*40.6*	27	**34.8**	*19.9*	1.78	29	.09
Gains N treated words	7	**17.7**	*15.8*	41	**16.2**	*8.5*	-0.07	45	.94

*. T-test is significant at the 0.05 level (2-tailed).

**. T-test is significant at the 0.01 level (2-tailed).

****Where studies reported multiple baselines, mean baseline value was taken*

Table 3. Pearson r correlations among therapy parameters.

		Cumulative Dosage (N hrs)	Therapy duration (N days)	Frequency (N hours per week)	N words treated	N words treated per hour
Cumulative Dosage (N hrs)	Correlation	1				
	Sig.					
	N	38				
Therapy duration (N days)	Correlation	.22	1			
	Sig.	.19				
	N	38	41			
Frequency (N hours per week)	Correlation	**.66****	**-.46****	1		
	Sig.	<.001	.004			
	N	38	38	38		
N words treated	Correlation	.18	**.48****	-.25	1	
	Sig.	.28	.002	.13		
	N	38	41	38	48	
N words treated per hour	Correlation	**-.58****	-.12	**-.52****	**.38***	1
	Sig.	<.001	.49	.001	.02	
	N	36	36	36	36	36

** = Correlation is significant at the 0.01 level (2-tailed)

* = Correlation is significant at the 0.05 level (2-tailed

Table 4. Pearson r correlations between therapy parameters and therapy outcomes.

		% Gains in treated words	**% Gains in untreated words**	**N words gained**
Cumulative Dosage (N hrs)	Correlation	-.21	-.04	-.02
	Sig.	.21	.84	.90
	N	38	30	38
Therapy duration (N days)	Correlation	**-.52****	-.19	.02
	Sig.	<.001	.30	.92
	N	41	32	41
Frequency (N hours per week)	Correlation	.10	.14	-.20
	Sig.	.54	.46	.24
	N	38	30	38
N words treated	Correlation	**-.55****	**-.39***	**.68****
	Sig.	<.001	.02	<.001
	N	48	37	48
N words treated per hour	Correlation	.01	-.25	**.51****
	Sig.	.97	.19	.001
	N	36	29	36

** = Correlation is significant at the 0.01 level (2-tailed).
* = Correlation is significant at the 0.05 level (2-tailed).

Table 5. Review of outcomes for studies contrasting High Dosage (**HD**) aphasia therapy vs Low Dosage (**LD**) aphasia therapy. Dosage established in terms of number of hours of therapy (hr=hour).

Study authors	Therapy Type	Design	N of ppts	Aphasia type	High Dosage			Low Dosage			Result	Measure	*Notes*
					Schedule	**Duration**	**Total hrs**	**Schedule**	**Duration**	**Total hrs**			
Carpenter & Cherney (2016)	Usual care + CILT=HD; Usual care=LD	Between-subjects	HD=6; LD=7	Acquired, acute aphasia	1 hr SLT 5-6 days per week + 1hr CILT 5 days per week	2 weeks	**20**	1 hr SLT, 5-6 days per week	2 weeks	**10**	**HD >LD**	Naming of treated and untreated words	*Larger effect size with HD for oral sentence production, reading and naming of furniture.*
Denes et al. (1996)	Conversational Therapy	Between-subjects	HD=8; LD=9	Acquired, global aphasia	94 -160 45/60min sessions (av. 130 sessions)	6 months	**130**	56 -70 45/60min sessions (av. 60 sessions)	6 months	**60**	**HD >LD**	AAT	*More improvement with HD on all subtests, but only significant on written subtest*
Hinckley & Craig (1998) Study 2	Individual, group and computer SLT	Within-subjects	25	Acquired, chronic aphasia	23 hrs per week	12 weeks	**276**	up to 3 hrs per week	6 weeks	**18**	**HD >LD**	BNT & Discourse content unit analysis	
Basso & Caporali (2001)	Conventional SLT	Matched pairs	HD=3; LD=3	Acquired aphasia	2/3 hrs per day, 7 days per week	14-40 months	**1064-3041**	1 hr per day, 5 days per week	14-40 months	**304-869**	**HD >LD**	Qualitative reports from clinicians	

Bakheit (2007)	Conventional SLT	Between-subjects	HD=51; LD=46	Post-stroke aphasia	5 hrs per week	12 weeks	**60**	2 hrs per week	12 weeks	**24**	**HD =LD**	WAB	*Higher withdrawal from HD/high frequency therapy*
Hinckley & Carr (2005)	Context-based	Between-subjects	HD=8; LD=5	Moderate - severe non-fluent post-stroke aphasia	20 hrs individual therapy + 5 hrs group therapy per week	-	-	4 hrs individual therapy per week	-	-	**HD =LD**	Varied verbal tasks	*Both groups improved; LD >HD on CADL-2; HD>LD on written task.*
Brindley et al. (1989)	Conventional' SLT	Within-subjects, cross-over	10	Chronic Broca's aphasia	5 hours, 5 days a week	12 weeks	**300**	1-2 hours per week	12 weeks	**12-24**	**HD >LD**	FCP & LARSP 200 word analysis	*Evidence of maintenance following HD*
Stahl et al. (2018)	ILAT	Between-subjects	HD=15; LD=15	PPA	4 hr session, 3 days per week	4 weeks	**48**	2 hr session, 3 days per week	4 weeks	**24**	**HD =LD**	AAT & ACT (discourse measure)	*Both groups improved on AAT after 2 weeks. Only LD group continued to improve during the 2nd phase*

Note:

CIAT: Constraint Induced Aphasia Therapy;
CILT: Constraint Induced Language Therapy;
BNT: Boston Naming Test;
AAT: Aachen Aphasia Test
WAB: Western Aphasia Battery

CADL: Communicative Abilities in Daily Living
LARSP: Language Assessment Remediation and Screening Procedure
ACT: Action Communication Test
FCP: Functional Communication Profile

Table 6. Review of outcomes for studies contrasting High frequency (**Massed**) aphasia therapy vs Low frequency (**Distributed**) aphasia therapy. M=massed practice; D= Distributed practice; sess=session; hr=hour.

					High frequency (Massed)			**Low Frequency (Distributed)**					
Study authors	**Therapy Type**	**Design**	**N of ppts**	**Aphasia type**	**Schedule**	**Duration**	**Total hrs**	**Schedule**	**Duration**	**Total hrs**	**Result**	**Measure**	**Notes**
Pulvermuller et al. (2001)	CILT vs. SLT	Between-subjects	M=10; D=7	Chronic, post-stroke aphasia	-	**3-5 weeks**	20-54	-	**10 days**	23-33	**M>D**	AAT and CAL;	*CILT (massed) compared to SLT (distributed); CILT group improved overall; SLT group only on naming subtest*
Marcotte et al. (2018)	Phonological Components	Between-subjects	M=1; D=1	Post-stroke Broca's aphasia	10 3-hrs sess	**2.5 weeks**	30	30 1-hr sessions,	**10 weeks**	30	**M>D**	Naming of treated words	*Both participants maintained gains at 4 and 8 weeks follow-ups*
Raymer, Kohen & Saffell (2006)	Computerised training	Within-subjects, cross-over	5	Post-stroke aphasia	3-4 times per week	**3-4 weeks**	12	1-2 times per week	**6-12 weeks**	12	**M>D**	Picture naming	*Better generalisation following massed schedule. No difference on comprehension. Differences not maintained over time*
Martins et al. (2013)	Multimodal Stimulation	Between-subjects	M= 15; D=15	Post-stroke aphasia	2 hrs per day, 5 days per week	**10 weeks**	100	2 hrs per week	**50 weeks**	100	**M>D (trend)**	Aphasia quotient scores	*Patients stratified for severity and clinical variables*
Mozeiko et al. (2016)	CILT	Between-subjects	M=4; D=4	Chronic, post-stroke aphasia	3-hrs sess, 5 days per week	**2 weeks**	30	1-hr sess, 3 days per week	**10 weeks**	30	**M>D (trend)**	WAB-AQ and CADL	*M>D for 4 participants and D>M for 2 participants*

Ramsberger & Marie (2007)	Computer-based anomia therapy	Within-subjects, cross-over	4	Post-stroke aphasia	5 45-60-mins sess per week	**3-4 weeks**	15-20	2 45-60-mins sess per week	**7.5-10 weeks**	15-20	**M>D**	Naming of treated words	*D >M in one participant*
Harnish et al. (2008)	A variety of word production tasks	Within-subjects, cross-over	1	Chronic conductio n aphasia		**2 weeks**	15	15 sessions	**7.5 weeks**	15	**M>D**	BNT and WAB	*More improvement following massed schedule.*
Dignam et al. (2015)	Impairment, functional, computer & group-based therapy	Between-subjects	M= 16; D=18	Chronic, post-stroke aphasia	16 hrs per week	**3 weeks**	48	6 hrs per week	**8 weeks**	48	**D>M**	BNT	*Differences immediately post-therapy and at 1 month follow-up. No differences in functional communication and confidence*
Sage, Snell & Lambon-Ralph (2011)	Cued picture naming	Within-subjects, cross-over	8	Chronic, post-stroke aphasia	10 sess	**2 weeks**	10	10 sess	**5 weeks**	10	**D>M**	Naming of treated words	*No differences immediately after therapy, but D better at 1 month follow-up.*

Note:

CIAT: Constraint Induced Aphasia Therapy;
CILT: Constraint Induced Language Therapy;
BNT: Boston Naming Test;
AAT: Aachen Aphasia Test
WAB (AQ): Western Aphasia Battery (Aphasia Quotient)
CADL: Communicative Abilities in Daily Living
CAL: Communicative Activity Log

Appendix A - List of studies included in the review of picture naming therapy studies with therapy parameters

Study authors	Therapy Type	Group vs. One to one	N of ppts	Aphasia type	Control condition	Multiple Baseline	Cumulative Dosage (Total hrs)	Duration (days)	Frequency (N hrs per week)	N words treated
Attard, Rose & Lanyon (2013) A	CIATplus	Group	2	Severe Broca's aphasia	No	Yes	26	14	13	66
Attard, Rose & Lanyon (2013) B	M-MAT	Group	2	Severe Broca's aphasia	No	Yes	26	14	13	67
Carpenter & Cherney (2016)	CIAT	Group	5	Acute aphasia	Usual care condition	Yes	10	14	5	25
Kirmess & Maher (2010)	CIAT	Group	3	Non-fluent	No	No	25	10	18	180
Romani, et al. (2018)	Game Therapy	Group	12	Severe to moderate	No	Yes	12	28	3	30
Nickels & Osborne (2016)	CIAT	Group	4	Various	Non-word reading task	Yes	54	42	9	180
Kurland et al. (2016)	CIAT (ILAT)	Group (pairs)	12	Various (fluent and non-fluent)	No	Yes	30	14	15	40

Adrian, Gonzalez & Buiza (2003)	Confrontational	One to one	1	Fluent aphasia, anomia	No	No	9	12	5	60
Best & Nickels (2000)	Confrontational	One to one	4	Anomia	No	Yes	5	35	1	50
Best et al. (2013)	Confrontational	One to one	16	Various	No	Yes	8	56	1	100
Best, Howard, Bruce & Gatehouse (1997)	Confrontational	One to one	1	Anomia	No	Yes	-	21	-	36
Conroy et al. (2018)	Confrontational	One to one	20	Chronic post-stroke aphasia	No	Yes	8	42	1	80
Conroy, Sage & Lambon-Ralph (2009)	Confrontational	One to one	7	Various (fluent and non-fluent)	No	No	7.5	35	2	40
Davis &Harrington (2006)	Confrontational	One to one	1	Fluent aphasia	No	-	25	28	6	24
Deloche et al. (1992)	Confrontational	One to one	18	Various (fluent and non-fluent)	Orthographic condition	No	-	-	-	80

Fillingham, Sage & Lambon Ralph (2005)	Confrontational	One to one	7	Various	No	Yes	10	35	2	40
Filllingham, Sage & Lambon Ralph (2006)	Confrontational	One to one	11	Various	No	Yes	10	70	1	60
Francis, Clark & Humphreys (2002)	Confrontational	One to one	1	Anomia	Reading & synonym judgements	Yes	5	28	1	20
Gravier et al. (2018)	Semantic Feature Analysis	One to One	17	Chronic Aphasia	Untrained words	Yes	72	28	18	22.5
Greenwald, Raymer, Richardson & Rothi (1995) Treatment 2	Confrontational	One to one	2	Anomic aphasia	No	Yes	16.5	28	4	10

Harnish et al. (2013)	Confrontational	One to one	8	Various (fluent and non-fluent)	No	Yes	8	14	4	20
Herbert, Webster & Dyson (2012)	Confrontational	One to one	6	2 with non-fluent, agrammatic aphasia, 4 with fluent aphasia	No	Yes	-	-	-	40
Hickin et al. (2002)	Confrontational	One to one	8	Anomia - range of aphasia severities	No	Yes	10	56	1	100
Leonard, Rochon & Laird (2008)	Phonological Components Analysis	One to one	7	Various	No	Yes	15	35	3	21
Lorenz & Ziegler (2009) A	Confrontational	One to one	10	Various	Unrelated control tasks	Yes	13.3	35	3	120
Lorenz & Ziegler (2009) B	Confrontational	One to one	10	Various	Unrelated control tasks	Yes	13.3	35	3	120
Marshall et al. (2012)	Confrontational	One to one	14	Severe aphasia	No	Yes	7.5	7.5	7	20
McKissock & Ward (2007)	Confrontational	One to one	5	Anomia	No	No	8	56	1	90

Neumann (2018)	Semantic Feature Analysis & Phonological Components Analysis	One to one	4	Various	Untrained words	Yes	-	-	-	40
Nickels (2002)	Confrontational	One to one	1	Fluent aphasia, anomia	Non-verbal written task	Yes	-	6	-	34
Off et al. (2016)	Confrontational	One to one	7	Chronic aphasia	No	Yes	12.5	30	3	40
Palmer et al. (2019)	CAT – Step by Step	One to one	83	Anomia, mixed	Usual care + puzzle	No	28	180	1	100
Ramsberger & Marie (2007)	Confrontational	One to one	4	Various	No	Yes	29	86	2	80
Raymer & Kohen (2006)	Confrontational	One to one	2	Various (fluent and non-fluent)	No	Yes	10	35	2	40
Raymer, Singletary et al. (2006)	Confrontational	One to one	9	Various (fluent and non-fluent)	No	Yes	-	42	-	40
Raymer et al. (2007)	Confrontational	One to one	8	Various, moderate to severe	No	Yes	35	21	12	40
Rider et al. (2008)	Confrontational	One to one	3	Non-fluent	No	Yes	-	-	-	30

Robson, Marshall, Pring & Chiat (1998)	Confrontational	One to one	1	Jargon aphasia	No	No	13.3	182.5	1	24
Rose & Sussmilch (2008)	Confrontational	One to one	3	Broca's	No	Yes	20	35	4	60
Rose, Douglas & Matyas (2002)	Confrontational	One to one	1	Mild conduction aphasia	D2 Test of Attention	Yes	-	-	-	60
Schuchard & Middleton (2018)	Confrontational	One to one	10	Chronic aphasia	No	No	-	-	-	88
Silkes, Dierkes & Kendall (2013)	Confrontational	One to one	1	Moderate-severe, non-fluent aphasia	Non-linguistic control task	Yes	16.5	28	4	20
Van Hees et al. (2013) A	Confrontational	One to one	8	-	No	Yes	6.75	14	3	30
Van Hees et al. (2013) B	Confrontational	One to one	8	-	No	Yes	6.75	14	3	30
Wambaugh & Ferguson (2007)	Confrontational	One to one	1	Moderate anomic aphasia	No	Yes	12	28	3	40

Wambaugh, Mauszycki & Wright (2014)	Confrontational	One to one	4	Various: conduction, anomic and broca's	No	Yes	-	-	-	20
Woolf et al. (2016) A	Confrontational	One to one	10	-	Attention control condition (normal conversation)	Yes	8	28	2	50
Woolf et al. (2016) B	Confrontational	One to one	5	-	Attention control condition (normal conversation)	Yes	8	28	2	50
One to one studies	**Confrontational**	**Mean**	**8.5**	-	-	-	**14.7**	**41.6**	**3.4**	**50.5**
		SD	13.0	-	-	-	13.0	39.2	3.5	29.6
Group studies	**CIAT, CIAT + Usual Care, M-MAT**	**Mean**	**5.7**	-	-	-	**26.1**	**19.4**	**10.8**	**84.0**
		SD	4.4	-	-	-	14.5	11.5	5.3	67.5
All studies		**Mean**	**8.1**	-	-	-	**16.8**	**37.8**	**4.8**	**55.4**

SD	12.1	-	-	-	13.8	36.9	4.8	38.4
Median	5.5	-	-	-	12.0	28.0	3.0	40.0

Note:
CIAT: Constraint Induced Aphasia Therapy;
M-MAT: Multi-Modal Aphasia Therapy;
ILAT: Intensive Language Action Therapy

Appendix B - List of studies included in the review of picture naming therapy studies with study outcomes

Study authors	% gains treated words	% gains untreated words	Serlin's d	Cohen's d	Gains N treated words	Efficacy - N of treated words gained /N of hours	Generalisation: other tasks	Improv standardized assessments
Attard, Rose & Lanyon (2013) A	15	-5	3.7	-	9.9	0.38	WAB-AQ only at follow up	No
Attard, Rose & Lanyon (2013) B	9	10	4.4	-	6.0	0.23	WAB-AQ only at follow up	No
Carpenter & Cherney (2016)	17.8	23.4	-	-	4.5	0.45	BNT; BDAE	Some
Kirmess & Maher (2010)	19.1	-	-	0.7	34.4	1.38	PALPA, TROG	Yes
Romani, Thomas, Olson & Landers (2018)	26.7	10	-	-	8.0	0.67	CAT; BNT (No) Narrative speech (Yes)	Some
Nickels & Osborne (2016)	25	0	-	1.1	45.0	0.83	BNT; TROG	No
Kurland (2016)	40.4	23.4	10.3	1.7	16.2	0.54	BNT; BDAE; PICA	No
Adrian, Gonzalez & Buiza (2003)	17	33	-	-	10	1.13		-
Best & Nickels (2000)	33	29	2.7	2.5	16.5	3.30		-
Best et al. (2013)	18.5	2.5	-	0.9	18.5	2.31		-

Best, Howard, Bruce & Gatehouse (1997)	32	0	-	-	11.5	-		-
Conroy et al. (2018)	35	1	-	-	28	3.50		-
Conroy, Sage & Lambon-Ralph (2009)	56	0	-	0.7	22.4	2.99		-
Davis, Harrington & Baynes (2006)	50	48	-	-	12	0.48	BNT; WAB repetition, naming & comp.	Yes
Deloche et al. (1992)	16	5	-	-	12.8	-		-
Fillingham, Sage & Lambon Ralph (2005)	40	-	20.5	3.9	16	1.60		-
Filllingham, Sage & Lambon Ralph (2006)	37	-	11.7	2	22.2	2.22		-
Francis, Clark & Humphreys (2002)	40	16	-	-	8	1.60		-
Gravier et al. (2018)	37	11	-	-	8.3	0.12		-
Greenwald, Raymer, Richardson & Rothi (1995) Treatment 2	25	0	1.8	1.3	2.5	0.15		-
Harnish et al. (2013)	39	0.6	-	-	7.8	0.98		-
Herbert, Webster & Dyson (2012)	27.5	10	-	1.7	11	-		-

Hickin et al. (2002)	20	-	-	1	20	2.00		-
Leonard, Rochon & Laird (2008)	76	-	-	4.9	16	1.06	Some improvement on PNT (4 participants)	Some
Lorenz & Ziegler (2009) A	24	1	-	-	28.8	2.17		-
Lorenz & Ziegler (2009) B	18	10	-	-	21.6	1.62		-
Marshall et al. (2012)	25.7	7.8	-	-	5.1	0.69		-
McKissock & Ward (2007)	50.5	-	-	-	45.5	5.68		-
Neumann (2018)	47.7	46.5	4	-	19	-		-
Nickels (2002)	18	-	-	-	6	-		-
Off et al. (2016)	21	1	6.41	0.7	8.4	0.67		-
Palmer et al. (2019)	16.4	-	-	-	16.4	0.59	Treated words in conversation, naming untreated words, quality of life	No
Ramsberger & Marie (2007)	27	-	5.31	1.0	21.6	0.74		-
Raymer & Kohen (2006)	43	13	-	-	17.2	1.72		-

Raymer, Singletary et al. (2006)	40	7	5.66	-	16.0	-	WAB-AQ (Yes) BNT; ANT (No)	Some
Raymer et al. (2007)	17	3	4.29	-	6.8	0.19	WAB; ANT; BNT (Yes) CETI; FOQ (No)	Some
Rider et al. (2008)	48	-	-	-	14.4	-		-
Robson, Marshall, Pring & Chiat (1998)	50	39	-	-	12	0.90	PALPA naming & repetition	Some
Rose & Sussmilch (2008)	40	10	4.29	-	24	1.20	2 ppt: VAST, OANB, pict. description and conversation; all 3 ppt better on LCQ	Some
Rose, Douglas & Matyas (2002)	45	33.5	4.8	3.6	27	-	Reduced naming errors in conversation	Some
Schuchard & Middleton (2018)	34.7	-	-	-	30.5	-		-
Silkes, Dierkes & Kendall (2013)	32	21	3.42	-	6.4	0.39	WAB-AQ; Discourse measures (Yes) BNT (No)	Some
Van Hees et al. (2013) A	62	17	4.22	-	18.6	2.76	BNT (Yes for 1 ppt)	Some
Van Hees et al. (2013) B	45	17	2.76	-	13.5	2.00	BNT (Yes for 1 ppt)	Some

Wambaugh & Ferguson (2007)		26	17	-	-	10.4	0.87		-
Wambaugh, Mauszycki & Wright (2014)		40	10	8.1	-	8	-	No improvement in production of CIUs in narrative discourse	No
Woolf et al. (2016) A		41	21	-	-	20.5	2.56		-
Woolf et al. (2016) B		46	18	-	-	23	2.88		-
One to one studies	**Mean**	**21.9**	**10.3**	**6.1**	**1.2**	**17.7**	**0.6**		-
	SD	10.1	11.7	3.6	0.5	15.8	0.4		-
Group studies	**Mean**	**33.5**	**13.8**	**6.0**	**1.8**	**16.4**	**1.5**		-
	SD	14.1	13.4	4.4	1.3	9.6	1.2		-
All studies	**Mean**	33.9	10.0	4.3	1.3	16.0	1.1		-
	SD	**21.9**	**10.3**	**6.1**	**1.2**	**17.7**	**0.6**		-
	Median	10.1	11.7	3.6	0.5	15.8	0.4		-

Note:
ANT: Action Naming Test;
BDAE: Boston Diagnostic Aphasia Examination;
BNT: Boston Naming Test;
CAT: Comprehensive Aphasia Test;
LCQ: La Trobe Communication Questionnaire;
OANB: Object and Action Naming Battery;
PALPA: Psycholinguistic Assessment of Language Processing in Aphasia;
PICA: Porch Index of Communicative Ability;
TROG: Test for Reception of Grammar;
VAST: Verb and Sentence Test;
WAB(-AQ): Western Aphasia Battery (Aphasia Quotient)