A feasibility randomised controlled trial of elaborated Semantic Feature Analysis delivered in the virtual world, EVA Park

Volume 1

by
Niamh Devane

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Language and Communication Science,
School of Health and Psychological Sciences,
City, University of London

April 2023

Table of Contents

List of tables:

List of figures:

List of images:

List of boxes:

Appendices (Available in Volume 2)

COVID-19 Impact Statement

This statement is provided for the aid and benefit of future readers to summarise the impact of the COVID-19 pandemic on the scope, methodology, and research activity associated with this thesis. The academic standards for a research degree awarded by City, University of London and for which this thesis is submitted remain the same regardless of this context.

Title of the research project: VESFA | Virtual Elaborated Semantic Feature Analysis

1. Summary of how the research project, scope or methodology has been revised because of COVID-19 restrictions

As a result of the COVID-19 pandemic all data collection was moved online. 36 participants were due to receive three testing visits in their homes. This data collection was carried out via the videoconferencing technology, Zoom. This resulted in an additional inclusion criterion; that participants had someone in their home who could support if there were technical difficulties. This in turn reduced the population from which the study could recruit e.g., participants who fitted the language criteria would not be contacted if they lived alone and were not tech confident. It also added to the burden of carrying out the tests, additional materials were needed to administer the tests and further student training was required to equip them for testing online.

2. Summary of how research activity and/or data collection was impacted because of COVID-19 restrictions, and how any initially planned activity would have fitted within the thesis narrative

Research activity and data collection was impacted in two ways.

I. The COVID-19 restrictions closed schools. Parents were provided with activities to deliver the curriculum to children. As a mother with three school aged children (in March ‘20 they were aged 5yrs, 9yrs, and 11yrs) the time available for research activity was reduced. During this time, I was working on: Systematic review of the literature, ethical approval, recruitment, methods and background chapters, therapy manuals, recruitment and training of testers.
II. Data collection was planned through the networks of community stroke groups. In these groups the target population meet in person weekly in community spaces. These groups were closed in March 2020 and have yet to reopen. This significantly

reduced the recruitment channels available to me. I have been able to visit groups that transferred to online meetings. However, not all groups made this transfer. As a result, there has been less access to the target population and recruitment has been slow.

3. Summary of actions or decisions taken to mitigate for the impact of data collection or research activity that was prevented by COVID-19

Online data collection: All testing sessions delivered via zoom
Online recruitment: All communication regarding recruitment to the study is via email or videoconferencing technologies. This will filter out all those from the target population who do not feel confident using these or do not have support. I email the staff from charity organisations and community groups who support people with aphasia at each recruitment point. I periodically share recruitment information via twitter. Recruitment is slow and there is less access to the target population.
Delayed upgrade exam: Initially planned for Dec/Jan this is now planned for March/April

4. Summary of how any planned work might have changed the thesis narrative, including new research questions that have arisen from adjusting the scope of the research project

There are additional feasibility questions. Specifically, I will ask about the acceptability of online testing.

Date of statement: 11.03.2021

Acknowledgments

Thank you to my supervisors Katerina Hilari, Jane Marshall and Stephanie Wilson. You have provided a lovely balance of support and encouragement to stand alone.

Thank you to colleagues in the School of Health and Psychological Sciences for their support of this work. Thank you to Shashi Hirani for his time and assistance with statistical analysis. Thank you to Aparna Ramachandran for all her thought and time on the systematic review. Thank you to Nick Behn his work on the systematic review and guidance on treatment fidelity. Thank you to Rachel Holland and Madeline Cruice for their thoughtful upgrade examination and the time given to talk through specific theory afterwards. Thank you to Sarah Northcott and Becky Moss for their conversations and resources on qualitative research methodology.

I am grateful to the School of Health Sciences doctoral studentship at City, University of London, whose funding made this project possible.

Thank you to the 34 participants with aphasia who gave their time to VESFA project and the 12 participants who gave their time to the focus groups. It was such a pleasure to work with you all.

Thank you to the trial advisory group Adrian Cumberworth, Chris Greenhough, Barry McIllroy and Paul Stocken. I was very grateful to have your input on decisions throughout the trial. You had such belief in the project, I felt very supported with you on the team.

Thank you to all those who volunteered on this project. I am very grateful to the many students who gave their time to test participants. Thank you Leema Miah, Sarah El-Fodil, Eavan Goodman, Ella Thorne, Kiera Thomas, Ann Mason, Daniella Stead, Julia Gainsford, Charlotte Gribbin and Victor Piotto. Thank you to the placement students who rated the fidelity of the therapy videos: Sereana Chima, Lisa Gracias, Bekki Graham, Sobana Rathasethupathy and Shannen Turner.

Thank you to the students who worked with me on their dissertation projects and strengthened elements of the work. Thank you to Deirdre Staunton, Shannon Buxton, Jennifer Whiddett and Chelsie Fox for their work on the qualitative study of meaningful topics of conversation for people with aphasia. Thank you to David Budworth, Cemaliye Burdane, Marie-Louise Watson, Emma Carter, Oisin Twomey Brenner and Jusveen Khurana for their work testing participants. Thank you to Sofia Mazzoleni and Aiysha Yusaf for their work assessing treatment fidelity. Thank you to Beth Glover, Natalie Douglas and Niamh Goodwin for their work on the post therapy interviews.

For taking time to talk to me about semantic feature analysis, I thank Professor Mary Boyle. For giving time to explain experience in virtual reality, I thank Professor Paul Cairns.

Finally, thank you to Alex for your steadfast belief. Thank you to Leo, Aidan and Flynn. You may not realise it, but you inspire me to be better.

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

Background: Aphasia affects 350,00 people in the UK with negative consequences to a person's social connectedness and quality of life. Speech and language therapy interventions aim to mitigate these consequences, but more therapies are needed that address the language impairment and its impact on conversations and quality of life.
The use of Virtual Reality (VR) is emerging in aphasia rehabilitation with potential for providing ecological validity for language interventions. This study aimed to develop and test the feasibility and acceptability of an intervention, Virtual Elaborated Semantic Feature Analysis (VESFA), that addresses language and conversation through the virtual world, EVA Park.

Methods: The intervention was developed through mapping out the research evidence, articulating the underlying theories and a programme model, Public Involvement workshops with key user groups, and a qualitative study to investigate meaningful therapy targets.
The intervention was tested in a feasibility randomised controlled trial comparing VESFA + usual care with a usual care control. People with word finding difficulties as a result of chronic aphasia were recruited from the community. Feasibility parameters were the percentage of eligible participants who consented to the trial, the percentage of participants available at follow up, rates of cancelled sessions, rates of missing data, the acceptability of the research procedures to participants and the acceptability of the intervention to the participants. Clinical outcomes provided preliminary findings of the impact on language, communication and quality of life.

Results: VESFA is a theory and evidence based intervention. It was found to be feasible against pre-specified feasibility criteria. 91% of eligible participants consented to the trial. 85% of participants were available at follow up. Less than 6% of sessions were cancelled. Participant interviews showed that, despite a testing burden, both the trial protocol and the intervention were acceptable. Preliminary clinical outcomes suggest a future definitive trial of VESFA may show benefit to the retrieval of treated and untreated words and improved quality of life.

Discussion: Good feasibility and acceptability outcomes suggest that this treatment would be a good candidate for a definitive trial. Delivery in the virtual world EVA Park provided opportunity for situated conversations. This was valued by participants but not captured by the outcome measures. A future trial could use an alternative measure for conversations. Preliminary findings suggest that the VESFA intervention shows promise to improve outcomes that are not consistently found in word retrieval therapies.

Chapter 1 | BACKGROUND

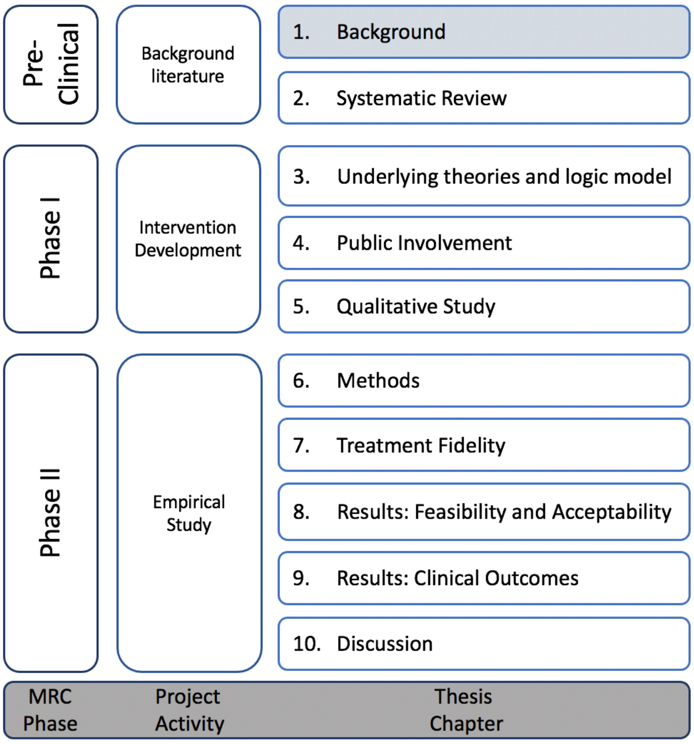

Figure 1.1: The thesis in a figure

This thesis explores the feasibility of delivering a specific speech and language therapy intervention for aphasia via a virtual reality platform. The intervention sought to address both the linguistic and social consequences of aphasia following stroke by making use of virtual reality technology.

This background chapter describes stroke, aphasia and their impact within the framework of the World Health Organisation (WHO) International Classification of Functioning and Disability (World Health Organization, 2001), and existing approaches to remediate that impact. It describes the emergence of virtual reality as a rehabilitation tool and its potential for optimising generalisation of learnt language functions for people with aphasia.

This study is presented using the Medical Research Council (MRC) framework for developing complex interventions at the end of Chapter 1. Chapter 2 systematically describes the literature on the use of virtual reality in the rehabilitation of aphasia and cognitive communication difficulties. Chapters 3, 4 and 5 describe the intervention development. Chapter 3 outlines the underlying theory and proposes a programme model of the VESFA intervention, Chapter 4 outlines the public involvement[1] activities carried out to inform the intervention and Chapter 5 reports on a qualitative study that informed the treatment targets. Chapter 6 outlines the methods of this phase II feasibility trial. Chapter 7 describes the methods used to assess treatment fidelity in the phase II feasibility trial and the fidelity results. Chapter 8 reports the feasibility and acceptability results. Chapter 9 reports the clinical outcomes and Chapter 10 discusses the findings and concludes the thesis.

Stroke

Someone experiences a stroke every 5 minutes, leaving 1.2 million stroke survivors in the UK (Stroke Association, 2018). A stroke is an injury to the brain that occurs when blood flow is interrupted or reduced, preventing the brain tissue from getting oxygen and nutrients (Mayo Clinic, 2021). A quarter of stroke survivors will leave hospital with moderate-severe disability (Royal College of Physicians Sentinel Stroke National Audit Programme, 2015) indicating a need for community-based rehabilitation programmes.

Aphasia

One of those disabilities is aphasia. A quarter of people who survive a stroke will experience persistent aphasia (Ali et al., 2015). There are approximately 350,000 people in the UK living with aphasia (Stroke Association, 2018). It is defined as 'a communication disability due to an acquired impairment of language modalities caused by focal brain damage' (Berg et al., 2020, p.7). Aphasia has negative social and emotional consequences: the interrupted ability to communicate masks competence, threatens identity, affects relationships and leads to reduced social networks (Berg et al., 2020, Northcott & Hilari, 2011; Kagan, 1995; Shadden, 2005). Low public awareness of the disability causes additional barriers to participating in society (Code, 2020). Social isolation is linked to premature death, and lower wellbeing (Cornwell & Waite, 2009; House, 2001; Brummett et al., 2001). For these reasons, aphasia is

[1] There are a number of terms used to describe the involvement of end users in research development: Patient and Public Involvement (PPI), user involvement, co-design, co-production. In this thesis the term in the National Institute for Health and Care Research (NIHR) guidance is used; Public Involvement (PI) (Kok, 2018)

a public health concern (Simmons-Mackie & Cherney, 2018). There is a need for therapies that address both the aphasia and the impact of aphasia on people's lives.

International Classification of Function, Disability and Health

The International Classification of Function, Disability and Health (ICF, WHO 2001) provides a common framework for talking about health conditions and their impact on people's lives. The ICF plots how an impairment of body structures and functions limits activities and opportunities in society, and how our environments and personal belief systems influence those activities and opportunities, see Figure 1.2.

Using the ICF, aphasia can be framed as damage to the brain functions and structures (impairment) that affects how a person functions in everyday activities and their societal roles (activities and participation). Personal factors are the internal context from which a person with aphasia approaches activity and participation, including age, gender, education, beliefs, coping strategies, emotional state, personality, self-esteem. Environmental factors are the external context, the 'entire background' (Mitre & Shakespeare, 2019, p.337) in which a person with aphasia approaches activity and participation including the attitudes of a family, a workplace, or systems, services and policies of governments. A stroke survivor's personal factors and environmental setting can positively or negatively impact recovery. Health related quality of life (HRQoL) is a connected and important construct that is not directly addressed by the ICF.

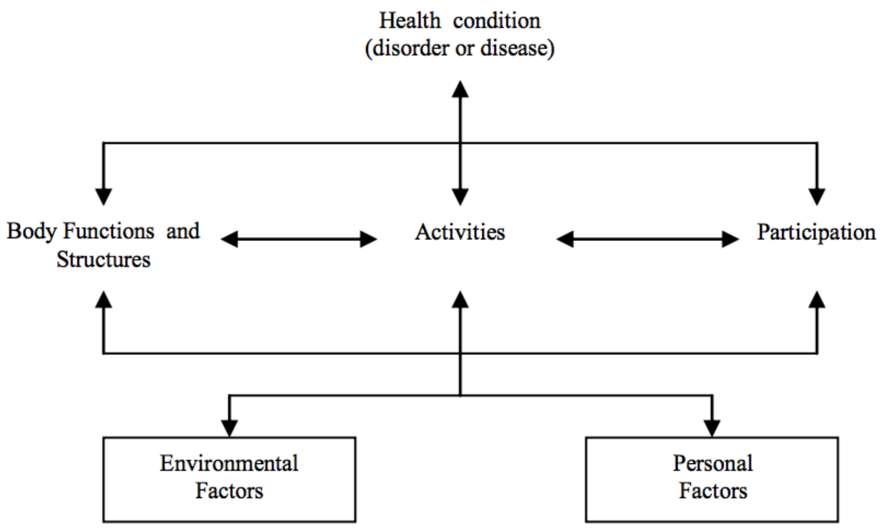

Figure 1.2: ICF framework

The World Health Organisation (WHO) defines Quality Of Life (QOL) as an individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns (WHO, 2012). HRQoL is defined as the factors of QOL that relate to health (Karimi & Brazier, 2016). QOL is particularly at risk in people with aphasia. People with aphasia report worse QOL than stroke survivors without aphasia (Hilari, 2011). When quality of life was compared across a number of health conditions, including cancer and progressive neurological conditions such as Parkinson's Disease, aphasia was found to have the lowest QOL ratings (Lam & Wodchis, 2010). QOL is an important consideration in aphasia interventions. The Living with Aphasia: Framework for Outcome Measurement (A-FROM) model pulls together both the construct of QOL and the domains of the ICF to create a framework for outcomes specifically for people with aphasia (Kagan et al., 2008). Thus, this thesis and the VESFA trial explore outcomes in terms of the ICF and QOL.

Speech and language interventions for aphasia can target all areas of the ICF. Worrall (2019) proposes that Speech and Language Therapists (SLTs) should aim for a 'smorgasbord' of treatments; intervention packages that target language function, activity, participation, and personal factors to be available to people with aphasia to choose what meets their needs at that time. Research exploring the opinions of people with aphasia has found a wish for therapy to target all areas of the ICF but with a particular interest in goals influencing everyday activities and participation (Wallace, et al., 2017; Worrall et al., 2011). There is some debate about how to achieve change at the level of activity and participation. For example, interventions that target language function (the body function and structures domain in the ICF, see Figure 1.2) might bring about improvements in functional communication (the activity domain) (Carragher, et al., 2012). Although language impairment and communication activity are correlated, linguistic skills (impairment) and functional (activity) communication skills have been shown to respond differently to treatment (Meier et al., 2017). This makes the case for treatments that explicitly target the different domains of the ICF.

The current study addresses both impairment, by aiming to increase the word retrieval capabilities of participants, and activity by practicing the use of those words in functional conversations in simulated environments. The subsequent sections of this chapter will introduce word retrieval difficulties in aphasia, outline the theory underlying word retrieval processes, discuss word retrieval breakdowns and therapies and the potential for virtual world technologies to support communication activities.

Word Finding Difficulties

A common linguistic consequence of aphasia is difficulty retrieving words (Goodglass & Wingfield, 1997). People with aphasia describe knowing what they want to say but being unable to think of the word. This can range in severity from the occasional tip-of-the-tongue experience to a repertoire of only a few single words. Often people with aphasia have good underlying sematic information (Jefferies & Lambon Ralf, 2006) and can be cued into remembering the word with semantic context or a sound reference. Our current understanding of the process of retrieving words, from semantic knowledge to word production, is explained by models of word retrieval from cognitive neuropsychology.

Models of word retrieval

Models of word retrieval provide the theoretical explanation for word retrieval difficulties in aphasia. Cognitive neuropsychology conceptualised a model of word retrieval in the 1980s (Patterson & Shewell, 1987). This box and arrow system describes the retrieval of the word meaning (semantics) from the cognitive system, the subsequent retrieval of the relevant phonological form (phonological output lexicon) and a buffer where that form is held before being produced, see Figure 1.3. The model was developed from key case studies that demonstrated disassociated processing errors that indicated separate systems (Kay, & Ellis, 1987; Howard & Orchard-lisle, 1984; Caplan, Vanier & Baker, 1986). The model formed the basis of some SLT assessments and interventions, such as the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) (Kay, Lesser & Coltheart, 1996).

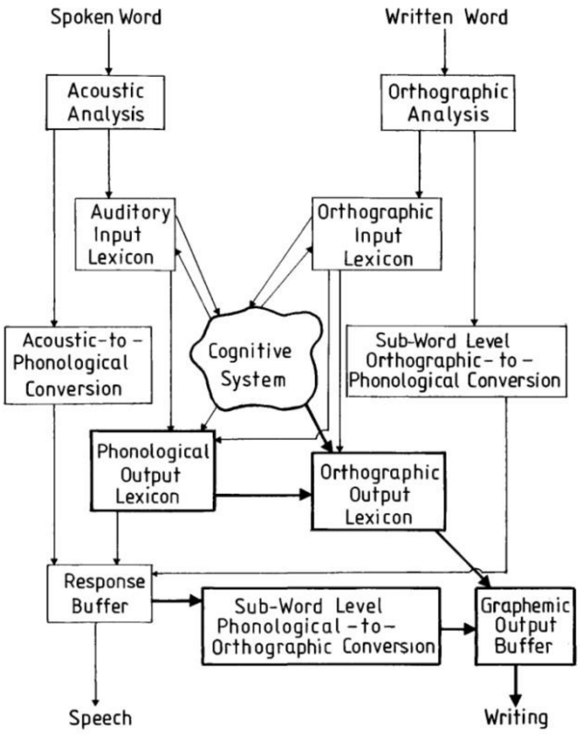

Figure 1.3: Cognitive neuropsychological model of language processing (Patterson & Shewell 1987)

An alternative model describes a dynamic, interactive activation of language (Dell, 1986). This model describes how words are primed and retrieved through spreading activation networks (Foygel & Dell, 2000). In stage one (lexical selection), the semantics, for example a domestic animal is mapped to the word 'cat'. In stage two (phonological encoding) the sound pattern for that word is retrieved e.g., the specific phonemes for the specific form required; 'cat' vs 'cats', see Figure 1.4.

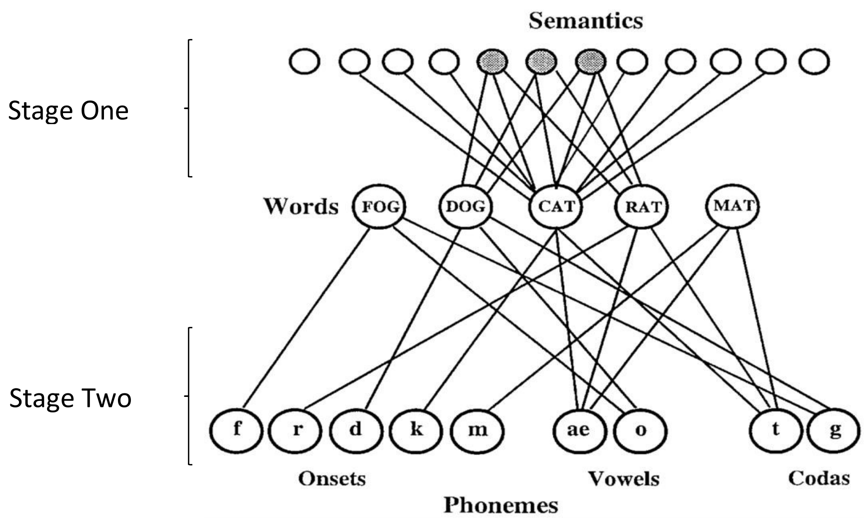

Figure 1.4: Model of word retrieval (Dell et al. 1997)

Words that share semantic features with the target (such as dog and rat) receive some activation from the semantic level and, in turn, relay activation to their corresponding phonemes. This causes spreading activation, affecting words other than the target. When the system is working normally the target receives the highest activation and is therefore selected. However, even in healthy language processing this does not always occur, leading to occasional lexical selection errors. This cascading activation from the target to relating concepts also flows back up the system. Activated phonemes feedback to all word forms that share that phoneme and again, this can be a source of error production (Dell & Reich, 1981). This spread of activation to the target and connected concepts decays over time (Dell et al., 1997). More recently retrieved concepts will be easier to retrieve again due to some latent activation, however the longer time passes the less this will be. Evidence for this priming in spreading activation theory can be seen in repetition tasks and error patterns. If you have just named a picture, and you name it again, the second repetition is quicker (Oppenheim, Dell & Schwartz, 2010). This is called repetition priming and demonstrates a learning effect from the first naming task. However, this phenomenon also has a negative effect, it will interfere with the subsequent naming of a semantically related item and a semantic error may occur (Oppenheim, Dell & Schwartz, 2010). The concept of spreading activation is central to current theories of word production (Wilshire, 2008).

Models of single word processing have been highly influential in the development of word retrieval therapies in aphasia. Firstly, they help to identify the level of impairment an individual is experiencing. For example, the production of semantic errors (labelling a picture of a cat ‘dog’) and difficulties comprehending differences between words that are semantically related would indicate a likely impairment at ‘stage one’ in sematic processing, Figure 1.4. Conversely, a tendency for phonological errors (labelling a cat ‘cot’) would suggest a likely impairment in ‘stage two’, phonological encoding. Secondly, these models have influenced the content of therapy. Understanding the level of impairment supports the choice of intervention. Interventions for word retrieval difficulties tend to focus on semantic or phonological processing and/or their connections, with some interventions explicitly imitating the spreading activation process by naming a target and its semantic or phonologically related items (see Semantic Feature Analysis and Phonological Components Analysis described below).

Word retrieval therapies

Interventions to remediate word finding deficits target the semantic and phonological systems and the interactions between them. Some treatment approaches focus on improving links between meaning and the word (Boyle, 2017; Kiran & Bassetto, 2008). Other approaches focus on the phonology of lexical items such as Phonological Components Analysis (Leonard, Rochon & Laird, 2008) and Phonomotor Therapy (Kendall et al., 2015; Silkes et al., 2020)) and some explicitly target both (Doesborgh et al., 2004). It has been argued that all word finding treatments where a person retrieves and produces a word entail both semantic and phonological processing to some degree (Nickels, 2002).

Evidence for the effectiveness of word retrieval therapies is predominantly single case level evidence. A recent umbrella review of aphasia intervention studies found 18 single case studies for lexical interventions compared to four randomised control trials (Dipper et al., 2021). As a result, the meta-analyses of word finding treatments are based on single case designs. Word retrieval therapies have been shown to improve the retrieval of treated words i.e., words that are practiced in therapy tasks (Nickels & Best, 1996; Nickels, 2002; Wisenburn & Mahoney, 2009; Sze et al., 2020). Effect sizes vary across different approaches. A meta-analysis of word finding treatments found that effect sizes are largest for the three months following therapy and then decrease with time (Wisenburn & Mahoney, 2009). Although effect sizes are largest for treated words, there is an effect that is greater than spontaneous recovery for untreated items, results are complicated by high variation with large standard deviations (Wisenburn & Mahoney, 2009). Thus, there is evidence of some generalisation beyond words specifically targeted in treatment. To explore what the components of successful word finding therapy might be, Sze and colleagues (Sze et al., 2020) used a different methodology, random forest, to explore predictors of success from a large dataset identified through a systematic search of recent studies. They extracted therapy ingredients using a proposed framework called RITA: Regime, Items, Techniques, and Application of techniques with their Adjuncts. The research team looked for predictors of success at three weeks post and ‘four weeks and more’ post therapy (p.17). They found orthographic cues provided by SLT predicted good outcomes in post treatment and follow up, as did total times an item was named. Semantic focussed treatments had the most evidence for generalisation to untrained items (Sze et al., 2020). This finding that semantic tasks promote generalisation has been found across a number of studies (Wisenburn & Mahoney, 2009; Sze et al., 2020).

Semantic Feature Analysis

One semantic approach to word finding therapy is Semantic Feature Analysis (SFA) (Ylvisaker & Szekeres, 1985; Boyle, 2004). SFA has a clear theoretical basis in the spreading activation model of word retrieval (Foygel & Dell, 2000) and a suggestion that repeated methodical stimulation of the semantic system in SFA improves lexical retrieval overall, more than treated and semantically related words (Boyle, 2004). This approach is based on the understanding that the meanings of words are organised in networks of associated items. Lemon, orange, apple etc. are in a network of fruit. Lemons, bananas, daffodils and rubber ducks are in a network of things that are yellow. Semantic errors in naming occur when someone retrieves the wrong item from the right network. As outlined by the theories above, retrieving a word spreads activation to its related semantic networks; in retrieving 'lemon', fruit, yellow and sour are also partially activated.

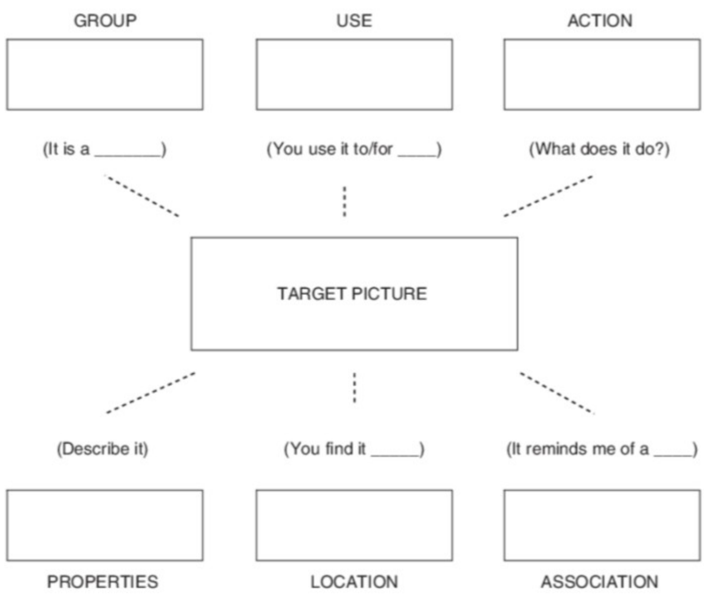

Figure 1.5: SFA chart (Boyle, 2004)

SFA follows a structured process to stimulate semantic information of target items. Using a visual chart (Figure 1.5) clients are encouraged to identify the target's supraordinate category, its use, its action, its properties, its location, and any personal associations.

Proponents of SFA argue that identifying the features of the target increases spreading activation through the lexical system to phonology and output (Boyle, 2004). The therapy is also thought to promote spreading activation to related concepts. Therefore, generalisation to untreated words is also hypothesised (Efstratiadou et al., 2018).

The evidence for SFA interventions is also predominantly single case studies. A systematic review (Efstratiadou et al., 2018) found only single case studies but in recent years two randomised controlled trials (RCT) of SFA have been carried out (Kendall et al., 2019; Efstratiadou et al., 2019). The systematic review of SFA interventions reported that naming improved for 81% participants, with 40% participants showing generalisation to untreated words (Efstratiadou et al., 2018). The two recent RCTs explored the ways in which SFA can impact communication beyond the treated words. Kendall and colleagues (Kendall et al., 2019) compared SFA with Phonomotor Treatment (PMT). They were interested to know which treatment would yield most gains to untreated items and lead to maintenance of those gains at three months post treatment. Thirty participants received SFA, and 28 participants received PMT. All participants received 56-60 hours of treatment over 6-7 weeks. Both interventions led to improved naming of related untreated items, that is words that were not targeted in therapy but shared features with treated words; semantically related items improved in the semantic therapy and phonologically related items improved in the phonological therapy. However unrelated untreated items did not improve. Despite no direct training of activity domain tasks, the authors explored the ecological validity of the interventions through the use of the self-reported Comprehensive Aphasia Test (CAT) Disability Questionnaire (Swinburn, Porter & Howard, 2004) and the Functional Caregiver Questionnaire (Glueckauf et al., 2003). Both treatments yielded a small change on these measures, for example both arms achieved an effect size of 0.36 on the CAT disability questionnaire (Kendall et al., 2019).

Efstratiadou and colleagues (Efstratiadou et al., 2019) investigated Elaborated SFA (ESFA). ESFA extended the SFA protocol to include using the target word in a phrase or sentence. Participants receiving ESFA (n=26) were compared to a waiting list control (n=12). This study also compared two different approaches of delivering ESFA, i.e., a combination of individual and group therapy (n=14) compared to individual ESFA (n=22). Outcome measures captured naming, functional communication, wellbeing and quality of life. Results demonstrated a significant improvement in naming for the ESFA group compared with the waiting list control group with a large effect size (η_p^2=.21) and encouraging improvements on the measure of quality of life. Authors found a significant difference between the ESFA group and the waitlist control on measures of naming. No significant difference was found between therapy type (individual therapy compared with a combination of individual and group therapy), even with respect to functional communication. This is an interesting finding. The authors had expected an increased benefit of group delivery. However, the group they described was highly structured, much like an individual ESFA session with others present, with no opportunity for naturalistic conversations. Could opportunities for conversation with ESFA improve communication in the domains of activity and participation? The next section will explore other studies that have used groups with SFA to improve communication activities.

Semantic Feature Analysis and the ICF

The goal of aphasia intervention is to impact change across multiple domains of the ICF. SFA can improve word retrieval. This demonstrates improvement at the body function domain of the ICF (see Figure 1.2). Some authors have explored whether SFA can also improve communication in the activity domain of the ICF. This has been explored by delivering SFA within discourse tasks employed in groups (Falconer & Antonucci, 2012; Antonucci, 2009). Rather than work on specific items, participants engaged in connected speech tasks and used SFA strategies when they encountered a word finding difficulty. Antonucci reported on six single case studies in the two papers using a multiple baseline as a control (Falconer & Antonucci, 2012; Antonucci, 2009). Change in word retrieval in discourse was measured using the Nicholas and Brookshire protocol (1993). All participants improved in their word retrieval in discourse, with one participant, P1, approaching a clinically meaningful change (4.7 where a change of 5 is meaningful) on the Western Aphasia Battery Aphasia Quotient score (Falconer & Antonucci, 2012). These small studies provide an encouraging indication that using SFA in discourse can improve informativeness or efficiency of discourse. More recently Boyle (in preparation) has investigated a protocol of four sessions of SFA naming followed by 8 sessions of SFA in discourse with three PWA. She reports encouraging findings in improved Western Aphasia Battery Aphasia Quotient scores and reduced word finding behaviours (see German, 1991).

These studies are promising. They demonstrate that discourse can be more informative or efficient following SFA treatment and that there is potential to change quality of life with SFA interventions. However, there is a need for further studies that attempt to treat and measure change in the activity domain of the ICF.

Generalisation

Many agree that aphasia therapy should target change in the activity domain (e.g., Brady et al., 2012). In other words, we need to improve everyday communication. This might be achieved by remediating the language impairment, i.e., therapy may address a specific aspect of language processing with the expectation that effects will generalise to communication activities (Carragher et al., 2012). This describes 'stimulus generalisation', where the target is used in a different context, and differs from 'response generalisation', where untrained language emerges (Thompson, 1989). However, there is a stronger theoretical argument that communication activities should be targeted to expect a change within communication activity (Thompson, 1989; Webster et al., 2015). Webster et al. (2015) propose a framework to conceptualise generalisation. They argue that generalisation can be expected within linguistic levels e.g., word level, sentence level or discourse level. For example, outcomes can be predicted at the discourse level if discourse has been targeted, but a word level treatment should not expect to change discourse. This has led to research groups developing multi-level treatments (Dipper, et al., 2021; Webster, Whitworth, & Morris, 2015; Dipper et al., 2020). In these examples, treatment targets word retrieval and sentence processing skills and encourages application of those skills in a narrative communication activity. In some cases, generalisation is also promoted by changing the therapy format. For example, group therapy may be employed to offer a medium in which participants can call upon trained language skills for the purpose of communication (Elman, 2006).

Group conversation treatments

Group treatments for aphasia can benefit both communication and the psychosocial consequences of aphasia (Hoover, DeDe, & Maas, 2021). Group interventions are used for different purposes (Lanyon, Rose, & Worrall, 2013). Language interventions can be delivered in a group to improve the ecological validity (Antonucci, 2009; Efstratiadou et al., 2019; Falconer & Antonucci, 2012; Pulvermüller et al., 2001). In these language focussed groups, activities can look similar to individual sessions but delivered in a group (Efstratiadou et al., 2019). Community aphasia groups (CAGs) can be used to build community and social networks and renegotiate identity following stroke (Ross et al., 2006; Shadden & Agan, 2004; Vickers, 2010). Activities in CAGs most often focus on conversations or leisure activities e.g., music or photography (Pettigrove et al., 2021). An exploration of possible mechanisms that appear to improve well-being in CAGs identified 1) opportunities for support, 2) opportunities for learning, and 3) opportunities for communication (Attard et al., 2015).

The VESFA intervention evaluated in this project included groups for both ecologically valid use of language skills and to provide social support and shared experiences. As a result, the VESFA group sessions gave time to both language games and conversations that share experiences. The VESFA individual sessions worked on topic-based words that can support the group conversations. Participants were invited to share stories from their life experiences on the target topics (travel, food and drink, daily routines, nature and gardening). The topic-based conversations were situated in a simulated environment that was congruous to the topic e.g., talk about travel was conducted on a simulated ship, gardening was discussed in a virtual greenhouse.

The current study addressed lexical skills in one-to-one therapy sessions and the use of those skills in group conversations. We hypothesised that clinical outcomes would indicate improved naming of treated words and generalisation to untreated but related words, as a result of spreading activation in semantic networks. In line with some previous studies, SFA will be augmented with conversation practice. However, a novel addition will deliver SFA via a virtual reality platform. It will therefore explore whether VR can promote generalisation of words into real world contexts through the use of conversations in simulated environments.

Virtual reality

'Virtual' describes a simulated experience that is 'almost real'; something we perceive to exist and yet know it has no physical properties beyond the screen (Girvan, 2018). Virtual worlds are 'shared, simulated spaces which are inhabited and shaped by their inhabitants who are represented as avatars. These avatars mediate our experience of this space as we move, interact with objects and interact with others, with whom we construct a shared understanding of the world at that time' (Girvan, 2018 p.13). There are a variety of terms to describe the use of 3D technology and how a representation of the user, the avatar, interacts with that technology, see Table 1.1. Virtual reality (VR) is typically used to describe experiences that involve a first-person viewpoint with additional technology to create a sense of immersion such as a headset, see Image 1.1. 'Virtual worlds' describe a 3D environment on a 2D computer screen. The camera is just behind head of the user's avatar and can leave the avatar to zoom in on objects, see Image 1.2.

Image 1.1: Oculus Rift headset (left) and Image 1.2: the view of EVA Park, where the camera is behind the head of the user's avatar

In healthcare, the term 'virtual reality' appears to be used more flexibly to include the use of any screen-based 3D technology. Some use 'semi-immersive' to denote that no headset is used (Giachero et al., 2020). However, others have used 'immersive' when talking about 3D images on a 2D screen (Konnerup, 2015).

Technology terms	
Virtual World	3D world represented on a 2D screen. Users are represented by avatars. They experience a third-person viewpoint, usually just behind their avatar's head. The camera view can also be moved away from the avatar. Example: Second Life, EVA Park
Multi-User Virtual Environment (MUVE)	A virtual world with many users represented by avatars. Virtual worlds are multi-user, therefore MUVW and virtual worlds synonymous terms. Example: Second Life, EVA Park
Virtual Reality (VR)	Virtual Reality is a 3D system which allows users to experience simulated environments (such as virtual worlds) from the first-person viewpoint e.g., looking through the user's own eyes. VR involves a headset that takes over the user's field of vision to replace it with a virtual experience. Example: Oculus Rift, PlayStation VR
Massively Multiplayer Online Role-Playing Game (MMORPG)	A role-play multi-user online video game. Players take on the role of a character. The world is online, so persists when a player logs out. Example: World of Warcraft
Virtual Environment (VE)	Virtual spaces where users are represented but do not inhabit an avatar with agency. This is a supraordinate term to virtual world. Example: Facebook
Virtual Learning Environment (VLE)	A virtual environment specifically designed for education. Example: Moodle
Augmented Reality (AR)	Augments but does not replace the user's field of vision. For example, projecting an overlay onto what you already see. Example: Pokémon Go
Mixed Reality	Brings together real world and virtual elements. Virtual objects are integrated into and respond to the real world. Example: Magic Leap
User experience terms	
Immersion	The experience of loss of awareness of the physical world. More likely engrossed/engaged (see Cairns 2004)
Presence	The experience of being there in a shared space.
Co-presence or Social Presence	The experience of being there in a shared space with others with whom the user can interact

Table 1.1: Virtual Reality Terminology (Girvan, 2018)

‘Immersion’ and ‘presence’ are terms that describe the user experience not the technological specification (Cairns, Cox, & Nordin, 2014). Immersion is a cognitive experience related to attention, a feeling of being in the virtual world. It describes being engrossed in the activity. This term can be used in tasks unrelated to virtual technologies i.e., you can be immersed in a book. ‘Presence’ describes that feeling that you are there in the simulated space and ‘co-presence’ or ‘social presence’ that you feel you are there together with others. The degree to which you commit your attention (immersion) and feel that the simulated space is real (presence) appear to be important elements of the therapeutic benefit of virtual reality.

Therapeutic benefit of virtual worlds

Virtual reality aims to give the user the experience of ‘being there’. The degree to which a person feels they are ‘there’ has been defined and measured via self-rating scales about responses such as empathy, suspense and interest in the simulation (Jennett et al., 2008). The mechanisms for change in therapy are the theory-driven reasons why a therapeutic benefit occurs (Kazdin, 2006). In the field of VR for therapeutic benefit the mechanisms of change are not yet well defined but there is a suggestion that presence and immersion contribute. For example, virtual reality for exposure-based intervention for phobias seems to work because it gives the user the feeling of being there (Maples-Keller et al., 2017). The simulated environment appears to be ‘real enough’ to stimulate the fear associated with the phobia.

Virtual reality and stroke

In stroke rehabilitation there is growing empirical evidence for the use of virtual reality for upper limb rehabilitation (Kiper et al., 2018) and cognitive rehabilitation (Faria et al., 2016) with a number of systematic reviews of intervention studies. The meta-analyses indicate that VR interventions produce therapy gains that are significantly larger than conventional upper limb rehabilitation (Mekbib et al., 2020; Karamians et al., 2020). There are also reviews of the use of technology in aphasia rehabilitation (Repetto et al., 2020) and one systematic review and meta-analysis of the outcomes of VR treatments in aphasia rehabilitation (Cao et al., 2021).

Virtual reality and aphasia

The literature on the use of VR for aphasia rehabilitation will be systematically reviewed in Chapter 2, with respect to how VR is being applied and the outcomes achieved. VR systems in aphasia are briefly introduced here.

Two studies have investigated the use of the virtual world, Second Life, with people with aphasia. Konnerup (2015) argued that virtual reality has the potential to support psychosocial adjustment as well as functional communication. She was building on the work of the Stanford Virtual Human Interaction Lab which investigated the effect of an altered self-representation on behaviour, coined the Proteus Effect (Yee & Bailenson, 2009). The Proteus Effect describes how positive behaviour in a virtual environment influences positive behaviour in real life. For example, people who are obese but have avatars that engage in physical exercise are more likely to engage in physical activity in real life (Dean et al., 2009). Konnerup investigated renegotiation of identity in aphasia through the use of avatars in Second Life. She concluded that the social, embodied interactions in Second Life impacted positively on the participants' sense of self (Konnerup, 2015). Galliers and Wilson carried out an exploratory study into the accessibility of Second Life for two people with aphasia (Galliers & Wilson, 2013). They found that 'off the shelf' VR technologies were not accessible to those with language difficulties due to text-based menus and drop-down structures. They reported the participants enjoyed their VR experiences and gave advice for adjustments that would increase accessibility.

More commonly, research teams use bespoke VR technologies in aphasia rehabilitation. 'AphasiaScripts', developed by a team in Chicago, used a bespoke virtual therapist to deliver scripts therapy to people with aphasia (Cherney et al., 2008; Cherney, Halper, & Kaye, 2011: Manheim, Halper, & Cherney 2009). The virtual therapist was programmed to produce naturalistic speech with the correct corresponding lip shapes. A 'real' SLT developed a functional script together with the client. The script was typed into the program and recorded by the SLT. Script training was then carried out without the therapist but with stepped support from the virtual therapist. This team also designed a virtual therapist as part of a self-management tool, Web-ORLA, to support their face-to-face reading program, ORLA (Cherney et al., 2021; Cherney & van Vuuren, 2012).

In Barcelona, a stroke rehabilitation team developed a Rehabilitation Gaming System (RGS; Grechuta et al. 2016; 2017; 2019; 2020). Most of the team's research has focussed on upper limb rehabilitation. For a review see Cameirão et al. (2009). The RGS also delivered a language rehabilitation program, which has been reported for two participants with Broca's aphasia (Grechuta, 2016). The therapy utilised Intense Action Language Training (ILAT) with a virtual room, virtual objects and two people with aphasia. The virtual objects were requested by one user and passed by the other. Kinect gloves allow the movement of the virtual arms to mimic the real reach for objects. Participants sat opposite each other in the room so that they could talk to each other with the therapist present.

In Sicily, a team has developed a virtual reality rehabilitation system on a tablet (VRRS-Tablet; Maresca et al., 2019). This aimed to provide a self-management tool. This system integrates different rehabilitation modules (motor, cognitive, linguistic, and orthopaedic). The linguistic module was described as mainly 2D tasks such as speaking or writing the name of an object from a list of features, and semantic link tasks. They described '3D virtual scenarios' and 'immersive objects through a magnetic localization sensor generally positioned on the hand (which allows a detection of the final effector's 3D position)' (p.4) although how the 3D world is utilised is not described. More recently this team described an intervention using a virtual rehabilitation platform BTS-Nirvana (https://www.btsbioengineering.com/nirvana/) that works on discrete cognitive and language functions (De Luca et al., 2021).

The Chicago, Barcelona and Sicily teams use virtual technologies to recreate language tasks that remediate the language impairment e.g., barrier games and word finding treatments. They do not describe multi-user experiences. In the Barcelona study, the user sees one other user in the virtual space, but both users are in the same physical room (Grechuta et al., 2016). In contrast, the platform used in this study, EVA Park, was developed to create a virtual world for simulated, situated, functional communication practice. For example, participants walk into a virtual restaurant, choose a table and practice ordering dinner with an avatar of the speech and language therapist acting as the waiter. The EVA Park studies (Marshall et al. 2016; 2018; 2020) have used the virtual world to deliver synchronous therapy remotely where the clients and therapists are in different locations.

EVA Park

EVA Park is a multi-user virtual world. To our knowledge it remains the only virtual world developed with and for people with aphasia. EVA Park was co-designed at City, University of London in 2012 with a multi-disciplinary team of paid consultants with aphasia, a speech and language therapist and Human Computer Interaction (HCI) researchers (Wilson et al. 2015). EVA Park is a virtual island with a range of spaces for functional conversations; houses, a town square (hairdressers, clinic, café, pizza restaurant), an island bar, and green spaces (for example lounge chairs overlooking a lake). It runs in a 3D browser via an internet connection allowing for remote delivery. Users are represented by avatars and can speak to other users in real time via the computer microphone or a headset. There is also an instant messaging function for written messages.

EVA Park was developed with the aim of creating a simulated real world environment that would allow people with aphasia to practice real world communication tasks in a safe space before attempting them in the 'wild'. The first study published by Marshall et al. (2016) described this work. Twenty people with aphasia received 5 weeks of treatment that involved meeting a communication support worker for an hour every day (total 20hrs). They set functional goals and practiced conversations in the simulated environment, for example asking for a hair cut in the barber shop. This quasi-randomised group study had a waitlist design, 10 participants received the intervention immediately and 10 received the intervention after a 5-week delay. Outcomes demonstrated good compliance with the intervention with no participants lost to follow up and participants receiving, on average, 88% of the intended dose. The intervention demonstrated a significant improvement on a measure of functional communication, the Communication Activities of Daily Living -2nd Edition (Holland, Frattali & Fromm, 1999). Nested interview and human computer interaction studies investigating the experience of receiving therapy via a virtual world demonstrated the intervention was acceptable and enjoyable (Amaya et al., 2018; Galliers et al., 2017). Although the numbers were small, these findings suggest that situated functional conversations in multi-user virtual world improved functional communication and are enjoyable, acceptable and feasible.

The multi-user aspect of the virtual world was explored in a subsequent feasibility study of social support groups delivered online in EVA Park (Marshall et al., 2020). A waitlist randomised controlled design was used with two groups immediately receiving 14 sessions of intervention over 6 months and two groups receiving the intervention after a 6-month delay. Feasibility outcomes were good; 72.34% of those eligible elected to take part, 85.3% of the participants completed the intervention and 79.4% completed all testing. Participants received 81.6% of the intended dose. Preliminary outcomes were explored but there was no significant change to wellbeing or language. This study demonstrated that it is possible to run groups of 8 people in a multi-user virtual world.

Delivery of specific language interventions in EVA Park has also been explored (Carragher et al., 2020; Marshall et al., 2018). A single case study of SFA in EVA Park delivered 20 hours of therapy over 5 weeks. The treatment protocol followed cued naming from Woolf et al. (Woolf et al., 2016) with SFA from Boyle (Boyle, 2004). At the end of each session 10 minutes were given to situated conversations in EVA Park e.g., requesting daffodils from the flower stand. Outcomes showed gains in treated items but no generalisation to functional communication (Marshall et al., 2018). This is comparable to face-to-face SFA interventions (Efstratiadou et al., 2018). It is possible that the dose of only 10min a session for situated conversation (a total time of approximately 3 hours) was not sufficient to bring about change to functional communication skills. These findings suggest that it is the 20 hours of practice in the simulated environment from the first study that supports functional communication and the specific language work of the SFA that increased naming ability. To deliver SFA together with conversation practice may yield both impairment and functional gains. The current study proposes to train words using ESFA and use the functional spaces and multi-user capabilities of EVA Park to embed words targeted in one-to-one sessions into group conversations.

Overview of the current study

This doctoral research developed a complex clinical intervention aimed at supporting the gains from an impairment level speech and language intervention, ESFA, to generalise into functional communication though the use of conversation groups in a simulated world by utilising virtual world technology (see Appendix 1 for the research proposal).

The Medical Research Council (MRC) framework for developing complex interventions (Campbell et al., 2000) outlines the pathway to best evidence for clinical interventions that include several components, such as those employed by speech and language therapists.

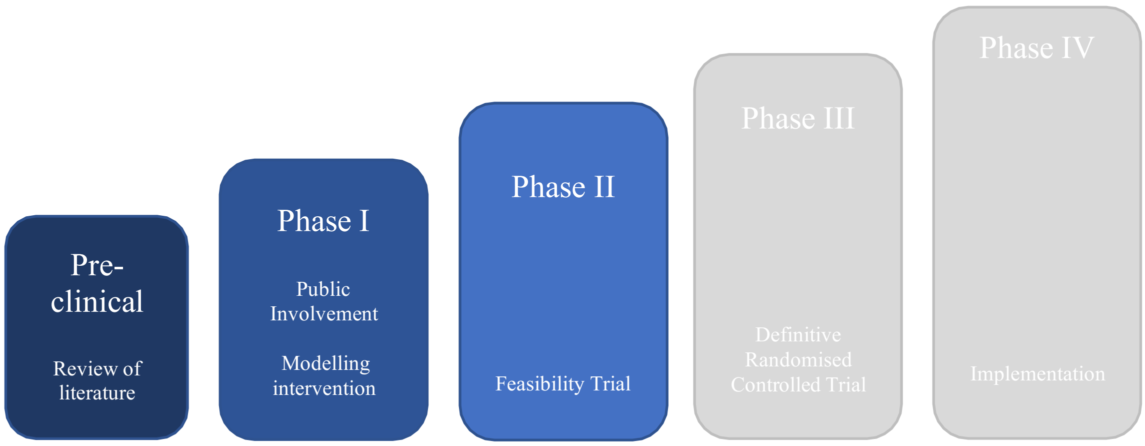

Figure 1.6: Elements of the current study within the MRC framework for developing complex interventions

The framework outlines pre-clinical work and four phases, see Figure 1.6: developing the intervention (phase I), running a feasibility trial (phase II), running a large, randomised control trial (phase III) and finally implementation (phase IV). Developing a complex intervention includes identifying the evidence through a systematic review of the literature, identifying and/or developing the theory for why this intervention is likely to cause a change and modelling the process (pre-clinical work). Single case and case series designs are Phase I studies. Phase II trials assess the feasibility of running a large scale trial, optimise procedures and explore candidacy and outcome measures. If feasibility outcomes are positive, a Phase III large scale RCT can be employed to assess clinical and cost effectiveness. Finally, Phase IV trials encompass implementation of the intervention in real life conditions. In aphasia research, large phase III or phase IV clinical trials (> 100 participants) are rare (Bowen et al., 2012; Breitenstein et al., 2017; Godecke et al., 2021; Palmer et al., 2019; Rose et al., 2022).

SFA and VR interventions can be plotted in this framework. SFA has a clear underlying theoretical basis, a growing body of evidence comprising good single case studies (phase I), small group evidence, two a randomised controlled efficacy trials (phase II) (Efstratiadou et al., 2019, Kendall et al. 2019), and a systematic review and meta-analysis demonstrating 80% of participants who receive SFA improve albeit with a small effect size (Efstratiadou et al., 2018). Evidence for VR interventions includes some single case studies (phase I) and exploratory small group controlled studies (phase II). Thus, this study's systematic literature review on VR (pre-clinical), public involvement (phase I) and feasibility randomised controlled trial (phase II) will contribute to the level of evidence for the use of SFA and VR in aphasia rehabilitation.

There is further published guidance to support the development of complex interventions for health. The brief description of Phase I of the MRC framework has been expanded to provide actions to consider (O'Cathain et al., 2019). The guidance was developed through a process of literature review and a consensus discussion with key stakeholders. The resulting actions are proposed as principles to guide a creative, iterative process, not sequential tasks. Table 1:2 shows the framework of actions outlined by O'Cathain and colleagues and corresponding actions in the development of the VESFA intervention.

	Framework of actions	VESFA intervention development actions
1	**Plan the development process** Including determining the need for the intervention, obtaining funding and producing a protocol.	This was outlined in the proposal for doctoral funding (2018). See Appendix 1
2	**Involve stakeholders** Identify the relevant stakeholders, plan involvement, and identify the best ways of working with each group.	Established a trial Advisory Group of people with aphasia and identified expert speech and language therapists to advise on implementation. This activity is described in Chapter 4, Public Involvement.

3	**Bring together a team** Include individuals with relevant expertise. Establish a decision-making process	The PhD candidate and supervisory team are experts in ESFA and clinical trial methodology, EVA Park interventions and Human Computer Interaction Design.
4	**Review published research evidence** Understand the evidence for each component of the intervention	The evidence for SFA, groups and the use of virtual worlds for aphasia therapy is reviewed. The review is described in Chapters 1 (SFA and groups) and 2 (virtual worlds for aphasia).
5	**Draw on existing theories** Identify the theories that underlie the therapy components	The theories that underlie the therapy components are described in Chapter 3, Intervention Development
6	**Articulate programme theory** Articulate the theory, inputs actions that lead to the outcomes and impact (logic model)	A logic model and Template for better Intervention Description and Replication (TIDieR) description are outlined in Chapter 3, Intervention Development
7	**Undertake primary data collection** Use a range of research methods to understand the context and intermediate outcomes	A qualitative research project asked the question: What topics are meaningful to people with aphasia? This informed the topics for treatment. This study is described in Chapter 5. The VESFA feasibility trial is described in Chapters 7, 8 and 9
8	**Understand the context** Consider factors that influence implementation	Focus groups were run with aphasia specialist SLTs from both national health service and independent practice, described in Chapter 4: Public Involvement
9	**Pay attention to future implementation** Understand the barriers and facilitators of future use of the intervention	Advisory group and SLT focus groups discussed acceptable regimes and implementation in clinical practice, described in Chapter 4: Public Involvement. The barriers to faithful delivery of the intervention are delivered is described in Chapter 7: Treatment Fidelity.
10	**Design and refine the intervention** Generate content, format and delivery with stakeholders. Refine the intervention in an iterative process	Work with the Advisory Group to design the intervention, experience a taster session, and give feedback is described in Chapter 4. The intervention was reviewed after the first set of 3 participants received the 8-week intervention. This is described in Chapter 3.

11	**End the development phase** Describe the intervention so that it can be delivered by individuals outside of the project team. Write up the intervention development process	The outcome of the intervention development is the VESFA therapy manual. This is available in Appendix 8. This thesis describes the intervention development process: Chapters 1 and 2 cover the background literature, Chapters 3, 4 and 5 outline the processes involved in the development of the manual, Chapter 7 describes fidelity and Chapters 6, 8 and 9 the feasibility randomised control trial

Table 1.2: Actions for developing a complex intervention to improve health with corresponding actions in the development of VESFA

Summary and study aims

Word finding difficulties are widespread in post stroke aphasia. There is good evidence that semantic feature analysis can improve the retrieval of treated words in this client group, impacting the impairment domain of the ICF. However, benefits for everyday language use, the activity domain, have not been well established. This study will explore if delivery of an elaborated SFA in a simulated environment is feasible. It will also collect outcome data to explore indicative benefits for naming, functional communication and quality of life.

This study evaluated existing evidence through a pre-clinical systematic literature review (Chapter 2), highlighted the underlying theories and articulated the programme[2] theory for VESFA (Chapter 3), consulted users in phase I public involvement activities (Chapter 4), ran a qualitative study to identity the treatment targets (Chapter 5), checked treatment fidelity (Chapter 7) and answered questions about the feasibility of a definitive phase III randomised control trial of clinical effectiveness of elaborated semantic feature analysis delivered in a virtual world (Chapters 6, 8, 9 and 10):

[2] A note on spelling: 'program' is used for a computer program, and 'programme' for a schedule of events.

Overall aim of this doctoral research is:

To explore how speech and language therapy can exploit virtual simulation, through the use of the virtual world EVA Park, for rehabilitation gains in situated language for people with aphasia.

To that end, this thesis will:

1. Systematically review how virtual reality is currently being used in the rehabilitation of aphasia
2. Create an intervention protocol, VESFA (Virtual Elaborated Semantic Feature Analysis), that addresses word retrieval in EVA Park and the use of word retrieval in situated conversations.
3. Evaluate the feasibility of running a definitive trial on the efficacy of the VESFA intervention.

Primary objectives of the feasibility trial are to evaluate:

a) the feasibility of recruitment and retention
b) the feasibility of delivering the assessment and intervention remotely
c) the acceptability of research procedures to participants
d) the acceptability of the intervention to participants

Secondary objectives are to:

e) evaluate the appropriateness of outcome measures
f) evaluate treatment fidelity

Clinical outcomes will provide preliminary evidence of the potential of the intervention to be effective in terms of:

g) Naming treated words (study specific naming task)
h) Naming words in general (Boston Naming Test, Kaplan, Goodglass, & Weintraub, 1983)

i) Retrieving words within discourse (Nicholas & Brookshire discourse analysis protocol, 1993)

j) Functional communication (The Scenario Test, UK, Hilari et al., 2018)

k) Reducing emotional distress (GHQ-12, Goldberg, 1972)

l) Improving health related quality of life (SAQOL-39g, Hilari et al., 2009)

m) Improving aphasia language profile (Western Aphasia Battery, Aphasia Quotient, Kertesz, 2007)

Chapter 2 | The use of virtual reality in the rehabilitation of aphasia: a systematic review

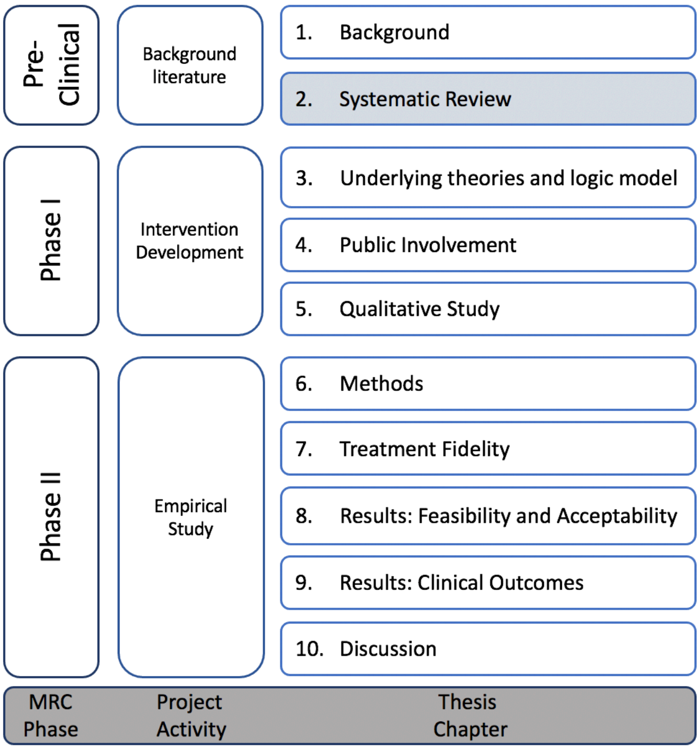

Figure 2.1: The thesis in a figure

Chapter 2 addresses aim 1 and reports on a systematic literature review of the use of VR in the rehabilitation of aphasia and cognitive communication difficulties. As this is an emerging field the review was not restricted to aphasia but broadened to include communication difficulties following brain injury. The systematic literature review has been peer reviewed and published in the journal Disability and Rehabilitation (Devane et al. 2022), therefore it is presented here in the format of a paper[3].

3

Statement of Contribution: Niamh Devane (ND) drafted the SLR protocol which was iteratively revised and finalised with Katerina Hilari (KH), Jane Marshall (JM) and Stephanie Wilson (SW). ND carried out the searches. ND and Aparna Ramachandran (AR) independently double screened the title and abstracts, full text articles and data extraction. ND, AR, Nicholas Behn (NB), KH, and JM completed the quality assessments. ND wrote the paper and NB, JM, AR, SW and KH iteratively revised the manuscript and approved the final version.

ABSTRACT

Purpose: This systematic review explored how virtual reality (VR) has been used to rehabilitate aphasia.

Materials and Methods: Empirical studies were included where VR was used to target language, wellbeing or quality of life in adults with acquired language impairment. Degenerative communication disabilities were excluded. Seven health databases were searched in October 2021. Risk of Bias was assessed using published checklists and completeness of intervention reporting evaluated. Narrative synthesis described forms of VR, rationales given, outcome measures, communication functions targeted, characteristics of interventions and outcomes achieved within the framework of impairment, activity and participation.

Results: 14 studies, involving 229 participants, met criteria. The studies employed four forms of VR with various rationales given. Interventions used published and novel protocols. Primary outcomes targeted language impairment (12/14), activity (1/14) and wellbeing (1/14) and achieved positive outcomes in impairment and activity. All studies were exploratory. Risk of bias was high. Findings are discussed in the context of gains achieved by VR in other health contexts and the multi-user gaming literature.

Conclusions: Uses of VR in aphasia rehabilitation described in the literature are limited. Most applications target the remediation of language impairments. Opportunities to address activity, participation and wider aspects of wellbeing are rare.

INTRODUCTION

Aphasia is a neurological condition that affects a person's ability to use language (Berg et al., 2020). The most common cause of aphasia is a stroke. At least a quarter of those who survive a stroke will experience aphasia (Ali et al., 2015). It affects .1-.4% of the population worldwide (Code, Chris & Petheram, 2011) and 350,00 people in the UK (Stroke Association, 2018). As stroke survival rates improve, more people are living with this lifelong disability (Royal College of Physicians, 2015).

Aphasia has a negative impact on people's lives. People lose their friends (Northcott & Hilari, 2011) with negative emotional effects (Worrall et al., 2016). Far reaching consequences have been reported for social inclusion, social connectedness, access to information and services, equal rights, and wellbeing in family, community and culture (Berg et al., 2020). Social isolation is linked to premature death, and poorer wellbeing (Brummett et al., 2001; Cornwell & Waite, 2009; House, 2001). For these reasons, it has been argued that aphasia is a public health concern (Simmons-Mackie & Cherney, 2018). There is a need for therapies that address both the aphasia and its impact on people's lives.

Treatments for aphasia can focus on all levels of the International Classification of Functioning and Disability (ICF) framework (World Health Organization, 2001) and go beyond the ICF to focus on wellbeing and Quality of Life (QOL). The language impairment (the body structure and function domain) has been targeted in treatments for words, sentences or narratives (for reviews see Dipper et al., 2020; Mehri & Jalaie, 2014; Wisenburn & Mahoney, 2009). Communication activity has been targeted in functional approaches (Wilkinson & Wielaert, 2012) as has societal participation (Horton, Lane, & Shiggins, 2016; Shrubsole et al., 2021). Aphasia is known to have a particularly negative impact on wellbeing, leading to depression (Hilari, et al., 2010; Kauhanen et al., 2000) and reduced QOL (Hilari, 2011; Hilari, Needle, & Harrison, 2012; Lam & Wodchis, 2010). Therefore, these constructs should form part of the focus of aphasia rehabilitation. The ICF framework, with the addition of wellbeing and QOL (hereafter referred to as ICF+) provides a structure for describing a wide range of potential rehabilitation outcomes in this review.

A key priority for people with aphasia is to improve communication in activities (Worrall et al., 2011). Using communication in a conversational context has been described as ‘situated language use’ (Doedens & Meteyard, 2018) and is key to this aspect of rehabilitation. It places the language functions (naming, syntax, narrative structure) in the context of the environment, the number of people in the conversation, interpersonal history and the multimodal (facial expression, gesture, tone) nature of conversations (Doedens & Meteyard, 2022). Multiple people and multiple environmental settings can be difficult to recreate in speech and language therapy sessions. There is a need for treatments that address this communication in context.

Virtual reality (VR) is the technology that allows one or many users to experience a three-dimensional space on a computer (Cambridge University Press, 2021). Multi-user virtual environments may be uniquely placed to treat communication in context. The potential to create faithful, simulated experiences has been harnessed for learning in a range of contexts. Examples include an island where you interact only in German (Thomas, Cinganotto, & Heike, 2018) and recreations of surgical procedures for the training of medics (Sutherland et al., 2006). The simulation allows for practice with minimal risk. The safe practice space that VR offers has been explored in other communication disabilities, notably autism (Bryant, Brunner, & Hemsley, 2019). This review will outline the ways the opportunity for simulated context has been used in aphasia rehabilitation.

The opportunities to interact with multiple users of VR may bestow social and emotional benefits. Indeed, such benefits have been reported in the gaming community where a sense of belonging and warm relationships are cited (Lin, Y., Lin, H., & Yang, 2017). Multi-user gaming has been embraced by people with disabilities. Interviews exploring the value of gaming with this group have highlighted why gaming was important to them (see Box 1). In addition to benefits cited by the general gaming community, people with disabilities highlighted the benefits of a space where they can be on an equal footing with other users and practice skills and showed an appreciation for the creativity in design and storytelling (Cairns et al., 2021).

Connecting:	A way of bringing people together both as friends and family but also to build communities.
Diverting:	A distraction from problems and a way to relax and unwind from day-to-day stresses and to enter different worlds.
Beneficial:	Playing games can bring about benefits to players outside of the world such as developing skills or learning about the world.
Art:	Games are of intrinsic value to players because they are a form of creative expression for both developers and players.
Fun:	Games are to be enjoyed.
A way of life:	Players play games because that's something that they have always done and always want to do.
Universal:	Players felt that games have something for everyone.
Enabling:	For the players with disabilities, games were a way to be on an equal footing with everyone else.

Box 2.1: Why gaming is important to players with disabilities (Cairns et al. 2021)

VR can replicate real world spaces or create novel environments. Some parts of the gaming community have embraced the development of novel creative spaces. There can be dream-like spaces (https://youtu.be/21FaS_bxReo) and worlds where the graphics are inspired by famous artists (Yarwood, 2015). Experiences of fun and diversion (Cairns et al., 2021) may have positive implications for mood and wellbeing. The potential for social and emotional benefits are notable in the context of negative consequences of aphasia.

In stroke rehabilitation there is a growing evidence base for the benefits of VR interventions in upper limb rehabilitation (Karamians et al., 2020; Mekbib et al., 2020), balance and gait (Ghai, S., Ghai, I., & Lamontagne, 2020), cognitive function and activities of daily living (Chen et al., 2022). There is even some evidence that physical gains following VR rehabilitation may be accompanied by cortical changes (You Sung et al., 2005). A number of reviews have examined the use of VR to improve motor outcomes after stroke using the ICF framework (Alt Murphy et al., 2015; Aminov et al., 2018; Lohse et al., 2014; Palma et al., 2017). The most recent review identified 34 trials with impairment level outcomes, 17 trials with activity outcomes and eight trials with a focus on participation (Palma et al., 2017). This illustrates that VR is used in physical rehabilitation to address all levels of the ICF, with most emphasis on impairment.

Synthesis studies of VR in aphasia rehabilitation have been published since 2020 (Cao et al., 2021; Picano et al., 2021; Repetto et al., 2020). In 2020, Repetto and colleagues investigated what innovative technologies (smartphones, tablets and VR) were effective in post stroke aphasia (Repetto et al., 2020). This systematic review included three studies which used VR. They were Marshall et al. (Marshall et al., 2016) with EVA Park and Grechuta et al. (Grechuta et al., 2017; Grechuta et al., 2019) with the Rehabilitation Gaming System (RGS). Outcomes were descriptive with effect sizes reported for one study (Marshall et al., 2016). The authors concluded that the field was in its infancy.

Picano and colleagues carried out a review that sought to understand 'existing unconventional approaches' (p.2, Picano et al., 2021) to aphasia rehabilitation in 2021 (Picano et al., 2021). They included eight studies that used VR. The review gave a narrative description of EVA Park (Carragher et al., 2020; Marshall et al., 2016; Marshall et al., 2018; Marshall et al., 2020), RGS (Grechuta et al., 2016; Grechuta et al., 2020), the Virtual Reality Rehabilitation System (VRRS) tablet (Maresca et al., 2019) and Giachero and colleagues' use of VR for functional communication situations (Giachero et al., 2020). The authors concluded that VR has the potential to increase treatment dose, maximise sensorimotor stimulation and, overall, improve ecological validity of aphasia treatment (Picano et al., 2021).

Cao and colleagues carried out a systematic review and meta-analysis of the effects of VR in post stroke aphasia in 2021 (Cao et al., 2021). They explored whether VR interventions had an effect on communication activity and language function compared to a control condition. The five studies included were EVA Park (Marshall et al., 2016), RGS (Grechuta et al., 2019), the VRRS tablet (Maresca et al., 2019), VR for communication situations (Giachero et al., 2020) and a conference paper exploring a virtual reality panoramic helmet (Zhang et al., 2017). The review found a borderline effect of reducing language severity and no difference between VR and control for communication activity, word finding or repetition outcomes.

The control conditions were both an alternative SLT treatment and no treatment. The conclusions of this review were supported by meta-analyses; however, these were based on limited data (2 studies per meta-analysis) and combined studies that employed different treatments (e.g., naming therapy combined with a conversation therapy) and different outcome measures (Communication Activities of Daily Living combined with the Communication Activities Log).

The current review updates and broadens the scope of these previous reviews. Firstly, it places greater emphasis on the rationales for using VR, the therapy goals, how they were measured and how VR was employed to enhance the therapy experience. The quality of reporting is also explored. Secondly, this review is not restricted to post-stroke aphasia. Aphasia can be caused by other brain pathology, e.g., a brain injury, tumour or surgery. We did not restrict the underlying cause of aphasia. Moreover, given that VR use is an emerging field in this area we broadened the scope of the review to acquired non-progressive language disorders. This review sought to identify the ways in which VR has been used to support language and communication rehabilitation, particularly in reference to the domains of the ICF. Previous reviews highlight that VR has particular value in providing ecological validity, aligning with the activity domain of the ICF (Picano et al., 2021). However, the use of VR in rehabilitation is a recent innovation, meaning that applications in aphasia were likely to be limited. The authors were therefore interested in innovations in related disorders that could inform the development of VR for aphasia rehabilitation. Thus, we included cognitive communication disorder, a related disorder where the communication deficit is due to impaired cognitive functions rather than language (Togher et al., 2014).

This review aimed to find out how VR has been used in the rehabilitation of acquired communication disorders. Specifically, it explored the following research questions:

i. What forms of VR were used?
ii. What rationale(s) were given for the use of VR?
iii. What outcome measures were used?
iv. What communication functions were targeted?
v. What were the characteristics of the interventions?

vi. What outcomes were achieved?

METHODS

The reporting of this review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline (Appendix 2). The protocol was prospectively registered with PROSPERO (CRD42020196285).

Eligibility criteria

Studies were eligible if VR was used in an intervention study targeting any of the following for people with aphasia or acquired cognitive communication disorder: language, communication activity, participation, wellbeing, quality of life. There were no language restrictions and dates were the earliest available within each database.

The population was defined as adults (>18years) with aphasia and/or cognitive communication disorders following stroke or traumatic brain injury. Mixed population studies were only included if outcomes were reported separately for people with aphasia or cognitive communication disorder. Studies of motor speech disorders were excluded. Degenerative language disorders such as dementia, progressive neurological conditions and primary progressive aphasia were excluded. In degenerative conditions the aims and methods of rehabilitation are different, and it would not be valid to conflate progressive and non-progressive participants.

P	Adults (>18yrs) with aphasia or acquired cognitive communication disorder following stroke or traumatic brain injury
I	Therapy delivered by immersive or non-immersive virtual reality
C	Experimental control: multiple baselines or control group of alternative treatment, usual care, placebo, or no treatment
O	Changes at any level of the ICF (language function, activity, participation) or wellbeing and quality of life

Box 2.2: Population, Intervention, Comparison and Outcome (PICO) framework

For this review we defined virtual reality as a set of images and sounds produced by a computer that seem to represent a real or imagined place or situation that a person can take part in (Cambridge University Press, 2021). Both immersive (using equipment such as a head mounted display) and non-immersive (interacting with an image on a screen) 3D environments were included. 2D applications were excluded. Studies had to report on empirical data with experimental controls and be published in a peer-reviewed journal. Beyond this, there were no constraints on study design i.e., experimental single case and case series designs were included. New data had to be reported, so review papers were excluded. Box 2.2 summarises the Population, Intervention, Comparison and Outcome of interest (PICO) in this systematic review.

Information sources

Seven electronic databases were searched following consultation with the subject librarian: CINAHL, Communication source, MEDLINE, Academic Search Complete, PsycINFO, Embase, and Ovid Emcare. Citation tracking from eligible articles was carried out using Scopus. Searches were run on 30th June-3rd July 2020 and repeated on 19th, 20th and 21st October 2021.

Search strategy

Search terms were variations on three concepts: acquired language impairments, rehabilitation and virtual worlds. Search strings varied slightly depending on the MESH terms within each database. Truncation (*) was used to capture variations in terms e.g., aphasia / aphasic. Box 2.3 illustrates an example search string. Full searches are available in Appendix 3.

“aphasi*” OR “dysphasi*” OR “cognitive communication”

AND

“rehabilitation” OR “speech therap*” OR “intervent*” OR “treat*” OR “train*” OR “program*” OR “language therap*” OR “social support” OR “stimulat*” OR “speech patholog*” OR “language patholog*”

AND

“virtual world*” OR “virtual reality” OR “virtual environment” OR “video games” OR “computer simulat*” OR “virtual” OR “augmented reality” OR “augmented virtuality” OR “mixed reality” OR "virtual reality exposure therapy" OR cyberspace OR "immersive environment" OR "multi-user virtual environment"

Box 2.3: Search string example

Study selection

Screening on title and abstract used the following hierarchy:

1. Participants were adults with aphasia or acquired cognitive communication disorder following stroke or other acquired brain injury
2. Virtual reality was used
3. Intervention studies to remediate language impairment, communication activity, participation or quality of life were reported

Full text articles were included if:

1. Participants had aphasia or acquired cognitive communication disorder.
2. Intervention targeted language impairment, communication activity, participation or quality of life
3. Empirical data was reported with an experimental control e.g., across time or a comparator group.
4. Immersive or non-immersive virtual reality was used.
5. The publication was peer reviewed.

Data collection process

Search results were double screened on title and abstract, and full text articles were double screened independently by the first (ND) and fourth author (AR). A data extraction table was developed in Microsoft Excel to cover study characteristics, participants, intervention, outcomes, and VR aspects. Data were extracted independently by ND and AR and any discrepancies were resolved by discussion between reviewers. If consensus could not be reached, a third senior researcher (KH) had the deciding vote.

The aim of the intervention was determined by the primary outcome measure used. Outcome measures of selected studies were mapped onto the ICF+ categories. For example, a language test as primary outcome measure (e.g., The Western Aphasia Battery – Revised

(Kertesz, 2007) indicated the use of VR to change language impairment (ICF domain: impairment), whereas a communication test (e.g., The Communication Activities of Daily Living (Holland, A. et al., 1999) indicated the use of VR to change activity (ICF: activity). All outcome measures (primary and secondary) and ICF+ domains were independently mapped by two authors (ND & KH). These decisions were subsequently checked against the categorisation published by Wallace et al. (Wallace et al., 2022) and found to be in agreement, with the exception that Wallace collapsed activity and participation into one category. Secondary outcome measures were recorded to indicate intervention aims that targeted additional levels of the ICF. Discrepancies were resolved by discussion between reviewers. If consensus could not be reached, a third senior researcher (JM) had the deciding vote.

Data items

Data items described (i) the forms of virtual reality employed, (ii) the theoretical basis given by authors for employing virtual reality, (iii) the primary and secondary outcome measures used, (iv) the ICF domain targeted by these measures, (v) the intervention characteristics, and (vi) the outcomes achieved /changes reported on outcome measures. Additional variables collected were participant number, age, sex, aphasia type and time post onset, study setting and country. If data was missing it was indicated as not reported.

Risk of bias

Completeness of the intervention reporting was explored using the TIDieR framework (Hoffmann et al., 2014). The framework outlines 12 items that should be reported. A complete TIDieR framework indicates a high quality of reporting that provides enough information for researchers and clinicians to replicate the intervention. Information from each study was extracted to complete the TIDieR framework by the first author (ND) with 35% (5/14 studies) independently extracted by the second author (NB). Discrepancies were resolved by discussion. Each study was given a point if the item was present in the report, to give each study a rating out of 12 for completeness of reporting.

The Physiotherapy Evidence Database (PEDro) scale (Moseley et al., 2015) was used to rate the methodological quality of randomised and non-randomised controlled trials. This is an

11-item checklist which gives a total score out of 10 (item 1 does not contribute to the total score). Quality is considered excellent if a study scores 9-10, good if a study scores 6-8, fair if a study scores 4-6 and poor if a study scores less than 4.

The Risk of Bias in N-of-1 Trials (RoBiNT) (Tate et al., 2015) scale was used to assess single-case experimental designs. This 30-item checklist addresses the internal and external validity of studies. The RoBiNT authors subsequently published an algorithm to qualify the methodological rigor of the internal validity (Perdices, Tate, & Rosenkoetter, 2019). A flow chart is followed to arrive at one of 10 grades from 'very low' to 'very high'. A point to note is that the RoBiNT tool was designed for studies that have dramatic "on-off" effects" (p.621, (Tate et al., 2013) where a decline in performance is hypothesised when treatment is withdrawn. Conversely, in speech and language therapy intervention studies the very aim of treatment is for lasting effects. Participants are not expected to revert completely to pre-treatment levels when the stimulus (treatment) is removed. Nevertheless, this design is considered 'nonwithdrawal' and described as an AB+ maintenance design in RoBiNT and scores 0 (Tate et al., 2015). If a study scores a 0 for design it can only score as 'very low' for quality in the algorithm (p.12, Perdices, Tate, & Rosenkoetter, 2019).

National Heart, Lung and Blood Institute (NIH) Quality Assessment tool for Before After studies with no control group was used to rate before-after studies with condition rather than group control. This 12-item checklist addresses the internal validity of a study. Quality descriptions followed published guidance (American Occupational Therapy Association, 2020), based on number of items in the quality tool that were not present: a score of 0–3 N (N=not present) indicates a low risk of bias, a score of 4–8 N indicates a moderate risk of bias, and a score with 9–11 N indicates a high risk of bias.

Each study was rated for risk of bias independently by two authors (RoBiNT: ND, AR, KH and JM; PEDro: ND, AR and NB; NIH Quality Assessment tool: ND and AR). Allocation ensured that authors did not rate their own publications. Disagreements were resolved by discussion.

Summary measures and data synthesis

Descriptive statistics were used to summarise participants. Descriptive information and a narrative synthesis described the focus and detail of interventions, outcome measures used, VR used and underlying theory. These were tabulated in Microsoft Excel. The TIDieR was used to summarise completeness of reporting. Low quality studies are known to provide biased results (Jüni et al., 2001). Therefore, only studies of adequate quality were included in the synthesis of outcomes. These were studies that scored 4 and above on the PEDro scale, graded as fair to very high on the RoBiNT scale, and moderate to low risk of bias on the NIH quality assessment tool. Where data permitted, effect sizes were calculated (d= 0.2, medium is d=0.5 and large d=0.8, (Cohen, 2013). A meta-analysis was planned for eligible group-level studies that used the same outcome measures using standardised mean differences (SMD) and a fixed-effect model. However, heterogeneity in the outcome measures was high and a meta-analysis was not feasible.

RESULTS

Study selection

The search of seven databases identified 639 articles with a further 13 identified through citation tracking on Scopus. All 652 records were imported into the evidence synthesis software, EPPI Reviewer (Thomas et al., 2020). The software removed 232 duplicates. The title and abstract of the remaining 420 articles were screened against the inclusion criteria by ND and AR. 392 were excluded from the review because they did not involve adults with aphasia or acquired cognitive communication disorder (n=244), did not use virtual reality (n=68) or they were not intervention studies (n=78), and two further duplicates were found by reviewers. Full text articles were retrieved for the remaining 28 records. Two were not available as they were abstracts from conference proceedings. The full text was reviewed of the remaining 26 articles by ND and AR. 12 more were excluded at this stage as they did not meet the inclusion criteria (see Figure 1). Fourteen articles met the criteria and were included in the review.

The review process had good agreement between raters, with 95% agreement on title and abstract decisions and 89% agreement on full text (25/28). Disagreements were resolved with discussion.

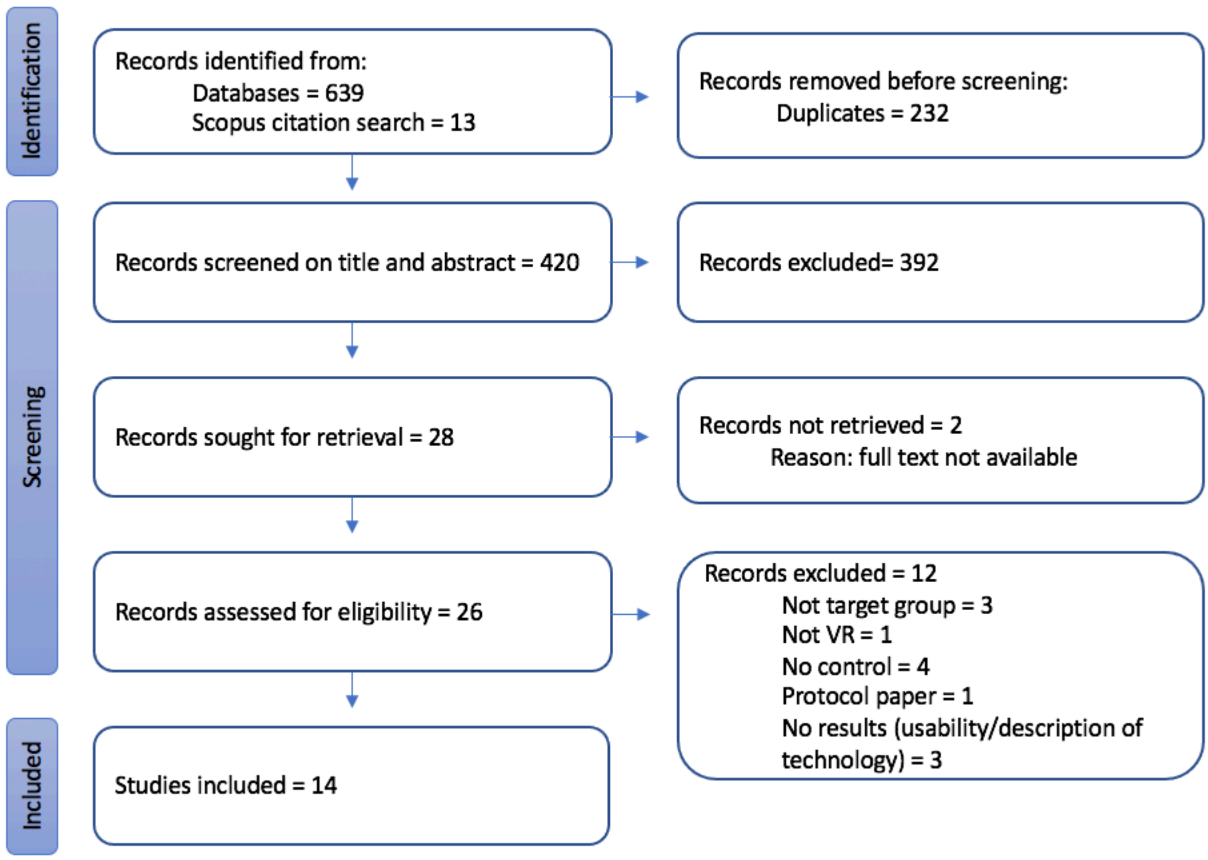

Figure 2.2: PRISMA flow diagram of study selection process

Study characteristics

Fourteen articles were included in the review. These articles represent 14 different studies investigating seven different examples of virtual reality in aphasia rehabilitation. They represent the work of seven research teams, working in the UK (n=1) (EVA Park) including a collaboration with Australia, Spain (n=1) (Rehabilitation Gaming Software (RGS)), Italy (n=2) (NeuroVR 2.0 and Virtual Reality Rehabilitation System (VRRS-tablet)) and the USA (n=3) (Sentactics, AphasiaScripts/Web ORLA and the Virtual Human toolkit (Hartholt et al., 2013). No articles were found of empirical research investigating the use of VR to rehabilitate the language of people with acquired cognitive-communication disorder.

Seven studies were randomised controlled trials, five were single case studies and two were before-after studies with no control group, where the experimental control was different conditions within the group e.g., two different cuing methods. All studies were described as exploratory: they described feasibility, pilot or efficacy studies or reported a sample size too small to be a definitive effectiveness study. Study characteristics are summarised in Table 1.

Participant characteristics

The fourteen studies reported on a total of 229 adults with aphasia, 95 female and 134 males. Almost all were in the chronic stage post stroke (>6m), with only one participant reported as less than six months post stroke. Where mean age was reported (9 studies), all means were younger than 60 years old and age ranged from 40-71 years. Where reported (11 studies), participants were predominantly people with non-fluent aphasia, n=119, vs n=26 with fluent aphasia. The studies were carried out in three settings: community, hospital and research laboratory. Table 2.1 summarises the studies and study participants.

Study	Design	Country / Setting	Participants (sex F/M) Mean age (SD) in years	Time post onset in months*	Aphasia type
Carragher et al., '20	Case series	Australia and UK/ community	n=3 (3F) 52, 64, & 68yrs	18, 79, 94	Non-fluent =2 Fluent=1
Cherney et al., '19	Before-After study with no control group	USA/not reported	n=20 (6F, 14M) 56.9 (8.4) years,	*55.1 (6.4–396.4)*	Non-fluent=17 Fluent=3
Cherney et al., '21	Randomised Control Trial	USA/hospital	n=32 (13F, 19M) EG: 58.27 (13.55) CG: 55.19 (11.56)	EG: 39.75, CG: 60.97	Non-fluent=18 Fluent=14
Giachero et al., '20	Randomised Control Trial	Italy/research lab	n=36 (12F, 24M) 59.75(11.21)	42.75	Non-fluent=36 Fluent=0
Grechuta et al., '17	Case series	Spain/hospital	n=4 (2F, 2M) 40, 58, 62, & 63yrs	5, 7, 31, 46	Non-fluent=4 Fluent = 0
Grechuta et al., '19	Randomised Control Trial	Spain/hospital	n=17 (8F, 9M) 54.6 (9.9)	59.94 (47.83)	Non-fluent=17 Fluent=0
Grechuta et al., '20	Group Controlled Trial	Spain/hospital	n=10 (5F, 5M) 57.6 (9.9)	69.9 (48.7)	Non-fluent=10 Fluent=0
Kalinyak-Fliszar et al. '15	Case series	USA/research lab	n=4 (1F, 3M) 49, 49, 51, & 51yrs	12, 44, 63, 103	Non-fluent=2 Fluent=2
Maresca et al., 2019	Randomised Control Trial	Italy/hospital and community	n=30 (16F, 14M) 51.2(11.3)	Not reported	Not reported
Marshall et al., 2016	Quasi Randomised waitlist controlled	UK/community	n=20 (9F, 11M) 57.8 (11.58)	62.10 (53.56)	Not reported
Marshall et al., 2018	Case series	UK/community	n=2 (2M) 60 & 54yrs	36, 60	Non-fluent=1 Fluent=1
Marshall et al., 2020	Randomised Control Trial	UK/community	n=34 (17F, 17M) 53.5, IQR: 48.75-71	*46.5 (15-83.75)*	Not reported
Snell et al. 2017	Case Series	USA/research lab	n=5 (1F, 4M) 49, 55, 56, 62 & 65yrs	25, 98, 99, 172, 175	Non-fluent=0 Fluent=5
Thompson et al., 2010	Group Controlled Trial	USA/research lab	n=12 (2F, 10M) 49.5, (10.96)	59.8	Non-fluent=12 Fluent=0

M=male, F=female, EG=experimental group, CG=control group, SD=standard deviation, IQR=Interquartile Range. * Case series: Individual values. Group studies: Means (SD) or *medians (IQR)*

Table 2.1: Study (n=14) and participant (n=229) characteristics

Risk of Bias

The seven randomised control trials were quality assessed using the PEDro scale, see Table 2.2. Two studies were rated as good quality (6-8/10) (Cherney et al., 2021; Grechuta et al., 2019), three studies were rated as fair (4-5/10) (Giachero et al., 2020; Marshall et al., 2016; Marshall et al., 2020) and two were rated poor (<4/10) (Maresca et al., 2019; Thompson et al., 2010).

PEDro Item	Cherney et al. '21	Giachero et al '20	Grechuta et al. '19	Maresca et al. '19	Marshall et al. '16	Marshall et al. '20	Thomp-son '10
1. Eligibility specified*	Y	Y	Y	Y	Y	Y	y
2. Random allocation	1	1	1	1	0	0	1
3. Concealed allocation	0	1	1	0	0	0	0
4. Similar at baseline	1	0	1	0	1	0	0
5. Blind subjects	0	0	0	0	0	0	0
6. Blind treating therapist	1	0	0	0	0	0	1
7. Blind assessor	1	1	1	0	0	1	0
8. Retention >85%	1	0	1	0	1	1	0
9. Intention to treat analysis	0	0	1	0	1	1	0
10. Between group comparisons	1	1	1	1	1	1	1
11. Point measures and variability measures	1	1	1	1	1	1	1
Total	**7/10**	**5/10**	**8/10**	**3/10**	**5/10**	**5/10**	**4/10**
Quality	**Good**	**fair**	**good**	**poor**	**fair**	**fair**	**poor**

Scoring: 1= reported; 0= not reported. **Quality**: 9-10/10 excellent, 6-8/10 good, 4-5/10 fair, <4/10 poor

*Item does not contribute to the score

Table 2.2: Randomised controlled trials rated using the PEDro

The five single case studies were rated using the Risk-of-Bias in N-of-1 trials (RoBiNT) tool, see Table 2.3. These studies received scores ranging 7-18/30 on the RoBiNT 30 item scale, with four studies scoring 15-18/30 and one 7/30. All five studies scored higher on external validity than internal validity. As expected, all five studies were scored as of very low quality, when the RoBiNT algorithm was applied, despite the variability in scores. A score for blinding in intervention was given for one study only where stimuli were computer delivered (Snell et al., 2017). None of the studies received points for replication.

RoBiNT items	**Carragher at al. '20**	**Grechuta et al. '17**	**Kalinyak-Fliszar et al. '15**	**Marshall et al. '18**	**Snell et al. '17**
1. Design with control	0	0	0	0	0
2. Randomisation	0	1	1	0	0
3. Sampling of behaviour	1	2	2	0	1
4. Blinding in intervention	0	0	0	0	1
5. Blinding of assessors	2	2	0	2	0
6. Interrater agreement	1	1	2	0	0
7. Treatment adherence	0	2	0	2	0
Internal validity score:	4/14	8/14	5/14	4/14	2/14
8. Baseline characteristics	0	1	2	1	0
9. Setting	2	2	1	2	1
10. Dependant variable	2	2	2	2	1
11. Independent variable	2	2	2	2	1
12. Raw data record	2	2	2	2	1
13. Data analysis	1	1	1	1	1
14. Replication	0	0	0	0	0
15. Generalisation	2	0	1	2	0
External validity score:	11/16	10/16	11/16	12/16	5/16
Total	**15/30**	**18/30**	**16/30**	**16/30**	**7/30**
Quality	**very low**	**very low**	**very low**	**very low**	**very low**

Scoring: each item is scored on three-point scale 0-2, where 2 represents the highest quality

Table 2.3: Single case designs rated using the RoBiNT

The final two studies were assessed for bias using the NIH Quality Assessment tool for Before After studies with no control group, see Table 2.4. Both studies were judged to have a moderate risk of bias. Neither study reported whether all participants who were eligible were enrolled, gave a rationale for the sample size, blinded assessors or reported a follow up rate.

NIH Quality Assessment tool Item	Cherney et al. '19	Grechuta et al. '20
1. Objective clear	Y	Y
2. Eligibility specified	Y	Y
3. Participants representative	Y	N
4. All eligible enrolled?	N	N
5. Sample size	N	N
6. Intervention	Y	Y
7. Outcome measures	Y	N
8. Blinded assessors	N	N
9. Follow up rate	N	N
10. Statistical analysis	Y	Y
11. Multiple outcome measures	Y	Y
12. Individual level data to determine group effect	N/A	N/A
Total	**7/12**	**5/12**
Quality	**Moderate**	**Moderate**

Y= yes; N= no or not reported; N/A not applicable

Table 2.4: NIH Quality Assessment tool for Before After studies with no control group

Completeness of intervention reporting was assessed using the TIDieR checklist (Hoffmann et al., 2014), see Table 2.5. Complete reporting allows for replication of interventions for research and clinical practice. The 14 studies scored from 6-10 on the 12-point scale. The rationale, materials and procedures, mode of delivery, schedule and dose (items 2,3,4,6 & 8) were most consistently reported. Tailoring, modification and treatment fidelity (items 9-12) were rarely reported.

Author	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
Carragher et al., '20	1	1	1	1	1	1	1	1	0	0	0	0	**8**
Cherney et al., '19	1	1	1	1	0	1	0	1	0	0	0	0	**6**
Cherney et al., '21	1	1	0	1	1	1	1	1	0	0	0	0	**7**
Giachero et al., '20	1	1	1	1	1	1	1	1	0	0	0	0	**8**
Grechuta et al., '17	1	1	1	1	1	1	1	1	0	0	0	0	**8**
Grechuta et al., '19	1	1	1	1	1	1	1	1	0	0	0	0	**8**
Grechuta et al., '20	1	1	1	1	1	1	1	1	0	0	0	0	**8**
Kalinyak-Fliszar et al. '15	1	1	1	1	0	1	0	1	0	0	0	0	**6**
Maresca et al., '19	1	1	0	0	0	1	1	1	1	0	0	0	**6**
Marshall et al., '16	1	1	0	1	1	1	1	1	1	0	0	0	**8**
Marshall et al., '18	1	1	1	1	1	1	1	1	0	0	0	0	**8**
Marshall et al., '20	1	1	1	1	1	1	1	1	0	0	1	1	**10**
Snell et al. '17	1	1	0	1	0	1	0	1	0	0	0	0	**5**
Thompson et al., '10	1	1	1	1	1	1	1	1	0	0	0	0	**8**

TIDieR items are: (1) brief name (2) rationale of essential elements of intervention (3) what—materials (4) what— procedures (5) who provided (6) how (7) where (8) when and how much (9) tailoring (10) modification (11) how well—planned (12) how well—actual

Table 2.5: TIDieR ratings of intervention description in studies (n=14)

Forms of VR used

Various forms of VR were used in a range of different ways. The treatment was delivered by the computer program or a therapist, the user was represented in the VR or not, the VR was single user or multi-user, the navigation was user controlled or not, and finally the VR environment was of a virtual clinician, a scenario or a virtual world. These are summarised in Table 2.6 and described in this section.

Study	Treatment summary	Rationale for VR	Software	Environment	Users	Independent	Represented
Carragher et al., 2020	Narrative structure training	Accessibility Cost-effective Stimulating	EVA Park	Virtual world	Multi-user	Yes	Yes
Cherney et al., 2019	Computer delivered script training	Treatment integrity	AphasiaScripts	Virtual clinician head	Single user	No	No
Cherney et al., 2021	Computer delivered oral reading treatment	Intensive Increased dose Compliance	Web ORLA	Virtual clinician head	Single user	Yes	No
Giachero et al., 2020	Conversations in functional environments	Ecological validity Embodied theory	NeuroVR 2.0	Virtual scenarios	3 people in a room with therapist	No	Not reported
Grechuta et al., 2017	Requesting objects with silent visuo-motor cues	Intensive	RGS	The virtual scene represented the actual setting: the table and participants	Pair, sat at a table in the same room	No	Yes
Grechuta et al., 2019	Requesting objects	Socially embedded, Intensive	RGS	The virtual scene represented the actual setting: the table and participants	Pair, sat at a table in the same room	No	Yes
Grechuta et al., 2020	Requesting objects with different cues	Not reported	RGS	The virtual scene represented the actual setting: the table and participants	Pair, sat at a table in the same room	No	Yes
Kalinyak-Fliszar et al. 2015	Computer delivered dialogue training	Cost-effective Intensive Home-based Simulations of ADLs	Not reported	Virtual clinician head	Single user	No	No
Maresca et al., 2019	Computer delivered VRRS-Tablet vs. usual linguistic treatment	Promotes functional recovery Motivating Intensive Rich environments	VRRS-Tablet	Virtual scenarios	Single user	Yes	Not reported
Marshall et al., 2016	Conversations in functional environments	Motivating Intensive Generalisation Social networks	EVA Park	Virtual world	Multi-user	Yes	Yes

Study	Treatment summary	Rationale for VR	Software	Environment	Users	Independent	Represented
Marshall et al., 2018	Noun and verb retrieval	Ecological validity Motivating and stimulates conversation	EVA Park	Virtual world	Multi-user	Yes	Yes
Marshall et al., 2020	Social support groups	Ecological validity Motivating and stimulates conversation	EVA Park	Virtual world	Multi-user	Yes	Yes
Snell et al. 2017	Virtual clinician dialogues	Ecological validity	The Virtual Human toolkit	Virtual clinician in a virtual environment	Single user	No	No
Thompson et al., 2010	Computer delivered sentence therapy	Increased dose	Sentactics	Virtual clinician head	Single user	No	No

Multi-user: multiple people can log into and interact within the same virtual space at the same time
Independent: user logs in without a professional to help them
Represented: the user is represented in the virtual space by an avatar
ORLA=Oral Reading for Language in Aphasia, RGS=Rehabilitation Gaming System, VRRS=Virtual Reality Rehabilitation System

Table 2.6: How VR is being used in the included studies

In six of the studies reported, the treatment was delivered by the computer program (Cherney et al., 2021; Cherney et al., 2019; Kalinyak-Fliszar et al., 2015; Maresca et al., 2019; Snell et al., 2017; Thompson et al., 2010). These include the programs using a virtual clinician; AphasiaScripts, Web ORLA, Sentactics, the Virtual Human toolkit, as well as the VRRS-tablet which uses virtual scenarios. The Web ORLA/AphasiaScripts and Sentactics VR applications are based on the virtual agent software from the Centre for Spoken Language Research at the University of Colorado (Thompson et al., 2010). The VR depicts the moving head and shoulders of a clinician. The virtual clinician is pre-programmed to deliver the treatment. Her lips move in a naturalistic way to give visuomotor prompts. Snell et al. (Snell et al., 2017) used the Virtual Human toolkit software (Hartholt et al., 2013). The VR shows the full body of a virtual clinician avatar standing outside a café. In the five programs using a virtual clinician, the user is not represented in the virtual world. These five programs are for a single user. In the studies, the navigation of the program was controlled by the participant (Cherney et al., 2021; Thompson et al., 2010), or a physically present researcher who moved the script forward (Cherney et al., 2019), or selected a pre-programmed response (Kalinyak-Fliszar et al., 2015; Snell et al., 2017). In two studies this was arranged using a 'wizard of oz' paradigm where the researcher controlled the virtual clinician from behind a curtain to give the illusion of independent use (Kalinyak-Fliszar et al., 2015; Snell et al., 2017).

The VR element of the Virtual Reality Rehabilitation System (VRRS-tablet) intervention was delivered by the computer program (Maresca et al., 2019). VRRS was single use and independently navigated by the user. It is not reported if users are represented in the VRRS 3D module. The paper also reports that "the linguistic module with 2D was mainly used" (p4) and the content of the 3D module was not further explained.

The four EVA Park studies used a multi-user virtual world (Carragher et al., 2020; Marshall et al., 2016; Marshall et al., 2018; Marshall et al., 2020). Virtual worlds are defined as "shared, simulated spaces which are inhabited and shaped by their inhabitants who are represented as avatars" (p.1099, (Girvan, 2018)). Users had EVA Park set up on a laptop in their own home. Users were represented by personalised avatars. Users viewed the world from a third person viewpoint just behind their avatar's head. They could see and move around the EVA

Park island and see and speak to the avatars of other users. Users could independently navigate and click on objects within the island to interact with them, e.g., they could click on the diving board and the avatar dived into the lake. The island was made up of a town square, two houses, a disco, a lakeside picnic area and a tiki bar on a smaller island, with all areas linked by green spaces. Interventions were delivered in real time by a therapist, communication support worker or group coordinator who was also represented in the world by an avatar.

The Rehabilitation Gaming System (RGS) technology provided a representation of the physical room in which the users sat (Grechuta et al., 2016, 2017, 2019 & 2020). Two users were positioned facing each other at the same physical table with two monitors between them. The monitor showed a virtual representation of the user's arms, a virtual table and a virtual peer across the table. The representation of the arms on the screen corresponded to the movement of the user's arms through the use of Microsoft Kinect technology. Users watched themselves select and pass virtual objects that had been requested. A rehabilitation assistant was present in the room to resolve technical or communication difficulties.

The virtual space in NeuroVR 2.0 (Giachero et al., 2020) represented functional scenarios such as a station, a hotel or a supermarket. Three participants with aphasia sat with the SLT in the same room with a large 50" curved screen showing the virtual scene. Participants had tasks within each scenario. For example, they bought a train ticket, checked the platform, and responded to an unexpected request for help from someone who had been mugged. Participants indicated their choice verbally to the SLT, and the SLT controlled and navigated the technology.

What rationale(s) were given for the use of VR

Researchers cited a variety of reasons for employing VR, which are summarised in Table 2.6 and mapped out for ease of comparison in Appendix 4. In Australia, where large, sparsely populated geographical regions make accessibility of services particularly pertinent, the accessibility of a remote online telehealth platform has been highlighted (Carragher et al., 2020). Cherney et al. (Cherney et al., 2019) argued that a computer delivered intervention removed human variation and therefore increased treatment integrity. She also argued for more efficiency as human clinicians could potentially 'detract from critical treatment time' (Cherney et al., 2021). Two studies (Cherney et al., 2021; Thompson et al., 2010) additionally argued that computer program delivered treatments allowed for an increased dose without the additional cost of therapist's time. Giachero et al. (Giachero et al., 2020) and Marshall et al. (Marshall et al., 2016, 2018 & 2020) highlighted the ecological validity of setting treatments in simulated real world situations. Settings are considered ecologically valid if they reflect how people behave in a real world setting. Giachero et al. (Giachero et al., 2020) also cited embodied theory – the idea that semantics and language are multimodal (Fernandino et al., 2015) and delivering language therapy in a virtual simulation creates a multimodal learning environment. Grechuta et al. (Grechuta et al., 2019) similarly discussed the value of delivering a 'socially embedded' protocol by using peer interactions (p.1). Six studies (Cherney et al., 2021; Grechuta et al., 2017; Grechuta et al., 2019; Kalinyak-Fliszar et al., 2015; Maresca et al., 2019; Marshall et al., 2016) cited the opportunity for increased intensity as a rationale for VR treatments.

What outcome measures were used

The 14 studies used 14 different primary outcome measures, see Table 2.7. Even when studies targeted the same impairment, e.g., object naming, the outcome measures used were different which made the data too heterogeneous for a meta-analysis. For example, Grechuta and colleagues (Grechuta et al., 2017 & 2020) reported a bespoke naming latency measure to demonstrate improved naming while Marshall and colleagues (Marshall et al., 2018) reported naming correctness, as measured by the Action and Object Naming Battery (Druks & Masterson, 2000).

Outcome measures used to measure a change in the language impairment were the Aachen Aphasia Test (AAT) (Huber, Poeck, & Willmes, 1984), the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983), the Token Test (De Renzi & Vignolo, 1962), the Naming and Oral Reading for Language in Aphasia 6-point scale, NORLA-6 (Pitts et al., 2018), content of narratives (Nicholas & Brookshire, 1993), script accuracy, and therapy specific noun naming and verbal fluency tests (Grechuta et al., 2020; Marshall et al., 2016).

Outcome measures used to measure communication activity and participation were the Communication Activities of Daily Living – 2nd Edition, CADL-2 (Holland et al., 1999), the Conversation Analysis Profile for People with Aphasia (Whitworth & Perkins, 1997), and the Psychosocial Impact of Assistive Devices Scale (PIADS) (Jutai & Day, 2002), the quantity and quality of information in a dialogue (Kalinyak-Fliszar et al., 2015) and the Communication Activities Log (Pulvermüller et al., 2001).

In terms of ICF+ outcomes, wellbeing was measured using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) (Tennant et al., 2007), self-esteem was measured using the Visual Analogue Self Esteem Scale, (VASES) (Brumfitt & Sheeran, 1999), confidence was measured using the Communication Confidence Rating Scale for Aphasia (CCRSA) (Cherney, Halper, & Kaye, 2011), depression was measured using the Aphasic Depression Rating Scale (ADRS) (Benaim et al., 2004) and quality of life was measured using the World Health Organisation Quality of Life Scale (WHOQOL) (WHO, 2012) the Euro-Qol-5D (EQ5D) (Balestroni & Bertolotti, 2012), the Psychosocial Impact of Assistive Devices Scale (PIADS) (Jutai & Day, 2002) and The Stroke and Aphasia Quality Of Life Scale-39 generic stroke version (SAQOL-39g) (Hilari et al., 2003).

The robustness of the measures used was variable. Eight studies used well validated measures as a primary outcome measure (Cherney et al., 2021; Giachero et al., 2020; Grechuta et al., 2017; Grechuta et al., 2019; Grechuta et al., 2020; Maresca et al., 2019; Marshall et al., 2016; Marshall et al., 2018; Marshall et al., 2020; Thompson et al., 2010). Three studies used measures that required rater reliability checks (Carragher et al., 2020; Kalinyak-Fliszar et al., 2015; Snell et al., 2017), with two reporting the interrater reliability agreement (Carragher et al., 2020; Kalinyak-Fliszar et al., 2015). In one study, the computer automatically recorded the primary outcome (Grechuta et al., 2017) and two studies presented their own specially developed outcome measures as their primary measure (Cherney et al., 2019; Grechuta et al., 2020).

Study	Primary outcome measure	ICF domain	Secondary outcome measures	ICF domain
Carragher et al., 2020	Content of narratives story grammar, story content, argument structure	Impairment	CADL-2	Activity
Cherney et al., 2019	NORLA-6	Impairment	Script accuracy	Impairment
Cherney et al., 2021	WAB-R LQ	Impairment	-	-
Giachero et al., 2020	AAT	Impairment	CAPPA, VASES, and WHO QOL	Activity, Self Esteem* and Quality of Life*
Grechuta et al., 2017	Correlation between a vocabulary test and the time between the selection and collection of an object	Impairment	Comparing the correlation in early vs late sessions	Impairment
Grechuta et al., 2019	BDAE	Impairment	Communication Activity Log	Activity/Participation
Grechuta et al., 2020	Vocabulary test (VocabT) of trained items	Impairment	Interaction time: Time of successful goal-oriented peer-peer interaction	Activity
Kalinyak-Fliszar et al. 2015	Discourse measure (Nicholas & Brookshire, 1993)	Impairment	Content of scripts	Impairment
Maresca et al., 2019	Token Test	Impairment	ENPA, ADRS, EQ-5D, PIADS	Impairment, Depression*, HRQOL*, Well-being*
Marshall et al., 2016	CADL-2	Activity	Verbal Fluency, Narrative measures, CCRSA & Friendship Scale	Impairment, Activity, Confidence*
Marshall et al., 2018	AONB	Impairment	CADL-2, Narrative measures	Activity, Impairment
Marshall et al., 2020	WEMWBS	Wellbeing*	CADL-2, Social Connectedness Scale, WAB-R, SAQOL-39g	Activity, Participation, Impairment, HRQOL*
Snell et al. 2015	Number and type of gesture	Impairment	Range of movement of gesture	Impairment
Thompson et al., 2010	NAVS	Impairment	Narrative measures	Impairment

*Outcome domains beyond the ICF

CADL-2=Communication Activities of Daily Living-2nd edition, NORLA-6=Naming and Oral Reading for Language in Aphasia 6-Point Scale, WAB-R LQ= Western Aphasia Battery – Revised, Language Quotient, AAT= Aachen Aphasia Test, BDAE= Boston Diagnostic Aphasia Evaluation, AONB= Action and Object Naming Battery, WEMWBS= Warwick-Edinburgh Mental Well-Being Scale, NAVS= Northwestern Assessment of Verbs, CAPPA=Conversation Analysis Profile for People with Aphasia test, VASES= Visual Analogue Self Esteem Scale, WHO QOL= World Health Organisation Quality of Life Scale, ENPA= Esame Neurologico Per l'Afasia, ADRS= Aphasic Depression Rating Scale, EQ5D= Euro-Qol-5D, PIADS= Psychosocial Impact of Assistive Devices Scale, CCRSA= Communication Confidence Rating Scale for Aphasia, SAQOL-39g=Stroke and Aphasia Quality of Life scale, HRQOL= Health Related Quality of Life

Table 2.7: Outcome measures used in included studies and ICF+ classification

What communication functions were targeted

This review explored how VR was used to treat language and communication within the framework of the ICF+, as determined by the primary outcome measure, see Table 2.7. The majority of studies (n=11) were primarily influencing the language impairment in aphasia (Carragher et al., 2020; Cherney et al., 2021; Cherney et al., 2019; Giachero et al., 2020; Grechuta et al., 2016; Grechuta et al., 2019; Grechuta et al., 2020; Maresca et al., 2019; Marshall et al., 2018; Snell et al., 2017; Thompson et al., 2010). In two studies, the primary outcome was communication activity (Kalinyak-Fliszar et al., 2015; Marshall et al., 2016), and one aimed to improve well-being (Marshall et al., 2020). Secondary outcome measures had a broader spread, in that they addressed impairment and activity/participation domains of the ICF as well as well-being, depression, self-esteem, confidence and QOL.

Characteristics of the interventions

The VR interventions addressed word finding (4 studies), sentence structure (1 study), narratives (1 study), communication activity (4 studies), script training (1 study), oral reading (1 study), comprehension (1 study) and social support (1 study). A summary of the intervention characteristics is in Table 2.8.

Some studies took existing protocols into the novel VR environment and some studies created a new protocol for the new environment. Nine studies used published intervention protocols. One naming therapy study (Marshall et al., 2018) used the protocols of Woolf (Woolf et al., 2016) and Edmonds (Edmonds, Nadeau, & Kiran, 2009). The RGS papers (Grechuta et al., 2017; Grechuta et al., 2019; Grechuta et al., 2020) used the principles of Intensive Language Action Therapy (Difrancesco, Pulvermüller, & Mohr, 2012). The sentence treatment followed the Treatment of Underlying Forms (TUF) protocol (Thompson et al., 2010). The narrative treatment described an adaptation of the Interactive Storytelling Therapy (Carragher, Sage, & Conroy, 2015). The script training (Cherney et al., 2019) used the script protocol from the team's previous research (Cherney et al., 2008), and WebORLA used the protocol from Oral Reading for Language in Aphasia (Cherney, Merbitz, & Grip, 1986). Marshall et al. (Marshall et al., 2020) developed a group social support intervention that drew on elements of a number of published approaches (Attard et al., 2015; Holland, Nelson, & Goldberg, 2014; Seligman et al., 2005; Shadden & Agan, 2004).

Two studies developed novel interventions that exploited the potential of simulated real world environments available in VR technology (Giachero et al., 2020; Marshall et al., 2016). Giachero and colleagues created functional scenarios for communication activity practice, e.g., check in to the hotel, decide how long to stay, book breakfast. There were additional, unexpected events to navigate, e.g., a forgotten suitcase. Marshall and colleagues (Marshall et al., 2016) used the EVA Park virtual world to address participant led communication activity goals, e.g., requesting a hair cut in the barbers, ordering dinner in a restaurant, sharing biographical stories.

Of the remaining three studies, two described exploratory work with a view to developing a novel protocol for dialogue practice tool (Kalinyak-Fliszar et al., 2015; Snell et al., 2017), and the final one described an 'experimental linguistic treatment' (p.3) that was delivered using paper and pencil for the control group and via the VRRS-tablet for the experimental group (Maresca et al., 2019).

Total treatment hours (dose) ranged from less than one hour to 100 hours (Maresca et al., 2019) with mean treatment dose of 19.59 hours, and both a median and mode of 20 hours. The duration of treatment ranged from one session to six months (Giachero et al., 2020; Maresca et al., 2019; Marshall et al., 2020). The frequency of the interventions (sessions per week) was mostly once a day, with 4 or 5 sessions per week (Carragher et al., 2020; Giachero et al., 2020; Grechuta et al., 2017; Grechuta et al., 2019; Grechuta et al., 2020; Maresca et al., 2019; Marshall et al., 2016; Marshall et al., 2018; Thompson et al., 2010) with one study delivering one session every fortnight (Marshall et al., 2020). Treatments were delivered by qualified SLTs, professionals in aphasia support services (e.g., aphasia group co-ordinators) and computer delivered. Treatments were delivered in one to one, peer interaction activities and group contexts. The VR interventions were set in hospitals, research laboratories, and community settings e.g., participant's home

Study	Provider	Hours	Frequency	Treatment summary	Comparator	Primary Outcome	Post-Therapy Results	Maintenance
Carragher et al., '20	SLT	20hrs	5 days a week for 5 weeks	*Aim*: To improve participant's story planning, production and content *Content*: Narrative structure training from video retells	Time	Essential story content	Large effect size (Cohen's d = 2.06) for producing essential content between mean pre-therapy score 5.33, (SD 1.03) and mean post-therapy score 12.5 (SD 4.8)	Follow up scores were combined with post therapy scores to give a mean post therapy score
Cherney et al., '19	Virtual clinician	1hr	1 session	*Aim*: To improve accuracy and production of scripts *Content*: Single dose of computer script treatment with and without a rest in the training	Rest vs. no rest	NORLA-6	Baseline to immediately post-treatment increase in per cent accuracy (10.4%; p < 0.0001) and rate (7.8 words/min; p= 0.004)	Baseline to 2 weeks post-treatment significant for accuracy (5.0%; p= 0.006) and rate (10.9 words/min; p < 0.0001)
Cherney et al. '21	Virtual clinician	53hrs	6 days a week for 6 weeks	*Aim:* To improve language performance *Content:* 6 weeks of ORLA (repeated choral and independent reading aloud of sentences with a virtual therapist) or a control game, Bejewelled.	VC vs. control game	Western Aphasia Battery – Revised Language Quotient (WAB-R LQ)	There was no significant difference in the gain from pre-treatment to post-treatment for the Web ORLA versus control group.	The difference in the gain from pre-treatment to follow-up for Web ORLA vs Control groups = 2.70 (SD = 1.01); P= 0.013; effect size=1.92
Giachero et al., '20	SLT	48hrs	4 hours a week for 6 months	*Aim:* To explore: 1) does conversational therapy delivered via VR environments enhance language recovery? 2) do therapy benefits generalise to measures of communication efficacy and psychological well-being? 3) Is VR therapy equivalent or more effective than conventional training? *Content*: Conversations in virtual functional environment (n=18) compared to f2f topic-based conversation Tx (n=18)	VR vs. face to face	Aachen Aphasia Test (AAT)	No difference between the groups All improved from 'moderate' to 'mild' on AAT.	none
Grechuta et al., '17	SLT	27hrs	40min 5 days a week for 8 weeks	*Aim*: To test whether multisensory signals improve naming *Content*: 5 dyads received both silent cueing (lip shape) and acoustic semantic cueing e.g., engine revving for car.	Silent video cues vs. no cue	Interaction time: Time of successful goal-oriented peer-peer interaction	Significant, negative, correlation between interaction time and VocabT scores. Range: rs =−.88 (p < .05) to rs = −.98 (p < .005)	none

Study	Provider	Hours	Frequency	Treatment summary	Comparator	Primary Outcome	Post-Therapy Results	Maintenance
Grechuta et al., '19	EG: peer to peer VR CG: SLT	20hrs	8 weeks, 5 days a week, 30-40min sessions	*Aim*: outcomes from RGS training will be comparable to face-to-face SLT *Content:* The Rehabilitation Gaming System for aphasia provides lexical and syntactic training in a multimodal, goal-oriented manner within a context of dyadic peer-interaction	VR vs. face to face	Boston Diagnostic Aphasia Evaluation (BDAE)	No significant differences between the groups.	Gains in BDAE scores for all ppts were maintained at 16 weeks post baseline
Grechuta et al., '20	Peer to peer	23hrs	5 weekly sessions (30-40min) for 2m	*Aim*: Visual priming will facilitate word retrieval *Content:* 50% of stimuli presented with silent videos of correct pronunciation and was compared to stimuli without a visual cue	Silent cueing vs acoustic semantic cueing	Interaction time (time between hearing the cue and naming the object).	Difference between cued and non-cued trials was statistically significant during the early therapy sessions (N = 15) both for SVC ($p = .002$) and SAC ($p = .001$). But no differences were found in the late sessions for either SVC ($p = .73$) or SAC ($p = .53$)	Follow up at 16 weeks measured. VocabT scores maintained for all participants but no between condition results reported
Kalinyak-Fliszar et al. '15	Virtual clinician	2-3hrs	4x 30-40min session	*Aim*: To develop a dialogue practice tool with a virtual clinician and virtual simulations of activities of daily living. Researchers determined the differences between interactions with a human clinician (HC) and a virtual clinician (VC) *Content*: PWA participate in four dialogues, 2 with HC and 2 with VC	VC vs human clinician	Accuracy and content of dialogues	No difference between conditions for 3 cases e.g., equally inclined to interact with a HC or a VC. For 1 participant the quality of information conveyed in dialogues suggested a preference for the HC condition.	none
Maresca et al., '19	Clinician monitored computer system	100hrs	2x 12weeks 5 days week 50min	*Aim*: To compare the use of VRRS-Tablet to usual linguistic treatment. *Content:* Two phases: inpatient Tx followed by community services	VRRS tablet vs. usual linguistic treatment	Token Test	Statistically significant difference between the groups post treatment. Linear mixed effects analysis results showed that the scores of the Token Test ($X2(3) = 33.78$; $P < .001$; $R2 = .92$) were affected by the type of the rehabilitative treatment.	none
Marshall et al., '16	Communi-cation support workers	25hrs	1hr 5 days week 5 weeks	*Aim*: To improve functional conversation through situated conversations *Content:* Goal directed 1:1 intervention with a weekly conversation group.	VR vs. no treatment	Communication Activities of Daily Living-2nd edition (CADL-2)	Significant interaction effect, with better functional communication for EG ($F(1,18) = 5.236$, $p = .034$, $\eta p2 = .225$)	Gains were maintained at week 13

Study	Provider	Hours	Frequency	Treatment summary	Comparator	Primary Outcome	Post-Therapy Results	Maintenance
Marshall et al., '18	SLT	20hrs	1hr 4x week 5 weeks	*Aim*: To improve noun and verb retrieval *Content:* Semantic Feature Analysis or VNeST	Time	Action and Object Naming Battery	Nouns: A significant improvement in the naming of treated words (T2 vs T3, McNemar $\chi2$ $p < .001$). Naming of the untreated words showed no change. Verbs: A small increase following therapy (T3), which was maintained at T4. The change was not significant (T2 vs T3, McNemar $\chi2$ $p = 0.18$; T2 vs T4, McNemar $\chi2$ $p = 0.18$).	5 weeks. Noun naming well maintained at T4 (T2 vs T4, McNemar $\chi2$ $p < .001$).
Marshall et al., '20	Aphasia group co-ordinators	21hrs	14 sessions, once a fortnight for 6 months	*Aim*: To determine if social support groups in a virtual world are feasible *Content*: 14 social support groups delivered in EVA Park.	VR vs. no treatment	CADL-2 and Warwick Edinburgh Mental Well-Being Scale (WEMWBS)	No significant difference between the groups (treatment vs waitlist control)	none
Snell et al. '17*	Virtual Clinician	≥3min	3x ≥1min	*Aim:* To determine whether a virtual clinician that produced gestures facilitated the use of meaningful gestures in individuals with aphasia. *Content*: The VC demonstrates three narratives; 1) without using gestures (pre-gesture), 2) using gesture (gesture), 3) without gesture (post-gesture). The participants spoke after each narrative.	Exposed to gesture vs. not exposed	A gesture count and a range of movement measure	The gesture count was not significant between conditions. The range of movement was larger in the gesture condition compared to the pre-gesture condition ($p \leq 0.01$, $Z=-3.56$)	none
Thompson et al., '10	EG: Virtual Clinician vs CG: SLT	EG: 13.67hr CG: 14.25hr	1hr session 4x per week	*Aim*: To improve agrammatic sentence deficits *Content*: 14 sentences trained and compared computer delivered (VC) vs SLT delivered	VC vs. SLT delivered	NAVS	No significant differences between the two conditions	none

EG=experimental group, CG=control group, SLT=Speech and Language Therapist, NAVS=Northwestern Assessment of Verbs and Sentences. *Snell et al. 2017 reports on two studies, (A) a gesture taxonomy study and (B) a gesture production study. We have reported on study B only. Highlighted (grey) studies are those with a quality rating of fair or above

Table 2.8: Outcomes of VR interventions in included studies

What outcomes were achieved

Changes demonstrated on the outcome measures for all studies are presented in Table 2.8. Only studies rated as fair quality or above are included in this section (n=7) and are highlighted in grey in Table 2.8 (Cherney et al., 2021; Cherney et al., 2019; Giachero et al., 2020; Grechuta et al., 2019; Grechuta et al., 2020; Marshall et al., 2016; Marshall et al., 2020).

As reflects the preliminary nature of the research, two trials included feasibility outcomes. They compared VR to a no treatment control and were rated as fair on the PEDro quality measure (Marshall et al., 2016; Marshall et al., 2020). They demonstrated that virtual worlds showed promise for delivering a communication activity intervention (Marshall et al., 2016) and that online social support groups of up to eight people with aphasia with four additional support staff, each logging in from their own home, were feasible (Marshall et al., 2020).

Five studies employed a measure of language impairment as their primary outcome with all five reporting significant treatment induced improvement (Cherney et al., 2021; Cherney et al., 2019; Giachero et al., 2020; Grechuta et al., 2019; Grechuta et al., 2020). Three studies explored change in communication activity either as a primary or secondary measure, and all three reported positive change (Giachero et al., 2020; Grechuta et al., 2020; Marshall et al., 2016). Two studies (Giachero et al., 2020; Marshall et al., 2020) explored quality of life as a secondary outcome with one reporting positive change (Giachero et al., 2020). One study explored change in activity/participation as a secondary measure and reported positive change (Grechuta et al., 2019). One study employed a measure of wellbeing as their primary outcome measure and reported no significant changes (Marshall et al., 2020)

Four trials compared VR with face-to-face speech and language therapy (Giachero et al., 2020; Grechuta et al., 2019; Maresca et al., 2019; Thompson et al., 2010). Only two of these studies were rated fair or good quality (Giachero et al., 2020; Grechuta et al., 2019). They both demonstrated no difference between the groups on the primary outcome measures. These findings offer preliminary evidence of equivalence between VR and face-to-face therapy.

DISCUSSION

The use of VR in the rehabilitation of aphasia is in the exploratory stages of research. In this review, four forms of VR were seen, none of which were immersive. Rationales for employing VR varied and despite the ecological validity offered by multi-user virtual environments, VR interventions predominantly targeted language impairments and used language impairment measures as the primary outcome. Most interventions used previously published protocols with two technologies making use of the simulated real world environments available in VR. In terms of outcomes, improvements in language impairment (n= 5 studies) and communication activity (n=3 studies) were achieved through the use of VR for aphasia rehabilitation. When compared to face-to-face therapy there was a suggestion that VR interventions achieve equivalent outcomes to face to face therapy (n=2 studies), with one study reporting added benefits to communication activity/participation.

Four forms of VR have been used in aphasia therapy research to date. They sit on a spectrum that ranges from a constrained, pre-programmed task with a static view to an open virtual space that can be used by multiple people for multiple activities. The computer programmed virtual clinician, often only the clinician's head, is a static view with computer delivered tasks (Cherney et al., 2021; Cherney et al., 2019; Kalinyak-Fliszar et al., 2015; Snell et al., 2017; Thompson et al., 2010). The replicated table with objects from RGS widens the lens to the table, the arms and the peer sitting opposite but the view is still static (Grechuta et al., 2017; Grechuta et al., 2019; Grechuta et al., 2020). In the virtual scenarios of everyday communication situations, a whole environment is represented, e.g., a train station. The view is moveable and unexpected communication tasks are presented (Giachero et al., 2020). Finally, EVA Park presents a whole island environment where some elements of the environment are interactive. Participants are represented by a personalised avatar. They moved around the island, chose how they would be represented in the virtual space and meet multiple other users. Images 1-4 (Appendix 5) depict examples of the forms of VR used in aphasia rehabilitation.

Recently, a framework for describing situated language has been proposed (Doedens & Meteyard, 2022). It defines language use as: (1) interactive, (2) multimodal, and (3) contextual. As the spectrum of VR spaces opens up to encompass multi-users within rich environments, they offer an opportunity for embedding situated language use into aphasia rehabilitation. When VR is selected as the mode of treatment delivery, the possibilities are wide ranging. VR is not bound by the constraints of geography, physical laws (e.g., in EVA Park you can fly), or physical impairments. There is a potential to develop novel, creative multi-user environments that address the particular issues pertinent in aphasia rehabilitation: how to improve situated language use, mitigate the loss of social networks, and support the renegotiation of identity (Konnerup, 2015). This review has demonstrated that this potential is not yet being fully realised. With the exception of EVA Park, where experiences outside of the bounds of reality are possible (e.g., avatars can ride on a turtle underwater), uses of VR in aphasia rehabilitation currently replicate reality, or even the constraints of the aphasia clinic room.

Turning to rationales given for using VR, researchers cited a variety of rationales. Some aligned with the rationale for 2D online remote delivery; that it is accessible and provides increased intensity and dose (Hall et al., 2013; Palmer, Enderby, & Paterson, 2013).). A number of the studies use VR to replicate the clinical context and deliver treatments either remotely and/or independently of therapist input (Cherney et al., 2021; Cherney et al., 2019; Grechuta et al., 2017; Grechuta et al., 2019; Grechuta et al., 2020; Thompson et al., 2010). Receiving interventions to your home via your computer with the option to practice an unlimited amount in your own time has been shown to be acceptable (Des Roches & Kiran, 2017; Palmer et al., 2019). Other rationales were cited that relate to the unique properties of VR, such as the opportunity to situate practice within multimodal simulations of the real world, and these were cited by seven studies. The rationale for the unexpected nature of fantasy elements in EVA Park was that it motivated genuine conversational exchange.

Some rationales were specific to multi-user VR. Gaming research shows users play multi-user games because of 1) warm relationships with others, 2) a sense of accomplishment, 3) a sense of belonging, and 4) fun and enjoyment (Lin, Y., Lin, H., & Yang, 2017). These align with speech and language therapy aims to improve social networks and social connectedness (see Marshall et al., 2020). Participant views on the multi-user EVA Park interventions similarly included fun, humour and warm relationships (Amaya et al., 2018; Galliers et al., 2017). Seven of the studies in this review could be described as multi-user. For one study the interaction with the technology was as a single user but the therapeutic experience was in a group of three (Giachero et al., 2020). Perhaps unsurprisingly, the studies with multi-user technology were those that cited socially motivated rationales e.g., ecological validity, socially embedded, social networks.

It is interesting to reflect on the rationales that were not cited in the papers. A theme in gaming literature is the *agency* of the user. The degree to which the user can effect change in the virtual space is considered important (Cole & Gillies, 2021) because agency is said to support users in feeling like they are really there; a concept called immersion (Cairns, Cox, & Nordin, 2014). There is a suggestion that immersion is a potential mechanism of change in VR for health. For example, in virtual reality exposure therapy it is proposed that the treatment works because the virtual representation is *real enough* to elicit the anxiety (Maples-Keller et al., 2017). Seven studies in this review reported that the user was represented in the world. In the four EVA Park studies the users have agency of movement in the virtual space. Additionally, users in EVA Park can interact with objects in the virtual space and can effect some limited changes. None of the studies alluded to concepts of immersion or presence in their rationales. Similarly, none referred to the experiences of other disability groups with VR. In summary, only some studies are citing rationales that are unique to VR environments. There is a potential that has not yet been explored.

In relation to outcome measures, all studies but one reported multiple outcome measures. In the study that reported a single score, the WAB LQ (Cherney et al., 2021), the score is calculated from multiple language tasks (reading, writing, speech, comprehension, repetition and naming). The outcome measures covered all domains of the ICF with language impairment most represented. Additionally, QoL, confidence communicating, wellbeing, self-esteem and depression were measured. The wide range of measures used has implications for evaluating the efficacy of these treatments. For meta synthesis to be carried out, the measurement of outcomes needs to be rationalised. A core outcome set

(COS) for aphasia research was agreed in 2018 (Wallace et al., 2018). Eight studies from this review were published after 2018 and yet only one measure from the COS was used in the included studies, the SAQOL-39g (Marshall et al., 2020). The Scenario Test was recently named as the measure of communication activity in the core outcome set for aphasia (Wallace et al., 2018). The Scenario Test was not used by any of the studies in this review.

With regards to the communication functions targeted, VR in aphasia rehabilitation has been used predominantly to rehabilitate language impairments (12/14 studies reviewed). This finding is consistent with the physiotherapy literature (Palma et al., 2017). Communication activity is cited as a priority for both people with aphasia and clinicians (Wallace et al., 2017; Worrall et al., 2011) and multi-user simulated environments are uniquely placed to target this (Picano et al., 2021). This could have driven a rise in use of VR for communication activity, however, communication activity was the primary outcome in only one study in this review.

The interventions often used familiar therapies delivered in the novel format of VR. This was the goal in some treatments. For example, virtual clinicians freed up therapist time while increasing patient dose. It is interesting that, to date, VR has not taken aphasia therapy in radical new directions. In some cases, (Giachero et al., 2020; Marshall et al., 2016) there was an attempt to exploit the virtual environment to promote generalisation of skills or to address multiple levels of the ICF, but this is not possible on all platforms or not attempted in all studies.

The completeness of intervention description in this review was comparable to an umbrella review of intervention descriptions in aphasia (Dipper et al., 2022). The umbrella review, 50% of studies scored 8 out of a possible 12 items, and in this review 9/14 (64%) studies scored 8 or more out of the possible 12 items. The location of the intervention, the fidelity plan and adherence were the missing items in both reviews. This makes it difficult to replicate therapies, with implications for developing the evidence base and implementing therapies in clinical practice.

Positive language impairment outcomes have been achieved in VR based interventions. When these outcomes were compared to face-to-face delivery, in two studies of ‘fair’ and ‘good’ quality, they were equivalent. This finding mirrors what is known about non-VR computer delivered speech and language treatments where online and face-to-face delivery were equally effective (Spaccavento et al., 2021; Woolf et al., 2016).

There was some preliminary evidence that VR delivered therapy can achieve change in other dimensions of the ICF, with changes seen in communication activity (n=3, fair or moderate quality) and participation (n=1, good quality), and improved quality of life (n=1, fair quality). Changes beyond the language impairment were seen only when interventions were delivered in multi-user virtual environments where there were opportunities to converse with peers with aphasia (Grechuta et al., 2019; Grechuta et al., 2020) and/or therapists (Giachero et al., 2020; Marshall et al., 2016). This finding adds to the argument that therapy activities must target areas beyond impairment if we are to see these gains (Webster et al., 2015).

Feasibility outcomes from the feasibility studies (n=2, fair quality) in this review were positive (Marshall et al., 2016; Marshall et al., 2020). This review was confined to experimental studies. However, qualitative investigations linked to these studies have explored the acceptability of VR studies to people with aphasia (Amaya et al., 2018; Galliers et al., 2017) and service providers (Caute et al., 2021) show that VR interventions are acceptable.

Outcome evidence is only indicative but suggests that VR treatments are feasible and can achieve similar gains to those reported from face-to-face therapy. However, the strength of evidence is weak and many issues relating to the potential outcomes from VR therapy remain unexplored.

Limitations

A limited number of studies met the criteria for this review. Despite opening the criteria to include acquired cognitive communication disorder, no articles were found that reported use of VR to remediate language in this area. This finding was similar to a recent review (Brassel et al., 2021) where authors expanded their search after finding no examples of designing VR for traumatic brain injury. None of the studies included in the review were definitive studies, demonstrating how new this field of research is. The criterion to include only intervention studies may have excluded some of the more creative and novel developments of VR. These are published as user testing or platform development articles. For example, a head worn display that provides vocabulary cues in context (Williams et al., 2015). Qualitative literature about the experience of people with aphasia using VR was also beyond the scope of the review, but this literature points to good acceptability (Amaya et al., 2018; Galliers et al., 2017). Moreover, only half of the studies were rated as of fair or higher quality with the risk of bias high amongst single case studies.

The participants were younger than is typical for the stroke population. The national average age for a first stroke is 68 years for males and 73 years for females (Public Health England, 2018). The relevance of age, e.g., with respect to technology uptake, is not clear cut (Menger, Morris, & Salis, 2020). Nevertheless, it would be desirable to test VR treatments with participants who more closely reflect the age of typical stroke survivors. Wider demographic data were not always available and were not extracted from the studies. It would be important to explore such variables, again to ensure that participants reflect the intended user population.

Another problem for this review was the disparate primary outcomes. This meant meta-analysis was not a sensible option. A previous review carried out a meta-analysis of a maximum of two studies (Cao et al., 2021).

Directions for future research

VR has the potential to create novel, multi-user spaces that engender fun and a feeling of belonging. They can mimic the real world or extend into fantasy. Such spaces may help to address aphasia and its negative social and emotional consequences. There is a need for well-designed empirical studies that explore the impact of multi-user VR interventions on the ICF domains of activity and participation and beyond. Potential variables to be explored are a) the impact of immersion and presence, to investigate whether this a mechanism for change in aphasia rehabilitation as it is suggested in the VR exposure therapy literature, b) the impact of autonomy/agency in using VR and/or creating within VR e.g., personalising avatars, and c) the impact of receiving treatments in beautiful, playful spaces. Collaborations between the gaming community, human-computer interaction design and speech and language therapy may achieve this potential.

There is a need for larger studies and more studies with a low risk of bias to provide definitive evidence. Designs should also explore a wider range of questions such as whether VR shows equivalence to face-to-face therapy and whether there are added benefits of VR, for example with respect to generalisation and maintenance of change. Additional outcome issues (e.g., cost effectiveness) could also be explored.

Consistent use of the Core Outcome Set for aphasia rehabilitation (Wallace et al., 2018) will support future meta-analyses in this field.

CONCLUSION

VR for the rehabilitation of aphasia is being used for predominantly impairment level interventions with non-definitive evidence of positive outcomes. The rationales for using VR vary across studies, from releasing SLT time to creating ecologically valid environments. There is a need for future studies to strengthen the evidence and explore the particular benefits of VR over other technologies. The opportunity to create novel multi-user spaces for communication activity gains has not been exploited.

Chapter 3 | Intervention Development: Underlying theories and programme modelling

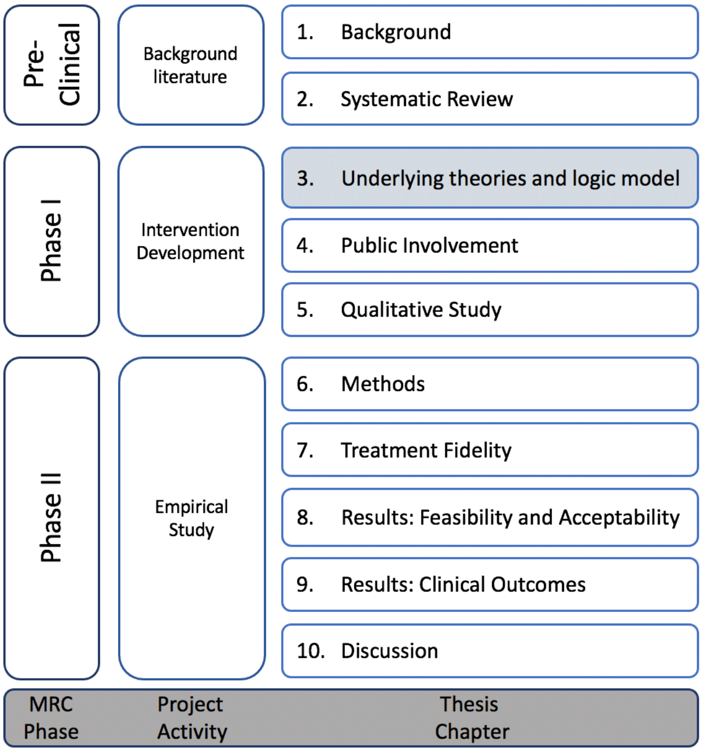

Figure 3.1: The thesis in a figure

The next three chapters address the development of the VESFA intervention, see Figure 3.1. The VESFA intervention aimed to tackle some of the issues highlighted in Chapter 2. Namely to develop an intervention that explores the impact of situated language therapy on language impairment, communication activity and quality of life. In this chapter the theories that underlie the therapy components are described. The actions that drive the outcomes of the intervention are articulated in an intervention model. This chapter, and Chapters 4 and 5, addresses doctoral objective 2:

To create an intervention protocol, VESFA (Virtual Elaborated Semantic Feature Analysis), that addresses word retrieval in EVA Park and the use of word retrieval in situated conversations

Background

A well-articulated development process breaks down a complex intervention to make clear the reasons why an intervention should work. Articulating the intervention development for VESFA specifies what each component of the intervention set out to achieve. It reviews previous evidence to build on current knowledge and minimise duplication. It ensures that the therapy is theoretically based and that rationales for activities are sound. Public involvement allows the therapy to be relevant and acceptable to those who will receive and deliver the intervention and supports future implementation. Clearly articulated inputs linked to outcomes support the testing of the intervention and ultimately the adopting of these interventions in clinical practice.

There are a number of published approaches to developing interventions. A systematic review explored these in healthcare settings and published a taxonomy of approaches (O'Cathain et al., 2019). The authors identified eight approaches that included co-producing interventions with end users (Voorberg, Bekkers & Tummers, 2015), following a stepped approach to development (Wight et al., 2016) and an 'evidence and theory-based approach'. This is the approach most used in the development of interventions for aphasia and the approach taken in the development of VESFA. In an evidence and theory-based approach interventions are developed by combining research evidence and published theory. An example of such an approach to intervention development is the Medical Research Council (MRC) framework for developing complex interventions (Craig et al., 2008). It has five phases: review, development, feasibility testing, evaluation and implementation, see Figure 1.6. The framework supports researchers, policymakers and clinicians in developing complex interventions. Complex interventions include multiple components and may have multiple outcomes. Behavioural speech and language therapy interventions are complex, they occur within a therapeutic relationship, can target multiple levels of the International Classification of Functioning (ICF), and can include a number of active ingredients aiming for a number of outcomes.

There are few accounts of intervention development for aphasia in the literature (Kirkevold et al., 2014; Pitt et al., 2019; Tarrant et al., 2016; Trebilcock et al., 2021; Volkmer et al., 2021; Wray et al., 2021). All accounts use an evidence and theory-based approach. Two accounts (Trebilcock et al., 2021; Wray et al., 2021) used the behaviour change wheel, which is based on behaviour change theory (Michie, van Stralen & West, 2011). The behaviour change wheel outlines eight steps that identify what needs to change, how to make the change and how to deliver an intervention. Most accounts of intervention development in aphasia rehabilitation use the MRC framework for developing complex interventions (Kirkevold et al., 2014; Pitt et al., 2019; Tarrant et al., 2016; Volkmer et al., 2021). Kirkevold (2014) outlined two steps from the MRC framework: 1) identifying evidence and theory and 2) modelling processes and outcomes. Pitt (2019) followed three stages in the development of teleGAIN: 1) identification of existing literature and relevant theories to determine the core components, 2) consideration of expert opinion and 3) a feasibility trial. Volkmer (2021) described a 6-stage process: 1) examining literature 2) consolation and co-production 3) initial draft 4) consensus work with speech and language therapists 5) focus groups with people with aphasia and 6) refinement of the manual. In summary, intervention developments in aphasia use an evidence and theory-based approach, most commonly the MRC framework. However, there is no clear precedent or consensus on the number of stages to include from that framework.

Intervention development falls into phase I of the MRC framework. However, information on phase I in the MRC framework is limited (Craig et al., 2008; 2013). In response to this, a consensus study was carried out to provide more guidance on phase 1 of the MRC framework (O'Cathain et al., 2019). Pulling together reviews, published empirical accounts and a consensus discussion with key stakeholders, a set of actions to guide intervention developers was agreed (Table 1.2). This chapter will describe how two of those actions were applied in the development of VESFA, outlining the underlying theories and articulating the mechanisms of the intervention (Table 1.2 signposts where other actions are described).

The end point of intervention development work is a therapy manual that outlines the activities and mechanisms that are “reasonably expected to have a worthwhile effect” (p.9

Craig 2008). Outcomes of this intervention development are a logic model (described below), the VESFA TIDieR and the VESFA Therapy Manual (see Appendix 7 and 8).

Review of research evidence

The core components of the VESFA therapy were the retrieval of known words, the production of words and phrases and the use of those words and phrases in situated conversations, a full description is available in Tables 3.1 and 3.2. The VESFA intervention refined existing protocols to create an intervention that addressed word retrieval and word use in situated conversations. It drew on the ESFA protocol from Efstratiadou and colleagues (Efstratiadou et al., 2019), the conversation therapy of Marshall and colleagues (Marshall et al., 2016) and the call to situate language treatments in functional situations from Doedens and Meteyard (Doedens & Meteyard, 2018; 2022). This research evidence for ESFA was described in Chapter 1. The case for using virtual worlds for ecological validity was presented in Chapter 2.

This chapter describes the research evidence for further elements of the VESFA therapy; the use of reflection, the theory behind the feedback used and the evidence for dose. Each session finished with a reflection on what had gone well. This activity is based on Positive Psychology research. The feedback provided in sessions is based on the research for Metacognitive Therapy. Evidence for these approaches has not yet been presented, so is outlined here. Finally, this section covers the research evidence for dose and regimes.

Positive Psychology

Positive psychology is the study of positive experiences and positive personal traits that allow individuals to flourish (Seligman & Csikszentmihalyi, 2000). It contrasts with the illness focus in health and psychology. There is now a large body of evidence for positive psychology interventions, with six meta-analyses published over the last 10 years (see Carr et al., 2020 for a summary). A recent synthesis of 347 studies (over 72,000 participants) demonstrated small to medium effects in a number of domains including wellbeing, quality of life and reduced stress and depression (Carr et al., 2020). Positive Psychology Interventions are many and can be delivered on their own or as multi-element interventions. They include practicing gratitude, practicing forgiveness, identifying and using signature strengths, solution focussed coaching, appreciating nature, humour through recalling funny things and encouraging a state of flow.

Positive Psychology approaches underlie some interventions for people with aphasia, specifically those that identify how to live well with aphasia (Holland 2007; Manning et al 2020; 2022b; Sather et al., 2017; Shiggins et al., 2020; Worrall et al., 2016). These include identifying and using signature strengths (Holland, Nelson, & Goldberg, 2014) regulating emotions by drawing on positivity, determination and gratitude (Manning et al. 2022a), and exploring flow, a state of intense, focussed concentration where self-consciousness is lost and the experience of time is distorted (Sather et al., 2017).

In an early review of positive psychology interventions, ‘Three Good Things’ showed the biggest gains (Seligman et al., 2005). In this intervention participants wrote three good things every day and why they were good. This gratitude practice has demonstrated gains on the happiness index (Mongrain & Matthews, 2012; Seligman et al., 2005) and was adapted in the VESFA treatment. At the end of each group participants were encouraged to identify what is going well. The prompt for this activity used the wording from Solution Focussed Brief Therapy, ‘what were you pleased to notice?’ (Burns, 2006).

Metacognitive therapy

Metacognitive therapy is used to increase self-awareness, self-reliance and develop compensatory strategies (Wadams et al., 2022). It increases the client’s awareness of their errors and teaches them strategies to manage them, with the aim that they can notice and manage them independently. A recent review found only five studies with participants with aphasia that used metacognitive therapy (Wadams et al., 2022), which suggests limited use. The outcomes synthesis was unclear, with approximately half the participants improving. However, the review only included studies where metacognition was the primary outcome. Aphasia treatments often increase self-awareness and develop strategies to produce independence (Wray, Clarke & Forster, 2018; 2020), but may not be labelled as metacognitive therapy. It is likely this review did not capture the extent of the use of metacognitive strategies in aphasia therapy.

In VESFA encouraging the employment of strategies to support a task is used across sessions. The participant is encouraged to use the phrases developed in the ESFA session in the group conversations. The description of a word by using the features from the SFA task is encouraged as a strategy if the word itself cannot be named. The reason is twofold; it could self-cue and it could cue the listener. Both these strategies engage metacognitive skills. They require the participant to see the need for the sentence or the features strategy and use the sentence or features practiced at the relevant time. In addition, rationales were given for session activities and specific feedback was given to highlight the strategies employed that worked well. For example, it was explained to participants that the SFA chart activity strengthens the networks around the word, making word retrieval easier; or that the description of a word, as is practiced in the 'Articulate!' game, is a useful strategy if you can't think of the word.

Evidence on dose and regime

Dose describes how much of a therapy is given. The term comes from drugs trials but is also used in complex behavioural interventions where dose is harder to define. Most often dose is described in hours of treatment (Harvey et al., 2021) but this gives no information about the potency of the intervention delivered. For example, in 1 hour of semantic feature analysis one person could complete 3 target words using the SFA chart and another person 6 target words. A recent meta-analysis described interventions in terms of intensity (hours/week), dosage (total hours), frequency (days/ week), and duration (weeks). The individual data of 959 participants from 25 randomised controlled trials revealed the optimal dosage by outcome (REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators, 2022). The greatest gains for overall language (as measured by the WAB AQ) were seen with a dose of 20-50 hours (1200-3000mins) of therapy, and a frequency of 3–5+ days/week. The evidence on intensity for overall language was not clear; up to 2 hours was similar to 9+ hours and marginally less than 3-4 hours. These findings were based on the individual patient data (IDP) of 11 RCTs, >480 participants. The greatest gains for functional communication were for 14-20 hours of therapy, at a frequency of 5 days per week. This was based on data from 14 RCTs, >520 IPD. Notably, there were no gains for auditory comprehension if the therapy was for less than 20 hours, less than 4 days per week and less than 3hrs per week (based on data from 16 RCTs, 540 IPD).

A review that explored dose in SFA interventions found that the treatment duration ranged from 2-12 weeks, sessions per week from 2-4 sessions and length of sessions from 45mins to 2hrs (Efstratiadou et al., 2018). All included studies were a single case design. Higher doses had better outcomes. In low dose interventions (315-729mins, approx. 5-12hrs) almost all participants made changes to treated items but not untreated items. In high dose interventions (1260-1470mins, 21-24.5hrs) all participants made gains on treated items and in addition 9 of the 10 participants generalised gains to untreated items (Efstratiadou et al., 2018). This suggests gains spread to generalisation with a larger dose. Harvey and colleagues made a similar suggestion regarding functional communication in a meta-analysis of dose (Harvey et al. 2022). They suggested 'functional communication may have a higher threshold to show an effect of treatment due to increased demands on multiple levels of linguistic processing and cognitive skills' (p.2555).

Drawing on the reviewed literature, proposed VESFA treatment regimes were drawn up that met the criteria of a total dose of between 20-50 hours, over 3-5 days per week. These regimes were brought to the trial advisory group for consultation. As a result, a regime of 4 sessions a week, for 8 weeks at a total of 40 hours was selected. This decision is detailed in Chapter 4: Public Involvement.

Underlying theories

The core elements of VESFA, retrieval of words and use of words in situated conversations, are based on published theories of language, communication and learning, Figure 3.2. Specifically, this section describes how behaviour affects the neural language network (Hebbian Learning), our understanding of how language is processed (the hub and spokes model of semantic processing and the spreading activation theory), how communication functions in everyday life (the framework for situated language use), and how we learn new skills (scaffolded tasks and social cognitive theory).

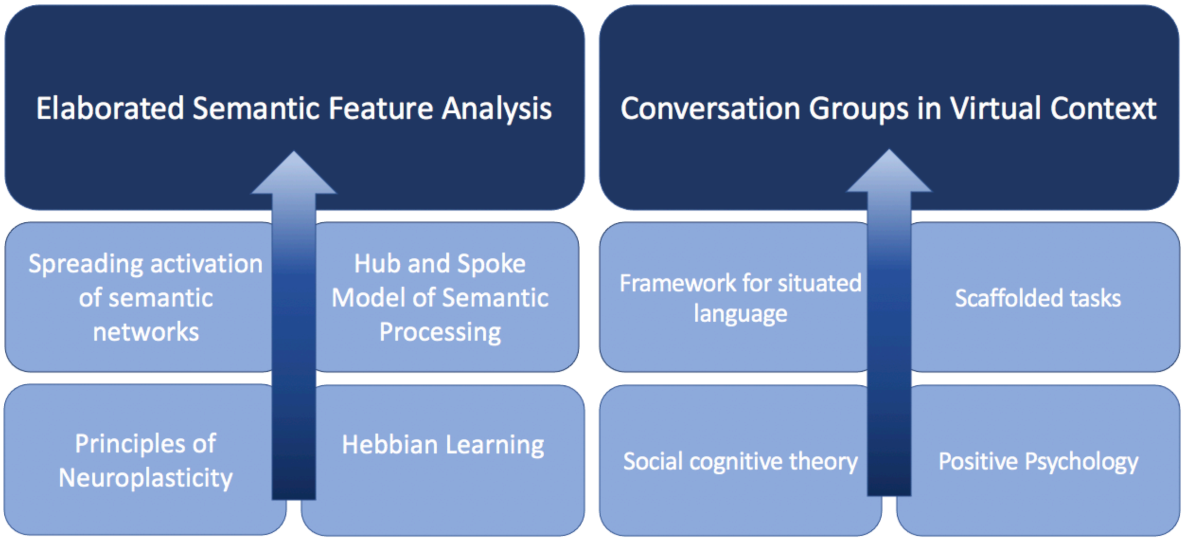

Figure 3.2: The theories that underlie the individual ESFA activities and conversation group activities in VESFA

Hebbian Learning and neuroplasticity

The principle of Hebbian learning was summarised by 'what fires together, wires together'. Hebb (1949) proposed that if a neuron B is repeatedly activated immediately after neuron A, then over time activating A will activate B (Keysers & Gazzola, 2014). Behaviourally this means that practice of a specific task improves our performance of that task because the neural firing becomes more efficient. The principles of neuroplasticity expand on this concept to outline the conditions that optimise the reorganisation of neural networks (Kleim & Jones, 2008). These principles are derived from animal studies. Further work explored these in the context of aphasia research to identify six principles of neuroplasticity in the rehabilitation of aphasia (Kiran & Thompson, 2019):

1. Use, improve, or lose it
2. Specificity rebuilds targeted networks
3. Salience is essential
4. Repetition and intensity promote learning and consolidation
5. Promote generalisation, avoid interference
6. Complexity enhances learning and generalisation

In the VESFA protocol, the repeated practice of word production in the context of related items is hypothesised to strengthen the networks that link semantically related items, in line with Hebbian learning. This use of the semantic network improves the semantic network (principles 1, 2 and 4). The words chosen for topic lists were nouns with good naming agreement, beyond that the most complex items were chosen; the least imageable. Words were selected from databases to ensure naming agreement and imageability ratings (Snodgrass & Vanderwart, 1980; Bates et al., 2000; Roach, et al. 1996). The use of the target words in the conversation contexts promotes generalisation and brings the complexity of real world communication (principles 5 and 6). The participants share stories from their own lives within the topic conversations e.g., 'tell us where you went the last time you travelled'. This ensures the target conversations are salient; the conversations are based on sharing personal experiences with each other (principle 3). The novelty of the environment of EVA Park also provided salience. Interference (principle 5) is minimally explored in aphasia (Kiran & Thompson, 2019). It refers to the possibility that a similar function might interfere with a therapeutic activity to impede recovery, for example a bilingual participant had interference from a second language during treatment (Keane & Kiran, 2015). The concept of repetition priming was introduced in Chapter 1 (Oppenheim et al., 2010). In repetition priming latent activation supports a second repetition of a word to be quicker but can negatively impact subsequent naming of semantically related items. The priming of the recently named item interferes with the new task, increasing the chance of semantic errors. It is possible that this represents the interference discussed by Kiran and Thompson (2019). This account suggests that the suppression of errors during therapy might be desirable, to reinforce activation of the target. A case could be made for an errorless learning approach, with a hierarchy that moves from maximal to minimal cueing. In fact, in the VESFA therapy the SLT used a minimal to maximal cueing hierarchy. This was to allow participants to self-generate the word. While semantic interference remains a theoretical possibility, the literature to support the self-generation of the target and its features as a potent element of the intervention is extensive (see p.10 Penaloza et al. (2022) and Boyle (2010)).

The cognitive-linguistic model of language processing

The cognitive-linguistic model of language processing is described in Chapter 1 (Figure 1:3). It visualises the pathways and cognitive components involved in the recognition of word meaning (comprehension) and the production of words (expression) via a diagram of boxes and arrows (Whitworth, Webster & Howard, 2013.). It represents the steps to understand a word by hearing sound, recognising it as a known word, and accessing its semantics. It also depicts steps to produce a word from semantics, accessing a phonological plan and producing the word. However, this model gives no detail of what happens within central semantics. How semantic knowledge is stored is described by the hub and spokes model of semantic processing. SFA is underpinned by such a model.

Hub and spokes model of semantic processing

The hub and spokes model of semantic processing outlines how word meanings are processed within the semantic system (Patterson, Nestor & Rogers, 2007; Woollams, 2012). The model proposes that each semantic representation is made up of a modality-independent core concept (the hub) and its associated modality specific features (the spokes), see Figure 3:3. The core knowledge of an apple is seated in the 'hub', and its features (e.g., its colour, smell, texture, the actions/verbs associated with it, and linguistic label) are the 'spokes' connected to the concept. This model proposes that semantic representations, our semantic knowledge, are made up of multi-model connected networks.

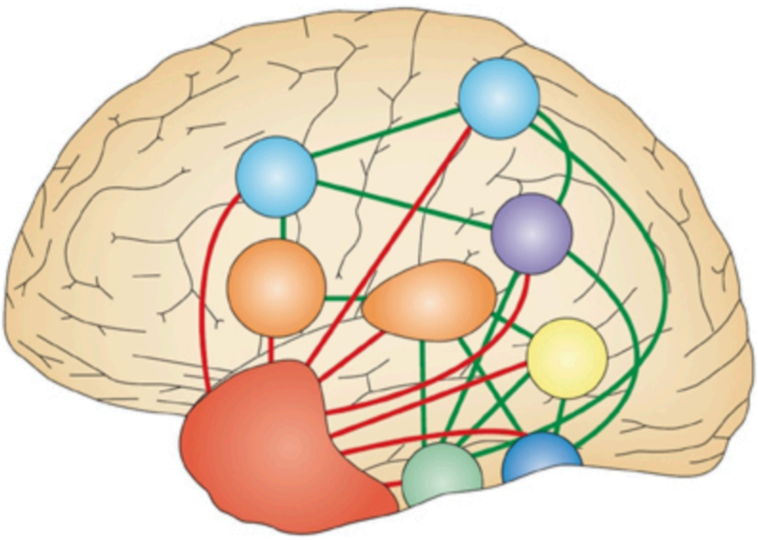

Figure 3.3: The hub and spokes model of semantic processing: The modality-independent core concept is represented in red, motion represented in yellow, colour (dark blue), shape (green), names (orange), actions (light blue) and task (purple). Image from Woollams, 2012 (p.2).

Studies comparing the semantic knowledge of people with aphasia and people with semantic dementia support the hub and spokes model (Jefferies & Lambon Ralf, 2006). These show that people with aphasia and semantic dementia both exhibit problems in naming and comprehending words, which may be comparable in terms of severity. However, patterns vary across the groups. For example, people with aphasia are often helped to name an object with semantic or phonological cues, whereas those with semantic dementia are not. This suggests that a semantic deficit in aphasia is due to damage to the spokes. You can prod the system with various cues, and you will get to the core concept. This is in contrast to a semantic dementia where the hub is damaged, and cues don't help (Jefferies & Lambon Ralf, 2006). In the individual sessions in the VESFA protocol, the actions of the SFA task, specifically naming a target and all its related features (supra-ordinate category, action, properties, location, personal association) systematically activate the 'spokes' around a core concept. The principle of Hebbian learning suggest that this repeated activation of the spokes and hub will strengthen these neural networks.

This model explains how a core concept is represented but not how we get from the semantics to the word. The spreading activation theory explains this.

Spreading activation theory

Spreading activation theory explains how we produce a target word through a process of spreading activation to prime and then retrieve words (Dell, 1986). Activation, the spark along the neural network, begins at the semantic concept and spreads, like water running down channels, to activate the word and then the phonemes needed to produce the word, see Figure 1.4. The activation, like water escaping down nearby routes, will prime related but incorrect words and phonemes, potentially resulting in naming errors. In an undamaged system, the majority of the activation will go to the target, making such errors rare. However, when the flow of activation is weak, as in aphasia, word retrieval failures and errors proliferate. This model is interactive. Activation can ricochet back up from the phoneme level. This converges on the target but will also relay some activation to words that share phonology with the target, thus providing a further mechanism for the production of phonological errors. Within this theory, correct word production relies on the contribution of semantic control. Such control is defined as the precise activation of semantic features and the inhibition of related items, so ensuring that the target word gains ascendancy over potential errors.

Some treatment approaches focus on semantic processing and others phonological processing (see Chapter 1 the 'word retrieval therapies' section), however any treatment that uses an activity in which a person retrieves and produces a word entails both semantic and phonological processing (Nickels 2002). Indeed, research that hypothesises that phonological cueing supports those who make phonological errors and semantic cueing supports those that make semantic errors remains unconfirmed (Lorenz A. & Ziegler W., 2009, Meteyard L., & Bose A., 2018). People with aphasia usually have deficits in both the semantic and phonological systems (Lorenz, A. & Ziegler, W. 2009), word finding treatments such as ESFA use both semantics and phonology in the naming of pictures and therefore likely strengthen the links between the semantics and the phonology of a word. In the VESFA therapy manual both semantic and phonological cueing were allowed, starting minimally (category or first phoneme) and increasing cueing support until the item was named.

In the VESFA protocol the target words are named every session. We can hypothesise that activation is sent from the semantic concept to the corresponding phonemes to produce words. Cuing from the SLT supports participants to activate the correct phonemes. Hebbian learning theory suggests that repeated practice of the pathway from concept to word for target words will strengthen that pathway.

Framework for situated language use

The model for situated language use describes how using language in a natural context, such as everyday conversations, is a more complex task than using language in an isolated cognitive linguistic task such as naming a picture (Doedens & Meteyard, 2018). Situated language is *interactive*, *multi-modal* and takes account of different *contexts* (Doedens &

Meteyard, 2022). Taking each of these concepts in turn, conversations are interactive because they involve two or more people who react to each other and the previous conversation turn. Conversations are multimodal because the meaning is derived from facial expression, gesture, intonation as well as the words used. Conversations have a number of contexts; there is the conversational context (what has been said already within this conversation), cognitive context (what you know this person knows already) and environmental context (where you are, what you can both see). Situated language is therefore complex. Neuroplasticity research advises that complex tasks may generalise to simpler tasks, but simple tasks will not generalise to complex one. The Complexity Account of Treatment Efficacy, CATE (Thompson, 2007), argues that targeting a more complex structure in therapy will lead to generalisation to a related, more simple structure. However, targeting a simple structure will not generalise to a related, more complex structure. This suggests that working on word production in isolation, as occurs in pure SFA, will not generalise to a complex task such as conversation. The CATE hypothesis cites both sentence and word level literature. For example, sentence processing therapy in agrammatism (Thompson et al., 2003) and word level treatments that treated atypical exemplars within a semantic category to drive generalisation to typical exemplars (Gilmore et al., 2020; Kiran, 2008; Plaut, 1996).

However, there are contradictions in the aphasia literature. For example, Springer and colleagues argue that the errors in aphasia could be compensation attempts, not the deficit, and simplifying language, as purposed in Reduced Syntax Therapy (REST), frees up processing resource (Springer et al., 2000). Developing a treatment on best available evidence is therefore not an easy task for a clinician. The VESFA treatment took the approach that retrieving words in the complexity situated language should be targeted in treatment if a change to situated language is what is hoped for (Webster et al., 2015).

VESFA includes group conversation in a related virtual context; conversations about food and drink occur in the virtual café, conversations about gardening occur in the virtual greenhouse. This simulates the complexity of real world situated language: navigating multiple people in a related environmental context. It is hypothesised that situating conversation in the simulated environment of EVA Park can achieve more complexity than a conventional treatment setting.

Learning theories: zone of proximal development

Writing in the early 20th century about learning, Vygotsky proposed that child learners had a Zone of Proximal Development (ZPD). This represented the space beyond what someone could learn by themselves, the potential someone could gain with support from a knowledgeable other. People can be supported into this learning space by ‘scaffolding’.

Figure 3.4: The zone of proximal development (ZPD)

Scaffolding has been used beyond child development in the fields of learning disabilities (Stone, 1998), cognitive therapies (O’Neill, Moran & Gillespie, 2010) and communication support in aphasia (Gillespie & Hald, 2017). In VESFA, the concept of scaffolding was used in the conversation groups in tasks of stepped complexity. People with aphasia confirmed that retrieving words in open conversation was more difficult than retrieving words in a constrained picture naming task (in a PI workshop described in Chapter 4). In response to this feedback, tasks of increasing difficulty were scheduled in the conversation groups in the VESFA protocol. The first activity was to remember the words worked on that week, a recall task that required participants to give a single noun response. The second activity was to describe these words in the ‘Articulate’ game. This required words or phrases including verbs, adjectives and nouns. These activities served to activate the core vocabulary for the next task (Creet et al., 2019; Nickels, 2002). The third activity was the topic-based conversation. Here participants shared something about themselves on the topic of the group. This was the least scaffolded and most complex of the activities in the group. The final activity increased the scaffolding again. This involved a bingo game in which participants had to name pictures of the target vocabulary. Thus, the open conversation was sandwiched by more supported activities.

Learning theories: Social cognitive theory

Social cognitive theory (SCT) was first published in 1986 (Bandura, 1986). It proposes that learning occurs in a social context in the intersection between behaviours, reactions and interactions. Essentially, we learn from doing the task ourselves, watching others do a task and talking about it. The SCT includes the concept of self-efficacy. This refers to a person's belief that they can achieve the task in hand. There are a number of influences on self-efficacy. That you have done it before feeds the belief that you can do it again (mastery experiences), that you see someone else do it supports the belief that you can do it (vicarious experience), that others believe you can do it persuades you of your capability (social persuasion) and a positive mood influences your belief that you can do it (physiological arousal and effective states) (Bandura, 1997). Setting treatment in a group brings in these positive social influences to learning.

The importance of self-efficacy for rehabilitation was explored in a study of self-efficacy and stroke survivors (Szczepańska-Gieracha & Mazurek, 2020). Participants (n=99) were tested using the Generalise Self-Efficacy Scale (Schwarzer & Jerusalem, 1995) and other measures as they entered a rehabilitation unit and 3 weeks into their stay. Participants with strong self-efficacy had better outcomes on a range of measures covering wellbeing and functional skills (Szczepańska-Gieracha & Mazurek, 2020). This finding is in line with an earlier systematic review on self-efficacy in stroke self-management (Jones & Riazi, 2011).

A telehealth group intervention for people with aphasia cited SCT as the theory underlying the intervention (Pitt et al., 2019). In the remote groups, self-efficacy was supported through opportunities for communicative success (mastery experience), reinforcement of successful communication models by group members (vicarious experience), encouraging positive feedback from peers (social persuasion) and promoting opportunities for comradery and humour (affective states) (Pitt et al., 2019).

In VESFA, target words are practiced in group conversations to situate the retrieval in a relevant context but also to have the learning benefits of the group dynamic highlighted above. Stepped tasks and supported conversations (Kagan, 1998) aim to create a space for participants to experience communication success. Observing others share stories from their lives gives vicarious experience. Feedback from peers supports members to share stories. EVA Park provides a safe environment in which to gain self-efficacy. The risk of failure is less than in the real world, there is no ‘loss of face’, and exchanges are experienced as playful and fun (Amaya et al., 2018; Galliers et al., 2017).

Modelling Planned Process

‘Programme theory’ outlines the mechanisms through which the programme is expected to drive change (Rogers, 2008). The logic model in Figure 3.5 shows how the underlying theories are the assumptions that feed into the resources used. In turn, the activities are what lead to the intended outcomes (short term) and impacts (long term).

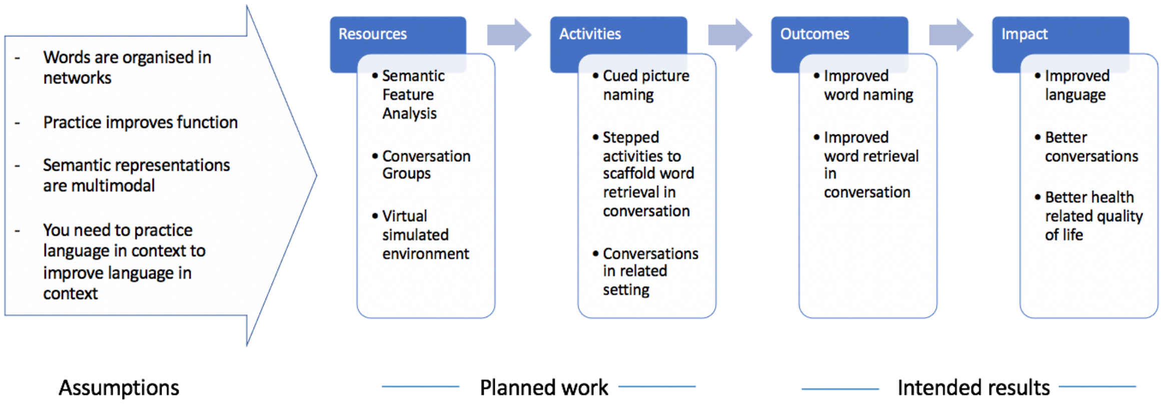

Figure 3.5: Logic model for the proposed intervention, VESFA

For example, the underlying theory tells us words are organised in networks (Patterson, Nestor & Rogers, 2007) and practice will strengthen connections (Kiran & Thompson, 2019). SFA activity leads the participant to generate words in their networks repeatedly. We can hypothesise that this will lead to improved word retrieval for treated words, those in the networks around the treated words and potentially unrelated words through stronger semantic systems.

The Template for Intervention Description and Replication (TIDieR) (Hoffmann et al., 2014) provides a framework for describing interventions. The TIDieR outlines 12 items that need to be reported in order for interventions to be replicable. The VESFA TIDieR is provided in Appendix 7.

The therapeutic actions of both the participant and the therapist and their underlying theory or rationales are described in Tables 3:1 and 3:2. These tables include the full list of activities which were set after refinement when the first group was completed (see 'Design and Refine' below). The core activities listed in these tables will be explored in the assessment of treatment fidelity described in Chapter 7.

Activity	Client acts	Therapeutic acts	Underlying theory/rationale	Dose
Opening	Understand the purpose of the chart activity	SLT gives rationale for activity	Metacognitive approach	Rationale is given once per session (x16)
ESFA chart activity	Generate word for a picture	SLT provides hierarchy of cues. This scaffolds the task	Principles of neuroplasticity	Approximately x6 per session, x12 per week
	Generate features of target word: category, use, action, description, location, association	SLT provides hierarchy of cues. This strengthens semantic networks based on Hebbian theory	Theories of semantic representation: Hebbian learning, hub and spoke model, spreading activation Principles of neuroplasticity	Approximately x6 per session, x12 per week
	Use of target word in phrase	SLT recaps the components that can be used e.g., verb, target and related items. This scaffolds steps towards groups tasks	Scaffolded, stepped tasks support the use of items from the feature generation tasks to prepare potential phrases for the complexity of the conversation group.	Approximately x6 per session, x12 per week
Close	Recap all words	SLT invited participants to name all targets worked on to date.	This repetitive practice is based on our understanding of neuroplasticity	Every session (x16)

Table 3.1: VESFA ESFA Treatment Specification

Activity	Client acts	Therapeutic acts	Underlying theory/rationale	Dose
Opening	Feels acknowledged, a sense that they belong in group	SLT greets each person.	Social Cognitive Theory. This activity builds rapport	Every session (x16)
	Share news	SLT invites the sharing of news.	Social Cognitive Theory. Builds rapport	Every session (x16)
	Group structure and topic is understood	SLT shares the topic and activities of the group	Metacognitive approach: providing a clear expectation of what activities members will need to take part in	Every session (x16)
Activity 1: 'Articulate!'	Understand the Articulate game	SLT describes the game, can include a model turn.	Metacognitive approach: to be clear to all members what is required in this activity	Every session (x16)
	Retrieve target words	SLT can provide features/categories if participants struggle to remember or name items e.g., 'We had 3 buildings'	Principles of neuroplasticity In line with the framework for situated language retrieving words in a group is harder than 1:1. This task also primes the relevant topic vocabulary for the next task	Every session (x16)
	Describe target words	Step towards use in conversations	Theories of semantic representation. Principles of neuroplasticity	Approx. x3 words described each
	Understand strengths	SLT provides specific feedback about what works well	Positive psychology and metacognitive approach	At the end of every description, thus approx. x3

Activity	Client acts	Therapeutic acts	Underlying theory/rationale	Dose
Activity 2: Conversation	Orientated to conversation topic	SLT introduces conversation task	Clear expectation, metacognitive awareness	Every session (x16)
	The virtual setting is linked to the topic	Simulated physical context provides cues e.g., discussion about what they had for dinner last night is had at the virtual dining table.	Framework for situated language use: virtual setting provides a shared, topic-relevant environmental context	Every session (x16)
	Contribute personal information to the conversation	SLT can support if participants get stuck. For example, can prompt participants to use SFA strategies by asking 'what is it used for?' or 'where do we find it?'	Framework for situated language use: the use of the target vocabulary is practiced in an ecologically valid setting. Recent supported tasks support this e.g., use of words in sentences, recap has primed the topic vocabulary	Every session (x16)
	Understand strengths	SLT provides specific feedback about what works well	Positive psychology and metacognitive approach	Every session (x16)
Activity 3: Bingo	Understand the BINGO game	SLT described the BINGO game, including modelling an answer	Metacognitive approach. To be clear to all members what is required in this activity	Every session (x16)
	Have a turn playing BINGO	SLT can provide cueing hierarchy to support word retrieval if necessary	Scaffolded word retrieval in conversation task. Principles of neuroplasticity	Every session (x16), includes retrieving min x4 target words

Activity	Client acts	Therapeutic acts	Underlying theory/rationale	Dose
Close	Identify own strengths	SLT asks 'what have you been pleased to notice?'	Positive Psychology: signature strengths	Every session (x16)
	Carry out a challenge task independently before the next session	SLT invites participants to practice activity e.g., try this conversation with someone in your social network	Neuroplasticity principles: repetition, salience, complexity	Every session (x16)

Table 3.2: VESFA Situated Group Treatment Specification

Design and refine

The VESFA therapy manual was designed through a review of published evidence (outlined in Chapters 1 and 2), illuminating the underlying theories and creating a logic model and TIDieR description of the VESFA intervention (this chapter), a process of public involvement (see Chapter 4), and a qualitative study to identify valid therapy targets (see Chapter 5). The manual describes an 8-week intervention delivered in the virtual world, EVA Park, that targets word retrieval in ESFA tasks and situated conversations. Participants received 4 sessions per week, 2x 60 minute individual ESFA sessions and 2x 90 minute conversation groups. Three people with aphasia received the treatment concurrently to allow for conversation groups of three people with the speech and language therapist. The VESFA therapy manual is available in Appendix 8.

After the first three participants had finished the 8-week treatment (set 1) in February 2021, the study team reviewed the intervention protocol. The ESFA sessions worked well, and no changes were made. Adjustments were made to the groups. The participants were more engaged in activities that were scaffolded than the open conversation and groups were often finishing 15 minutes early. So, an activity was added to the beginning of the group, the word description game ‘Articulate!’. In this game someone describes a word without saying the word e.g., “This is part of the landscape, it is tall and often rocky, we climb or hike on it" (mountain). This activity is well aligned with semantic feature analysis as the features can be used as clues. It also primes the vocabulary that will be used in the conversations. Specific feedback was also added after each task. The treatment review highlighted the need for more consistent, specific feedback to participants after their turn in the Articulate, Bingo and conversation tasks. The feedback aimed to highlight the target words successfully retrieved or the strategies that worked well. The manual (Appendix 8), the TIDieR intervention description (Appendix 7) and the treatment specification tables (Table 3.1 and 3.2) include these added tasks.

Conclusions

The VESFA intervention is a complex speech and language therapy intervention that targets word retrieval in isolation (picture naming) and word retrieval in functional conversations situated in simulated environments. It builds on existing empirical treatment research and theory to exploit the ecological validity of the virtual reality setting. The ingredients that drive the treatment outcomes have been articulated in the treatment specification above and the TIDieR description.

We can reasonably hypothesise improved naming of words treated. Research arguing that the whole semantic network is strengthened through SFA tasks (Boyle, 2004) and evidence of generalisation in the literature (Boyle, 2010; Efstratiadou et al., 2018) lead us to hypothesise that an independent naming measure of untreated words should also show improvement. Treating word retrieval within situated conversations leads us to hypothesise an improvement in word retrieval in discourse. Finally, the positive psychology activities (Carr et al., 2020) and wellbeing benefits of group participation and learning (Attard et al., 2015; Szczepańska-Gieracha & Mazurek, 2020) can lead us to hypothesise an improvement in mood and quality of life.

The next two chapters address further elements of intervention development. Chapter 4 describes the Patient Involvement (PI) activities and Chapter 5 describes a qualitative study undertaken to explore the most meaningful conversation topics to address in the VESFA therapy.

Chapter 4 | Intervention Development: Public Involvement

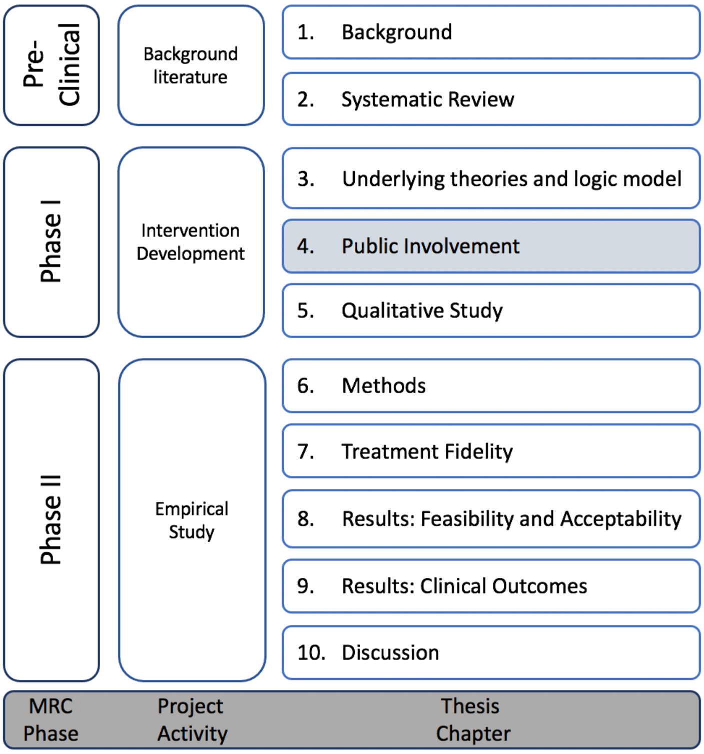

Figure 4.1: The thesis in a figure

Background

Public Involvement (PI) is the term used to describe research carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them (INVOLVE, 2012). The term 'public' describes patients and potential patients, carers and people who use health services. In this trial the researcher consulted people with aphasia to represent those who would receive the intervention and speech and language therapists (SLT) as key stakeholders in the delivery of speech and language therapy interventions.

Collaborating with end users enabled the researcher to plan procedures that were practically possible in the lives of people living with aphasia. This was particularly pertinent for a feasibility trial. PI activity aimed to ensure that developments were decided *with* end users and thereby better met the needs of people with aphasia and aphasia clinicians. Working with the groups that will implement and receive the intervention from the start is ethically indicated (Oliver et al., 2008) and is reported to have particular value in improving trial design in clinical trials (Staley 2009). The right of users to participate in the planning and implementation of care is published in the World Health Organisation (WHO) declaration (International Conference on Primary Health Care, 1978). The UK has established organisations such as NIHR Centre for Engagement and Dissemination and the NICE Public Involvement Programme (NICE, 2019) to support the active participation of users in research, demonstrating the commitment to public participation in research.

Public involvement with people with aphasia (PWA) poses specific challenges. Research context, discussions and decisions need to be accessible to users with a communication disability. Historically, this challenge meant that participants with aphasia were often excluded from research studies (Shiggins et al., 2022; Townend, Brady & McLaughlan, 2007) despite published methods to make information accessible (Kagan, 1998; Rose et al., 2012). In recent years, accounts of studies involving participants with aphasia as advisors, consultants or co-researchers are increasing (Cruice et al., 2021; Hilari et al., 2021; McMenamin et al., 2021; Roper & Skeat, 2022; Wilson et al., 2015).

This PI project explored the feasibility of VESFA. The planned therapy regime was a complex intervention to remediate word finding difficulties in people with aphasia. It had two levels: single words and situated conversations. It was proposed that participants in the study receive a total of 36 hours of individual and group speech and language therapy to improve the retrieval of words in 1:1 sessions, and to improve the use of words in group conversation sessions. Word retrieval was addressed using ESFA, an established semantic cueing therapy for word retrieval difficulties (Efstratiadou et al., 2019). The *use* of the target words was addressed by conversation groups where conversation topics included the words being targeted in the ESFA sessions, and the virtual environment corresponded to the topic i.e., gardening was discussed in the virtual greenhouse. The proposed regime was all online with testing on the Zoom videoconferencing platform (www.zoom.com), and both 1:1 ESFA and group conversations sessions being delivered in the online, multi-user virtual world, EVA Park.

Those who would deliver the therapy, SLTs, and those who would receive the therapy, people with aphasia, were consulted. SLTs were consulted on the feasibility of implementing the protocol in clinical practice and PWA were consulted on the content, format and delivery before the trial and interpretation of the results after the trial. Some elements in the VESFA protocol were not up for consultation; the study would include elaborated semantic feature analysis with conversation groups in EVA Park for a minimum of 36 hours. Elements open for discussion were 1) the content of the groups 2) how to deliver the 36 hours (how many sessions of what length over how many weeks) 3) experience an example session and provide feedback and 4) how best to disseminate findings. Involving people with aphasia and clinicians in the planning stages of the feasibility trial sought to improve the acceptability of the feasibility trial and the relevance of the research.

This group of four people with aphasia became the VESFA trial advisory group. They were consulted in the funded PI activities reported here but additionally they supported the training of testers (see Chapter 6), ratified the treatment fidelity checklists (Chapter 7) and explored the meaning of the results (Chapter 9).

Funding

The PI project was funded by the British Aphasiology Society (BAS) Initiatives in Aphasia seed fund. As such it met with BAS seed fund goals to promote knowledge of aphasia and aphasia treatments and make its evidence base accessible (see workshop 4). Workshops 1-4 focussed on promoting and drawing upon the expertise of PWA and on sharing, discussing and collaborating on research with PWA and their family members/carers (British Aphasiology Society, 2018). The funding covered travel to the workshops in the university, payment for time and refreshments.

Aims

1. Refine the goals and content of VESFA intervention in response to PWA and specialist SLT opinion
2. Develop acceptable research procedures: screening, testing, treatment activities and treatment dose

3. Explore the delivery of VESFA in the virtual environment
4. Explore the best ways to tell PWA and practicing clinicians about the findings
5. Understand how this treatment approach can be integrated and implemented into clinical practice

Methods

The Guidance for Reporting Involvement of Patients and the Public, GRIPP2 short form, informed the write up, see Appendix 9.

Workshops

The researcher collaborated with four PWA in five workshops that supported the planning of the feasibility trial. Workshops were chosen as a methodology to support a participatory, problem-solving approach where information flowed both from users to the researcher and from the researcher to the users. Workshops have been defined as an arrangement where 'a group of people learn, acquire new knowledge, perform creative problem-solving' (Ørngreen and Levinsen, 2017, p.71). This iterative approach allowed users to generate ideas on a topic and the researcher to share relevant literature and ideas to be revisited.

Workshop 1: Content	What goes into the therapy?
City, University of London	Explored and charted the content of group therapy
January 2020	Explored priorities for therapy
Workshop 2: Regime	How is the therapy delivered?
City, University of London	Explored delivery of assessment and treatment regimes
January 2020	
Workshop 3: Experience	People with aphasia experienced a taster VESFA session
City, University of London	Feedback on a taster VESFA session
February 2020	
Workshop 4: Tell	How best to disseminate findings to the aphasia community?
Online, March 2021	Explored dissemination methods

Table 4.1: Public Involvement workshops with people with aphasia

The four workshops met aims 1-4 (see Table 4:1) and the clinician focus group met aim 5. Topics were posed for each workshop: Content, Regime, Experience and Tell. Discussions were supported by pen and paper activities, see Image 4.1. Due to the COVID lockdowns workshop 5 was held online in March 2021.

In workshop 1, Content, PWA explored the content of the group therapy. Users generated examples of good/poor practice in conversation support groups from their experience. These were written on post-it notes. The group then sorted the post-it notes into three columns; 1) activities to keep, 2) activities to keep but change in some way and 3) activities to avoid, see Image 4:1.

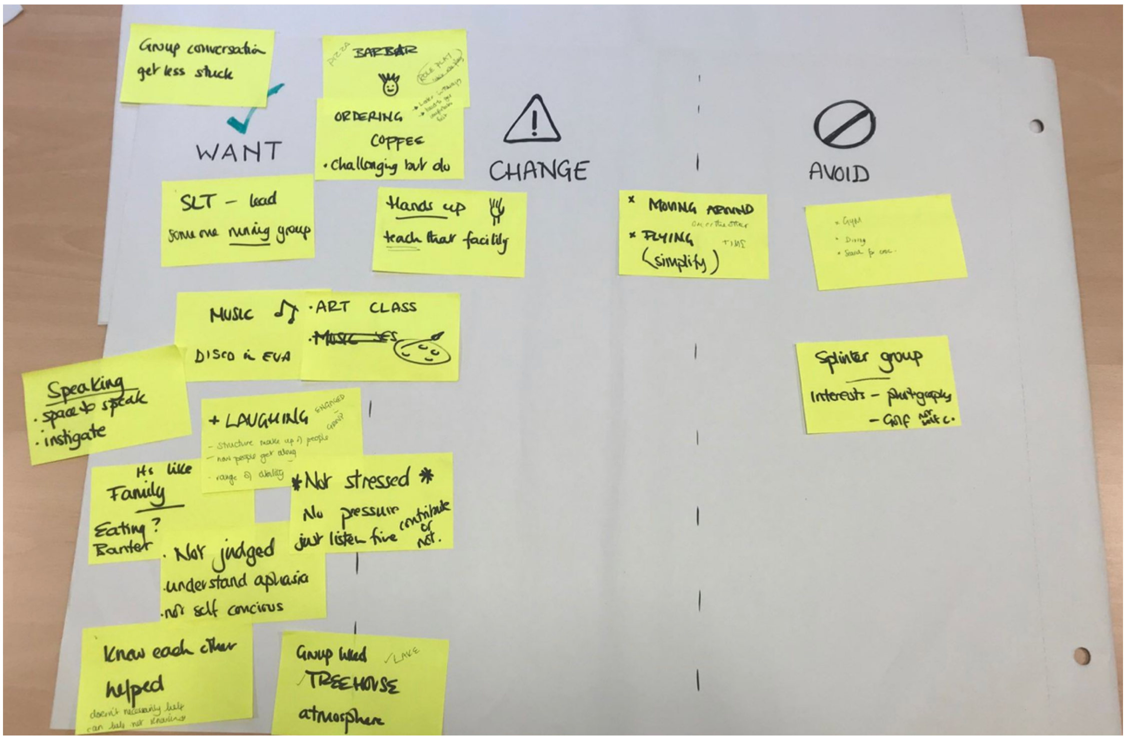

Image 4.1: Elements of group therapy to keep, change or avoid.

Users also discussed and agreed the percentage of the overall treatment time that should be dedicated to the naming activities and conversation activities. For example, should more time be given to naming tasks to remediate the impairment or to conversation tasks that might improve communication activity and participation? An A4 page of key research findings of relevant literature with visual supports was created (Wallace et al., 2017; Worrall et al., 2011) and this supported the discussion, see Appendix 10. The literature showed what a sample of people with aphasia and their families reported they wanted as therapy outcomes.

In workshop two, Regime, PWA were consulted on the delivery of assessment and treatment protocols, e.g., how long testing sessions should be, how often therapy sessions should occur and over what time period to make up the total of 36 hours. PWA were given a range of modelled regimes as a basis for the discussion, see Image 4:2. They were consulted on their own opinion and the opinion of a SWIM (Someone Who Isn't Me). This technique in a way doubles the pool of people being consulted. Each user thought of someone who was not themselves, but had aphasia, and how they might view the treatment regime (Wilson et al., 2015). Again, relevant literature was shared in an accessible A4 summary format. These covered the core outcome set for aphasia therapy (Wallace et al., 2018) to provide a rationale for the assessments proposed, and what was known about naming therapy dose (REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators, 2022).

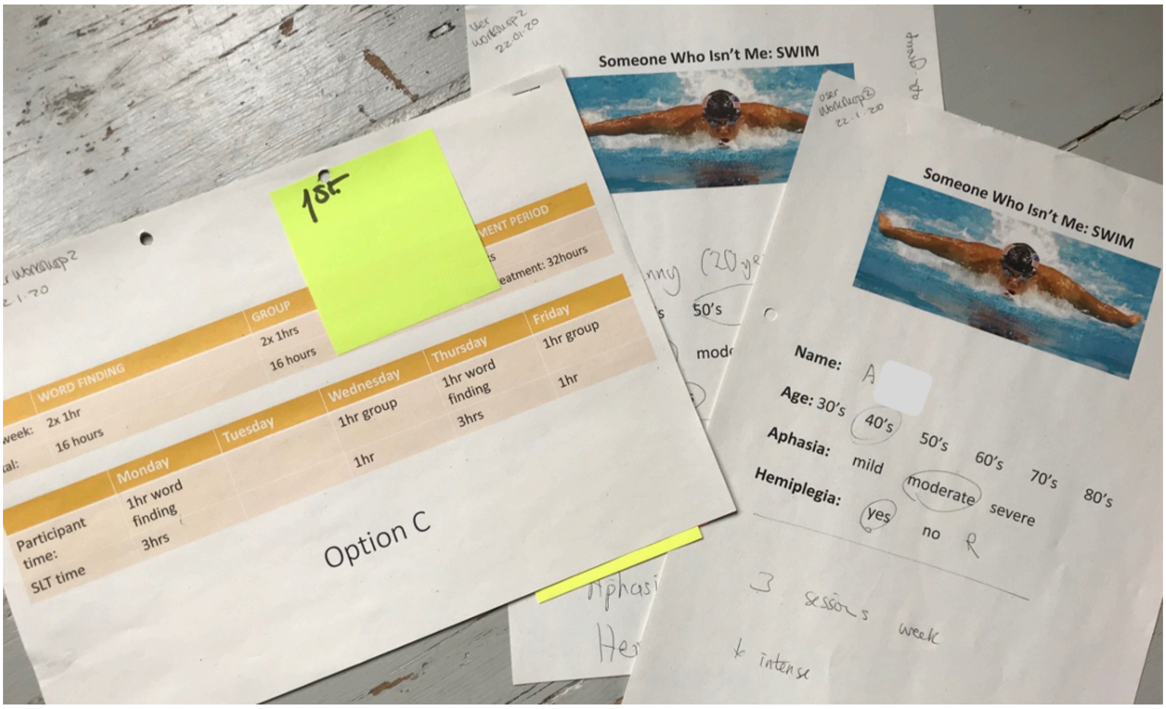

Image 4.2: Regime examples

In workshop three, Experience, the PWA experienced a taster VESFA session. They logged into EVA Park and experienced both 1:1 ESFA work and a situated group conversation. In the workshop, two topics were trialled: food and drink, and gardening. First, the ESFA activity worked on food and drink vocabulary and the group activity was a role play in the virtual cafe. Then, the ESFA activity worked on gardening vocabulary and the group was a conversation in the virtual greenhouse. After the taster session in EVA Park the PWA gave feedback on the session using a questionnaire format. This posed questions about how they enjoyed the session (Was there anything in the session you particularly liked? Anything you particularly didn't like?) as well as the cueing methods used and suggestions for alternative methods. They discussed the merits of the techniques used to scaffold the activities, e.g., having in the target images on a prompt sheet to support naming in conversation, and were consulted on alternatives.

In workshop four, Tell, the users shared what research they had heard about and how they had received the information. The benefits of different methods of receiving research findings were ranked. The PWA considered the views of Someone Who Isn't Me (SWIM) to widen the scope of opinions (Wilson et al., 2015).

Focus group

A focus group with aphasia clinicians aimed to understand how the VESFA treatment approach could be integrated and implemented into clinical practice. A focus group was chosen as it can identify consensus and agreement within a group (Tausch & Menold, 2016). The focus group was carried out online to adhere to COVID-19 lockdown guidance.

Clinicians were asked to share their experience of situated language therapy. Aspects of the situated therapies that could translate to virtual world therapies were discussed. They were presented with the planned VESFA protocol and discussed the feasibility of running the protocol within their clinical settings.

Participants

Four PWA took part in all four PI workshops, see Table 4:2. We purposely invited PWA who had previously received treatment in EVA Park to give their opinions because they had experienced the potential of the EVA Park space. This enabled the discussion to focus on the protocol for this project and not what EVA Park was or how it worked. Workshops 1-3 took place in January and February 2020, the 4th workshop was delivered on Zoom in March 2021 due to the COVID19 pandemic.

Identifier	Sex	Aphasia	Time post stroke	Relevant experience
C1	Male	Mild, fluent	9 years	Advisor on other research projects Supported an aphasia group delivered in EVA Park
C2	Male	Severe, non-fluent	10 years	Received both 1:1 and group interventions in EVA Park
C3	Male	Moderate, fluent	11 years	Received a group intervention in EVA Park
C4	Male	Mild, fluent	10 years	Received a group intervention in EVA Park

Table 4.2: PI participants with aphasia

Three experienced aphasia speech and language therapists took part in an online focus group in October 2020, see Table 4:3. They were invited to take part if they were aphasia specialists and had experience of EVA Park.

Identifier	Sex	Aphasia experience	EVA Park experience
SLT1	Male	15 years	Managed multiple EVA Park projects including screening and testing participants and delivering interventions
SLT2	Female	15 years	EVA Park Early Adopter in the pilot study and managed EVA Park early adopter's project
SLT3	Female	28 years	Supported two social support groups delivered in EVA Park

Table 4.3: Speech and language therapists

Outcomes

The PWA identified therapy content, a therapy regime, strengths and weaknesses of the proposed sessions, interpretation of the results and methods for disseminating results. The clinicians ratified the resulting treatment protocol.

Content of the VESFA intervention

PWA identified what worked well in group interventions they had attended both face to face and in EVA Park. They agreed that groups should be led by a facilitator and should be topic based. They felt role play was a positive activity but not in the first few sessions. They agreed it was acceptable to meet online, and a face to face 'meet and greet' before the remote sessions was not necessary. PWA highlighted things that should be included to facilitate delivery in the virtual world; moving in EVA Park should be simplified and the 'hands up' function should be explicitly taught. They also highlighted things to avoid, namely not splitting the group up into subgroups.

Consensus was reached on the division of time to naming activities and conversation activities which informed the VESFA therapy protocol. The PWA agreed that equal sessions should be given to both. This view persisted following discussion of the literature.

Acceptable research procedures

The second workshop consulted on testing session parameters and how to deliver a dose of 36 hours. It was agreed testing sessions should be a maximum of 2 hours. Discussion about therapy regime focussed on other life commitments and their relationship to time post stroke. The PWA felt that daily intervention sessions were difficult to accommodate with other life commitments. However, if participants were newly discharged home from hospital, then daily sessions might be desirable. Consensus was reached on a more distributed regime. The protocol was adjusted from 5 sessions a week over 6 weeks to 4 sessions a week over 8 weeks. Each week had two 60min ESFA sessions and two 90min groups. Total hours increased from 36 hours to 40 hours for this distributed regime to work.

Delivery of VESFA

In the third workshop the PWA experienced a taster VESFA session in EVA Park and responded to a feedback questionnaire. Responses highlighted that navigating in EVA Park was challenging for some participants and should be kept to a minimum in early sessions. Group size should be a maximum of four participants. Comments from the group highlighted that word retrieval in conversation is more challenging than word retrieval in an ESFA picture naming task. C1 summarised the discussion: "So, what are we saying? That it's difficult to include it in a conversation in a whole group of you, but is it easier if it's a one to one?" There was a discussion about positive challenge e.g., asking for a coffee in the coffee shop made one consultant anxious, but it was a good thing to practise. The PWA agreed that providing opportunities for situated conversations was a strength of the proposed VESFA protocol "by going to the gardening that gets you in the mood for gardening things. When you're in the coffee shop you're thinking about food and whatever. So that whole thing is really good" (C1). Negative comments focussed on the technical aspects of EVA Park; learning to walk and fly around the island, issues with sound and hardware e.g., not having headphones.

Dissemination

Methods for disseminating research findings to PWA were explored in workshop four. Due to the pandemic, the workshop was carried out online in March 2021. Different methods of disseminating research findings were ranked. Videos of research findings were praised but users questioned how they could come to know about a video's existence. Facebook, YouTube and WhatsApp groups were recommended as channels of alerting people with aphasia to research videos. Leaflets were considered an outdated method, described as 'last century'. One user highlighted that he was only interested in research that showed a functional difference to communication.

In summary, the PWA felt younger stroke survivors make good use of information on social media, but older survivors would only access face to face dissemination. A preliminary dissemination plan was agreed, see Table 4.4. The participants of the study, the aphasia community, family members, speech and language therapists and the general public should be informed. A video summary should be made and uploaded to YouTube with the search terms 'aphasia', 'research' and the study title, conference talks can alert people to the video. The key message should focus on communication gains and a famous spokesperson with aphasia, such as Emilia Clarke, Chris Ellison or Edwyn Collins, would improve the reach of the findings.

Who to tell	How to tell
Participants of the research	YouTube Video
Aphasia groups	YouTube Video
Families of people with aphasia	Website
Speech and language therapists	Conference Talk
General Public	TV

Table 4.4: Dissemination plan agreed in workshop four

Integration with clinical practice

Three SLTs took part in an online focus group.

Current practice

The SLTs were asked about their experience of working on language in context. Context-based language in *individual* therapy interventions included taking clients to functional settings in the community (coffee shops, swimming pools, fitness centres). Context-based language in *conversational* interventions included conversation partner training with family members in their homes. Context-based language in *group* interventions addressed life participation goals such as attendance at a concert or a pub night.

Translating current practice into virtual environments

SLTs were asked what elements of this current practice could translate into the virtual space. SLTs emphasised that the challenge of real world context should be maintained in the virtual world, that the virtual environment 'not be cleaned up' (SLT3). There was a discussion about the benefits of challenge. One SLT reported that it was helpful to get into the habit of problem solving. It was suggested that we lose the learning opportunity if things go too well. The carry-over of therapy tasks into situated settings through the use of 'challenge tasks' was thought to work well in face to face and virtual environments. In a 'challenge task' aspects of the therapy work are practiced in situated contexts in between therapy sessions.

Views on the proposed protocol and implementation

The clinicians valued the environmental context-based word retrieval in the proposed VESFA intervention. Going to a simulated environmental context to trigger relevant cues was viewed by clinicians as a strength e.g., talking about dream travel on the tall ship. Clinicians noted that vocabulary retrieved by people with aphasia in EVA Park is vocabulary that is unlikely to be accessed in a clinic environment. For example, a client with severe aphasia said 'porpoise' in response to seeing one in EVA Park. Holiday vocabulary was often elicited on the EVA island, for example sitting in the sunshine on deck chairs prompted a client to tell stories of her previous holidays. It was commented that the context of the unusual island supported word retrieval for low frequency items, as SLT1 said it 'promotes language a little bit out of the ordinary'. It was suggested that the playful, creative use of EVA Park should be promoted.

The virtual nature was reported to be 'freeing' (SLT2). You leave behind all physical attributes and disabilities and enter a world where you can fly. Your avatar represents you, so it is important to create it how clients want it to be. There was a conversation about the 'relief' when using EVA Park for therapy compared to alternative remote delivery where you have to see yourself in a little box throughout sessions.

One clinician suggested that EVA Park removes some of the real world chaos, things that divide your attention in a public space. The resulting controlled environment gives cognitive (the environment) and linguistic (the conversation) priming. Using target words in everyday conversations involves the client noticing when it is a good time to use practiced words. The real world is busy and chaotic with a high cognitive load and divided attention. Perhaps the focussed environment helps you to notice that this is the place to use the word you have worked on. She concluded that the supported nature of EVA Park is therefore an 'optimal environment' for therapy. But it is important to support clients to keep their word retrieval skills when the cognitive load increases. So, building in levels of challenge is important.

The clinicians reported biggest challenge in running groups was the logistics of arranging 3 or 4 people with aphasia to come together for the groups. This can be particularly difficult in rural settings where a large geographical area has reduced public transport. VESFA would address this challenge. VESFA using EVA Park could be implemented in independent practice without adjustments. The model of individual sessions plus a group was reported to work well in one independent aphasia clinic. In NHS services it might be more challenging to implement due to the varying cyber security rules across different NHS trusts.

Implications for VESFA intervention

As a result of involvement from the PWA, the VESFA protocol included the following elements: the therapy was delivered as a regime of 4 sessions a week over 8 weeks. Two sessions per week focussed on word retrieval in ESFA sessions and two sessions focussed on word use in conversation in a group conversation session. Conversation groups were topic based, SLT led and focussed on situated language. Training in moving the avatar was simplified to teach the on-screen controls for movement only. This removed the requirement for the user to switch their hand between a mouse and the keyboard. All interaction with EVA Park was therefore via the mouse. The 'hands up' function, that indicates a user would like to contribute to the session was taught in the technical set up. Sessions in the first week of the regime limited movement of the avatar around the island. Dissemination activities designed to reach the aphasia community should focus on a video abstract and talks at conferences and aphasia groups.

Reflections on the workshops with the PWA provided the user perspective that the word retrieval in conversation is harder than word retrieval in a picture naming activity. This highlighted the need for a hierarchy of scaffolded tasks in the conversation groups that would provide structured, supported word retrieval activities that build up to a more open topic based conversation.

The participating clinicians ratified the protocol as acceptable and feasible with some adjustments in clinical practice. Reflecting on the clinician focus group highlighted three key ideas:

1. The need for *positive challenge*; the therapy should provide scaffolded levels of challenge and the opportunity for the participants to develop problem-solving strategies.
2. The idea that EVA Park is a *focused* environment for delivering therapy where what

you see in the virtual environment inspires language use. Therefore, the virtual setting in EVA Park for conversation groups should be congruent to the topic.

3. The importance of creating an avatar that represents you as you wish to be represented, as 'your avatar speaks for you' within the virtual space.

Critical perspective

PI with a group of PWA has two priorities. Firstly, that we ask for input where the responses can genuinely be actioned. For example, research tells us that we need 20-50 hours of treatment to gain improvements in naming (REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators, 2022). We did not present options that were below this level of dose, therefore whatever regime came from the group could be implemented. Secondly, that the questions are scaffolded enough to be accessible to the group with aphasia and supports are provided for responding without leading a response. For example, the use of post it notes with group activities written on them that could be physically moved around a 'keep | change | avoid' framework worked well, see Image 4:1. Moving the workshops onto Zoom during the pandemic restricted the physical manipulation of concepts/activities in the workshops.

Presenting accessible research information was key to the BAS aims and worked well for some points e.g., dose parameters were clear. However, some concepts were possibly new to some members of the group and harder to follow. For example, Appendix 10 shows a simplified ICF framework. It was presented to discuss single word naming (impairment task) vs. word naming in conversations (activity/participation task). The ICF framework is a key concept in speech and language therapy, introducing a more social model of disability approach to therapy (Oliver, 2013). It is possible this concept needs to be presented multiple times with clear, concrete examples to be fully understood.

Conclusion

This PI project with people with aphasia and speech and language therapists addressed four of the actions in the guidance for developing complex interventions: 2) involve stakeholders, 8) understand the context, 9) pay attention to future implementation of the intervention in the real world and 10) design and refine the intervention (Table 1.2). The context of current practice was discussed in the clinicians' focus groups, with discussion about how virtual therapy could enhance practice and what it cannot replace. Real world implementation was also discussed. Specific decisions about content, format and delivery were made by the PI group of people with aphasia and they provided feedback on a taster session. As a result, the proposed VESFA intervention is informed by PWA and SLTs and thus has a greater chance of being acceptable and feasible.

Chapter 5 | Intervention Development: What topics are meaningful to people with aphasia? A qualitative study.

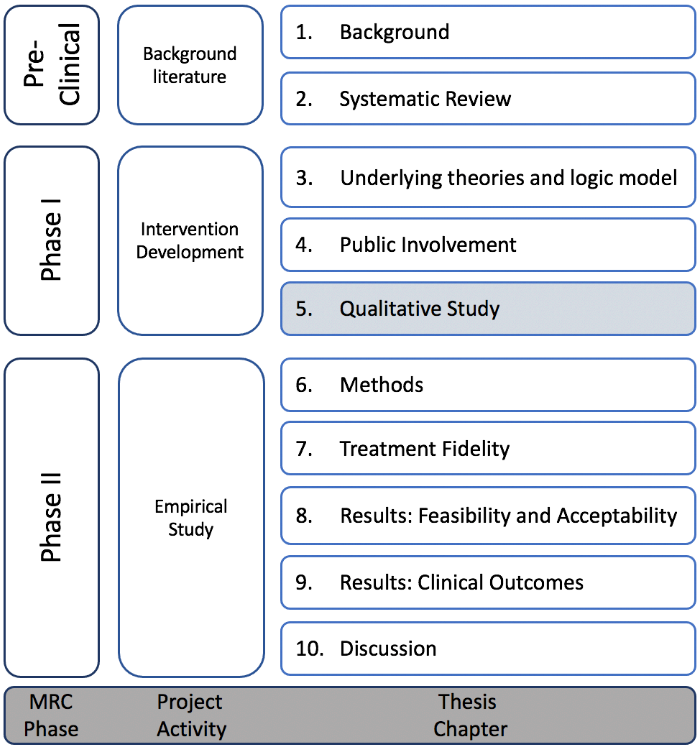

Figure 5.1: The thesis in a figure

Background

For many people with aphasia anomia is the most common feature of their aphasia (Goodglass & Wingfield, 1997). Speech and language therapists apply word finding therapies to address anomia with good outcomes on treated words but limited evidence of generalisation to untreated words (Wisenburn & Mahoney, 2009). As generalisation cannot be assumed, there is a need to select words for therapy that have the biggest impact. Previous research has explored the stimuli for naming therapy based on what words are most *frequent* (Renvall, Nickels, & Davidson, 2013a; 2013b) or prompted participants to choose words that are most *useful* (Palmer, Hughes & Chater, 2017). However, when studies observed what topics people with aphasia chose to talk about topics tended towards adjustment rather than function (Davidson, Worrall & Hickson, 2003; Holland, Halper & Cherney, 2010).

This project sought the views of a sample of 12 people with aphasia to explore topics for a meaningful word list for a word finding treatment. To this end, this qualitative research study asked the question:
What topics do people with aphasia find most meaningful to talk about and why?

Methods

Ethical approval was gained from the Language and Communication Science Proportionate Review within the School of Health and Psychological Sciences Research ethics committee, ETH1920-0148 (Appendix 11). See Appendix 6 for the Standards for Reporting Qualitative Research (SRQR) (O'Brien et al., 2014).

Focus groups were used to identify the most meaningful topics across a sample of twelve people with aphasia. Focus groups were chosen as they can identify and clarify what aspects are most important and how a variety of opinions should be prioritized (Tausch & Menold, 2016 p.8). Focus groups met the aim of the project as it provided a data that represented the group consensus that came as a result of a sharing and comparing process (Acocella, 2012).

Participants were recruited from community stroke groups by four speech and language therapy students, see Figure 5.2. The project was presented in the stroke group and interested people were given the participant information sheet (Appendix 12). Participants took a week to read the study information and gave informed consent to take part. Study information was made accessible by following published guidance (Rose et al., 2012). Participants were invited to take part if they were >18 years old, had aphasia as a result of a stroke that occurred more than 4 months ago, and were a fluent English speaker pre-stroke. Participants were excluded if they had additional cognitive impairments or neurological diagnoses that impacted on cognition. Two focus groups, with six participants in each group, were run in a room in City, University of London in January 2020.

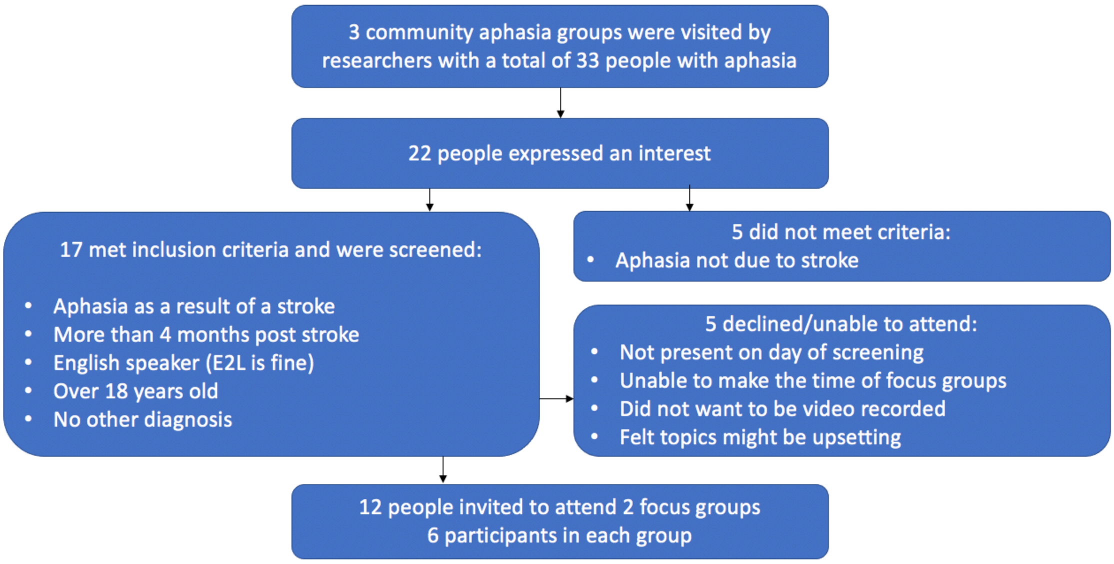

Figure 5.2: Recruitment flow chart

Speech and language therapy students moderated the focus groups. They received a two-hour training session on how to run focus groups from the doctoral candidate, who is an aphasia specialist speech and language therapist and researcher, with experience of focus groups. The training included information from the Social Research Association (https://the-sra.org.uk/) course, Conducting Focus Groups, and was informed by the work of Jane Lewis, Conducting Focus Groups (Ritchie & Lewis, 2003). The training covered the nature of focus groups and focus group data, the stages of group discussion (Ritchie & Lewis, 2003), how to elicit *breath* or *depth* of a topic with specific questions, and how to moderate the group e.g., use body language to include all participants, the use of neutral but encouraging responses. The focus group moderators had not met the participants prior to the recruitment.

A topic guide was developed to ensure both groups covered the same content (Appendix 13). In the focus groups conversation topics were generated through a discussion about *who* with and *where* people with aphasia have conversations and how those conversations make them *feel*. The resulting topics were then explored in two activities. First participants placed a piece of paper with written topics on a drawn line that represented a sliding scale from not meaningful to very meaningful/important. Participants placed topics on an individual scale before a group discussion about where topics should go on one agreed scale for the group. Analysis then identified those topics where there was a consensus that the topic was meaningful e.g., placed high on the scale.

The focus groups were videoed, transcribed, and then analysed by the students under supervision from the doctoral candidate, using framework analysis (Richie & Spenser, 1994) to identify dominant topics, see Figure 5.3. A further stage of analysis was undertaken by researchers to determine which topics had consensus. If more than half the participants in a focus group (>3) agreed on a rating it was stipulated that consensus was reached (Sirman, Beeke & Cruice, 2017).

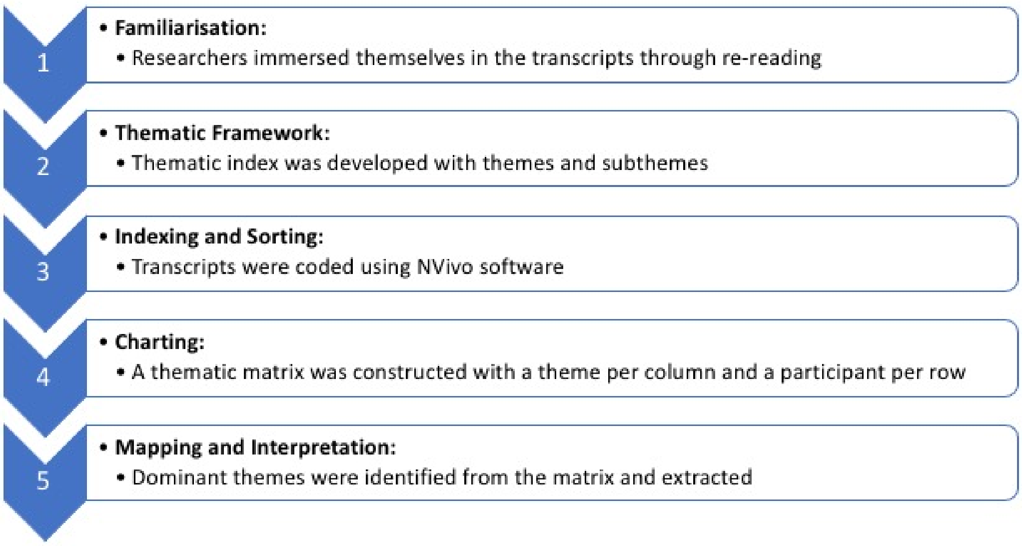

Figure 5.3: Framework analysis stepped process

To enhance trustworthiness 10% of the data was double coded to check reliability between coders.

Participants

Twelve people with aphasia volunteered to take part in the focus groups, see Table 5.1. The participant's aphasia was not formally assessed but they were observed to have moderate to mild aphasia, as determined by the clinical judgement of the therapist (ND). They were all in the chronic stage post stroke (>4 months)

		Focus Group 1 n=6	Focus Group 2 n=6	Total n=12
Age Range (years)	40-49	2	1	3
	50-59	0	1	1
	60-69	3	3	6
	70+	1	1	2
Gender	Male	6	4	10
	Female	0	2	2
Ethnicity	White	6	4	10
	Black	0	2	2
	Asian	0	0	0
Highest Education	GCSE	2	1	3
	A-level	2	0	2
	Undergraduate	1	4	5
	Postgraduate	0	1	1

Table 5.1: Participants

Results

Twenty-two topics were generated by the 12 participants with aphasia in the two focus groups. The themes from the framework analysis were the topics discussed. Eleven topics were rated as highly meaningful by the majority of participants. The three topics rated most meaningful were 1) Family and Friends, '*they are the most – they are to you the most meaningful people that live*' followed by 2) Food and drink, '*It's good … for everybody'* and 3) Living with Aphasia '*speaking about stroke and things are very important'*. Culture (arts, museums), Humour, Travel, Life experiences, News, Politics, Sports and Recovery also reached consensus as highly meaningful. Three topics reached consensus as having low meaning: Religion "*it's very personal*", Money "*no one wants to talk about it*" and Books, see Table 5.2.

Reliability of the coding was 'almost perfect', kappa=0.92 (McHugh, 2012).

Meaning	Topic and example quote	Participants		
		FG1	FG2	Total
High	**Family and friends** *"Ummm well I would go nuts if I didn't talk to my sister once a week"* (P3A) **"***But family are at the top" (*P6A) *"I'm quite proud of them and (.) so deep down it's important"* (P3B)	6	5	11
	Food and drink *"It's good (.) for everybody"* (P6B) *"It's as important as any of them food is very important that's my opinion."* (P5A)	5	4	9
	Living with aphasia *"Speaking about stroke and things are very important"* (P3A) "*But for me consistent ones are promoting aphasia,* promoting [name of aphasia group]" (P3A)	6	2	8
	Culture *"Well, I would say as important as food and drink"* (P1A) "*It's more important and meaningful than food and drink"* (P3A)	4	1	5
	Sharing jokes and humour *"I'd put it as one that's most meaningful"* (P1A)	5	-	5
	Holidays and travel *"they are everywhere oh my god yeah"* (P1B)	1	3	4
	Life Experiences *"I'm very interested in um what people did before they had their stroke 'cos everybody has an interesting story to tell"* (P5B)	-	4	4
	News *"Ok em how about talking about the news? Is that meaningful for you?" "Very very"* (R4/ P3B)	-	4	4
	Politics Asked 'Where would you put it on here? would you say it's very important, not so important?' *P1B points to P6B's scale and points to most meaningful* (R5/ P1B)	-	3	3

	Sports P6A indicated on the scale that sports are very high up but not exactly at the top. (P6A)	3	-	3
	Recovery "I couldn't talk (.) now I can talk I can (.) or I could I can! talk but I eh tend to get lost" (P2B)	-	3	3
Medium	No topics reached consensus for a medium ranking		-	-
Low	**Religion** *"Against it"* (800, P3B) *"it's very personally" (personal)* (P3A)	3	2	5
	Bank, money and numbers *"no one wants to talk about it!"* (P1A)	4	-	4
	Books Asked 'is talking about books meaningful to the group?' *"no"* (P2B)	-	3	3

Table 5.2: Topics that reached consensus (rated by >3 participants in each group)

Discussion

The participants reached consensus that eleven topics were meaningful. The three with highest agreement were: family and friends, food and drink and living with aphasia. This project builds on previous research into preferred treatment targets in word finding therapies.

Synonyms for meaningful are 'important' and 'worthwhile'. The agreement that family and friends are highly meaningful highlights the importance of a person's social network after their stroke. Social networks are at risk of shrinking in people with aphasia (Northcott, Marshall & Hilari, 2016) and are connected to a person's well-being. Community belonging is associated with better general and mental health (Michalski et al., 2020). The people you love and share your life with have been identified as a factor influencing quality of life (Cruice, Worrall & Hickson, 2006; Cruice et al., 2010) and living successfully with aphasia (Brown et al., 2010). Individual words within the topic 'family and friends' are predominantly proper names, people's names. This is a problematic word set. Proper names have distinct properties (Yasuda, Beckmann & Nakamura, 2000) that do not easily lend themselves to a therapy task like SFA and have rarely been addressed in therapy research (Robson et al., 2004).

The importance of conversations about food and drink is highlighted across a number of studies. It is considered ‘core’ vocabulary for communication devices, such as voice output aids (Carter, 1987; Graves, 2000). Food and drink was the most frequent topic of conversation for both people with aphasia and healthy older people in Davison and colleague’s observational study (2003). Equally, Palmer (2017), who asked people with aphasia to select the words they wanted to target in therapy, found most words selected in the food and drink topic. Individual words within this topic are often concrete, familiar items with high naming agreement that make good targets in SFA treatments. Food and drink can be the mundane everyday items we consume and those that represent special occasions e.g., birthdays, cultural holidays, celebrations.

Living with aphasia is a core theme in aphasia research (Hilari et al., 2012). It is linked to the renegotiation of a sense of self that occurs post stroke (Shadden & Agan, 2004). The need to share your stroke journey is well documented (Corsten et al., 2015; Frank, 1995; Strong & Shadden, 2020) and targeting vocabulary in this topic may support a person to talk about, and therefore process, what has happened to them. This sharing of personal stories, both the small stories of everyday and bigger illness narratives, is argued to be transformative (Strong & Shadden, 2020).

The small stories from your life, identified in the holidays and travel and life experiences topics, share something of who you are. Personal narratives have a dual purpose. Telling a story makes sense of events to the speaker, intra-personal, but also connects them to others through shared experience, inter-personal (Olness & Ulatowska, 2011). Conversations that allow you to reveal something of your identity, what has changed and what persists, can support adjustment to living with aphasia (Taubner, Hallén & Wengelin, 2020).

Culture as a meaningful topic sits with literature that highlights satisfying activities and doing things as factors that influence quality of life and living well with aphasia (Brown et al., 2010; Cruice, Worrall & Hickson, 2006; Cruice et al., 2010). Cultural activities, such as visiting museums and galleries, have known links to health and wellbeing (Camic & Chatterjee, 2013; Cuypers et al., 2012; Napier et al., 2014). They have been described as journeys of self-discovery (All-Party Parliamentary Group on Arts, Health and Wellbeing, 2017) and reported to connect our personal beliefs to universal truths (Dodd, Sandell & Scott, 2014), perhaps reminders that we belong in something greater than ourselves. The agreement that cultural activities are meaningful may have been influenced by the characteristics of the sample e.g., they were from London and 50% were graduates.

Focus group methodology was a strength of this study which enabled us to get consensus on what topics were meaningful to talk about across a range of individuals with aphasia. The age range of participants allowed us to gather both working age and retirement age perspectives. The use of structured written topics on a visual sliding scale enabled the opinions of different people with aphasia, including those with severe aphasia, to be collected.

A more in-depth exploration of meaningful conversations could be explored in individual interviews. This would allow for more probing about why these topics are important. The voices of women were underrepresented in this sample. The sample was recruited from London, so represents mostly educated people from an urban centre. It is likely there are views not captured. The focus groups gave consensus on meaningful topics of conversation but not individual words within those topics.

Conclusion

Previous research into the potential vocabulary for aphasia therapy has explored topics that are useful or frequently used. This project highlighted meaningful topics for conversation from a sample of twelve people with aphasia. Selecting naming stimuli from this topic list has the potential for impairment level therapies to focus on vocabulary that can impact participatory roles and well-being.

The findings of this qualitative study influenced the topics chosen for the VESFA therapy. A fuller account of how these findings were combined with the findings from relevant literature to determine the topics and subsequently the individual words is described in 'stimuli', Chapter 6.

This chapter concludes the report of the development of VESFA (thesis objective 3). The next five chapters address thesis objective 4: Evaluate the feasibility of running a definitive trial on the efficacy of VESFA. These chapters detail the methods and outcomes for treatment fidelity, the methods of the feasibility randomised control trial, the feasibility and acceptability outcomes, the clinical outcomes and a discussion of the results in the context of aphasia rehabilitation literature.

Chapter 6 | Methods

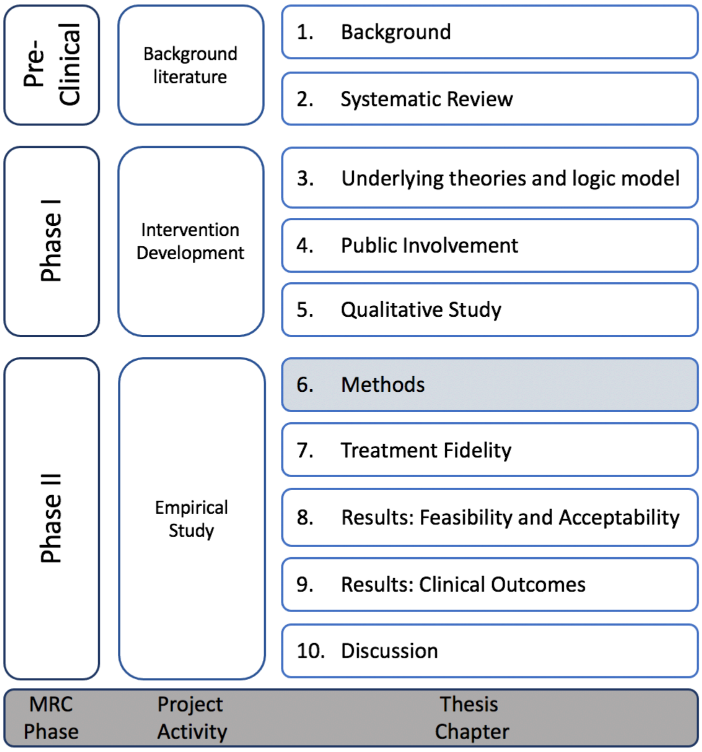

Figure 6.1: The thesis in a figure

This chapter outlines the methods employed in the phase II feasibility trial. Reporting of the methods follows the Consolidating Standards of Reporting Trials (CONSORT) guidelines for pilot and feasibility studies (Eldridge et al., 2016). See Appendix 14 for CONSORT checklist.

Trial Design

The Virtual Elaborated Semantic Feature Analysis (VESFA) trial was a single-blind, phase II feasibility randomised controlled trial comparing usual care + VESFA intervention with a usual care control.

Feasibility studies explore 'can it work?' and, as such, primary outcomes related to the trial processes. Feasibility outcomes of this trial explored recruitment and retention, willingness to be randomised, compliance with and acceptability of the treatment and of the outcome measures. Secondary outcomes explored preliminary findings from clinical outcome measures of word finding in picture naming and discourse, functional communication, language, mood and quality of life.

Funding

The study was funded by a School of Health Sciences doctoral studentship from City, University of London, awarded December 2018.

Ethics

The study received full ethical approval from the Senate Research Ethics Committee at City, University of London on 01/09/20 (ETH1920-1223). The ethics form is available in Appendix 15, participant information is in Appendix 16 and consent form is in Appendix 17. There were a number of ethical considerations in this study. Participants needed to be fully informed, aware that half of the recruits would not receive the intervention and the privacy of participants' data needed to be maintained. In addition, the lawful basis for holding health data and personal data needed to be explicit. How these issues were addressed is outlined below.

Informed consent

People with aphasia have the capacity to give informed consent, providing there are no co-morbidities that affect cognition and providing that information is presented in an accessible manner. Consent and information materials were designed to be accessible to people with aphasia (Rose et al., 2012). To be eligible to take part in the study potential participants' comprehension was screened (see 'eligibility criteria' below). This ensured that participants had capacity to understand the project information. It was possible that participants might not wish to undergo the testing burden without the benefit of treatment. The potential of being randomised to a 'usual care' control group was made clear in the conversations during screening and in the written and video information provided to participants.

Protection of privacy

Participant privacy is an important consideration in research; the university, as the sponsor, has a duty to protect the privacy of research participants and their data. Personal and health data were collected via the videoconferencing technology 'Zoom' (www.zoom.us). Zoom was chosen as it allows users to annotate a shared document. This makes language assessment possible, where comprehension is tested by items being selected. Zoom meets the privacy and security standards of the European Union's General Data Protection Regulation (https://support.zoom.us/hc/en-us/articles/360000126326-Official- Statement-EU-GDPR- Compliance) and the American Health Insurance Portability and Accountability Act (https://zoom.us/docs/doc/Zoom-hipaa.pdf). Zoom has been recommended by previous research investigating remote assessment (Dekhtyar et al., 2020) and has proved acceptable and accessible in a UK Group Pilot with people with language and communication needs as a result of dementia (Dementia Voices, 2019). All study data were hosted on a City, University of London secure network drive accessed only by the project team. The data was pseudo-anonymised, with each participant being identified by a unique number.

Lawful basis for collected personal and health data

City, University of London was the data controller of this study based in the United Kingdom. Thus, the university was responsible for looking after participant information and using it properly. City, University of London considers the lawful basis for processing personal data to fall under Article 6(1)(e) of GDPR (public task) as the processing of research participant data is necessary for learning and teaching purposes. All research with human participants by staff and students has to be scrutinised and approved by one of City's Research Ethics Committees. City, University of London considers the lawful basis for processing of special category data relating to health to fall under Article (9)(2) (a)of GDPR (Explicit Consent). The research participants have given their explicit consent for the processing of health information by volunteering to take part in the research and the completion of the consent form. The research participants are able to withdraw from the research project at any time.

Participants

Eligibility criteria

Participants were eligible for the study if they had a diagnosis of ischaemic or haemorrhagic stroke, were at least four months post stroke, 18 years old or over, presented with word finding difficulties as a result of aphasia and had adequate comprehension. Word finding was screened using the naming and word finding subtests of the Western Aphasia Battery (Kertesz, 2007) and participants were included if they scored <76/100. This mitigated against the risk of participants scoring at ceiling on the trial outcome measures. Auditory comprehension was screened using the Frenchay Aphasia Screening Test (Enderby et al., 1987) and participants had to score 6/10 or above to be included. A minimum comprehension score of 6/10 ensured participants could understand the participant information sheet, could access outcomes measures such as the SAQOL-39g, and could follow instructions in the virtual environment without the face-to-face context that supports auditory comprehension (e.g., facial expression, natural gesture).

In order to take part in a remote study, each participant needed to:

- Name a person, either living in the participant's home or within their COVID bubble, available to assist with technical glitches
- Have a computer with the outlined minimum specification (https://secondlife.com/support/system-requirements/) or be willing to receive a laptop posted from the University
- Allow the researcher to access their computer to provide remote support via the Zoom remote control function or TeamViewer software (www.teamviewer.com)

Participants were excluded if they had other diagnoses affecting cognition such as dementia, had severe uncorrected visual or hearing problems (that would prevent them from accessing computer-based stimuli), had severe or a potentially terminal co-morbidity on grounds of frailty, or were not fluent English speakers prior to the stroke (based on self or family report). Participants continued to receive usual care.

Setting

The feasibility trial was carried out in 2020 and 2021 during the COVID-19 pandemic and therefore used entirely remote recruitment and data collection methods. Assessment sessions were carried out on the videoconferencing technology, Zoom. The intervention was delivered remotely via the online virtual world, EVA Park. The location of the participants for both data collection and intervention was their own homes. The location of the researcher and testers was either a room in the university or their own homes.

Recruitment

Participants were recruited from the community. Methods of community recruitment included: virtual visits to stroke and aphasia groups; sharing information about the project on social media; distributing information about the project to third sector organisations; contacting people known to the university who have given permission for their details to be shared for this purpose; accepting self-referrals (e.g., where a potential participant learnt about the project from Twitter or word of mouth). Written informed consent was obtained from all participants. All information sheets and consent forms met aphasia-accessible principles e.g., presenting one idea at a time, using short simple sentences, presenting key ideas with a suitable pictorial image (Rose et al., 2012).

Consent

If a potential participants met the screening criteria, they received detailed information from the doctoral candidate with, if relevant, the participant's significant other present. Potential recruits were emailed the project information sheet (Appendix 16). They were advised to read it carefully and discuss it with others where appropriate. A video of the information sheets read aloud was shared. Additionally, a video explaining the overall project was shared with all participants. They were directed to the project website for further self-directed investigation, if required.

If a person met the participant criteria, had taken time to consider and wanted to take part, they were asked to sign the consent form (Appendix 17). Routinely one week was left between receiving the information and seeking consent. The consent form was explained by the SLT researcher. Participants received both an electronic document and a Qualtrics link to the form and they could choose to give consent via the online form or complete and return the electronic document.

Intervention

Intervention arm – Usual Care plus VESFA (UC+VESFA)

Participants randomised to UC+VESFA (n=18) received 40 hours of treatment; two one to one (60min) sessions and two group (90min) sessions per week (5 hours per week) for 8 weeks. All treatment was delivered remotely in EVA Park. The treatment was documented in a therapy manual (Appendix 8) with a corresponding Participants Handbook (Appendix 18). Participants in UC+VESFA continued to access health care, social care and charity support organisations. One to one sessions consisted of 50 minutes of naming practice with ESFA and 10 minutes planning for content for the conversation sessions. For example, 'mountain' was covered in the individual ESFA session and the subsequent group conversation focused on past travel experiences. The participant revisited the ESFA sentence in anticipation of sharing it with the group 'I climbed the mountain'. The treatment was delivered by the doctoral researcher, a specialist speech and language therapist.

Treatment was delivered in sets of three participants at a time, to allow for a small group per set. Six sets of treatment were delivered over the 14-month intervention period (6 sets of 3 participants, total = 18 participants). Set 1 began in December 2020 and set 6 completed in February 2022.

Four topics with 30 vocabulary items were the treatment stimuli (see 'stimuli' below). Each set of 3 participants chose three of the four topics to target in therapy. This allowed for some choice in target topics. Each one-to-one ESFA session focussed on one of the three topics for two weeks in sessions 1- 12. There was a recap session for each topic in sessions 13-15 and a final session covering all 90 items treated, session 16 (see Appendix 18). Using an SFA chart on a board in EVA Park the clinician and participant worked through the topic vocabulary, see Image 6.1. The target image was shown, and the participant attempted to name it. A minimal to maximal cueing hierarchy was used if the participant could not name the target. If the participant made semantic errors a hierarchy of semantic cues were given, if the participant made a phonemic error, phonemic cues were given. Then the SFA chart was shown, and each feature retrieved: superordinate category, use, action, physical properties, location and association. The cueing hierarchy was used where appropriate. The target was named again and then the participant was encouraged to write the target if they could and use the target in a phrase or sentence using the features identified. Subsequent sessions began where the previous session finished. For example, if session 1 worked on items 1-5 then session two worked on items 6 onwards. At the beginning of each session the items that were not named first time in the session before were presented to the participant again before moving on to new items. At the end of each session all items in the topic targeted to date were presented for naming. For example, in sessions 3 the participant worked in vocabulary items 18-24 on the topic list, at the end of the session they would run though items 1-24.

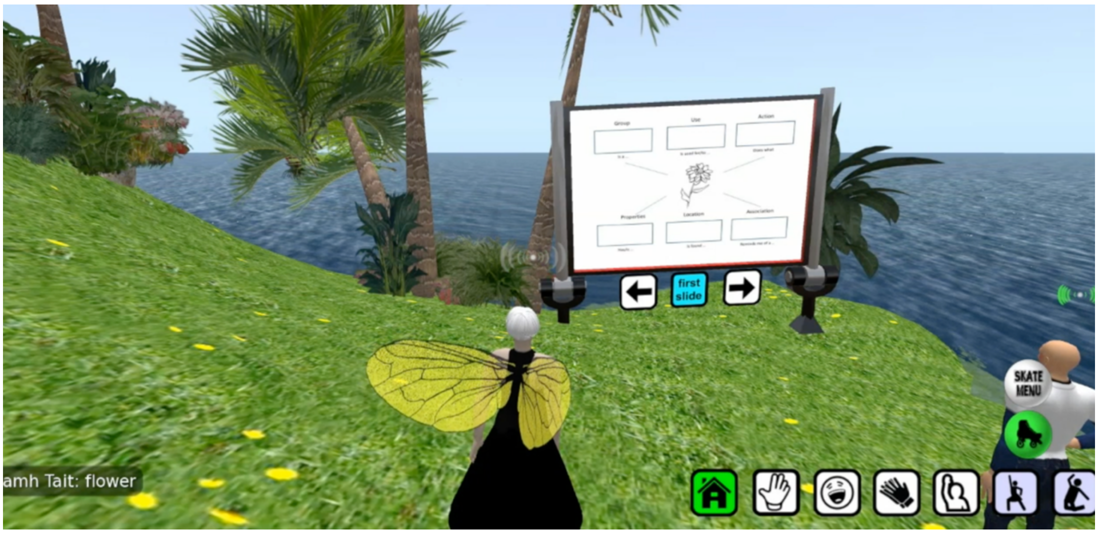

Image 6.1: ESFA session in EVA Park working on gardening vocabulary

Group sessions consisted of word games and topic-based conversations and role plays situated in EVA Park settings, see Table 3.2. The groups worked on the same topic as the ESFA sessions. Thus, two weeks focussed on topic one in groups 1-4, topic 2 in groups 5-8 and topic three in groups 9-12. There was a recap group for each topic in groups 13-15 and a party in the final session, group 16 (see Appendix 18).

Group	Travel	Place in EVA Park
1.	Past travel: Tell us about the last time you travelled	Camp fire

2.	Dream travel: Describe your dream holiday	Yacht
3.	Process: Directions from the nearest train station to your house	Cadillac, see Image 6.2
4.	Anecdote/ personal story about travel	Deck chairs

Table 6.1 Example of the four conversation activities for the topic of travel

Each topic had four conversations: a past example, a dream example, a process, and an anecdote. One group was given to each conversation so that the four conversations covered the two weeks that VESFA was dedicated to a particular topic, see Table 6.1 for the 'travel' example.

Image 6.2: The EVA Park Cadillac

Control arm – Usual Care Control (UCC)

Participants randomised to the Usual Care Control (UCC) (n=18) were not offered VESFA therapy. They continued to access health care, social care and charity support organisations. When they completed their final outcome measures for the study (week 18 of their involvement in the study) they were offered VESFA in the form of student placements, this was not part of the VESFA trial. A questionnaire after the final testing explored what usual care was accessed. Usual care was predominantly community aphasia groups. No one received targeted language therapy.

Intervention fidelity

The fidelity of the VESFA treatment was evaluated. Chapter 7 outlines the treatment fidelity methods and results.

Stimuli

The therapy stimuli were identified in the following way. Firstly, topics from existing literature about what people with aphasia want to talk about (Palmer, Hughes & Chater, 2017; Holland, Halper & Cherney, 2010) and the topics from the qualitative study (Chapter 5) were reviewed, see Table 6.2. Secondly, specific vocabulary for the 30 items within each conversation topic were chosen. Words were chosen from published word lists that had established good naming agreement.

Focus Groups	**Palmer et al. 2017**	**Holland et al. 2010**
Percentage that considered this meaningful:	*Percentage of words selected were in this topic:*	*Percentage of the stories on this topic:*
family and friends (91%)	food and drink (30.6%)	personal stories (68%)
food and drink (75%)	nature and gardening (10.3%)	conversations with family (21%)
living with aphasia (66%)	entertainment (9.4%)	seeking or providing information (18%)
culture, art, theatre (42%)	places (7.3%)	discussion of outside interests (14%)
jokes / humour (42%)	people (6.7%)	
holiday/travel (33%)	house' (6.5%)	
life experiences (33%)	clothes (5.2%)	
news (33%)	travel (3.5%)	

Colours highlight common topics across data sets.

Table 6.2: Topics identified in the literature and the focus groups

Selection of conversation topics

The final topics were chosen by drawing on the outcomes of the focus groups and existing literature and words that were practical for use in EVA Park. For example, there are different food and drink settings in EVA Park (restaurant, bar, kitchen) so this topic could easily be situated in relevant simulated settings. The four topics chosen were 1) food and drink, 2) nature and gardening, 3) daily living, and 4) travel. Names of friends and family were not chosen for the one-to-one ESFA stimuli despite being the importance highlighted in the focus groups (Chapter 5). Proper nouns are argued to have different semantic properties from common nouns. For example, names of people do not fall into categories but have unique referents (Robson et al. 2004). It is likely that proper nouns do not activate a spreading network, which is what this intervention is designed to simulate. Additionally, a word list was needed that was appropriate for all study participants. Personally relevant vocabulary (names of family and friends) and abstract vocabulary associated with the topic 'living with aphasia' (constancy/change, sameness/difference and agency/dependency, see (Taubner, Hallén & Wengelin, 2020) were encouraged in the conversation groups only. To this end personal anecdotes were encouraged in the conversation groups; the SLT sought personal connections to the topic vocabulary and expanded on personal associations.

Selection of words

Thirty nouns from each conversation topic were chosen as treatment targets for ESFA. Targets needed to be 1) an image with established naming agreement and 2) within the meaningful conversation topics. An excel spreadsheet was created with a sheet per conversation topic. Items with images were chosen from existing word lists to ensure good naming agreement. The published lists were the Snodgrass and Vanderwart list of 260 pictures (Snodgrass & Vanderwart, 1980), the 224 pictures available from the International Picture Naming Project (Szekely et al., 2003) and 175 items from The Philadelphia Naming Test, a list developed in the Language and Aphasia Lab of the Moss Rehabilitation Research Institute (Roach et al., 1996). Items were organised by conversation topic. Imageability values were sought for all words within the topic. The 30 items with the lowest imageability values were selected from each topic. This metric was chosen in light of the Complexity Account of Treatment Efficacy (CATE, Thompson 2007) that outlines that abstract items, which are more complex, may generalise to concrete items but there is no evidence of concrete items generalising to abstract items. Although low imageability items were targeted, all items had to be picturable, so that they could be represented on the SFA chart. For example, in the food and drink topic 'octopus' had a lower imageability score (score=531) than 'banana' (score=644). Frequency and syllable length were also collected but did not inform the final list. A full list of selected words is available in Appendix 18.

Primary Outcomes

As a feasibility study the main endpoints relate to feasibility outcomes. We outlined four primary and two secondary endpoints. Pre-specified criteria guided the decision as to whether to proceed to a future definitive trial. These are **highlighted in bold below**. The pre-specified criteria were based on published trials investigating complex behavioural interventions with people with aphasia (Northcott et al., 2019; Palmer et al., 2015; Thomas et al., 2013). The reported recruitment, retention and adherence rates in those studies have informed the criteria here.

These endpoints are a guide and will be considered together with qualitative evidence. After completing the last outcome measure, participants completed a questionnaire in an interview format to explore their views on the study processes, outcome measures and testing burden, and intervention (for those receiving the intervention). Interviews were carried out by independent researchers (MSc students).

Primary outcomes

a) Feasibility of recruitment and retention to the trial

This evaluation was based on the proportion of those screened that were eligible for the study and the proportion of participants eligible who consented to the trial. **The feasibility of recruitment was met if 60% of eligible participants consented to the trial.** The rate of participants randomised each month; attrition rates (overall, by stage and by study arm) and reasons for attrition will be reported if they were known. **The feasibility of trial retention was considered met if 70% were available at follow up**.

b) Feasibility of delivering the intervention remotely in a virtual environment

This was based on questionnaires with participants at the end of the study and rates of sessions cancelled due to technical difficulties. **Delivering intervention in a virtual environment will be considered feasible if the rate of cancelled sessions is no higher than NHS community clinics where 24% of appointments are cancelled** (NHS Benchmarking, 2019). The participant questionnaire included questions exploring satisfaction with the method of delivery and accessibility/usability of the technology, see Appendix 20, Post Therapy Questionnaire. Previous EVA Park studies have employed these methods (Galliers et al., 2017).

c) Acceptability of research procedures

The acceptability of the research procedures was evaluated based on data from the questionnaire, dropout rates and rates of missing data. Questions about the acceptability of online assessment, the length of the assessment sessions and the specific assessments used were asked in the Post Therapy Questionnaire (Appendix 20) and the Usual Care Questionnaire (Appendix 21). The primary criterion for testing procedures to be considered acceptable is **less than 15% missing data per scale/outcome measure**.

d) Acceptability of intervention to participants

The evaluation was based on the rates of adherence to the intervention where **participants were considered to have adhered if they received at least 80% of intervention** (32 of the 40 hours). Moreover, questionnaire data added detail on participants' views on the acceptability of the intervention (Appendix 20).

Secondary outcomes

e) Determine and evaluate measure of word retrieval in discourse

Evaluating the retrieval of words within conversation and/or meaningful measurement of discourse is multifaceted. Discourse requires a number of skills on multiple linguistic levels and therefore the range of potential assessments is large (Dipper & Pritchard, 2018). In addition, clinicians report a lack of skill in measuring discourse and find it burdensome (Cruice et al., 2020). In the first year of the study discourse measures were explored. The

VESFA treatment aims to improve word retrieval in conversation. Therefore, a decision was made to use the Nicholas and Brookshire (1995) protocol which measures the retrieval of words in discourse (see 'word retrieval in discourse' below).
The appropriateness of outcome measures was considered in terms missing data; floor or ceiling effects; whether the measure matched changes described in the post-therapy questionnaire; and the participants perspective on acceptability (Northcott et al., 2019).

f) Evaluate treatment fidelity

See Chapter 7 for a full report of the treatment fidelity evaluation.

Clinical outcomes

There were three testing points. Testing 1 (T1) in week 1, when participants were randomised. Testing 2 (T2) was in week 10 and Testing 3 (T3) was in week 18. This allowed for an 8 week intervention period for the VESFA+UC arm between T1 and T2, and an 8 week no treatment period between T2 and T3 (follow up). Outcome measures were administered by student speech and language therapists. They received two hours of training on administration and scoring of the measures from the doctoral candidate (ND).

Although the study was not powered to detect effectiveness of therapy, clinical outcomes were measured to evaluate the feasibility of outcome measurement and to explore whether the intervention showed promise (trends in the data) and was worth exploring further in a definitive trial. Outcome measures comprised of the core outcome set for aphasia research (the Scenario Test, the GHQ-12, the SAQOL-39g and the WAB-R) (Wallace et al., 2019), a treatment specific measure of word retrieval (the VESFA Naming Test) an independent word retrieval measure (the Boston Naming Test, Kaplan, Goodglass & Weintraub, 1983) and a measure of words in discourse (Nicholas & Brookshire, 1993). The core outcome set in aphasia trials is the result of a four-year programme to investigate important outcomes for key stakeholders (people with aphasia and family members, aphasia clinicians, aphasia researchers) and recommends outcome measures to be used as a core set in all aphasia intervention studies (Wallace et al., 2018). With outcome measures harmonised across phase I-IV trials in aphasia future meta-analyses of outcomes are possible.

The feasibility of the outcome measures will be established through reporting of the completeness of data, time taken to administer the tests, and acceptability of the outcomes to the participants. Questionnaires after T3 addressed the acceptability of the data collection procedures (Appendix 20 & 21).

Measures

g) VESFA Naming Test

A picture naming test of the vocabulary treated in the VESFA therapy was carried out. A total of 120 words were tested (30 items across four topics) in a confrontation naming test, see Appendix 19 for the record form. 90 of those words were practised in the one-to-one ESFA therapy tasks and 30 were not. The pictures were the same as those used in the treatment and had good naming agreement; they were drawn from the Snodgrass and Vanderwart pictures, the International Picture Naming Database and the Philadelphia Naming Test (Snodgrass & Vanderwart, 1980; Bates et al., 2000; Roach et al. 1996). Participants were presented with a picture and asked, 'what is this?'. They scored 2 for a correct response, 1 for correct after a cue and 0 for an incorrect or no response. There was an agreed list of synonyms e.g., both boot and welly would be accepted. Thus, the highest score 240/240 indicated an ability to name all 120 items.

Th VESFA Naming Test was administered online. The stimuli were presented in a PowerPoint and displayed to participants on Zoom via screen share.

h) The Boston Naming Test (BNT)

This 60-item naming test (Kaplan, Goodglass & Weintraub, 1983) served as an independent measure of naming ability i.e., it did not test specifically treated items. Internal consistency ranges from r=.78 to .96. Good test-retest reliability has been reported in neurologically healthy adults (from r=.59 to .92). Additionally, the BNT correlates with other naming tests (Pedraza et al., 2011). The items increase in difficulty from item 1, 'tree', to item 60, 'abacus'. As a result, a discontinue rule exists where the test can be abandoned if there are 6 consecutive failures to name. This incremental difficulty has been criticised, as has the ambiguity in how to administer the discontinue rule (Pedraza et al., 2011; Ferman, Ivnik & Lucas, 1998). Reliability in scoring was improved by explicitly applying the lenient discontinue rule with testers in the VESFA trial, that is, if the participant names the item with a phonemic cue, it does not represent a failure (Ferman et al, 1998). Scores range 0-60 with higher scores indicating better naming abilities.

A systematic review explored the reliability of administering assessments via videoconferencing technology (Brearly et al., 2017). They found four studies of that compared face to face and online administration of the BNT (Cullum et al., 2006; 2014; Vestal et al., 2006; Wadsworth et al., 2016). This review found consistent small but significant differences between scores of face-to-face and online delivery. This needs to be considered when planning a future trial.

i) Word retrieval in discourse

A test of words used in discourse was carried out following the protocol from Nicholas and Brookshire (Nicholas & Brookshire, 1993). This protocol was developed to address a need to quantify informativeness and efficiency of the connected speech of adults with aphasia. In a review of 58 discourse information measures this protocol was found to be one of the most reliable and valid (Pritchard et al., 2018). Four discourse samples were elicited. Four samples have been shown by the authors to demonstrate the best balance between test-retest reliability and time required to transcribe (Brookshire & Nicholas, 1994). Participants were asked 1) to produce a procedural discourse (how would you go about ordering a drink in a café) 2) share personal information (tell me what you do on Sundays, tell me where you live – describe it to me) and 3) describe an event depicted in 6 pictures. The final sample was 4) the picture description from the WAB.

The samples were audio recorded, transcribed, and the words counted. The following information was extracted for analysis: words per minute (WPM), percentage of words that were Correct Information Units (%CIU) and CIUs per minute (CIU/min). WPM indicates the fluency of the persons speech but not the content. CIUs indicate correct content and percentage of CIUs can indicate the informativeness of the sample e.g., a high %CIUs suggests that most words used conveyed meaning. CUI/min indicate both content and fluency.

j) The Scenario Test – UK

The Scenario Test – UK (Hilari et al., 2018) is an adaptation of the Dutch Scenario Test (van der Meulen et al., 2010) and is a measure of functional communication. It measures how a person with aphasia conveys everyday messages, verbally and/or non- verbally, in an interactive setting. It comprises six scenarios e.g., at the doctor's, in a restaurant, a social visit, that elicit 18 messages with a maximum total score of 54, where a high score indicates good functional communication. A qualitative checklist allows the administrator to note the clients preferred mode of communication, i.e., verbal, written or gestural.

The Scenario Test UK was validated on a sample of 74 people with aphasia and 20 controls and showed high levels of reliability. Internal consistency was excellent (Cronbach's α = 0.92) showing that the items within the test measure the same construct. The reliability of scoring between testers was excellent with an inter-rater reliability interclass correlation (ICC) of 0.95. Additionally, test–retest reliability was excellent with ICC of .96 (Hilari et al., 2018). Sensitivity to change has been explored in the Dutch original version in a sample of 22 people with aphasia. An 8-point increase in scores was considered clinically significant (van der Meulen et al., 2010).

The Scenario Test was delivered via Zoom. The scenarios were presented via PowerPoint and screen share. This administration method has not been validated.

k) The General Health Questionnaire -12 item

The General Health Questionnaire -12 item (GHQ-12) (Goldberg, 1972) is a quick screen of mood and can identify depression. It asks the participant to rate their current state compared to what is usual for them and therefore will pick up recent changes in mood. The GHQ-12 has been used extensively in research with validity studies in nine countries (Goldberg et al., 1997). A validity project across 15 international centres and 5,438 participants showed good sensitivity (83.4%) and specificity (76.3 %) (Goldberg et al., 1997).

Scores on the GHQ-12 range 0-12, with higher scores indicating higher distress. The GHQ-12 also provides categorical data (distressed vs not distressed) using a cut-off score. We used a cut-off of >/=3 (rather than the commonly used >/=2 in the general population) as an indicator of distress in people with stroke and aphasia, as it has been argued that older people, and those with physical conditions may require higher cut-off thresholds (Hackett et al., 2005). Previous aphasia trials have used a cut off of >/=3 (Hilari et al., 2010; 2019; Northcott et al., 2019).

The GHQ-12 was adapted for people with aphasia by putting one question per page, instead of 12 questions per page. This was administered via Zoom by screen sharing a document with the questions. Online administration has not been validated.

l) The Stroke and Aphasia Quality of Life Scale

The Stroke and Aphasia Quality of Life Scale-39 item generic version (SAQOL-39g, Hilari, et al., 2009) is a self-reported health related quality of life (HRQL) measure. It asks about functions and feelings within three domains: physical, communication and psychosocial. Each item is scored on a 5-point scale. Scores are added and divided by the number of items within subdomains. Total scores are calculated by adding all items and dividing by 39. The overall score is out of 5, where a higher score indicates better HRQL. The psychometric properties of the SAQOL-39g were established on a sample of 87 people with stroke (including 32 with aphasia). Test-retest reliability was good (interclass correlation = 0.96 overall, 0.92—0.98 domains) and the test showed good internal consistency (α = 0.95 overall score, 0.92—0.95 domains); convergent (r=0.36–0.70 overall, 0.47–0.78 domains) and discriminant validity (r=0.26 overall, 0.03–0.40 domains); and responsiveness to change (d=0.35–0.49).

The SAQOL-39g has been shown to be robust to different modes of administration, with telephone interviews and postal surveys yielding similar results to face to face delivery. The authors concluded that researchers and clinicians may employ alternative modes of delivery (Caute et al., 2012). This study used the online version of the SAQOL-39g (https://cityaccess.org/tests/saqol).

m) The Western Aphasia Battery-Revised (WAB-R)

The WAB-R (Kertesz, 2007) is a comprehensive diagnostic language test. It classifies aphasia subtypes and rates the severity of the aphasia via an aphasia quotient (AQ). An AQ score of 0-25 represents very severe, 26-50 represents severe aphasia, 51-75 moderate aphasia and a score of 76-93 mild aphasia and >93.8 represents not aphasic by the WAB.

There are two parts. The subtests in part 1 cover speaking and understanding language and the subtests in part 2 address reading, writing and cognitive tasks. For this study only part 1 was carried out. This gave a score for speech, comprehension, repetition, naming and word finding and an AQ score.

The WAB was standardised on a sample of 150 people with aphasia and 59 controls. Internal consistency is good, with a coefficient of .97. Intra-rater reliability ranged from .79 - .99, and inter-rater reliability averaged at .98 indicating strong (>.70) scoring reliability (Boyle, 2020). A high test-retest correlation confirmed reproducibility of the test (Kertesz, 2007). The WAB-R has been validated for online administration (Dekhtyar et al., 2020).

Investigations into significant change after rehabilitation indicates a 5.05 point difference (1.64-8.46, P=004) in the WAB AQ in a between-groups design represents a significant change (Gilmore, Dwyer & Kiran, 2019).

Qualitative outcomes

The perspectives of the trial participants were explored in two questionnaires: a post-therapy questionnaire for participants in the VESFA arm and a usual care questionnaire for the participants in the usual care arm (UCC). Both questionnaires had the same questions regarding testing. Responses from the questionnaires are reported in Chapter 8 (the acceptability of the research procedures and the acceptability of the treatment) and Chapter 9 (experiences of change in communication).

Post-Therapy Questionnaire

The post-therapy questionnaire contained Likert scales and free comments sections regarding the experience of the testing sessions, the experience of the treatment and remote delivery of assessment on Zoom and intervention via EVA Park (Appendix 20). The interviews were co-constructed with the person with aphasia and the interviewer and the participant's partner/carer if they wished. The questionnaire was built in Qualtrics. The interviewer shared the questionnaire on the screen and wrote the participants comments into the shared document, see Image 6.3.

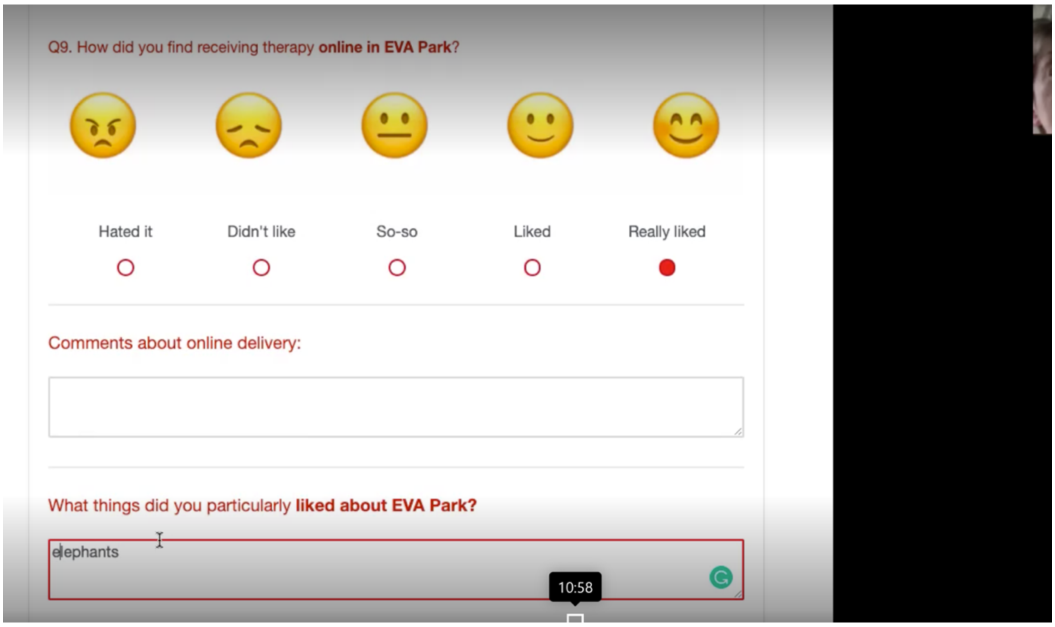

Image 6.3: Screengrab illustrating the shared record of the interview on zoom

The written answers and quotes below were therefore agreed between the interviewer and the participant. Interviewers were student speech and language therapists who were independent of the VESFA trial. They received a 2 hour training on qualitative interviewing techniques, including avoiding leading questions and to use neutral prompts such as 'Can you tell me more about that?' to elicit more information.

Usual Care Questionnaire

The usual care questionnaire was developed by adapting the Client Service Receipt Inventory (Forster et al., 2013) and testing questions from the post-therapy questionnaire. The questionnaire contained Likert scales and free comments sections regarding the testing and questions that probed what usual care had been received (Appendix 21). The questionnaire was developed in Qualtrics and carried out by sharing the screen with the participant and going through the shared document. The usual care questionnaire was carried out by the doctoral candidate (ND), who had only met the UCC group for screening.

Sample size

The recruitment target of 36 participants was set with 18 participants allocated to each arm of the study. Assuming retention rates of 85%, this would allow 30 participants to be followed up at 19 weeks post randomisation. This sample is in line with recommended sample sizes for feasibility studies i.e., 24-50 participants (Julious, 2005) and will allow for parameters of a definitive trial to be estimated, such as recruitment rates, consent rates, completion rates, acceptability, and standard deviation of outcome measures for sample size calculation of a future trial.

Randomisation

To control for individual differences, participants were randomly assigned to the VESFA+UC or UCC. When 6 participants had been recruited, they were randomised into two groups of three participants. Allocation was concealed. A researcher blind to the screening and testing process (JM) allocated participants by entering participant numbers into the random list generator on www.random.org. It was stipulated that the first three numbers on the list would be treated and the last three would be controls. Randomisers use one of two methods, a mathematical formula that generates a string of random numbers (this formula can be replicated and therefore the random list can be repeated) or by extracting a truly random pattern from atmospheric phenomena e.g., the numbers generated by the static in lightning discharges. Random.org uses atmospheric noise as the basis for the list generator.

Blinding

The participants and the doctoral candidate were aware of group allocation. Outcome measures were administered and scored by speech and language therapy students blinded to group allocation. The participants were asked not to reveal group allocation to the testers during testing sessions. Testers were asked to inform the doctoral researcher if they thought they were unblinded.

Data analysis

Participant characteristics and the primary and secondary outcome measures' scores were summarised using descriptive statistics (means (SDs), medians (IQRs), counts (%), rates as appropriate), for the entire trial population and by trial arm, at each trial time point. Rates of missing data were reported. Potential imbalances between the groups at baseline were reported.

Despite this this being a feasibility trial, quantitative analysis compared the clinical outcomes of the two groups to reveal whether the intervention shows promise. These analyses were strictly exploratory. Data was restructured to a long format for analysis. Multilevel linear modelling (MLM) was used to look at the effect of intervention (UC+VESFA vs. UCC) across time (from baseline to T2 to T3) on language, communication and wellbeing/quality of life outcome measures. MLM allowed for baseline differences to be accounted for in the analysis, by including baseline as a covariate. As such, the difference between the groups at T2 is where any effects of treatment can be seen. No particular change is expected between T2 (post treatment) and T3 (follow up). Differences at T3 showed whether changes were maintained for 8 weeks. Additionally, MLM uses all available data, it doesn't use listwise deletion that occurs in wide format data, so can calculate outcomes with some missing items. The statistical analysis was completed using SPSS.

Data from questionnaires comprised of quantitative rating scale responses and qualitative free text responses. Descriptive statistics was used as appropriate to summarise quantitative responses. Free text responses will be transcribed verbatim and analysed using thematic analysis to identify themes within the data (Braun & Clarke, 2006) using NVivo 12 software.

Additionally, the advisory group of PWA were presented with the feasibility results, the acceptability findings and the clinical outcomes in turn. The group gave their views on the most important findings, the main message of the research and suggested next steps for this research. They had each result on a piece of paper to manipulate findings and support expression where needed.

Chapter 7 | Treatment Fidelity

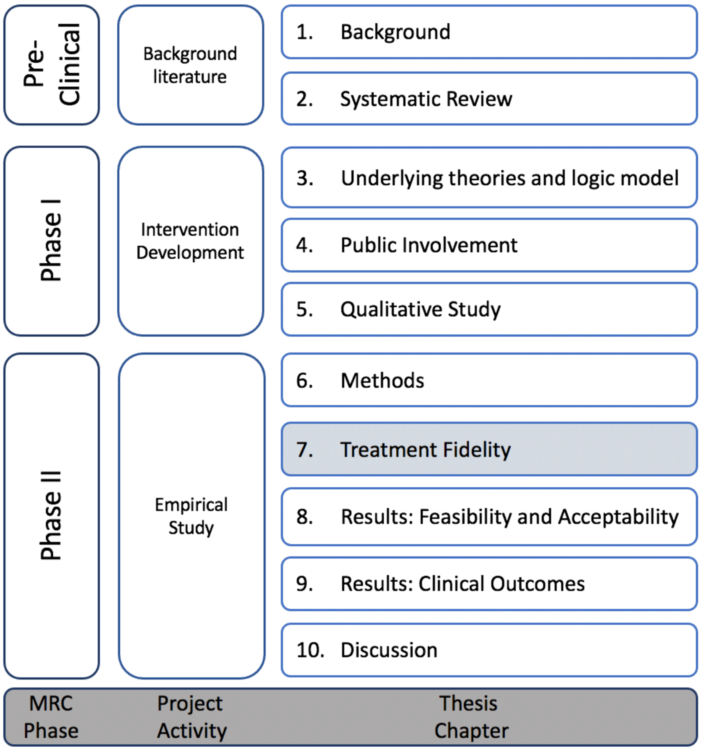

Figure 7.1: The thesis in a figure

Background

Treatment Fidelity (TF) is defined as the methods used 'to monitor and enhance the reliability and validity of behavioural interventions' (Bellg et al., 2004 p.443). Behavioural interventions, such as VESFA, are developed based on underlying theory and programme logic. Then they are tested. Confident conclusions about the outcomes cannot be drawn unless we know that the intervention was delivered as intended. Thus, well designed TF methods improve internal validity. Monitoring TF also leads researchers to identify issues in the delivery of the intervention, such as components with poor adherence. This may lead to refinements of the treatment protocol or to the methods used for training intervention providers. TF monitoring can enable hypotheses to be developed about which core components drive change. This supports the replication of the intervention, the external validity (Borrelli, 2011).

There are five areas where guidance exists to enhance TF: 1) the design of the study, 2) the training of intervention providers, 3) the delivery of treatment, 4) the receipt of treatment and the 5) enactment of treatment skills (Bellg et al., 2004). Taking each in turn, *study design* can ensure the treatment is in line with the underlying theory and equal doses are planned within each condition. *Training intervention providers* aims to ensure the participants receive the same intervention. The *delivery of treatment* monitors that the delivery is adhering to the treatment protocol. This is often assessed by watching intervention sessions and rating them against a fidelity checklist of core components. The advice on the percentage of sessions to be rated is 15-40% (Heilemann et al., 2014) and most studies report 10-20% (Hinckley & Douglas, 2013). Direct observation of sessions using a priori coding categories is considered the gold standard (Brogan, Ciccone, & Godecke, 2019). The *receipt of treatment* checks that participants can perform the targeted skill following treatment. This is most often achieved through administration of relevant outcome measures before and after an intervention. Receipt of treatment is also addressed by the skill of the provider (e.g., providers of aphasia treatment are trained to facilitate communication), by inclusion/exclusion criteria which ensure participants can access the treatment and by ensuring learning is incremental to support achieving the relevant skills (Behn et al., 2022). Finally, the *enactment of treated skills* monitors that participants can perform the targeted skill following treatment in real world settings. Difficult to assess, this aspect can be explored through post therapy interviews, e.g., with treatment participants and/or their family members. A good example comes from the Big Cactus study where participants were interviewed about factors that were associated with adherence (Harrison, 2019) and videos analysed of the use of the target words in conversation. It has been argued that following the Bellg guidance (Bellg et al., 2004) increases TF and, thereby improves 'the power to detect effects that might otherwise have been obscured by variance' (Spell et al., 2020) p.288).

Published accounts of TF in aphasia interventions are increasing. A review of TF reporting published in 2013 identified only 21 studies in the previous 10 years that had published TF (Hinckley & Douglas, 2013). A later review found 37 studies in the following 5 years (2012-2017), however, only 1 article in this review contained all five elements of TF (Brogan, Ciccone, & Godecke, 2019). Most recently, Behn and colleagues supplemented the Brogan review and explored seven RCTs that were either completed or ongoing in the years 2017-2021 (Behn et al., 2022). All five areas of fidelity were addressed in four RCTs. The least reported element in this review was treatment enactment. TF, key to our understanding of whether a treatment is valid, is now more consistently reported in aphasia studies.

TF has been monitored in two previous EVA Park studies (Marshall, 2020; Marshall in press) but not all (Carragher et al., 2020; Marshall et al., 2016; Marshall et al., 2018). In both studies, fidelity checking addressed study design, treatment delivery, and treatment receipt. In terms of treatment delivery, in a social support intervention in EVA Park, 32% of treatment sessions were rated for the delivery of core components by two raters independent to the study. Adherence ratings showed 81% of core components were fully present in the sessions and 12% were present to some degree. Inter-rater reliability was excellent (Marshall et al., 2020). In a scripting intervention, 27.5% of intervention sessions were rated for adherence to core components by two independent raters. 72.7%-91.6% of core components were present in the sessions. There was substantial agreement between raters (Marshall in press). Where reported, EVA Park studies have shown high adherence to core treatment components.

There are few reports of TF monitoring in previous studies of SFA (Evans et al., 2021; Gravier et al., 2018; Kendall et al., 2019; Kladouchou et al., 2017). SFA is highly prescribed making monitoring TF relatively straightforward. Indeed, where TF has been reported for SFA, adherence to the protocol was above 95% (Evans et al., 2021; Kendall et al., 2019). Elaborated SFA (ESFA) is more complex but adherence to the protocol remains high with a treatment adherence rating of >90% (Kladouchou et al., 2017).

Identifying the core components of SFA has been explored in a few studies (Boyle, 2010; Evans et al., 2021; Gravier et al., 2018; Quique, Evans, & Dickey, 2019; Sze et al., 2020). Feature generation, not feature repetition, appears to be a key driver of change in SFA studies (Boyle, 2010; Evans et al., 2021; Gravier et al., 2018). High dose has a positive impact on outcomes, for both treated and untreated items (Quique, Evans, & Dickey, 2019). The study gave 15 sessions as an example dose (Quique, Evans, & Dickey, 2019). A recent study explored all variables that might influence word finding outcomes. Providing the written form of the target as a cue was found to be a good predictor of outcomes, as was the provision of cues, dose parameters (number of sessions, number of times items were named) and the provision of feedback (Sze et al., 2020). This literature informed the development of the ESFA element of the VESFA intervention.

The core components of conversation groups are less well researched. One study proposes the mechanisms of change that support improved well-being within community aphasia groups (Attard et al., 2015). The authors suggest the opportunities for support, learning and communication are what make aphasia groups potent (Attard et al., 2015). This, and other fidelity monitoring of social support groups (Marshall et al., 2020), informed the development of the conversation groups in VESFA.

TF activities in this research trial aimed to answer three research questions:

1. Was the treatment delivered as planned?
2. Which components most influenced treatment adherence scores?
3. How reliable were the fidelity checklists?

Methods

Design

The VESFA trial was a feasibility randomised controlled trial with two arms: Virtual Elaborated Semantic Feature Analysis (VESFA) with a usual care control.
Full details of the VESFA intervention are reported in Tables 3.1 and 3.2, 'intervention' in Chapter 6, Appendix 7 (VESFA TIDieR description) Appendix 8 (VESFA therapy manual).

Fidelity Strategies

Treatment fidelity was supported in VESFA by strategies in four of the five fidelity areas identified by Bellg et al., (one, training providers, was not applicable) see Table 7.1. Study Design aspects are covered in more detail in Chapter 6, Methods.

Fidelity Strategy	How it was addressed in the VESFA trial
Study Design	• Pre-specified participant inclusion/exclusion criteria • Therapy manual specified 40 hours of intervention for all participants in treatment arm • Testing protocol specified 2 hours per testing session, with the same assessments to be delivered in the same order for both study arms • Planned to record sessions (1 session per participant per week and all groups) for later adherence checking
Training Providers	*No providers were trained because the treatment was delivered by the researcher who developed the intervention and wrote the manual.*
Delivery of Treatment	• Treatment delivery followed the treatment protocol outlined in the VESFA Therapy Manual (Appendix 8) • Adverse events were recorded (see Chapter 8, safety) • Session videos were rated for adherence to the protocol
Treatment receipt	• Provider was an experienced aphasia therapist, with skills in providing hierarchical cues, in supporting conversations etc. • Inclusion criteria ensured those receiving the treatment could participate (had a minimum level of comprehension) and had room to improve (anomia identified at screening) • A rationale was given for the ESFA activity when it was introduced • Feedback in the group was aimed to highlight how the use of strategies supported conversations • Practiced target words in naming, phrases and conversations in sessions
Treatment enactment	• Provided Challenge Tasks. Participants were asked to identify a real world situation where the conversation could be practiced in a real situation before the next group • Qualitative post therapy questionnaires specifically asked: 'have you used the words and phrases practiced in EVA Park in real world conversations?'

Table 7.1: Fidelity strategies and how they were addressed in VESFA trial

Development of the treatment delivery fidelity checklists

Two fidelity checklists were developed to monitor treatment delivery. Fidelity Checklist A (Figure 7.2) outlined the core components for the individual ESFA sessions. Fidelity Checklist B (Figure 7.3) outlined the core components for the group conversation sessions. The fidelity checklists were based on the core activities identified in the intervention development (Tables 3.1 and 3.2) and informed by published TF checklists for ESFA (Kladouchou et al., 2017) and TF checklist for a virtual group intervention (Marshall et al., 2020). The fidelity checklists were drafted by the doctoral researcher with input from a master's student (SM). They were reviewed by members of the supervision team and a member of the team at City with expertise in fidelity and finalised through an iterative process. The core components of the treatment outlined in the checklists were verified by a workshop with the trial advisory group. In this workshop the advisory group members generated what important activities drive change in therapies and identified what activities in VESFA were the important ones. The advisory group ratified all checklist items with the exception of two. Opinion was divided about whether it was necessary to provide a rationale for an activity. Additionally, advisory group members were very cautious about recommending feedback. They felt it was so reliant on the skill of the therapist to be sensitive, that there was a chance that feedback might be detrimental to participant confidence. Despite this, the rationale and feedback remained in the checklist as there was evidence for these to support the adoption of communication strategies (see Chapter 3, Intervention Development). The sensitivity of feedback should be addressed in any future training of treatment providers.

A) VESFA individual session					
Session ID:		Rater ID:		Date:	
Item		Component	Please tick		Comments
			Done	Not done	
At the start	1	SLT gives rationale for chart activity			
	2	SLT gives the opportunity to recap the items not named in the previous session			
For each target			Done >/=75%	Done <75%	
	3	SLT provides a naming opportunity for the target word			
	4	If the word cannot be named, the SLT follows a cueing hierarchy			

	5	SLT elicits a minimum of 4/6 SFA chart categories			
	6	SLT writes generated features			
	7	SLT provides the opportunity for the participant to produce a phrase or sentence			
			Done	Not done	
At the end	8	SLT asks the participant to name all the targets worked on within the topic			
	9	SLT provides specific feedback. *Examples: number of words correct, effective strategies, supportive cues*			
		Total:	/9		

Figure 7.2: Fidelity Checklist A for the VESFA individual sessions

B) VESFA conversation group					
Session ID:		Rater ID:		Date:	
Item		Component	Please tick		Comments
			Done	Not done	
At the start (10mins)	1	SLT acknowledges each person in the group			
	2	SLT provides the opportunity for participants to share news			
	3	SLT introduces the group structure and topic of the session			
Activity 1: Articulate (20min)	4	SLT describes 'Articulate': Includes the need to describe /give clues and guess			
	5	In the vocabulary recap, SLT provides the opportunity for participants to retrieve target words			
	6	SLT offers each participant a turn describing a target item			
	7	SLT provides specific feedback on 'articulate' descriptions to participants			
Activity 2: Conversa-tion (30min)	8	SLT introduces the group's conversation topic			
	9	The virtual setting is linked to the topic. *Example: recipes are shared in the kitchen.*			
	10	SLT provides the opportunity for each participant to contribute to the conversation *Example: invites a contribution from someone who has not yet spoken*			
	11	SLT provides specific feedback on the strengths of the conversation contributions			

		Examples: range of words, structure of story			
Activity 3: Bingo (15min)	12	SLT describes the BINGO game. Includes the need to say the words on the BINGO card			
	13	SLT provides the opportunity for participants to have a turn playing BINGO			
At the end (5-10min)	14	SLT provides the opportunity for participants to reflect on their strengths, by asking 'what have you been pleased to notice?'			
	15	SLT provides a challenge task or homework to be carried out independently before the next session			
Overall	16	There are more than three demonstrations of enjoyment of the activity *Examples: laughing, jokes*			
		Total:	/16		

Figure 7.3: Fidelity checklist B for VESFA group sessions

Data Sampling

A total of 96 ESFA sessions and 96 group sessions was delivered to 16 participants in the VESFA therapy (total sessions=192). Participants were treated in sets of 3 participants at a time thus, x6 8-week sets were delivered between December 2020 – February 2022. During the treatment period 174 sessions were videoed. A sample of 39 videos, representing 20% of total sessions, were selected for fidelity rating. The group and individual sessions were selected independently to ensure both session types were well represented. The list randomiser on www.random.org was used to randomise the list of session videos. The first 20 individual sessions and 19 of group sessions were selected from the list. A range of early and late sessions were represented in the sample, see individual session numbers (possible range 01-16) in Table 7.2. All 6 sets were represented in the group sessions, see Table 7.7 (possible range VESFA1-VESFA6). Of these 39 videos, 35 were used for intra-rater and inter-rater reliability checks.

Data Allocation

Seven raters carried out the ratings of treatment fidelity. They are identified here as raters A-G. Rater A was a qualified speech and language therapist, raters C-G were speech and language therapists in training and rater B was an undergraduate student in human communication. All raters were independent of the treatment study but familiar with the treatment to different degrees. Raters A and B watched videos of sessions to familiarise themselves with the treatment. Raters C-G had delivered the VESFA treatment in a student placement so had direct experience of the intervention.

			Rater:	
	Video	**Adherence**	**Inter-rater reliability**	**Intra-rater reliability**
Individual sessions	S01 042.S11	A	B	A
	S02-094.S07	A	B	A
	S03 015.S14	A	B	A
	S04 060.S02	A	B	A
	S05 115.S07	A	B	A
	S06 087.S11	E	D	E
	S07 009.S12	E	D	E
	S08 007.S16	E	D	E
	S09 098.S08	E	D	E
	S10 098.S09	E	D	E
	S11 113.S08	C	G	C
	S12 107.S15	C	G	C
	S13 053.S09	C	G	C
	S14 107.S06	C	G	C
	S15 087.S07	C	G	C
	S16 065.S09	D	C	D
	S17 021.S15	D	C	D
	S18 075.S09	D	C	D
	S19 088.S13	A		
	S20 115.S12	B		

Group sessions	G01 VESFA5grp01	B	A	B
	G02 VESFA1grp05	B	A	B
	G03 VESFA6grp10	B	A	B
	G04 VESFA5grp4	B	A	B
	G05 VESFA6grp03	B	A	B
	G06 VESFA5grp04	F	E	F
	G07 VESFA4grp05	F	E	F
	G08 VESFA4grp14	F	E	F
	G09 VESFA5grp03	F	E	F
	G10 VESFA2grp05	F	E	F
	G11 VESFA1grp10	G	F	G
	G12 VESFA4grp07	G	F	G
	G13 VESFA6grp08	G	F	G
	G14 VESFA4grp11	G	F	G
	G15 VESFA1grp08	G	F	G
	G16 VESFA3grp14	D	C	D
	G17 VESFA4grp09	D	C	D
	G18 VESFA1grp06	A		
	G19 VESFA3grp10	B		

Table 7.2: Allocation of individual session and group session videos to raters A-G

Raters A and B were allocated 17 videos to rate and raters C-G 15 videos to rate. This included videos that were watched twice for intra-rater reliability, see Table 7.2.

Training procedure

All raters attended a 1.5 hour training session. The training aims were for raters to understand the concept of treatment fidelity, to be familiar with all items on the checklist and to gain experience of rating a range of items on the checklist. Following the training session, raters rated an example individual and example group session and then met to share ratings, discuss discrepancies and agree ratings going forward. These example videos were not used in the actual scoring. Disagreements were identified in decisions about the cueing hierarchy, feedback and rationales. Following these disagreements further criteria were developed to guide raters. These were 1) if cueing was used and was hierarchical (started minimal and became more supportive) then this item can be marked as DONE, 2) feedback can only be considered done if it is specific e.g., mentions something about what the participant did. For example, 'good' 'excellent' are NOT DONE and 3) rationale needed to be specific to SFA. The SLT needed to refer to improving word finding/word retrieval and/or strengthening networks.

Procedure for rating adherence

Raters watched the videos of individual and group sessions with the relevant checklist and ticked ‘done’ or ‘not done’ adding notes where necessary. Raters were instructed to rate the videos in a private space and watch the full session with the checklist in front of them. For items where scores were above or below a certain percentage, raters were advised to keep a tally of each episode seen and calculate the score at the end of the video. In individual sessions, each video had a maximum score of 9. Group sessions in set 1 had a maximum score of 14, and group sessions sets 2-6 had a maximum score of 16. Two components were added to the group protocol after set 1 (see Chapter 3). Scores were either 1, for present or >/ =75%, or 0, for not present or <75%, see checklists in Figures 6.1 and 6.2.

Procedure for reliability

The scores of 35 session videos were compared between the same rater across different time points (intra-rater) and two raters (inter-rater) to determine the reliability of the checklists, Table 6.2 shows how these were allocated. A minimum of 10 days was left between ratings for intra-rater reliability (Streiner, Norman, & Cairney, 2015). This aimed to ensure that the new ratings were based on what was seen in the video and not remembering the previous score.

Data analysis

Adherence scores were calculated as a percentage: present items were divided by total items and multiplied by 100. An adherence percentage of 80% or more represented high fidelity (Heilemann et al., 2014). Intra- and inter-rater reliability was calculated using a Cohens Kappa coefficient in the software Statistical Package for Social Sciences (SPSS). A kappa value can range from -1 to +1. Reliability is considered almost perfect if the value of Kappa is above .90, strong if the value of kappa is .80-.90, moderate if the value of kappa is .60-.79, weak if the value is .40-.59, minimal if the value is .21-.39 and no reliability if the value is 0-.20, see Table 7.3 (McHugh, 2012).

Kappa value	Level of agreement	Percentage of the data that are reliable
0-.20	None	0-4%
.21-.39	Minimal	4-15%
.40-.59	Weak	15-35%
.60-.79	Moderate	35-63%
.80-.90	Strong	64-81%
Above .90	Almost perfect	82-100%

Table 7.3: Interpretation of Cohen's kappa (McHugh, 2012 p.7)

Results

Study Design

34 participants met inclusion criteria and equal numbers (n=17) were randomised to the VESFA treatment arm and the Usual Care Control. 16 participants completed the 8 week treatment. 94% of sessions ran as planned, with only 30/512 sessions cancelled. 75% of participants (12/16 participants) received over 90% (>36/40hours) of the intended dose.

Delivery of treatment

Of the 39 videos randomly selected for rating 36 were rated (18%, 35/192). Three videos were not rated due to technical problems with the recordings e.g., no sound on the video. Videos comprised of 18 individual session videos and 17 group session videos from the full range of 16 sessions.

Adherence to Individual session protocol

Individual sessions had nine core components. Sessions could therefore gain a maximum score of 9. Adherence scores ranged from 44% (4/9) to 89% (8/9) adherence. See Table 7:4 for scores for individual sessions.

	Individual session	Components delivered and components planned (*actual score /maximum score*)	Adherence $\frac{maximum\ score}{actual\ score} x\ 100$
1	Ppt042.S11	7/9	78%
2	Ppt094.S07	7/9	78%
3	Ppt015.S14	6/9	67%
4	Ppt060.S02	6/9	67%
5	Ppt115.S07	7/9	78%
6	Ppt107.S14	8/9	89%
7	Ppt107.S12	8/9	89%
8	Ppt060.S08	8/9	89%
9	Ppt098.S08	7/9	78%
10	Ppt098.S09	7/9	78%
11	ppt113.S08	8/9	89%
12	ppt107.S15	8/9	89%
13	ppt053.S09	6/9	67%
14	ppt107.S06	8/9	89%
15	ppt087.S07	6/9	67%
16	ppt021.S15	4/9	44%
17	ppt021.S01	8/9	89%
18	ppt115.S12	7/9	78%
	Total	**126/162**	
	Mean	**7/9**	**78%**

Ppt=participant, S=session e.g., ppt042.S11= participant 42, session 11

Table 7.4: Adherence scores for individual sessions

Adherence to group session protocol

Group sessions had 16 core components. Two components were added to the review after set 1, therefore all set 1 groups (VESFA1) had a total of 14 core components. Adherence scores ranged from 50% (7/14) to 100% (16/16). See Table 7.5 for adherence scores for group sessions.

	Group session	Components delivered and components planned (*actual score /maximum score*)	Adherence $\frac{maximum\ score}{actual\ score} x\ 100$
1	VESFA5.Group01	13/16	81%
2	VESFA1.Group05	7/14	50%
3	VESFA6.Group10	14/16	88%
4	VESFA5.Group04	12/16	75%
5	VESFA6.Group03	14/16	88%
6	VESFA4.Group05	16/16	100%
7	VESFA4.Group14	14/16	88%
8	VESFA5.Group03	13/16	81%
9	VESFA2.Group05	15/16	94%
10	VESFA1.Group10	11/14	79%
11	VESFA4.Group07	15/16	94%
12	VESFA6.Group08	15/16	94%
13	VESFA4.Group11	14/16	88%
14	VESFA1.Group08	11/14	79%
15	VESFA3.Group14	15/16	94%
16	VESFA4.Group09	15/16	94%
17	VESFA1.Group06	9/14	64%
18	VESFA3.Group10	12/16	75%
	Total	**235/282**	
	Mean	**13/16**	**84%**

VESFA# refers to the number of the set. There were 6 sets.

Table 7.5: Adherence scores for group sessions

Overall adherence

When the mean adherence for both individual and group sessions was combined ((126+235)/(162+282)=81.3), the adherence to the protocol in VESFA intervention was 81%.

Ratings by component

To understand what components of the sessions were driving variability in the adherence ratings we looked at the item responses. Figure 7.4 shows the frequency that items were present in the individual sessions. Two items scored low for adherence, (A1) rationale for the activity was only seen once in 18 sessions and specific feedback (A9) was seen 50% (9/18) of the time. All other components where present more than 72% of the time. The naming of the target word (A3) and the use of hierarchical cueing by the therapist (A4) were present in all rated sessions.

If we remove the item A1. Rationale, the adherence rating rises to an average of 88% across the individual sessions and an overall adherence of 86%.

A1. Rationale	5%
A2. Recap	72%
A3. Name target	100%
A4. Cueing	100%
A5. Features	94%
A6. Written name	94%
A7. Phrase	94%
A8. All targets	89%
A9. Specific Feedback	50%

Figure 7.4: Visualisation of the frequency that each checklist item was rated present (green) or not present (pink) in the individual sessions

The frequency of components present in the group sessions is shown in Figure 7.5. In the group sessions a description of the group structure (B3) was the least present component, seen in 59% of rated sessions (10/17). Specific feedback (B7, B11) was seen in 65% of the sessions rated (11/17). The introduction of the conversation topic, the opportunity for participants to take a turn in the conversation and evidence of enjoyment (B8, B10, B16) were seen in all group sessions rated.

B1. Acknowledge	88%
B2. News	88%
B3. Structure	59%
B4. Articulate: introduction	76%
B5. Articulate: targets	82%
B6. Articulate: describe	82%
B7. Articulate: feedback	65%
B8. Conversation: topic	100%
B9. Conversation: setting	82%
B10. Conversation	100%
B11. Conversation: feedback	65%
B12.Bingo: introduction	82%
B13.Bingo: turn	82%
B14.Reflect	94%
B15. Challenge task	71%
B16: Enjoyment	100%

Figure 7.5: Visualisation of the frequency that each checklist item was rated present (green) or not present (pink) in the group sessions

Reliability of the checklists

Fidelity Checklist A

Fidelity checklist A rated the individual sessions (Figure 7.2). Two raters independently rated 17 individual sessions, see Table 7.6. Inter-rater reliability was perfect for 65% of sessions (11/17) with Kappa=1, p<.001, moderate for 37% of sessions (4/17), Kappa=0.63-0.67, and weak for 12% sessions (2/17), Kappa=.44 and .52.

The 17 sessions were also rated by the same rater twice (Table 7:6). Intra-rater reliability was perfect for 94% of sessions (16/17). For the one session with disagreement reliability was moderate with Kappa=.70, p<0.001.

Fidelity Checklist A. Individual Sessions							
Inter-rater reliability				Intra-rater reliability			
Rater	Kappa	Sig.	95%CI	Rater	Kappa	Sig.	95%CI
A & B	1	<0.001	1-1	A	1	<0.001	1-1
A & B	0.63	<0.001	.31 - .95	A	1	<0.001	1-1
A & B	0.67	<0.006	.22- 1.12	A	1	<0.001	1-1
A & B	0.52	<0.002	.18 - .85	A	0.70	<0.001	.42 - .98
A & B	0.63	<0.001	.31 - .95	A	1	<0.001	1-1
E & D	1	<0.001	1-1	E	1	<0.001	1-1
E & D	1	<0.001	1-1	E	1	<0.001	1-1
E & D	1	<0.001	1-1	E	1	<0.001	1-1
E & D	1	<0.001	1-1	E	1	<0.001	1-1
E & D	1	<0.001	1-1	E	1	<0.001	1-1
C & G	1	<0.001	1-1	C	1	<0.001	1-1
C & G	0.44	0.048	-0.2-1.09	C	1	<0.001	1-1
C & G	1	<0.001	1-1	C	1	<0.001	1-1
C & G	1	<0.001	1-1	C	1	<0.001	1-1
C & G	1	<0.001	1-1	C	1	<0.001	1-1
D & C	0.62	<0.003	.32 - .91	D	1	<0.001	1-1
D & C	1	<0.001	1-1	D	1	<0.001	1-1
Mean:	**.85**			**Mean**	**.98**		

Sig.=significance, CI=Confidence Interval

Table 7.6: Results for inter- and intra-rater reliability rating for Fidelity Checklist A

Fidelity Checklist B

Fidelity checklist B rated the group sessions (Figure 7.3). Two raters independently rated 15 group sessions, see Table 7.7. Inter-rater reliability was perfect for 20% of the sessions (3/15), moderate for 40% of sessions (6/15), weak for 27% (4/15) and minimal for 13% (2/15).

The 15 sessions were also rated by the same person twice (Table 7.7). Intra-rater reliability was perfect for 60% of sessions (9/15), moderate for 20% of sessions (3/15), weak for 7% of sessions (1/15) and minimal for 13% of sessions (2/15).

Fidelity Checklist B. Group Sessions							
Inter-rater reliability				Intra-rater reliability			
Rater	Kappa	Sig.	95%CI	Rater	Kappa	Sig.	95%CI
B & A	0.54	<0.001	.20 - 88	B	0.72	<0.001	.45 - .99
B & A	1	<0.001	1-1	B	1	<0.001	1.0 - 1.0
B & A	1	<0.001	1-1	B	1	<0.001	1-1
B & A	0.62	<0.001	.32 - .92	B	1	<0.001	1-1
B & A	0.72	<0.001	.45 - .99	B	0.72	<0.001	.45 - .99
F & E	0.39	<0.001	.03 - .74	F	0.32	<0.001	-.03 - .67
F & E	0.72	<0.001	.45 - .99	F	0.72	<0.001	.45 - .99
F & E	0.72	<0.001	.45 - .99	F	0.54	<0.001	.20 - .88
F & E	0.65	<0.001	1-1	F	0.29	0.034	-.12 - .70
G & F	0.54	<0.001	.20 - .88	G	1	<0.001	1-1
G & F	0.54	<0.001	.20 - .88	G	1	<0.001	1-1
G & F	1	<0.001	1-1	G	1	<0.001	1-1
G & F	0.78	<0.001	.47- 1.08	G	1	<0.001	1-1
G & F	0.29	<0.034	-.12 - .70	D	1	<0.001	1-1
G & F	0.54	<0.001	.20 - .88	D	0.72	<0.001	1-1
Mean:	**.67**			**Mean:**	**.80**		

Sig.=significance, CI=Confidence Interval

Table 7.7: Results for inter- and intra-rater reliability rating for Fidelity Checklist B

Treatment Receipt

In order to support the receipt of treatment, all participants had a minimum level of comprehension (screening criterion >/=6/10 on FAST comprehension). This ensured that they could access the virtual world, EVA Park, and the focus on verbal language in VESFA. Three components of the checklists supported treatment receipt: rationale for activities (item A1), specific feedback on actions (item A9) and conversation practice (item B8). A rationale for the activity and feedback on actions were the least present in the VESFA therapy sessions rated (see Figure 7.4). However, the opportunity to practice the use of the target words (item B8, Figure 6.8) was present in all sessions rated, which meant participants used the target words during therapy sessions and in subsequent therapy sessions.

Treatment Enactment

Treatment enactment was supported through the provision of challenge tasks. Challenge tasks were presented to participants in 71% of the sessions rated.

Additionally, questions in the post therapy questionnaire probed treatment enactment through the question: 'have you used the words and phrases practiced in EVA Park in real world conversations?'. 94% of participants (14/15) answered positively to this question, Figure 6.9. Examples include how the practice in EVA Park supported talking to a waiter in a restaurant "a meal - me in a restaurant, sentences." (ppt21) and speaking about holidays in EVA Park inspired a conversation about holidays in the real word.

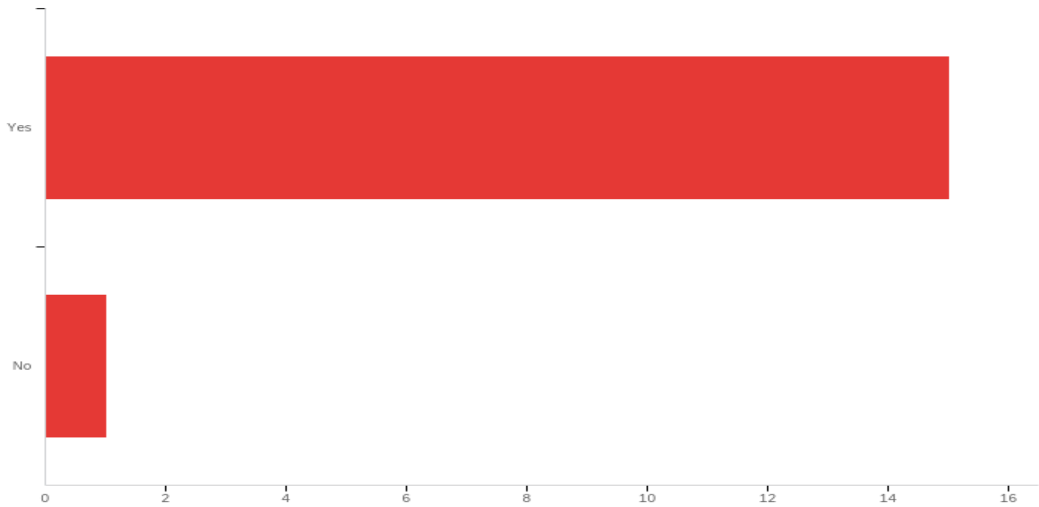

Figure 7.6: Answers to 'Have you used the words and phrases practiced in EVA Park in real world conversations?'

Discussion

In the VESFA trial TF strategies covered study design, treatment delivery, treatment receipt and treatment enactment. Study design strategies ensured 94% of sessions ran as planned and 75% of participants (12/16) received over 90% (>36/40hours) of the intended dose. The average TF across all sessions rated was 81%, demonstrating a high degree of fidelity in the delivery of the VESFA intervention (Heilemann et al., 2014). The TF for individual sessions was 78% and 84% for group sessions. In terms of treatment receipt strategies, a rationale for the activities and specific feedback for performance were not practised as intended by the therapist. Nevertheless, participants consistently practised target words both in individual session and in conversations in the group sessions, demonstrating treatment receipt. 94% of participants (14/15) reported the words and phrases practiced in EVA Park in real world conversations, indicating treatment enactment.

Adherence TF of 81% is comparable to other EVA Park interventions, with the scripts therapy reporting over 80% (Marshall in press) and the social support intervention reporting 81.9% (Marshall et al., 2020). However, adherence was lower than face-to-face ESFA interventions (Kladouchou et al., 2017) and other aphasia interventions (Bacon et al., 2021; Heilemann et al., 2014) where TF was reported at over 90%. Although the average TF score was strong, one session scored less than 50% adherence (ppt021.S15, Table 7.4). This was session 15/16. In sessions 13, 14 and 15 the therapy protocol dictates that the therapist recaps all 30 words in the topic (see Appendix 8). Thus, the SFA chart is not completed for each word. It is likely that items A5-A7 were not elicited in this session leading to the low score. A future trial could address how to check these different session types within the VESFA intervention.

A closer look at the ratings shows that item A1.Rationale was only seen once in session 021.S01. This is the only session 1 in the sample. It is likely that the activity rationale was given when the activity was first introduced but not in subsequent sessions. This raises the question: is providing a rationale for the activity in every session a core component of the VESFA therapy? Understanding why treatment works is a construct within the acceptability of an intervention (Sekhon, Cartwright, & Francis, 2017). ‘Intervention coherence’, the participant’s understanding of the intervention and how it works, and ‘perceived effectiveness’, the extent to which the intervention is perceived to work, contributes to acceptability, impacting adherence and therefore clinical outcomes. We know from clinical experience that information needs to be given more than once to be taken on, and this is confirmed in the literature (Kessels, 2003). This suggests participants would need to hear the rationale more than once to understand why an activity is done. Providing a rationale for the task fits with a metacognitive approach, described in Chapter 3.

The other component with low ratings was feedback (A9.Specific Feedback, B7.Articulate feedback and B11.Conversation feedback). This was marked as present if the feedback provided specific information related to the task. For example, ‘telling us the location helped us to guess the word' would be rated as present but 'well done' would be rated as not present. It has long been established that feedback can improve performance (Thorndike, 1927). In aphasia rehabilitation feedback is multifunctional (Simmons-Mackie, Damico J, & Damico H, 1999). It serves to shape target behaviours, encourage, boost confidence, maintain a partnership, set a tempo in the task, consolidate the therapist-client roles and communicate rules (Simmons-Mackie, Damico J, & Damico H, 1999).). This checklist only captured the feedback that shaped the target behaviours. Physiotherapists give more motivational than informational feedback during stroke rehabilitation activities (Stanton et al., 2015). It is likely that this is the case with SLTs too. Motivational feedback can highlight a sense of success and boost confidence leading to increased self-efficacy. Future studies could accept both forms of feedback.

It is interesting to note that the two items that were least present in the sessions, rationale and feedback, were the two that the advisory group suggested might not be core activities, and those most difficult to rate consistently during training. These items support the recipient of the treatment to understand why the treatment works and what they can do to benefit from treatment. There are interventions where the repetitive practice is the driver of change e.g., those that aim to reorganise neural networks. In behavioural interventions understanding why you should change your behaviour influences outcomes e.g., when is the right moment to use the strategy you have worked on. VESFA pulls on both these mechanisms; strengthening the semantic neural networks and using the features' descriptions to compensate for word finding occurrences in conversations. As such they are core components of the intervention. Adherence testing is carried out to pick up the issues with treatment delivery. Here it has revealed issues with the consistency of rationale and feedback. The VESFA therapy manual should be updated to place emphasis on a regular description of the task rationale and examples of specific task related feedback, and the timing of feedback. Future TF studies could give more opportunities for benchmarking in training.

Inter-rater reliability was moderate (average Kappa of .76) and intra-rater reliability was strong (average Kappa of .89) showing that the fidelity checklists were reliable. The mean inter-rater reliability for the individual session checklist was strong, but moderate for the group session checklist. A similar pattern was seen in intra-rater reliability where reliability of the individual session checklist was almost perfect and strong for the group session checklist. Interestingly the TF scores were higher for group than individual sessions, but reliability of the scoring of the group checklist was poorer. Low intra-rater reliability scores suggest a training issue. Some items on the fidelity checklist were objective e.g., participant produces the target word, and some were more subjective e.g., feedback was specific to the task. The latter example can be harder to judge. Raters had one opportunity to discuss differences in judgments in the training. They may have benefited from more opportunities to benchmark, particularly for feedback and cueing hierarchy items. Disagreements by item were predominantly for judgements about feedback in both individual and group sessions. This matches a comparable aphasia fidelity paper where disagreements were due to judgements about differential feedback (Bacon et al., 2021).

The number of raters was high. Seven students rated the session videos. They had a wide range of experience; one rater had no experience of rehabilitation, 5 were first year SLTs in training and one was a qualified SLT. These differences would have contributed variability to the ratings. This is evident in the range of reliability ratings. The ratings for the group sessions were less reliable. It is possible that the components in the groups required more judgement e.g., what constitutes specific feedback vs generic feedback.

Treatment receipt strategies are often carried out in the screening of a trial, and occasionally during a trial. For example, a checklist ensured the participant was attentive, understood and the attempted to do the behavioural task of interest (Spell et al., 2020). The VESFA trial addressed treatment receipt in screening only.

Measurement of treatment enactment is rare in aphasia studies. In a recent review two studies demonstrated treatment enactment (Breitenstein et al., 2017; Palmer et al., 2015) and two studies partially met the review criteria. In the VESFA trial treatment enactment was addressed by asking about enactment in an interview. This is method was used in a recent peer befriending study (Hilari et al. 2021). One study aimed to capture the enactment of therapy skills by observing participants in conversation (Palmer et al. 2019).

Planning for treatment fidelity before the intervention is delivered can increase the chance of high fidelity ratings (Behn et al., 2022; Brogan, Ciccone, & Godecke, 2019). In this study the fidelity checklists were not developed prior to the intervention. As a result, some core components were not delivered consistently across sessions. Completing a treatment provider self-assessment after each session can keep the emphasis on the core components, especially in long delivery periods such as in this study (18 months). Despite the long treatment period, there was no apparent effect of time on adherence (Table 7.4).

A limitation of this study was the retrospective development of the checklist. Prospective, iterative fidelity checking can pick up on issues and adjust them in a long trial. More benchmarking in training could have led to better reliability in the scoring.
Additionally, this study did not explore the fidelity of assessment. Although guidance now exists for treatment fidelity (Bellg et al., 2004), there is very little guidance for researchers on assessment fidelity (Richardson et al., 2016). In the VESFA trial the assessments were completed by a large number of testers. Sixteen undergraduate and masters speech and language therapy students conducted the testing and scoring of tests. They delivered the testing either as part of their research dissertation projects or as a volunteer to gain experience of working with people with aphasia. There were 7 assessments in the testing battery with differing administration guidelines, see Chapter 6 for a description of the measures. The complexity of many assessments and many testers leads to a higher chance of assessor drift or contamination between administration procedures. The study design ensured testing was a maximum length of 2hrs, although some participants had more than one testing session. Tester training and a testing protocol ensured all participants received the same assessments in the same order and testers were blinded to group allocation. But no strategies were in place to monitor adherence or consistency of assessment scoring beyond the initial training session. In future studies assessment sessions could be videoed and rated against a checklist and/or tester self-report checklists.

Training of intervention providers was not an issue in this trial, but a future trial with multiple sites would need consistent training for providers. Here fidelity of training might also be assessed.

Conclusion

A range of treatment fidelity strategies were embedded within the trial protocol (study design, treatment delivery, treatment receipt and treatment enactment). VESFA demonstrated high adherence to the core components of the intervention. The reliability of the checklist was moderate to strong. The components that elicited most drift were providing a rationale for activities and the provision of specific feedback. Training for fidelity raters should include more opportunities to benchmark. The therapy manual should be updated to include a task rationale and the provision of specific task-related feedback.

Chapter 8 | Results: Feasibility and Acceptability

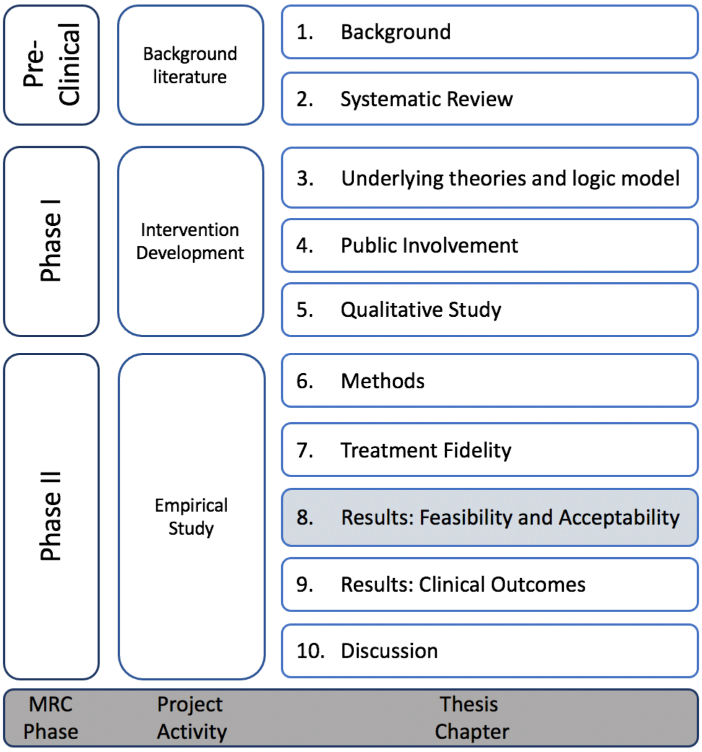

Figure 8.1: The thesis in a figure

This chapter reports on participant flow in the study, participant characteristics, feasibility and acceptability outcomes.

Participants

Participant flow through the study is described in the CONSORT diagram in Figure 8.2.

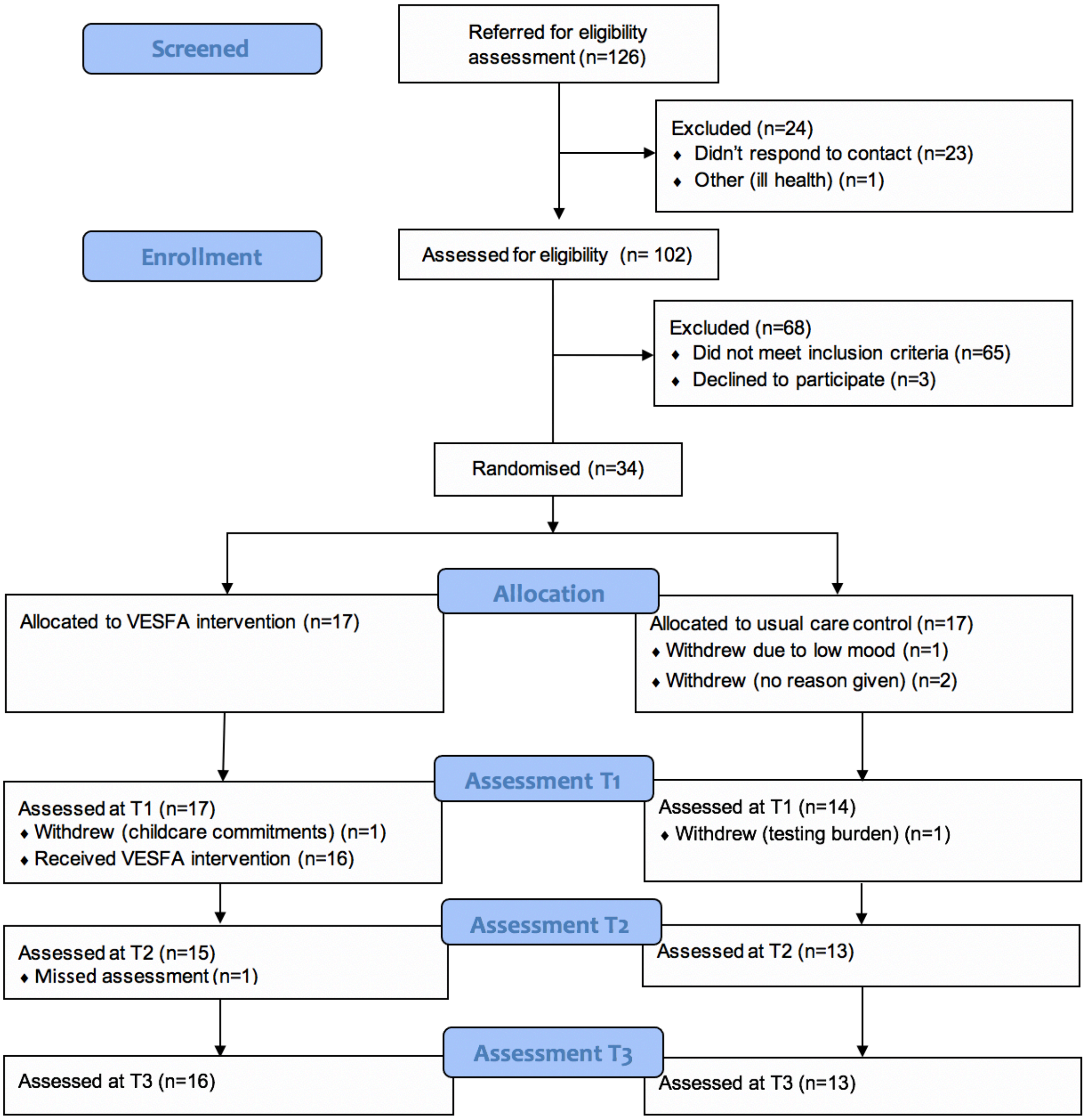

Figure 8.2: CONSORT flow diagram

Participant recruitment began in October 2020 and finished in December 2021. The first set had initial T1 in November 2020 and the final set finished T3 testing in April 2022.

Participant characteristics

Thirty-one participants underwent baseline testing, see Table 8.1. This table details the participant characteristics as an overall group and split by VESFA+ UC treatment (n=17) and Usual Care Control (UCC) (n=14) groups.

		VESFA n=17	UCC n=14	Total n=31
Age in years, mean (SD)		57.18 (11.67)	57.07 (8.06)	57.13 (10.04)
Sex	Female	4 (23.5%)	8 (57%)	12 (39%)
	Male	13 (76.5%)	6 (43%)	19 (61%)
Ethnicity		n=16*	n=13*	n=29
	White	13 (76.5%)	9 (64%)	22 (71%)
	Asian	2 (12%)	2 (14%	4 (13%)
	Black	1 (6%)	2 (14%)	3 (10%)
Marital status		n=17	n=13*	n=30
	Single	2 (12%)	3 (21%)	5 (16%)
	Lives with partner/married	12 (70.5%)	10 (71%)	22 (71%)
	Divorced/ widowed	3 (18%)	0	3 (10%)
Social Class		n=17	n=14	n=31
	Professional, managerial & technical occupations	11 (65%)	6 (43%)	17 (55%)
	Skilled occupations	4 (23.5%)	5 (36%)	9 (29%)
	Partly skilled & unskilled occupations	1 (12%)	3 (21%)	5 (16%)
Language Status		n=17	n=12*	n=29
	English as first language	14 (82%)	9 (64%)	23 (74%)
	Fluent but not native speaker	3 (18%)	3 (21%)	6 (19%)
Months Post Stroke *median* (IQR)		*35 (17.5-88.5)*	*50.5 (33.5-62.5)*	*41 (24-66)*
Aphasia Severity		n=17	n=14	n=31
	Mild	6 (35%)	1 (7%)	7 (23%)
	Moderate	9 (53%)	8 (57%)	17 (55%)
	Severe / very severe	2 (12%)	5 (36%)	7 (23%)

*includes missing data

Table 8.1: Participant baseline characteristics (T1).

The participants had an average age of 58 years and were predominantly white (71%), living with a partner (71%), with English as their first language (74%). The majority (55%) presented with a moderate aphasia, as measured by the Western Aphasia Battery-Revised (Kertesz, 2007).

The groups were well balanced in terms of age and had a similar pattern of ethnicity, marital status, social class and language status. Social Class was determined using the national office for statistics three class version of the National Statistics Socio-Economic Classification (NS-SEC) (Rose and Pevalin, 2003).

The UCC group had slightly more females (8/14, 57%), whereas the majority of the VESFA group were male (13/17, 76.5%); this difference was not significant (χ^2 (1)=3.66, p=.06). Participants in the UCC group seemed to be longer post stroke [median (IQR)= 50.5 (33.5-62.5)] versus [median (IQR)= 35 (17.5-88.5)] in the VESFA group, but the difference was not significant (U=94, p=.336). Only 1 (7%) person in the UCC group had mild aphasia versus 6 (35%) in the VESFA group, and 5 (36%) had severe aphasia in UCC versus 2 (12%) in the VESFA group. It was not feasible to compare this statistically as 66% of expected count cells in the chi-square analysis had values <5.

a) Feasibility of recruitment and retention to the trial

The pre-specified criteria stipulated that the feasibility of recruitment was met if at least 60% of eligible participants were consented to the trial. 102 people were screened and 37 were eligible (36%). Of the 37 who were eligible, 34 consented to the trial. Thus, **92% of eligible participants consented** to the VESFA trial, meeting this feasibility criterion.

Recruitment ran from 11/09/2020 until 09/11/2021. In that 14 month period 34 people consented and were randomised giving a rate of participants consented of 2.43 per month, see Table 8.2. Of the 34 randomised, 17 were allocated to usual care control (UCC) and 17 were allocated to the VESFA treatment.

Five participants (15%) withdrew from the VESFA trial. One participant withdrew from the intervention arm (6%) after baseline (T1) but before intervention began due to other commitments. Four participants (24%) withdrew from the usual care control arm. Three withdrew before baseline assessment (T1), one cited low mood and the others gave no reason. One withdrew after baseline assessment (T1), citing testing burden as the reason for withdrawal. The feasibility of trial retention was considered met if 70% were available at follow up. **85% of participants (29/34) were available for follow up (T3)**, meeting this feasibility criterion, see Figure 8.2.

Feasibility Criteria	Proportion % [CI]	Numbers
Proportion eligible of those identified	29% [.22-.38]	37/126
Proportion eligible of those screened	36% [.27-.46]	37/102
Proportion consented of those eligible	92% [.78-.98]	34/37
Rate of eligible/month	2.64 per month	37 in 14 months
Rate of consent/month	2.43 per month	34 in 14 months
Proportion of withdrawals		
- Before randomisation	0%	0
- After randomisation:	15%	5
o UUC	12%	4
o VESFA	3%	1
- Overall, of those consented	15% [.05-.31]	5/34

Table 8.2: Feasibility outcomes

b) Feasibility of delivering the intervention remotely in a virtual environment

The main indication of feasibility of delivering intervention in a virtual environment was for the rate of cancelled sessions to be no higher than NHS community clinics, where 24% of appointments are cancelled (NHS Benchmarking, 2019). There were 512 sessions held during the intervention phase (32 sessions for 16 participants**). Less than 6% (30 sessions) were cancelled,** meeting this feasibility criterion.

In terms of acceptability, the participant questionnaire included questions that explored satisfaction with the method of delivery. All participants were positive about receiving therapy online in a virtual world (Figure 8.3). The accessibility of an online treatment was valued by many participants, “For me they were good because of where I live” (ppt75). Additionally, the comfort of being in your own home was mentioned “definitely good, sit down at home – easy” (ppt65).

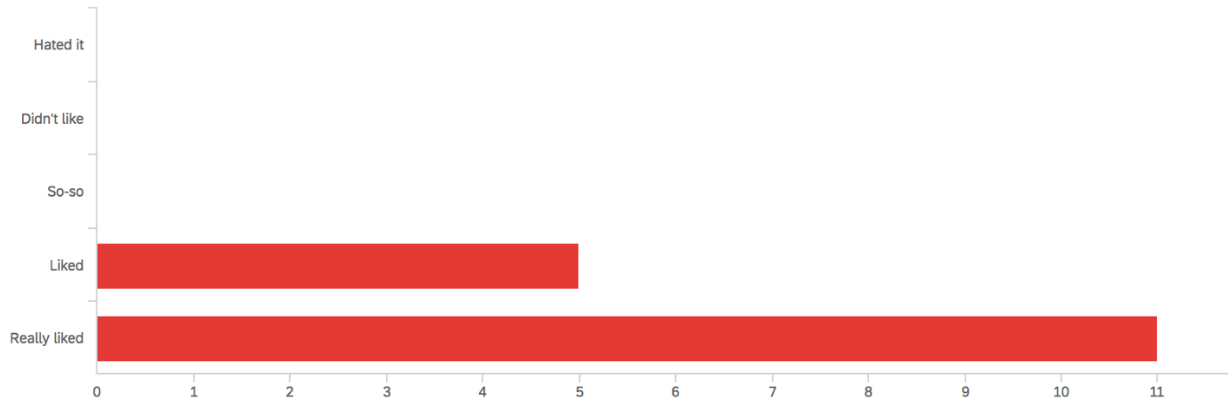

Figure 8.3: How did you find receiving therapy online in EVA Park?

Some participants needed help to get the software set up on their computer “Partner completed set up of EVA Park, as I am not very good at that type of thing, even pre-stroke” (ppt75). The number of steps were noted as a challenge “Frustrating - too many steps for set up” (ppt107). Others found the set up straight forward “excellent, good, easy to set up” (ppt42).

Once EVA Park was set up, logging in to join the treatment sessions was described as easy for all but one participant: “once you have user and password, every other session, you essentially open up laptop, get the application and log in” (ppt87).

Moving the avatar around the virtual island during sessions took some practice “Easy once I got used to it” (ppt88) and some functions were easier than others “walking was great. Flying a little bit tricky” (ppt21). Most participants (14/16) liked their avatars “I liked the hair colour and wings” (ppt53) and “Really liked it, big smile!” (carer of ppt107 describing ppt107’s response to his avatar).

For all participants being represented as an avatar was described as equalising “Same abilities as the other avatars - empowering and not limited!” (carer of ppt107). Participants enjoyed the physical opportunities presented by the avatar, such as dancing, “Liked being able to dance and do different things” (ppt75), and flying “quite liked it, could fly in EVA park but not in real life!” (ppt107).

Receiving treatment in a virtual world was described as novel "I thought the whole thing was amazing. It was so different from anything I had seen" (ppt75) and beautiful "It was very good, and I was happy to be there (Eva Park). It was beautiful to see." (ppt9). It was reported to be fun, "It felt like fun" (ppt98), and helpful "It wasn't only just fun it was helping us a lot" (ppt75).

Participants shared what supported them to take part in an online study as well as what the barriers were. The most cited support was the person at home with them, for example a sister (ppt60), wife (ppt9, ppt94), or partner (ppt7, ppt75), followed by the therapist "Not frustrating, SLT was patient and helpful" (ppt88) and the therapy handbook "the resource pack helped so much" (ppt94). The most discussed burden was the testing and the process of setting up EVA Park on the computer "setting up was difficult, a lot of options on the screen" (ppt107).

c) Acceptability of research procedures

The acceptability of the research procedures was evaluated through a post-intervention questionnaire, dropout rates and rates of missing data. The primary criterion for testing procedures to be considered acceptable was less than 15% missing data per scale/outcome measure, Table 8.3. **Rates of missing data ranged from 4.3% for the SAQOL-39g to 11.83% for the Scenario Test, with an average across all measures of 7.29%.** This met the testing procedure feasibility criterion.

Measure	Total	Missing Data Points			Percentage
	Test items x participants (31) x timepoints (3)	Not tested	Missing record form	Total	
VESFA Naming Test Items = 2 (treated and untreated totals)	186	10	4	14	7.53%
Discourse Items = 3	279	12	3	15	5.38%

Boston Naming Test Items=3 (total score)	279	2	6	8	2.87%
WAB-R Items =10 subtest totals	930	40	11	51	5.48%
SAQOL-39g Items = 39	3,627	156	0	156	4.3%
The Scenario Test, UK Items = 18	1,674	90	237	327	7.86%
GHQ-12 Items = 12	1,116	60	72	132	11.83%
				Mean percentage:	7.29%

Table 8.3: Percentage of missing data by outcome measure

The research procedures comprised three 2-hour testing sessions on Zoom at baseline (T1), 9 weeks (T2) and follow up at 18 weeks (T3). For those in the VESFA group, the treatment ran in the 8 weeks between T1-T2. Participants had four treatment sessions a week, typically 1 session a day from Monday to Thursday, two of which were 1hr of individual therapy and two of which were 1.5hrs group therapy.

Acceptability of the research procedures was explored in the post therapy questionnaire and the usual care questionnaire. Participants in the VESFA arm were interviewed after the intervention block (n=16). They were asked about 1) the testing protocol, 2) use of Zoom for testing 3) the treatment protocol and 4) the use of EVA Park for treatment delivery. For some participants their partner/carer was present, and they also expressed their opinions on the study. Participants in the UCC arm were interviewed after T3. They were asked about 1) the testing protocol, 2) use of Zoom for testing and 3) what usual care they received during the study. They were interviewed by the doctoral researcher (ND). They had received the testing from student volunteers or project students so had contact with this researcher only at screening. This section will address the acceptability of these procedures.

Acceptability of the testing protocol

Most participants (14/16) described the testing sessions as 'challenging', with one carer describing them as "long and difficult" (carer of ppt107). Some participants suggested breaking up the sessions, but others preferred to continue in order to finish sooner, "was offered breaks throughout, but just wanted to get through the number of tests" (ppt053). Despite this, overall participants (11/16) liked/really liked the testing sessions. They acknowledged the purpose of the tests "they have to do it, see progress" (ppt88) and the benefit of testing "useful though, challenging but good" (ppt107), "you learn about issues you have and performance you have" (ppt87), and only one participant in each arm rated the testing sessions negatively, see Figure 8.4.

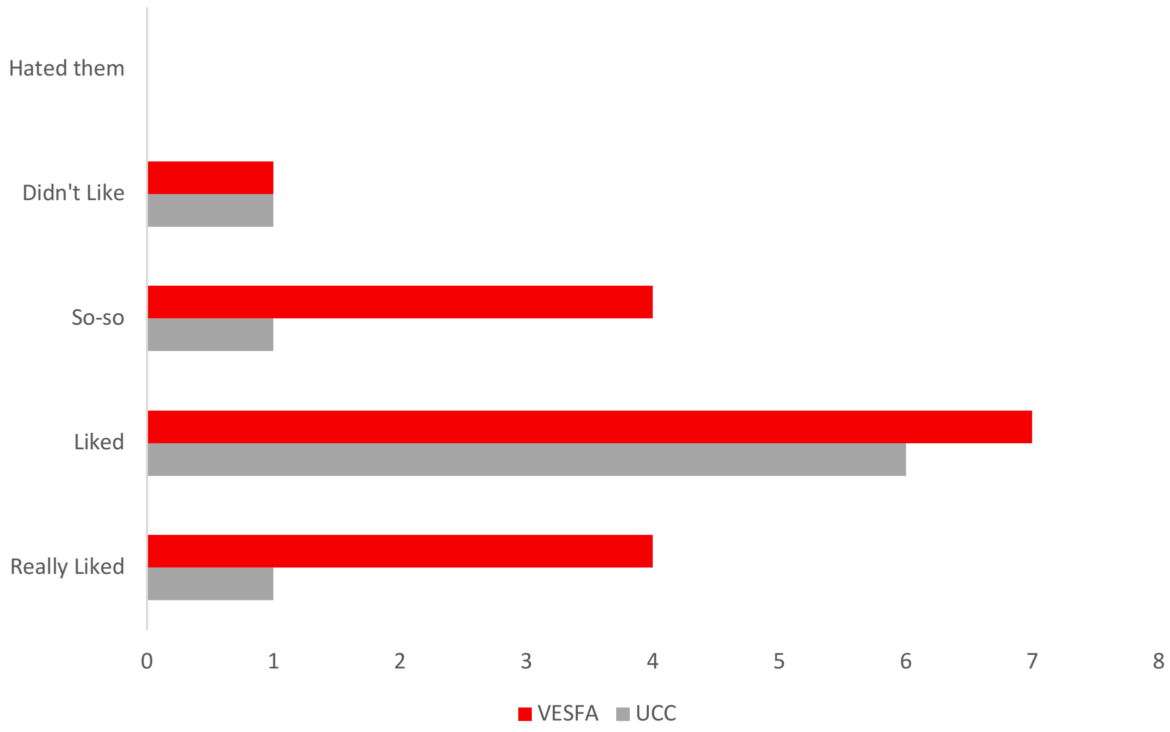

Figure 8.4: How did you find the testing sessions?

Acceptability of using Zoom for assessments

Just over half participants could use Zoom independently (56%), and just less than half required support from someone in their home, Figure 8.5. The passwords for Zoom were reported to be difficult "Passcodes. Quite tricky" (ppt107). For others Zoom had become part of their routine during lockdown "Had been on Zoom calls for last 18 months" (ppt87).

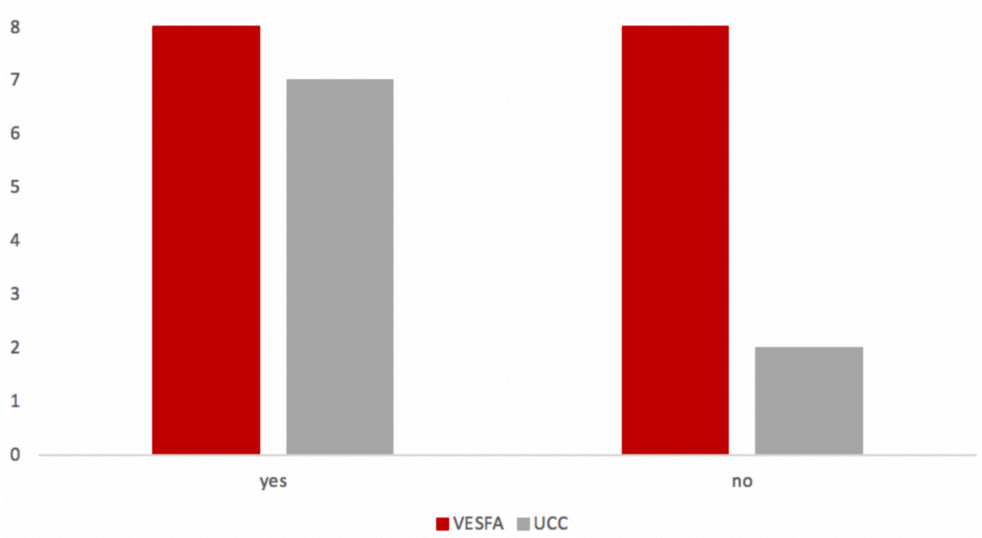

Figure 8.5: Did you use Zoom without help?

Acceptability of the treatment protocol

All VESFA participants found the number of sessions per week acceptable, with no-one reporting that they were dissatisfied and 15/16 satisfied or very satisfied, Figure 8.6. Participants spoke about enjoying the intensity, “I like to work hard and having lots of session per week” (ppt.42), and that the protocol had the “right amount” (ppt.07) of sessions. Even when the intensity was tiring, it was thought to be beneficial: “I felt tired after it. Intense. It seems like a lot but worthwhile - felt like getting something out of it” (ppt.98). One carer suggested that less sessions per week would be better, “3 sessions would have been better. Manageable, but would have preferred less” (carer of ppt107). One participant wanted the intervention to be longer than 8 weeks “I wished it went on longer and longer” (ppt15).

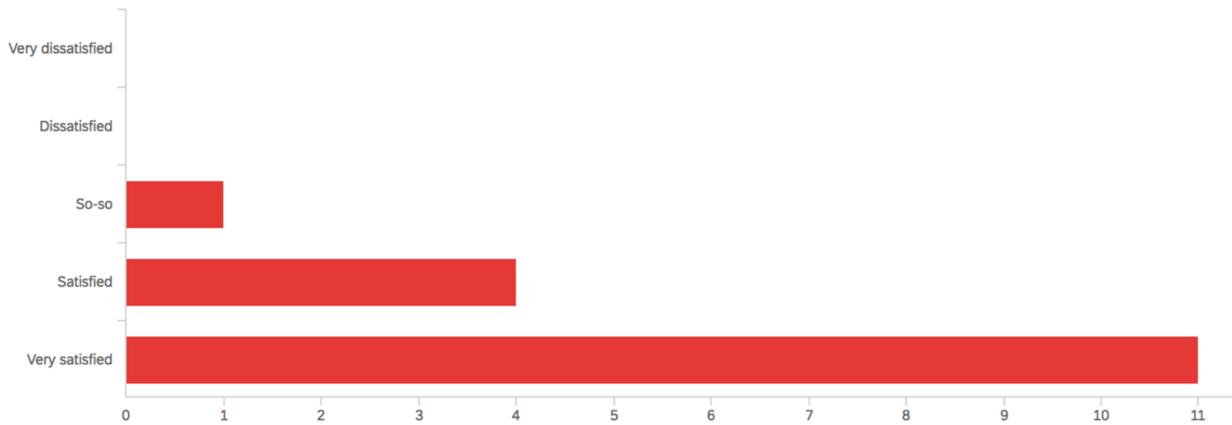

Figure 8.6: Responses to: How did you find the number of sessions per week?

d) Acceptability of treatment to participants

The main evaluation of treatment acceptability was based on the rates of adherence to the intervention where participants were considered to have adhered if they received at least 80% of intervention (32 of the 40 hours). Of the 16 participants who began the intervention all completed the 8 weeks, and **all received more than 82% of the sessions** (range: 33-40hrs), meeting this feasibility criterion.

The post therapy questionnaire data added detail on participants' views on the acceptability of the intervention. All participants were positive about the intervention overall, Figure 8.7.

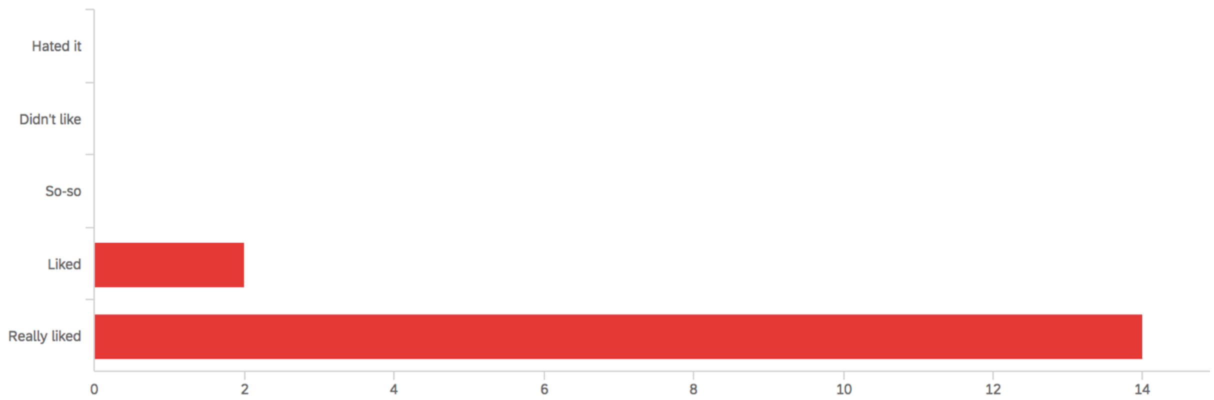

Figure 8.7: Responses to the question: How did you find the treatment overall?

The therapy was delivered as 16 individual sessions and 16 group sessions. This format provided the opportunity to work on language individually with the therapist before bringing it into the group; and this was seen as supportive "Better to do it by yourself and then have others" (ppt98).

The participants reported that they liked working on words and sentences in the individual sessions (Figure 8.8), "liked being in that work" (ppt07). Many people commented that the work was hard but good "It was good. It was hard work. It was helpful" (ppt115), "some was harder, but I like that some of it was harder" (ppt42).

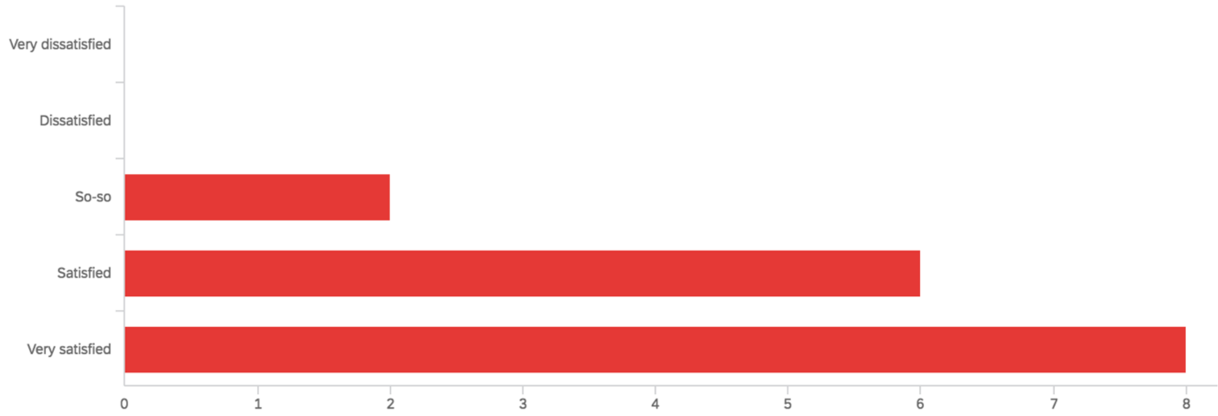

Figure 8.8: How did you find the 1:1 sessions at the boards?

Some found the group sessions more challenging than the individual sessions "The group was more challenging" (ppt15) and "struggle group" (ppt7). Nevertheless, all participants reported that they liked the groups (Figure 8.9). Two of the six groups had two instead of three participants and this was seen as less acceptable: "Supposed to be three people but ended up being two. Compared yourself with the other person in the small group" (ppt88). Many participants highlighted the social connections made in the groups "Really liked the groups, other people" (ppt107), "enjoyed it. Liked speaking with others" (ppt60) "Good, amazing people helped me. Excellent" (ppt94).

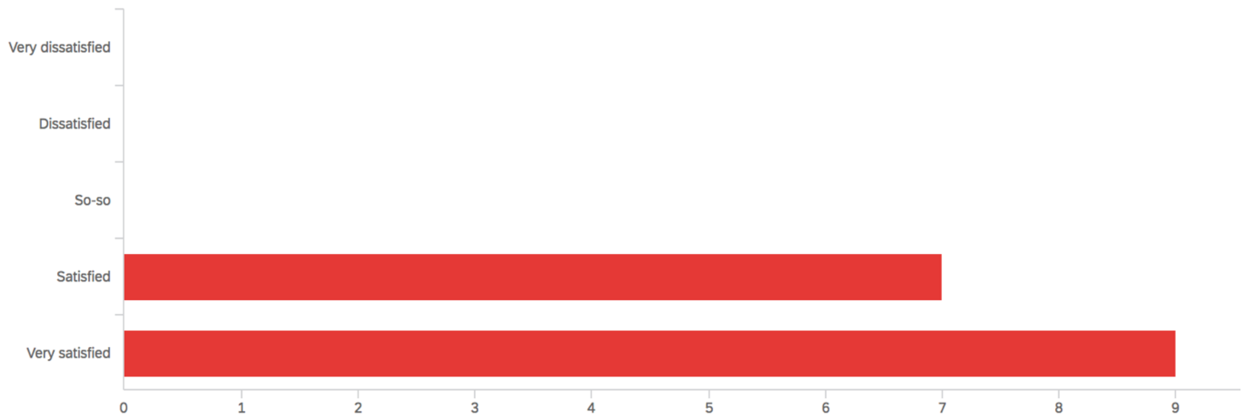

Figure 8.9: Responses to: How did you find the groups?

When asked about whether the treatment was an appropriate challenge most participants (14/16) reported that the treatment was 'at the right level', Figure 8.10. Participants welcomed the challenge, “It was helpful because I had to make the effort to do it” (ppt75), and perceived the challenge as useful, “All of it was challenging. But it got better and better. It helped” (ppt115).

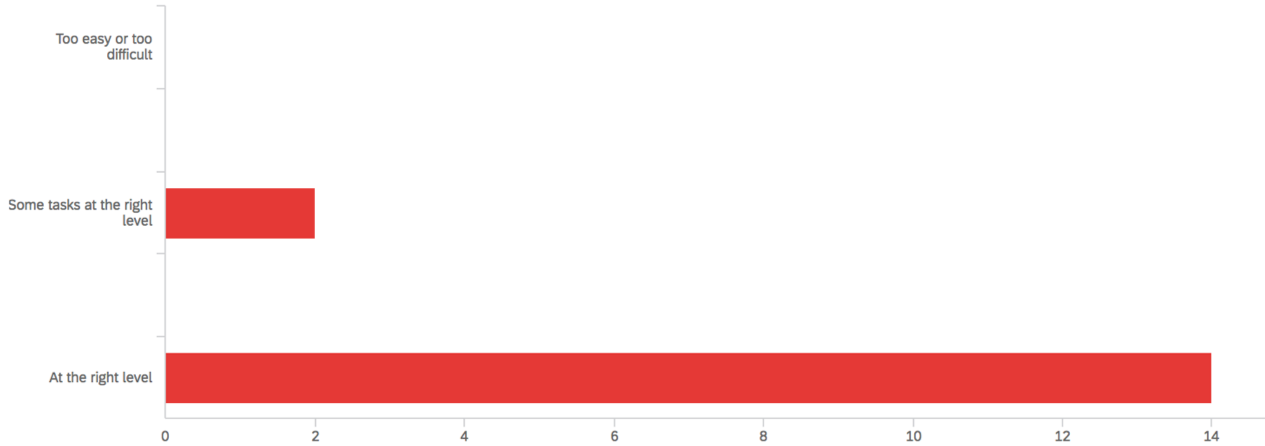

Figure 8.10: Was the therapy a good challenge?

Participants shared what they particularly liked and didn’t like. When asked what they didn’t like, 6/16 participants had things they didn’t like, and all raised individual aspects. One didn’t like logging in (this participant required help every time), one wanted a particular hat for their avatar, and one found navigating the avatar hard at first “Biggest thing: moving avatar at the start” (ppt87), one didn’t like the background noise in the Dr Who Tardis, and two participants commented on elements of the group. One reported he didn’t like silences in the group and the other said he struggled in the group. When asked what they did like 16/16 participants had things to say. The responses spoke about improvement, enjoyment of the work and the social connections. Many people talked about noticing improvement “I like that I improved” (ppt53), “I think I got better” (ppt9) and enjoying the individual sessions “One to ones. Liked being in that work” (ppt7), “The one on one were very good” (ppt15). Others liked the group sessions “Enjoyed it, liked speaking with others” (ppt60) and referenced the social connections made “Made friends in EVA Park, good to be together” (ppt65). Additionally, EVA Park featured as something people liked with comments about things on the island, “Animals, bars, restaurants, different boat sizes” (ppt21), the avatar “I liked the hair colour and wings” and environment “The trees and the plants” (ppt115).

e) Appropriateness of the outcome measures

The appropriateness of outcome measures was considered in terms missing data; floor or ceiling effects; whether the measure matched changes described in the post-therapy questionnaire; and the participants perspective on acceptability. All measures had less than 15% missing data. Participants were asked if any of the tests were problematic, see Figure 8.11. No test was considered problematic by more than two participants. Two tests were considered problematic by two participants: the Boston Naming Test and the narratives for the discourse score (Nicholas and Brookshire, 1993).

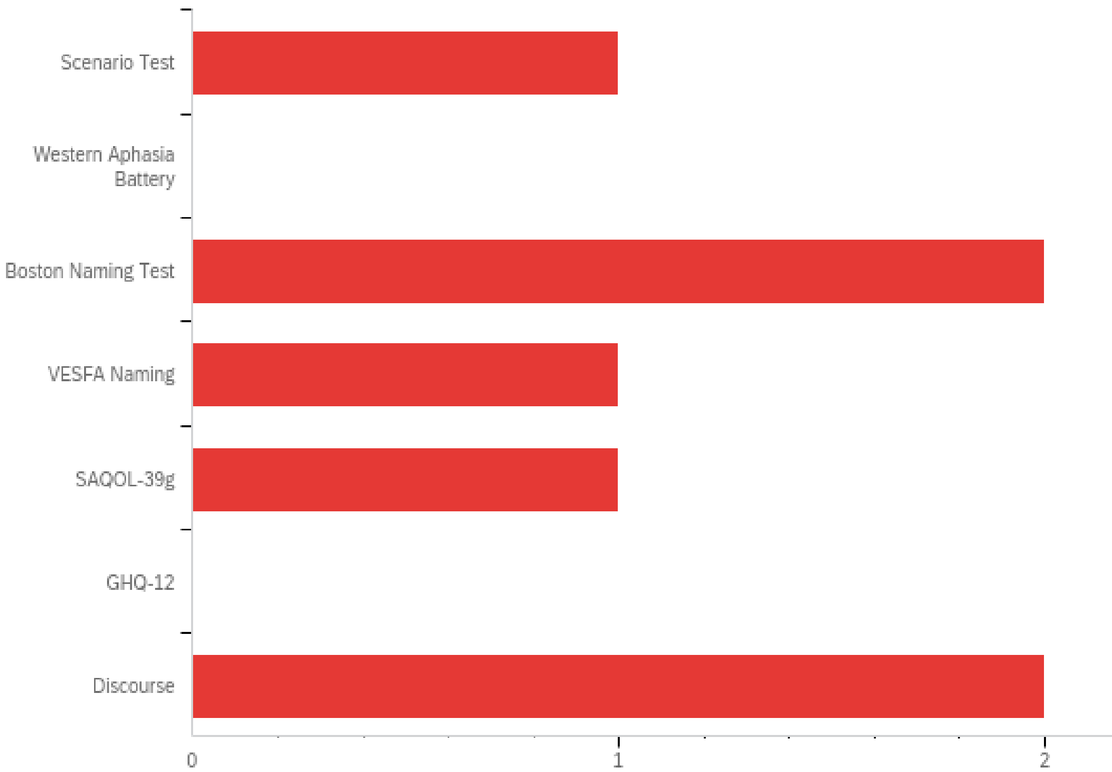

Figure 8.11: Were any of the tests problematic? If yes, which ones?

Safety

Five adverse events were reported during the feasibility trial (Table 8.4). Two transient ischaemic attacks (TIA) and a seizure occurred in participants within the treatment arm. One seizure and a marked deterioration in mood occurred in the UCC arm. These events represent known consequences within a post-stroke population (Camilo & Goldstein, 2004; Elkind, 2009) and were unrelated to the trial.

Adverse Event	Proportion %	Number
Transient Ischaemic Attack (TIA)	6%	2/34
Seizure	6%	2/34
Deterioration in mood (measured by the GHQ-12)	3%	1/34

Table 8.4: Adverse Events

Summary

The VESFA trial pre-specified feasibility criteria and met them all. Post-trial questionnaires with all participants explored acceptability of the trial protocol. The 2 hour testing session was experienced as a burden, but the purpose of testing was understood and only two participants didn't like the testing. A post therapy questionnaire explored the acceptability of the intervention and views were positive.

Chapter 9 | Results: Clinical Outcomes

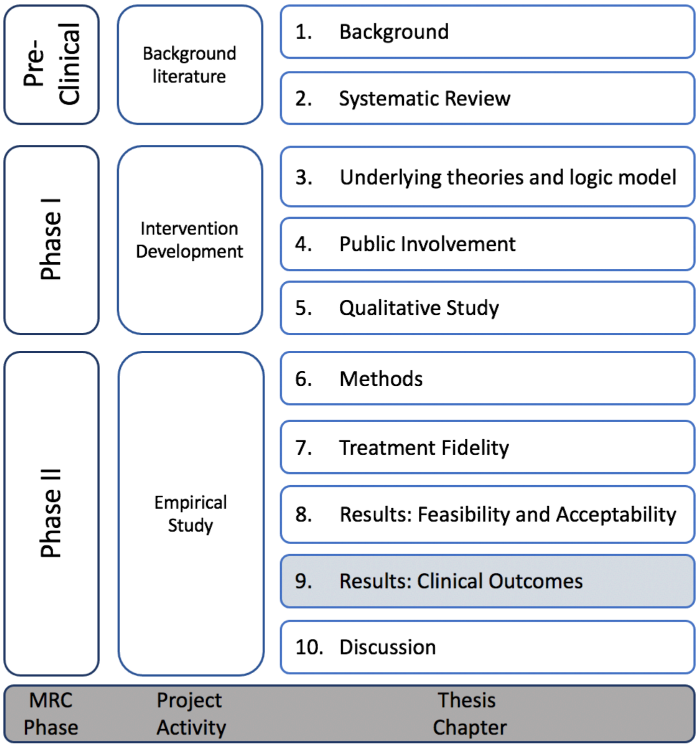

Figure 9.1: The thesis in a figure

This chapter reports the results from the clinical outcome measures. Clinical outcome measures provided preliminary evidence of the potential of the intervention to be effective in terms of:

- Naming treated words as measured by the study specific naming task (VESFA naming test)
- Naming words that were not targeted in the ESFA treatment as measured by the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983)
- Retrieving words within discourse, as measured by the Nicholas & Brookshire discourse analysis protocol (Nicholas & Brookshire, 1993)
- Functional communication, as measured by The Scenario Test, UK (Hilari, et al., 2018)
- Reducing emotional distress, as measured by the General Health Questionnaire, 12 items (GHQ-12) (Goldberg et al., 1997)

- Improving health related quality of life, as measured by the Stroke and Aphasia Quality of Life Scale, 39 items for a generic stroke population (SAQOL-39g) (Hilari, et al., 2009)
- Improving aphasia language profile as measured by the Western Aphasia Battery, Aphasia Quotient (WAB AQ) (Kertesz, 2007)

This chapter will also report themes that arose from the post therapy questionnaire that explored participants perceptions of change in language, everyday communication, mood and technology. These were probed using a questionnaire with multiple choice Likert scale responses and open discussion (Appendix 20).

Clinical Outcome Measures

Table 9.1[4] shows the means (SD) (and *medians (IQR)* for skewed data) of clinical outcome measures across all time points. There are some differences seen at baseline (pre-treatment) between the two groups. For example, for the VESFA naming test, the BNT, the GHQ-12 and the WAB AQ scores the treated group have better scores at baseline despite random allocation. All means for the VESFA group move in the expected direction. That is, scores increase post treatment or decrease in the case of the GHQ-12, where lower scores indicate better mood. In the UCC group, scores dropped over time for the VESFA Naming Test and words per minute (WPM), remained relatively stable for the BNT, the GHQ-12 and the SAQOL-39g, but increased over time for the percentage of Correct Information Units (%CIUs), CIU's per minute, the Scenario Test and the WAB AQ.

Measure [possible range]	**Baseline**			**Week 9**			**Week 18**		
	Mean (SD), *Median (IQR)*								
	VESFA n=17	**UCC** n=14	**All** n=31	**VESFA** n=15	**UCC** n=13	**All** n=28	**VESFA** n=16	**UCC** n=12	**All** n=28
VESFA Naming [0-240]	138.65 (72.98)	122.86 (67.47)	127.00 (70.59)	155.4 (71.37) *164 (148-208)*	113.85 (64.47)	136.00 (70.24)	161.67 (69.05) *169 (139-224)*	116.67 (72.81)	138.96 (73.05)
Boston Naming Test [0-60]	22.88 (14.42)	n=13 19.38 (14.5)	21.37 (14.31)	25.73 (15.39)	19.08 (13.04)	22.64 (14.48)	26.60 (15.76)	19.75 (16.65)	23.56 (16.22)

4 In all tables in this chapter results are given to 2 decimal places with .5 always rounded up.

Measure [possible range]	**Baseline**			**Week 9**			**Week 18**		
	Mean (SD), *Median (IQR)*								
	VESFA n=17	**UCC** n=14	**All** n=31	**VESFA** n=15	**UCC** n=13	**All** n=28	**VESFA** n=16	**UCC** n=12	**All** n=28
Discourse Measures:									
WPM	40.96 (36.04)	50.13 (36.65)	45.1 (36.4)	46.04 (35.80)	48.26 (40.28)	47.07 (37.24)	46.59 (32.90)	47.24 (41.34) *29.16 (19.00-76.40)*	46.88 (36.23) *30.43 (19-78.15)*
Percentage of CIU	49.82 (26.21)	50.77 (28.36)	50.25 (26.74)	58.71 (25.69) *64.50 (50.98-78.30)*	52.43 (23.51) *57.33 (33.25-6825)*	55.80 (26)	60.06 (25.36) *61.85 (44.82-77.28)*	59.3 (27.79) *61.6 (34.93-81.95)*	59.75 (26)
CIU per Minute	25.01 (23.48)	27.32 (26.12)	26.05 (24.31)	31.14 (26.80)	35.79 (29.41)	33.3 (27.61)	32.17 (31.08) *19.07 (7.07-51.58)*	32.13 (34.52) *15.00 (9.23-59.10)*	32.15 (32.07) *16.4 (7.41-52.03)*
The Scenario Test [0-54]	42.00 (10.31)	43.43 (8.91) *45.50 (38.75-49.50)*	42.65 (9.57) *44 (39-49)*	44.33 (9.74) *47.00 (40.00-51.00)*	44.23 (6.04)	44.29 *(8.1) 46.5 (39.25-50.75)*	43.81 (10.69) *46.50 (40.00-51.75)*	46.58 (6.75) *48.50 (42.25-52.00)*	45.00 (9.17) 46.5 *(42.25-52)*
Stroke and Aphasia Quality of Life Scale -39 items [0-5]	3.52 (.70)	3.35 (.81)	3.44 (.74)	3.91 (.43)	3.48 (.98)	3.7 (.75) *3.96 (3.32-4.18)*	3.85 (.52)	3.45 (.93)	3.67 (.74)
General Health Questionnare-12 [0-12]	2.18 (2.99) *1 (0-4)*	2.79 (2.19)	2.45 (2.6) *1 (1-4)*	1.2 (1.97) *0 (0-2)*	1.69 (2.29) *1 (0-3)*	1.43 (2.1) *0 (0-2)*	.5 (.82) 0 (0-1)	2.08 (3.29) *0 (0-3.75)*	1.18 (2.33) *0 (0-1)*
Western Aphasia Battery Aphasia Quotient	65.88 (17.4) *67.70 (61.20-77.3*	61.89 (14.55)	64.08 (16.04)	67.72 (18.62)	63.43 (13.58)	65.73 (16.32)	65.67 (17.19)	64.23 (16.90)	65.02 (16.53)

Table 9.1: Participant outcomes across timepoints

Table 9.2 shows the comparisons between the adjusted means of the two groups at T2 (post treatment) and T3 (follow up). The means are adjusted for the baseline differences seen in Table 9.1. Again, trends towards improvement post treatment can be seen. The adjusted means for the VESFA group are better than the UCC group for all measures, with the exception of the CIU per minute and GHQ-12, where the two groups are similar post-treatment but the VESFA group scores better at T3; and the Scenario Test and the WAB AQ where the differences are minimal.

Measure	**VESFA T2** Adjusted Means *(95% CI)*	**UCC T2** Adjusted Means *(95% CI)*	**VESFA T3** Adjusted Means *(95% CI)*	**UCC T3** Adjusted Means *(95% CI)*
VESFA Naming Totals	146.52 (133.17-159.87)	126.56 *(111.89-141.23)*	149.57 (137.56-161.58)	129.98 *(116.62-143.35)*
Boston Naming Test	25.02 *(22.16-27.78)*	20.13 *(17.04-23.22)*	24.56 *(21.52-27.59)*	21.40 *(18.02-24.78)*
Discourse Measures				
WPM	49.55 *(36.33-62.76)*	43.04 *(28.57-57.51)*	51.74 *(39.46-64.01)*	42.02 *(28.42-55.62)*
CIU	59.04 *(47.84-70.25)*	52.12 *(40.02-64.22)*	60.69 *(49.73-71.64)*	59.00 *(46.85-71.15)*
CIU per Min	32.46 *(24.15-40.76)*	33.51 *(24.46-42.57)*	34.64 *(25.02-44.26)*	29.86 *(19.19-40.52)*
The Scenario Test	44.41 *(40.73-48.09)*	43.60 *(39.62-47.59)*	44.60 *(41.46-47.75)*	45.71 *(42.12-49.29)*
SAQOL-39g	3.89 *(3.62-4.17)*	3.49 *(3.19-3.79)*	3.86 *(3.64-4.08)*	3.46 *(3.22-3.70)*
GHQ-12	1.45 *(.51-2.39)*	1.51 *(.49-2.53)*	.56 *(-.40-1.51)*	1.86 *(.76-2.96)*
WAB AQ	65.93 *(61.56-70.29)*	65.13 *(60.41-69.85)*	64.25 *(60.88-67.61)*	65.93 *(62.20-69.66)*

Table 9.2: Adjusted means table

These trends were explored using multilevel linear modelling. Table 9.3 illustrates the differences between the groups post treatment (T2 and T3) while controlling for baseline differences. Therefore, the group column is of most interest. Table 9.3 also gives us the time effects (T2 to T3) and interaction between group and time (T2 to T3) effects. However, we do not expect much change between post treatment and follow up. Therefore, a treatment effect in this table would be demonstrated by a significant group effect and no time or interaction effect.

Out of the seven measures taken, three showed a significant difference between the groups, see *p* values in bold in Table 9.3. The naming tests, i.e., the VESFA Naming Test [F(1, 25.82) = 5.11, *p*=.03) and the Boston Naming Test [F(1, 24.49) = 4.97, *p*=.04] and the quality of life measure (SAQOL-39g) scores [F(1, 25.85) =5.56, *p*=.03] were significantly higher for the VESFA group. Although the emotional distress scores, as measured by the GHQ-12, improved for the VESFA group between T2 and T3 but not for the UCC group, this difference was not significant [F(1, 26.54) = 2.76, *p*=.12]. All other differences between the groups were not significant.

Measure	**Group**			**Time**			**Interaction**		
	Stat	df	p	Stat	df	p	Stat	df	p
VESFA Naming Totals	F=5.11	1, 25.82	**.03**	F=1.11	1,24.29	.30	F=.00	1, 24.29	.952
Boston Naming Test	F=4.97	1, 24.49	**.04**	F=.13	1,23.98	.72	F=.59	1,23.94	.448
Discourse Measures									
WPM	F=.87	1,26.13	.36	F=.03	1,26.22	.86	F=.25	1,26.22	.625
CIU	F=.42	1,25.83	.52	F=.90	1,26.46	.35	F=.34	1,26.46	.565
CIU per min	F=.10	1,26.08	.75	F=.07	1,26.85	.79	F=1.11	1,26.13	.301
The Scenario Test	F=.01	1,25.88	.95	F=.84	1,25.91	.37	F=.58	1, 25.91	.453
SAQOL-39g	F=5.56	1,25.85	**.03**	F.32	1,26.00	.57	F=.00	1,26.00	.98
GHQ-12	F=1.36	1,26.87	.25	F=.53	1,26.52	.47	F=2.76	1,26.54	.12
WAB AQ	F=.04	1,25,02	.85	F=.08	1,25.81	.78	F=.66	1,25.81	.42

WPM=words per minute, CIU=content information units, SAQOL=Stroke and Aphasia Quality Of Life, GHQ=General Health Questionnaire, WAB AQ=Western Aphasia Battery, Aphasia Quotient

Table 9.3: Effects from week 9 (post treatment) to week 18 (follow up), controlled for baseline differences

The charts show adjusted mean differences between the groups at post treatment (T2) and follow up (T3) with error bars indicating the confidence intervals for outcome measures that reached statistical significance post treatment, Figure 9.2, and those that didn't reach significance post treatment, Figure 9.3.

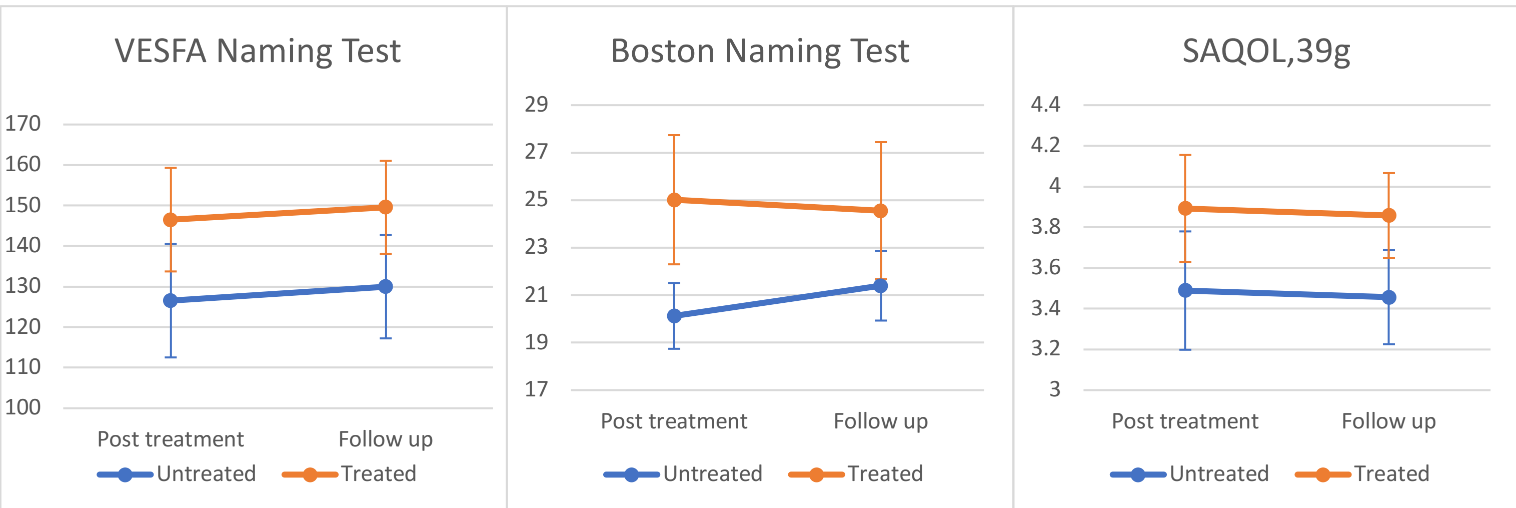

Figure 9.2: Chart of adjusted means and confidence intervals at post treatment and follow up for measures that reached a statistically significance difference between the groups at post treatment.

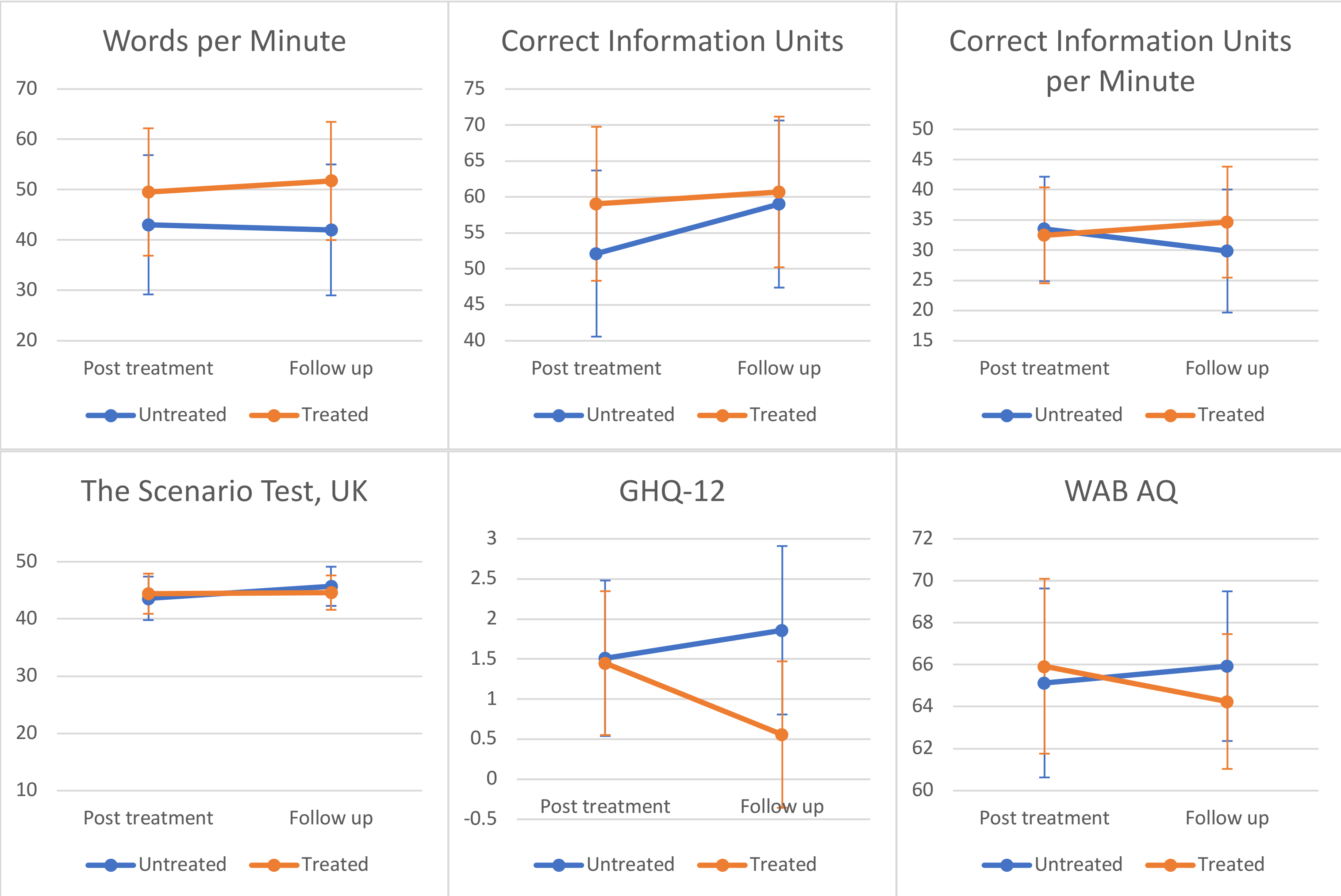

Figure 9.3. Charts of adjusted means and confidence intervals at post treatment and follow up for measures that did not reach a significant difference between the groups post treatment

Many of the error bars overlap, demonstrating an overlap in scores between the groups. The exception is the Boston Naming Test where groups are distinct post treatment.

Qualitative Experiences of Clinical Outcomes

The 16 participants who received the intervention completed a questionnaire after their treatment. Questions were asked with Likert scales and scale responses are reported in the charts below. Additionally, free comments responses were qualitatively analysed. A strong theme of change was seen in the data. Change was reported as improvements in communication, confidence, mood, technology use, and social networks.

Communication Changes

Experiences of a positive change in communication post treatment was well represented across the sample. Many participants spoke of change in word finding “Lots of words. Can say more words” (ppt60) and “increased range of words” (ppt53). Some reported an improvement in the use of sentences “Making very nice sentences now. Verbs have a little bit of progress” (ppt21) and other reported better conversations “They (family) noticed I was chatty, chatty, chatty” (ppt15).

Three participants reported that despite improved word finding, “I was getting words correct - 100% at the end of therapy” (ppt42) and others noticing changes to their talking “They have noticed me talking more”(ppt42) and “I went to see my doctor and he said your speaking has got better” (ppt75), they did not experience an overall change in their talking themselves “My talking has been similar”(ppt42) and “I've still got problems with myself and they tried as they could helping me” (ppt09). One participant commented that this is because the improvement does not represent her pre-stroke abilities “I think my speech has improved but I am never happy with it as it is not the same as pre-stroke” (ppt75).

Mood Changes

Participants were asked if they noticed any change to their mood (Figure 9.3). No one reported a change for the worse. About a third of the participants noticed no change and two thirds agreed or strongly agreed that they noticed a change in their mood. One commented that “it made my mood great” (ppt60) and others explained that talking more drove the change in mood “Talking more feels great” (ppt65) and “I am a little bit better and happy because I am out there talking to people when first I wasn't, a little bit more” (ppt09).

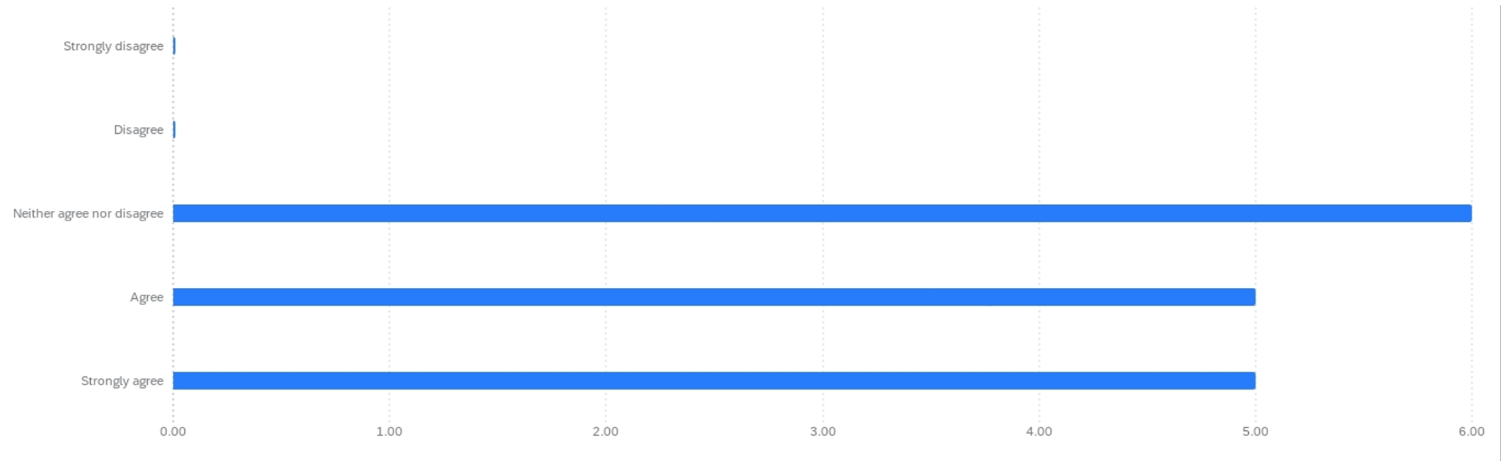

Figure 9.4: Have you noticed any mood changes?

Real World Changes

Participants were asked if they noticed a change in real world communication following this virtual intervention (Figure 9.4). Most people agreed or strongly agreed that they noticed a change in real world communication. Different examples were shared in the interviews. One participant described how he now used the phone to speak to his mum “Mum change telephone. Covid, ring, me, call up, chat. Me hardly phoned but chat, nice” (pp07). Another participant described how practicing situated conversations in the EVA Park restaurant supported the real world task: “a meal, me in a restaurant, sentences” (ppt21). A participant’s wife described the change she noticed in their conversations: “we can have better conversations now. We can lay in bed and discuss the news on television which we never could before” (wife of ppt09).

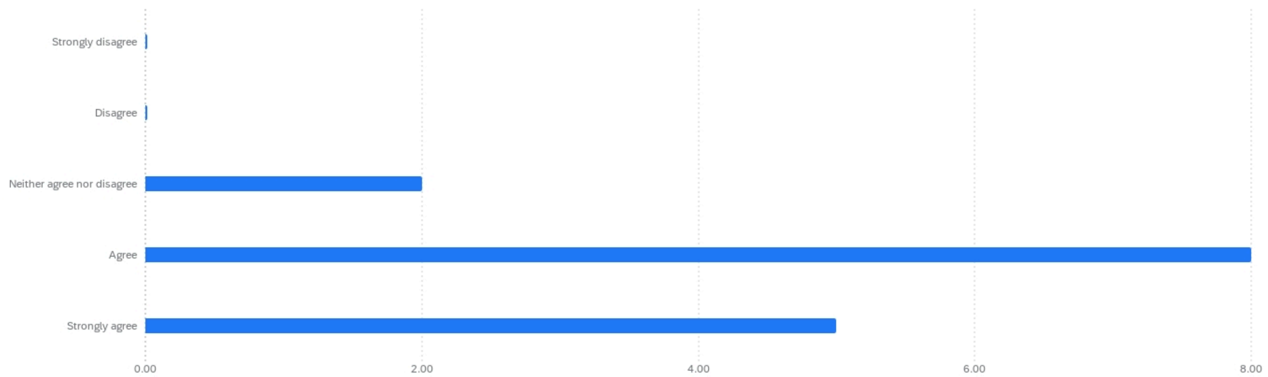

Figure 9.5: Responses to ‘I noticed a change to my talking in the real world

Technology Use Changes

Two participants did not see a change in the way they used technology and seven neither agreed nor disagreed on whether they noticed a change (Figure 9.5). Some talked about improvements in their use of EVA Park "I can now fly my avatar" (ppt42) and seven reported they noticed changes in general technology use "Better computer, email, internet banking" (ppt60), "learnt how to use a mac" (ppt87).

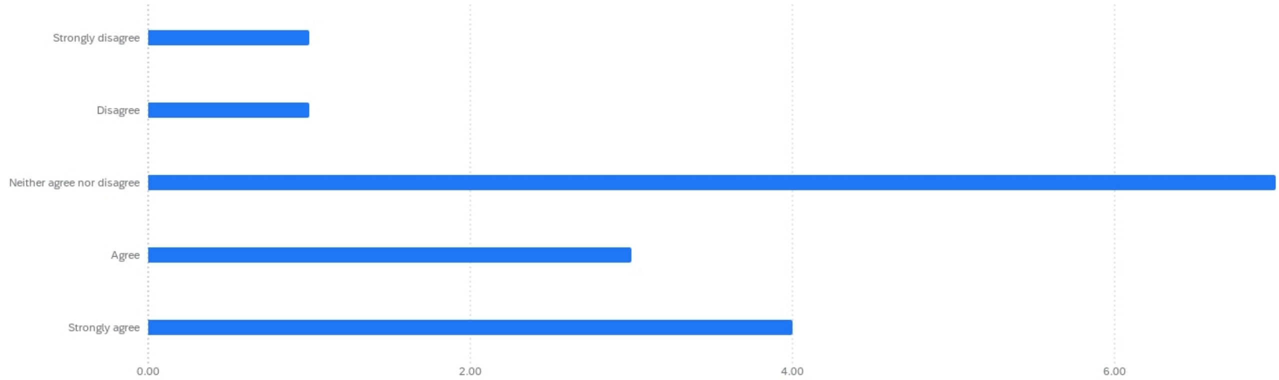

Figure 9.6: Responses to 'I noticed a change in how I use technology'

Confidence changes

Four participants reported more confidence in having a go at communicating: "Trying to talk more and get message across, more confidence" (ppt88).

Social network changes

Participants valued the social connections of the group "Made friends in EVA Park, good to be together" (ppt65) and suggested that came from meeting someone with the same lived experience "If someone else has similar problems it really helped him "(wife of ppt94). Some participants continued to meet as a group on zoom once the intervention ended "Since we came to the end of the 8 weeks I have agreed to meet up with the 2 people online once a week" (ppt15).

Reflections on the treatment

Many participants referred to the treatment as a welcome challenge. Comments included the idea that the intervention was "good difficult" (ppt07) or "nice but hard" (ppt65), "work hard, fantastic" (ppt94). One participant expanded on this to say, "it was helpful because I had to make the effort to do it" (ppt75), suggesting effort is required to make changes. The participants also reflected on how they felt at the end of the VESFA treatment "looking back at the eight weeks I think we did good, good, good" (ppt15) and "after it I was very happy with it" (ppt09). Others shared they "feel better, stronger" (ppt 94) and valued the chance to learn new skills: "It was fantastic for me to learn more myself" (ppt113).

The advisory group's interpretation of the results

The feasibility, acceptability and clinical outcomes were presented to the trial advisory group of PWA in October 2022. The positive change in quality of life was thought to be the most important outcome for people with aphasia. One member reported that 80% of stroke survivors are depressed as justification for the importance of this outcome. Changes in words, confidence and mood were also discussed as important. One member described why he believed word finding is easier in EVA Park "word finding is easier as well, I think…maybe it's because you are in a … the hairdressers or the garden. Because you are visualising garden, it brings back memories, things, words related to gardening." Sharing a message (as measured by the Scenario Test, UK) was thought to be an important outcome from therapy, group members acknowledged that this was not found in this study.

The discussion about the main message to give to PWA from this study was "Do it!" and for funders "absolutely continue with this project". This was expanded to list the benefits "Word finding, conversation, confidence, interesting and the fact that you're actually gaining… helping yourself. It's a God send". There was a discussion about when this might be beneficial. The consensus was that this type of intervention would work best after face-to-face NHS treatment; "SLT first, then move onto EVA Park".

The advisory group suggested the next steps should be to roll out EVA Park. They were not positive about more trials "I don't think a bigger trial – wheel it out". Although this topic was not invited, there was a discussion about paying for private speech therapy and paying for apps. The consensus was that it would be acceptable to have a small charge for an evidence-based programme.

Summary

Preliminary outcomes showed significant differences between the groups in treated words (the VESFA Naming Test), untreated words (the Boston Naming Test) and quality of life (SAQOL-39g) post therapy. The participants reported an increase in words used, better sentences and conversations. Positive impacts to mood, confidence and social connections were highlighted. For some, improvements in the use of technology were also noted.
The trial advisory group proposed the most meaningful outcome for PWA was the positive change in quality of life as a result of the VESFA intervention.

Chapter 10 | Discussion

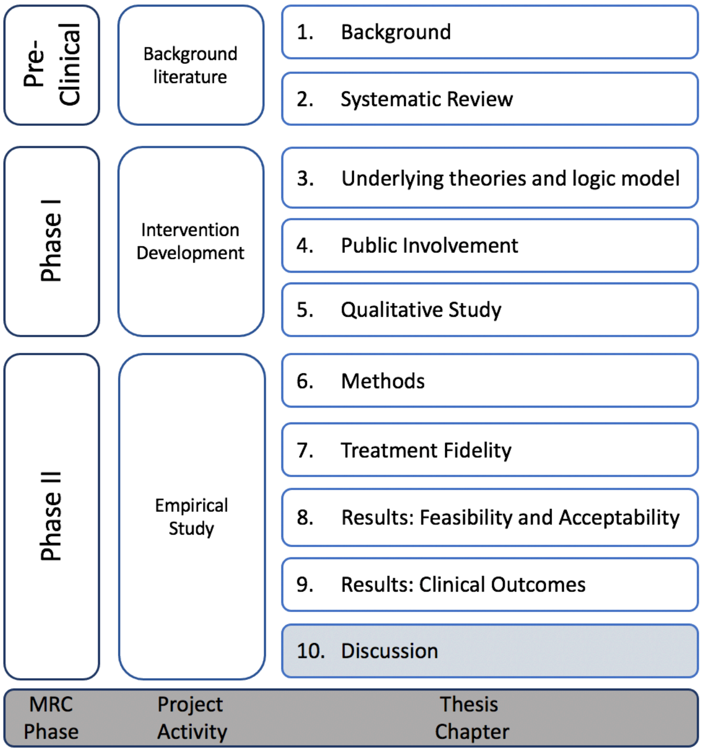

Figure 10.1: The thesis in a figure

The VESFA trial investigated the feasibility and acceptability of ESFA and conversation groups delivered in the virtual world, EVA Park. The trial developed an intervention through mapping out the research evidence, articulating the underlying theories and a programme model, Public Involvement workshops with key user groups, and a qualitative study to investigate meaningful therapy targets. The resulting therapy was therefore a user informed, evidence and theory based intervention. The fidelity of intervention delivery was tested, with good adherence to the protocol. The VESFA intervention was tested in a feasibility randomised controlled trial comparing VESFA + usual care with a usual care control. People with word finding difficulties as a result of chronic aphasia were recruited from the community. Feasibility parameters included the percentage of eligible participants who consented to the trial, the percentage of participants available at follow up, rates of cancelled sessions, rates of missing data, the acceptability of the research procedures to participants and the acceptability of the intervention to the VESFA arm. The pre-specified feasibility criteria were met, and participants' views on the assessment and intervention were positive. Clinical outcomes provided preliminary findings of the impact on language, communication and quality of life.

Recruitment to the VESFA trial was good. The percentage of those who were eligible who consented (91%) and the rate of participants recruited per month was higher than other remotely delivered aphasia word finding or group interventions, see Table 10.1. Rates were similar to the recruitment to a face-to-face ESFA study (Efstratiadou et al., 2019), where 80% of eligible participants consented. The percentage of those screened who meet eligibility criteria is less often reported. When it is, the percentage of participants with aphasia who meet study criteria is low. In TeleGAIN 15% of those screened met criteria (Pitt et al., 2019) and, when recruiting for remote naming intervention, Woolf et al (2016) reported that under 50% of those screened were eligible. In the VESFA trial 36% of participants screened were eligible. A systematic review of stroke randomised controlled trials explored the recruitment and drop out in trials (McGill et al., 2020). A synthesis of 512 studies (n= 28,804) found an average randomisation rate of 34%, and average recruitment rate of 1.5 participants per month and an average dropout rate of 6%. This shows that the VESFA trial was fairly average on randomisation and dropout rates, at 36% and 5% respectively. However, the VESFA recruitment rate of 2.4 participants a month is high for stroke trials. It is possible that the novelty of an EVA Park study was a draw. In a previous EVA Park study a participant reported that being involved in a virtual word study was exciting. Additionally, recruitment occurred during the pandemic lockdown when 81% patients reported they had less therapy and 62% of patients received no therapy (Clegg, O'Flynn & Just, 2021). This trial offered an opportunity for additional therapy in a context of reduced provision.

The recruitment rate indicates that recruitment to aphasia trials requires high numbers of participants referred and screened. If a definitive VESFA trial aimed to have 100 participants, over 300 referrals would need to be screened to achieve this number. Widening the recruitment sites with support from networks such as the National Institute of Health and Care Research’s (NIHR) Clinical Research Network would be recommended.

Retention to the VESFA trial was measured in the percentage of those available at follow up. Understandably, time commitment to the study affects study retention with shorter studies reporting higher retention (Table 10.1). The VESFA trial had 85% of participants available for an 18 week follow up. The VESFA trial had a usual care control arm, where half of the participants received testing but no treatment. A usual care arm can negatively affect retention. In the VESFA trial 4/5 withdrawals were from the usual care arm. Overall, the retention rates in the VESFA trial were similar to another aphasia intervention study with a no treatment control, Big Cactus.

	VESFA	**Cactus** Palmer et al., 2012		**TeleGAIN** Pitt et al., 2019	**Remote Naming** Woolf et al., 2016	**Face-to-Face ESFA** Efstratiadou et al., 2019	
Recruitment:							
Those eligible who consented:	91%	34%		10%	46.6%	80%	
Recruitment period:	14m	22m		6m	17m	37m	
Rate of recruits per month:	2.43	1.4		1.5	1.2	0.6	
Retention:							
Available at follow up:	85%	86%	79%	100%	95%	98%	17%
Time commitment to the study:	18wks	26wks	52wks	12wks	14wks	19wks	45wks
Remote delivery:							
Cancelled sessions:	6%	n/a		0 **	n/r	n/a	
Attendance:	94%	31%*		100%	100%	n/r	
Acceptability of Intervention:	16/16 liked (n=2) or really liked (n=14)	n/r		Satisfied? 5/5 replied yes or definitely yes	19/20 rated 'good' 1/20 'neutral'	n/r	

* Self-administered; average of 28hrs practice reported out of 90hrs practice recommended. ** No full sessions cancelled; technical issues occurred but were resolved. n/a=not applicable, n/r=not reported

Table 10.1: VESFA feasibility results and comparable aphasia intervention studies

Previous research has established good compliance with the remote delivery of aphasia therapy (Cacciante et al., 2021; Hall, Boisvert & Steele, 2013). The low rates of cancelled sessions in the VESFA trial support this finding; 6% of sessions were cancelled and 94% of all remote sessions were attended. Attendance speaks to the acceptability of an intervention. Therapist led remote sessions have better attendance than self-directed practice (Table

10.1). The VESFA trial had comparable attendance rates to other therapist led remote interventions (Pitt et al., 2019; Woolf et al., 2016).

Interviews with VESFA participants found that the testing, treatment protocol and treatment content were acceptable to participants. Acceptability is a ‘necessary but not sufficient’ condition of a successful intervention (p.1, Sekhon, Cartwright & Francis, 2017). Acceptable interventions are more likely to be adhered to, and therefore produce better outcomes. Acceptability is multifaceted. Whether an intervention is acceptable depends on how a person feels about the intervention (affective attitude), the perceived effort required to participate (burden), the ‘fit’ with a personal value system (ethicality), the extent to which the person understands the intervention and how it works (intervention coherence), the extent to which values must be given up to engage in the intervention (opportunity cost), the extent to which the intervention is perceived as effective (perceived effectiveness) and the persons confidence that they have the ability to participate in the intervention (self-efficacy) (Sekhon, Cartwright & Francis, 2017). In the VESFA trial, all participants reported that they liked the intervention (affective attitude). Many people commented that it was hard work but that they valued that; ‘work hard, fantastic’ (burden). Some spoke about the therapy leading to improvements in speech ‘The articulate game, the way to describe, search for words felt improved speech’ (perceived effectiveness). In summary, the participants’ attitude towards the intervention was positive, they acknowledged a burden, but the perceived effectiveness appeared to outweigh the burden (opportunity cost). Future post-therapy interviews could explicitly address the domains of acceptability to include ethicality and intervention coherence by adjusting the TFA Acceptability Questionnaire (Sekhon, Cartwright & Francis, 2022) to be accessible to people with aphasia.

The pre-trial consultation with a small sample of SLTs also suggested that the intervention was potentially acceptable to service providers, see Chapter 4. However, the acceptability to providers has not been tested. Strong levels of enjoyment were also reported by service providers in a previous EVA Park study (Caute et al., 2021). The acceptability of delivering VESFA could be included in a future trial.

Previous EVA Park delivered interventions have also been acceptable to participants (Amaya et al., 2018; Galliers et al., 2017; Marshall et al., 2018). A study that explored the views of participants who received a conversation intervention in EVA Park (Marshall et al., 2016) found overwhelmingly positive views of the EVA Park intervention (Amaya et al., 2018). A study observing participants using EVA Park highlighted incidences of pleasure/fun and playfulness/making jokes as the most frequently observed behaviours (Galliers et al., 2017). The role of playfulness in coping with stress in adults has gained research interest (Clifford et al., 2022; Farley, Kennedy-Behr & Brown, 2021; Magnuson & Barnett, 2013; Proyer, 2013). Playfulness in adults predicts well-being (Farley, Kennedy-Behr, & Brown, 2021) and serves as an adaptive function against stress (Magnuson & Barnett, 2013). In a population where rates of depression are high (Code, Hemsley & Herrmann, 1999) interventions that address wellbeing and elicit play are warranted. In the VESFA trial, fidelity checklists captured whether enjoyment was experienced in the groups. This could be marked as 'present' if there were more than three demonstrations of laughter/fun. Enjoyment was seen in all videos rated (Figure 7.5).

Despite being represented as avatars and never meeting face-to-face, the positive relationships that are built in EVA Park interventions are frequently highlighted in acceptability interviews (Amaya et al., 2018; Caute et al., 2021; Marshall et al., 2018). Likewise, in the VESFA post-therapy interviews the value of the other group members and the therapist was highlighted: 'all people involved. The way the therapist followed and helped bring us all together. Everyone worked together'.

This study was not designed to provide definitive clinical outcomes. That being said, the clinical outcomes in a feasibility trial can indicate which elements of the treatment show promise and can help predict the outcomes of a larger trial. The clinical outcomes in the VESFA trial suggested that a definitive trial is likely to improve naming of treated words, with generalisation to untreated words and improve health related quality of life (HRQOL).

SFA consistently finds improvements in treated words with smaller changes in untreated items (Boyle, 2010; Efstratiadou et al., 2018; Wisenburn & Mahoney, 2009). The VESFA outcomes suggest that a future definitive trial will have a similar pattern of gains with the additional gain in HRQL. The VESFA treatment group improved by an average of 25 words post treatment (164 - 139 = 25) and 17 words at follow up on treated items (156 - 139 = 17), as measured by the VESFA Naming Test (Table 9.1). Scores for untreated words improved by an average of 2.85 post therapy and 3.72 at follow up, as measure by the Boston Naming Test (BNT). A review of group effect sizes indicates that a meaningful change is greater than 3.3 points on the BNT (Gilmore, Dwyer & Kiran, 2019). This was achieved in the VESFA trial at follow up testing. The minimal clinically important difference for HRQL, as measured on the SAQOL-39, is an average group change of .21 (Breitenstein et al., 2022). This was achieved in VESFA at post therapy (3.91-3.52=.39) and follow up (3.85-3.53=.33).

Given the sequential difficulty of the BNT (Pedraza et al., 2011), some items of the BNT are likely to be less familiar and less imageable than the VESFA Naming Test. This could be argued to be a conservative approach to measuring stimulus generalisation. The most difficult items on the BNT are abacus (item 60), compass (item 50), yoke (item 56) (Pedraza et al., 2011). Difficulty was established though item response theory (Graves et al. 2004). As an example, yoke has an imageability rating of 514, as given by the MRC psycholinguistic database (Coltheart, 1981). The lowest imageability rating of an item on the VESFA Naming Test was 532 (tin/can), followed by 550 (rake). In essence, the independent measure of naming tested the participants on words more difficult than those that were treated, and still the treated group showed significantly more words post therapy and at follow up than participants the UCC arm.

The use of words in connected speech was measured using the Nicholas and Brookshire system for quantifying the informativeness (measured content words) and efficiency of speech (percentage of content words) (Nicholas & Brookshire, 1993). These measures did not reach statistical significance, but all measures showed an advantage to the treated group post treatment when baseline differences were controlled for (Figure 9.3). In an underpowered study this finding is encouraging.

The hypothesis that situating the conversation groups in simulated real world environments might drive change in functional communication was not confirmed in the clinical outcomes. Either the assessments used did not pick up the changes made, or the treatment did not induce functional change, potentially because the dose did meet the required threshold. In terms of measures, this study measured functional communication using the Scenario Test UK (Hilari et al., 2018) as this was recently added to the agreed measures in the Core Outcome Set (COS) for aphasia research (Wallace et al., 2021). Scores on the Scenario Test stayed relatively stable over time in both groups, see Figure 9.3. Scores were also high. Given that the Scenario Test was developed for people with severe aphasia (van der Meulen et al., 2010) we explored whether this represented a ceiling effect. A ceiling effect is defined as more than 15% of participants scoring the maximum score (Terwee et al., 2007). In the VESFA trial participants the maximum score was achieved by 10%, 7% and 11% of participants at baseline, post intervention and follow up respectively, so this was not a ceiling effect. Previous EVA Park studies (Carragher et al., 2018; Marshall et al., 2016; Marshall et al., 2018; Marshall et al., 2020) have used the CADL-2 (Holland et al., 1999) to measure functional change. The CADL-2 was validated on 175 people with neurological communication disorders from age 20-96: 131 following a stroke and 20 following traumatic brain injury (Hughes & Orange, 2007). It is possible this measure is a better match for the VESFA population of predominantly mild/moderate aphasia. A value of the Scenario Test is the assessment of conveying a message regardless of modality. VESFA therapy addresses spoken word production, in isolation and conversation. Furthermore, most communication in EVA Park is via speech. For example, there is no opportunity to employ natural gestures. Therefore, gains from VESFA therapy are most likely to occur in the spoken modality. The Scenario Test would not register a difference in the modality of the response. Future trials of VESFA should use a functional communication measure that is sensitive to improved word retrieval abilities.

The Scenario Test was chosen because it was the measure of communication in the Core Outcome Set for aphasia research (Wallace et al. 2022) and had been identified in a review as the measure that most captured situated language (Doedens & Meteyard, 2020). The ROMA COS, and other COS for specific health conditions, were developed to allow data sets across many research projects to be combined (Wallace et al. 2019). This is particularly relevant in a field like aphasia where sample sizes are typically small. Adding and adding small sets of data can allow for a much larger sample size over time and therefore more power. Trends that are not visible in a single study may be seen when the data from many studies are combined. Indeed, the Collaboration of Aphasia Trialists asks studies to share anonymised data with the CATs data set for this reason (Collaboration of Aphasia Trialists). The ROMA COS was developed using the guidelines from the COMET initiative (Prinsen et al., 2016). That is, in consultation with key stakeholder groups and including synthesis of all the literature. In aphasia the COS is a work in progress in a young field. For example, initially there was no consensus on a test for functional communication and the Scenario Test UK was only added in 2022 (Wallace et al., 2022). As a result, there are limited studies that have used the Scenario Test UK. A COS has benefit to the field, beyond the individual study but there are also drawbacks. The measures in the COS are not always the most appropriate measures for the study, the measures are not always accessible and adding measures adds to testing burden (Wallace et al., 2021). A recent implementation study using the Theoretical Domains Framework made recommendations to improve the uptake of the COS in aphasia research (Wallace Sarah et al., 2021). External incentives, encouragement from colleagues and monitoring systems were identified as areas to target. For example, as an external incentive funders, journal editors and trial registries could require COS to be used where they exist.

It has been argued that functional communication may need a higher dose than discrete cognitive linguistic tasks due to its cognitive and linguistic complexity (Harvey et al., 2022). This view was based on findings from a meta-analysis that a dose of 60 hours achieved better functional outcomes than 30 hours. However, data leading to this conclusion were drawn from only two studies. Additionally, 48 hours of ILAT showed no additional gains over 24 hours of ILAT (Harvey et al., 2022). The most recent findings on dose reviewed the individual participant data of 959 participants. This synthesis suggested functional communication needs a total dose of 20-50hrs of treatment (Harvey et al., 2022). In the VESFA intervention the group conversation therapy accounted for 24hours of the total dose. This is at the low end of this range. The dose hours were, however, comparable to Marshall et al. (2016) conversation therapy where participants received 20 hours of goal directed conversations and made gains in functional communication. Dose-response relationships are under researched in aphasia, so guidance remains inconclusive.

Whether change occurred in the functional communication of the participants can also be explored through participant views. Self-reported gains in post therapy interviews suggested the treatment made an impact on participants experiences of everyday communication. This is in line with previous EVA Park research where qualitative interviews revealed a perceived improvement to communication, with participants talking more and finding it easier to talk (Amaya et al. 2018). Similarly, comments in VESFA referred to talking more and better conversations. One carer valued that the participant 'could be part of a social exchange based on language and ideas, still frustrating, but those beginnings are hard to come by'. Interview data can raise concerns about over-compliance on the part of respondents (McClendon, 1991). For example, there may a wish to reward the researchers by expressing positive views. It is therefore reassuring that the VESFA participants were willing to make negative comments (testing burden) and identify areas in which there had been no change. This suggests that using interviewers unconnected to the study may have supported participants to provide frank accounts of their experience.

Although the difference between the mood scores for the different groups did not reach significance, Figure 9.3 shows GHQ-12 scores registered a small improvement for the VESFA group, against a marginal decline for the UCC group. In GHQ-12, a high score indicates high psychological distress, which may be indicative of depression (>3/12) (Hilari, 2011). This may suggest that being involved in the VESFA intervention can have a positive effect on the mood of participants. This finding was confirmed by participant reports of improved mood "Talking more feels great" (ppt65). The fact that this effect did not reach significance may be a true finding or may be because the study was underpowered.

The suggestion that the VESFA intervention can improve HRQL alongside language outcomes is an encouraging finding. The known benefits of play on wellbeing (Clifford et al., 2022; Farley, Kennedy-Behr & Brown, 2021; Proyer, 2013) may partly explain this finding, given the fantastical and whimsical features of EVA Park. Additionally, it is likely that sharing personal stories and hearing the stories of others has also influenced HRQL. The conversation groups shared life experiences e.g., where participants have travelled, the restaurants participants have eaten in, the value of nature, what has changed since the stroke. There is growing evidence for the need to tell the story of the stroke and renegotiate a sense of self to support living well with aphasia (Alawafi, Rosewilliam & Soundy, 2021; Corsten et al., 2015; D'Cruz, Douglas & Serry, 2020; Strong et al., 2018). A recent review confirmed that gains can be made in quality of life through tele-practice (Gauch et al., 2022).

Bringing the results to the trial advisory group gave the trial team an interpretation of the results from the perspective of those living with aphasia. The VESFA advisory group felt the change to quality of life was the most meaningful result. They also thought change to conversation was important but acknowledged that this was not indicated by the Scenario Test. Literature that addresses the outcomes that matter to people with aphasia highlight communication activity/participation as most meaningful with a need to address all areas of the ICF (Wallace et al. 2017). People with aphasia have also addressed what it is to live well with aphasia (Manning et al. 2019). Ongoing, flexible services in the chronic stage post stroke were reported to enable continued engagement with recovery. The James Lind Alliance Priority Sharing Partnership identifies the outcomes that matter to stroke survivors (https://www.jla.nihr.ac.uk/). Priorities are established through patient carer and clinicians working together equally to set priorities for research. The first priority in the top 10 priorities for rehabilitation and long-term care for people with stroke is interventions that support emotional wellbeing (James Lind Alliance, 2021). Interventions for communication difficulties is the third priority. This aligns with the views of the advisory group and the indicative outcomes of the VESFA feasibly trial.

An obvious question is the contribution of EVA Park to the success of this study. As discussed above, it seemed to add novelty and enjoyment for participants, so contributing to the positive acceptability and compliance data. Perhaps more intriguingly, the EVA Park environment may also have played a role in the observed naming gains. The semantic system is conceived of as a widely distributed sensory-motor system (Varley, 2011)**.** A number of recent advances in the neurology of semantics propose modality specific and supra-modal elements to our semantic knowledge (see the hub and spokes model of semantic processing in Chapter 3). The 'hub' is thought to be in bilateral anterior temporal regions (Ralph et al., 2017) and the 'spokes' (shape, colour, sound, word etc) are distributed across the cortex (Rogers & Lambon Ralph, 2022). The spokes include emotional, sensory and cognitive systems, sometimes described as embodied cognition (Meteyard et al., 2012). This suggests that rich experiences are what will create strong semantic representations, experiences that include emotional, physical and sensory stimuli. Imagine holding a fish in your hand. You see the glint of the scales, you feel the slippery scales, its wriggle, you experience the joy of catching it - the 'fishness' of the experience is strong. EVA Park aims to simulate this richness of experience. Travel can be discussed whilst standing on the boards of a tall ship in the sunset and talk about gardening can take place whilst sitting in a greenhouse with a cup of tea and a bunch of radishes. The richness of experience could be evident in comments from participants that reported a sensory experience, "It was beautiful to see" (ppt09). We could speculate that such experiences played a role in activating semantic representations and hence facilitating naming. Giachero and colleagues cited the embodied theory of semantics as the underlying theory for the potency of VR environments (Giachero et al., 2020). They also addressed the question of the added value of VR by comparing a VR conversation therapy with face-to-face conversation therapy. They found no significant difference between the groups. However, a detailed look at improvements showed that the repetition score drove the change in the face-to-face group whereas the VR group had improved across three functions: repetition, written language and oral comprehension (Giachero et al., 2020). Future research could explore whether a semantic therapy in EVA Park achieves different gains to a comparable face-to-face semantic therapy.

A dominant theme in the post therapy interviews was that the treatment was a good challenge, exemplified by comments such as "good hard work" and "hard work, fantastic". These comments suggest that the aim to incorporate the challenge of genuine conversation in the group treatment was achieved (see Chapter 3). The framework for situated conversations highlights the linguistic and cognitive complexity of situated language (Doedens & Meteyard, 2022). This concept of a good challenge may also speak to self-efficacy (Gangwani et al., 2022; Lo et al., 2022; Szczepańska-Gieracha & Mazurek, 2020). The challenge was such that participants were aware of their self-efficacy. Comments suggest participants felt they could not only tolerate the challenge but achieve the challenge "Intense. It seems like a lot but worthwhile- felt like getting something out of it" (ppt98) and "All of it was challenging. But it got better and better. It helped." (ppt115).

Strengths

The VESFA intervention was a user informed and evidence and theory based intervention. The trial pre-specified the feasibility and acceptability criteria. Although pre-specified criteria are a design strength a review of 227 studies found only 45 (19.8%) feasibility or pilot studies pre-specified progression criteria (Mbuagbaw et al., 2019). The inclusion criteria for the study were clear. This ensures that the participants in the trial can access the intervention activities. Participants were randomly assigned to the treatment groups, minimising confounding variables and randomisation was concealed, reducing the chance of selection bias (Sil et al., 2019). The testers were blind to treatment group, with no reports of unblinding. 85% of participants were tested at follow up. All participants' outcome scores, including those that withdrew, could be included in the statistical analysis through the use of MLM. Trial reporting in this thesis was informed by relevant reporting guidelines (Chapter 2: PRISMA, Appendix 2. Chapter 3: GRIPP2 short form, Appendix 9. Chapter 5: SRQR Appendix 6, Chapter 6,8 and 9: CONSORT Appendix 14).

Limitations

Pre-specified treatment fidelity criteria would likely have improved the treatment fidelity scores. Some studies ask that treating therapist complete a checklist of core activities after each session (Behn et al., 2022). This would also improve fidelity scores and increase the potency of the intervention, as the core activities that drive change are present in every therapy session.

The VESFA trial had a large number of testers with diverse experience. The VESFA trial had 16 testers, 10 were volunteers and 6 were dissertation project students. They were all speech and language therapy students with varying experience delivering standard assessments. They received comparable training but nevertheless the size of the testing team may have potentially threatened the reliability of assessment scores. A smaller team of qualified testers could mitigate this risk. Monitoring the fidelity of the assessment procedures could be added.

A strength of the study was the pre-specified feasibility and acceptability outcomes. However, there were no pre-specified criteria for the perspectives of the intervention recipients gathered in the post therapy questionnaire. Future studies could pre-specify acceptability criteria through the use of the Theoretical Framework of Acceptability questionnaire (Sekhon, Cartwright & Francis, 2022).

The 2020 coronavirus pandemic moved the trial testing online. Testing was planned to be face-to-face and therapy via the online virtual world, EVA Park, but all contact with participants went online. Some assessments were developed for online delivery (www.cityaccess.org/tests/saqol) and some were adjusted by the research team to deliver them online e.g., creating a PowerPoint of the Scenario Test scenarios for sharing via screenshare with permission from the authors. As such, not all assessments used have been validated for online delivery.

On a PEDro scale (Moseley et al., 2015) this study would score 7/10 indicating that the quality of the trial is 'good'. Three items were not present: 1) the groups were not the same at baseline, 2) the participants and 3) treating therapist were not blind to treatment group. Randomisation should control for individual differences in the sample (Tucker-Drob, 2011), however with a small sample of 34 participants by chance the groups differed at baseline on key criteria e.g., word finding ability. One solution would be to minimise the differences at randomisation through stratified or minimised randomisation. Stratification was considered in this trial, i.e., to balance aphasia severity in the two groups. However, this is only possible where randomisation is prospective (Suresh, 2011). In the VESFA trial participants were recruited in sets of 6 participants at a time until 6 sets of the 8 weeks treatment had been completed. Difference at baseline is a known problem with small samples (Sella et al., 2021), which is common in feasibility trials.

Future Research

Future research could explore a more appropriate functional communication outcome measure, one that is validated on people with mild/moderate aphasia and will pick up change in verbal communication, and remove The Scenario Test, UK.

The optimal dose still needs to be explored. Inaccurate dosage wastes resources (Baker, 2012). A more intensive program was in the research proposal: 5 sessions per week for 6 weeks (total hours = 36). An intensive dose was proposed because there is some evidence for better outcomes with intensive treatment (REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators, 2022). However, the PI group suggested a more distributed regime would be more acceptable. The VESFA regime delivered was, therefore, 4 sessions per week for 8 weeks (total hours = 40). This revised dose still met the described parameters for the greatest gains (REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators, 2022) but additionally considered the acceptability to people living with aphasia. High intensity treatments are not always tolerated (Harvey et al., 2020). What the proposed regimes do not specify is the number of episodes of treatment related activity (Harvey et al., 2022). In ESFA this would be the retrieval of the target word and the generation of the phrase/sentence. The number of times a participant names a target is, to some extent, dependent on the severity of the aphasia i.e., someone with milder aphasia will name more targets than someone with moderate/severe aphasia (Evans et al., 2020). So, although dose could be prescribed in terms of treatment activities, this would be more time for someone with a more severe aphasia. Dose parameters are not yet well understood in aphasia treatment with very few studies comparing dose (Harvey et al., 2020).

A definitive trial could compare the VESFA intervention and usual care or compare VESFA with face-to-face delivery of ESFA plus conversation groups to explore the potency of the virtual situated conversations. A future trial could have a larger team of intervention providers, to see whether fidelity can be maintained when more than one person is delivering therapy. Acceptability of the intervention to providers could also be explored. A long term objective would be to include implementation research, particularly exploring whether the VESFA intervention can be delivered from routine clinical settings.

The EVA Park platform was built in 2013 so would benefit from an upgrade. This could include the ability to run EVA Park on a tablet. This could explore the use of technology that captures real face/lip movements and assigned them to your avatar, such as lip synch technology, or arm movement for realistic, real time gesture, as used in the RGS technology (Grechuta et al. 2016).

Conclusion

This doctoral thesis reports on the development, feasibility, acceptability and preliminary clinical outcomes of an intervention that addressed multiple domains of the ICF and quality of life using a multi-user virtual world. Virtual Elaborated Semantic Feature Analysis is a user informed evidence and theory based intervention. Although not powered to report on clinical outcomes preliminary results show that a definitive trial would be recommended to determine whether the VESFA intervention is likely to improve naming of treated words, generalise to untreated words, and impact everyday communication and health-related quality of life. It was found to be feasible against pre-specified criteria and acceptable to treatment recipients, “Fantastic, enjoyed it! Thought it was going to be crap! And no, no it was really good, so much help” (ppt94).

References:

Acocella, I. (2012). The focus groups in social research: Advantages and disadvantages. Quality & Quantity: International Journal of Methodology, 46, 1125-1136. 10.1007/s11135-011-9600-4

Alawafi, R., Rosewilliam, S., & Soundy, A. (2021). An integrative review considering the impact of storytelling and sharing interventions in stroke. Behavioral Sciences, 11(6)10.3390/BS11060088

Ali, M., Lyden, P., Brady, M., & VISTA Collaboration. (2015). Aphasia and Dysarthria in Acute Stroke: Recovery and Functional Outcome. International Journal of Stroke, 10(3), 400-406. 10.1111/ijs.12067

All-Party Parliamentary Group on Arts, Health and Wellbeing. (2017). Creative Health: The arts for health and wellbeing. Available at: https://www.culturehealthandwellbeing.org.uk/appg-inquiry/ (Accessed: 10.08.22)

Alt Murphy, M., Resteghini, C., Feys, P., Lamers, I., Institutionen för neurovetenskap och fysiologi, sektionen för klinisk neurovetenskap och rehabilitering, Sahlgrenska akademin, Göteborgs universitet, Gothenburg University, Sahlgrenska Academy, & Institute of Neuroscience and Physiology, Department of Clinical Neuroscience and Rehabilitation. (2015). An overview of systematic reviews on upper extremity outcome measures after stroke. BMC Neurology, 15(1), 29. 10.1186/s12883-015-0292-6

Amaya, A., Woolf, C., Devane, N., Galliers, J., Talbot, R., Wilson, S., & Marshall, J. (2018). Receiving aphasia intervention in a virtual environment: the participants' perspective. Aphasiology, 32(5), 538-558. 10.1080/02687038.2018.1431831

American Occupational Therapy Association. (2020). Guidelines for Systematic Reviews. Available at: https://ajot.submit2aota.org/journals/ajot/forms/systematic_reviews.pdf (Accessed: 14.01.22)

Aminov, A., Rogers, J. M., Middleton, S., Caeyenberghs, K., & Wilson, P. H. (2018). What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. Journal of Neuroengineering and Rehabilitation, 15(1), 29. 10.1186/s12984-018-0370-2

Antonucci, S. M. (2009). Use of semantic feature analysis in group aphasia treatment. Aphasiology, 23(7-8), 854-866. 10.1080/02687030802634405

Attard, M. C., Lanyon, L., Togher, L., & Rose, M. L. (2015). Consumer perspectives on community aphasia groups: a narrative literature review in the context of psychological well-being. Aphasiology, 29(8), 983-1019. 10.1080/02687038.2015.1016888

Bacon, K., Marshall, J., Caute, A., Monnelly, K., Cruice, M., Moutou, C., & Woolf, C. (2021). Treatment fidelity of technology-enhanced reading therapy (CommuniCATE) for people with aphasia. International Journal of Language & Communication Disorders, 56(6), 1114-1131. https://doi.org/10.1111/1460-6984.12637

Baker, E. (2012). Optimal intervention intensity. International Journal of Speech-Language Pathology, 14(5), 401-409. 10.3109/17549507.2012.700323

Balestroni, G., & Bertolotti, G. (2012). EuroQol-5D (EQ-5D): an instrument for measuring quality of life. [L'EuroQol-5D (EQ-5D): uno strumento per la misura della qualità della vita] Monaldi Archives for Chest Disease = Archivio Monaldi Per Le Malattie Del Torace, 78(3), 155-159. 10.4081/monaldi.2012.121

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentice-Hall, Inc.

Bandura, A. (1997). Self-efficacy: The exercise of control. W H Freeman/Times Books/ Henry Holt & Co.

Behn, N., Harrison, M., Brady, M. C., Breitenstein, C., Carragher, M., Fridriksson, J., Godecke, E., Hillis, A., Kelly, H., Palmer, R., Rose, M. L., Thomas, S., Tippett, D., Worrall, L., Becker, F., & Hilari, K. (2022). Developing, monitoring, and reporting of fidelity in aphasia trials: core recommendations from the collaboration of aphasia trialists (CATs) trials for aphasia panel. Aphasiology, , 1-23. 10.1080/02687038.2022.2037502

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., Czajkowski, S., & Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association, 23(5), 443-451. 2004-18051-001

Benaim, C., Cailly, B., Perennou, D., & Pelissier, J. (2004). Validation of the Aphasic Depression Rating Scale. Stroke, 35(7), 1692-1696. 10.1161/01.STR.0000130591.95710.20

Berg, K., Isaksen, J., Wallace, S. J., Cruice, M., Simmons-Mackie, N., & Worrall, L. (2020). Establishing consensus on a definition of aphasia: an e-Delphi study of international aphasia researchers. Aphasiology, , 1-16. 10.1080/02687038.2020.1852003

Borrelli, B. (2011). The Assessment, Monitoring, and Enhancement of Treatment Fidelity In Public Health Clinical Trials. Journal of Public Health Dentistry, 71(s1), S52-S63. 10.1111/j.1752-7325.2011.00233.x

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A., Long, A. F., Watkins, C., Wilkinson, M., Pearl, G., Ralph, M. A. L., & Tyrrell, P. (2012). Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial. BMJ : British Medical Journal, 345(7868), 16. 10.1136/bmj.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. 10.1044/1058-0360(2004/025)

Boyle, M. (2010). Semantic Feature Analysis Treatment for Aphasic Word Retrieval Impairments: What's in a Name? Topics in Stroke Rehabilitation, 17(6), 411-422. 10.1310/tsr1706-411

Brassel, S., Power, E., Campbell, A., Brunner, M., & Togher, L. (2021). Recommendations for the Design and Implementation of Virtual Reality for Acquired Brain Injury Rehabilitation: Systematic Review. Journal of Medical Internet Research, 23(7), e26344. 10.2196/26344

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. 10.1191/1478088706qp063oa

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological Test Administration by Videoconference: A Systematic Review and Meta-Analysis. Neuropsychology Review, 27(2), 174-186. 10.1007/s11065-017-9349-1

Breitenstein, C., Hilari, K., Menahemi-Falkov, M., L. Rose, M., Wallace, S. J., Brady, M. C., Hillis, A. E., Kiran, S., Szaflarski, J. P., Tippett, D. C., Visch-Brink, E., & Willmes, K. (2022). Operationalising treatment success in aphasia rehabilitation. Aphasiology, ahead-of-print(ahead-of-print), 1-40. 10.1080/02687038.2021.2016594

Breitenstein, C., Dr, Grewe, T., Prof, Flöel, A., Prof, Ziegler, W., Prof, Springer, L., PhD, Martus, P., Prof, Huber, W., Prof, Willmes, K., Prof, Ringelstein, E. B., Prof, Haeusler, K. G., MD, Abel, S., Prof, Glindemann, R., PhD, Domahs, F., Prof, Regenbrecht, F., MA, Schlenck, K., PhD, Thomas, M., MA, Obrig, H., Prof, de Langen, E., PhD, Rocker, R., MD, . . . FCET2EC study group. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. The Lancet (British Edition), 389(10078), 1528-1538. 10.1016/S0140-6736(17)30067-3

Brogan, E., Ciccone, N., & Godecke, E. (2019). Treatment fidelity in aphasia randomised controlled trials. Aphasiology, 33(7), 759-779. 10.1080/02687038.2019.1576442

Brown, K., Worrall, L., Davidson, B., & Howe, T. (2010). Snapshots of success: An insider perspective on living successfully with aphasia. Aphasiology, 24(10), 1267-1295. 10.1080/02687031003755429

Brumfitt, S. M., & Sheeran, P. (1999). The development and validation of the Visual Analogue Self-Esteem Scale (VASES). The British Journal of Clinical Psychology, 38(4), 387-400. 10.1348/014466599162980

Brummett, B. H., Barefoot, J. C., Siegler, I. C., Clapp-Channing, N. E., Lytle, B. L., Bosworth, H. B., Williams, J.,R B., & Mark, D. B. (2001). Characteristics of socially isolated patients with coronary artery disease who are at elevated risk for mortality. Psychosomatic Medicine, 63(2), 267-272. 10.1097/00006842-200103000-00010

Bryant, L., Brunner, M., & Hemsley, B. (2019). A review of virtual reality technologies in the field of communication disability: implications for practice and research. Disability and Rehabilitation. Assistive Technology, , 1-8. 10.1080/17483107.2018.1549276

Burns, K. (2006). Focus on solutions: a health professional's guide. John Wiley & Sons.

Cacciante, L., Kiper, P., Garzon, M., Baldan, F., Federico, S., Turolla, A., & Agostini, M. (2021). Telerehabilitation for people with aphasia: A systematic review and meta-analysis. Journal of Communication Disorders, 92, 106111. S0021-9924(21)00034-4

Cairns, P., Power, C., Barlet, M., Haynes, G., Kaufman, C., & Beeston, J. (2021). Enabled Players: The Value of Accessible Digital Games. Games and Culture, 16(2), 262-282. 10.1177/1555412019893877

Cairns, P., Cox, A., & Nordin, A. I. (2014). Immersion in Digital Games: Review of Gaming Experience Research. In M. C. Angelides, M. C. Angelides, H. Agius & H. Agius (Eds.), (1st ed., pp. 337-361). Wiley. 10.1002/9781118796443.ch12

Cambridge University Press. (2021). Cambridge Dictionary. Available at: https://dictionary.cambridge.org/dictionary/english/virtual-reality (Accessed: 13.4.21)

Camilo, and Goldstien. (2004) Seizures and Epilepsy After Ischemic Stroke. *Stroke, 35*(7), 1769-1775. 10.1161/01.STR.0000130989.17100.96

Camic, P. M., & Chatterjee, H. J. (2013). Museums and art galleries as partners for public health interventions. Perspectives in Public Health, 133(1), 66-71. 10.1177/1757913912468523

Cao, Y., Huang, X., Zhang, B., Kranz, G. S., Zhang, D., Li, X., & Chang, J. (2021). Effects of virtual reality in post-stroke aphasia: a systematic review and meta-analysis. Neurological Sciences : Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, 42(12), 5249-5259. 10.1007/s10072-021-05202-5

Carr, A., Cullen, K., Keeney, C., Canning, C., Mooney, O., Chinseallaigh, E., & O'Dowd, A. (2020). Effectiveness of positive psychology interventions: a systematic review and meta-analysis. The Journal of Positive Psychology, , 1-21. 10.1080/17439760.2020.1818807

Carragher, M., Steel, G., Talbot, R., Devane, N., Rose, M. L., & Marshall, J. (2020). Adapting therapy for a new world: storytelling therapy in EVA Park. Aphasiology, 10.1080/02687038.2020.1812249

Carragher, M., Sage, K., & Conroy, P. (2015). Preliminary analysis from a novel treatment targeting the exchange of new information within storytelling for people with nonfluent aphasia and their partners. Aphasiology, 29(11), 1383-1408. 10.1080/02687038.2014.988110

Carragher, M., Talbot, R., Devane, N., Rose, M., & Marshall, J. (2018). Delivering storytelling intervention in the virtual world of EVA Park. Aphasiology, 32(sup1), 37-39. 10.1080/02687038.2018.1484880

Carter, R. (1987). Is There a Core Vocabulary? Some Implications for Language Teaching. Applied Linguistics, 8(2), 178-193. 10.1093/applin/8.2.178

Caute, A., Cruice, M., Devane, N., Patel, A., Roper, A., Talbot, R., Wilson, S., & Marshall, J. (2021). Delivering group support for people with aphasia in a virtual world: experiences of service providers. Disability and Rehabilitation, , 1-19. 10.1080/09638288.2021.2011436

Chen, X., Liu, F., Lin, S., Yu, L., & Lin, R. (2022). Effects of Virtual Reality Rehabilitation Training on Cognitive Function and Activities of Daily Living of Patients With Poststroke Cognitive Impairment: A Systematic Review and Meta-Analysis. Archives of Physical Medicine and Rehabilitation, 103(7), 1422-1435. https://doi.org/10.1016/j.apmr.2022.03.012

Cherney, L. R., Babbitt, E. M., Semik, P., & Heinemann, A. W. (2011). Psychometric Properties of the Communication Confidence Rating Scale for Aphasia (CCRSA): Phase 1. Topics in Stroke Rehabilitation, 18(4), 352-360. 10.1310/tsr1804-352

Cherney, L. R., Lee, J. B., Kim, K. A., & van Vuuren, S. (2021). Web-based Oral Reading for Language in Aphasia (Web ORLA®): A pilot randomized control trial. Clin Rehabil, 35(7), 976-987. 10.1177/0269215520988475

Cherney, L. R., Braun, E. J., Lee, J. B., Kocherginsky, M., & Van Vuuren, S. (2019). Optimising recovery in aphasia: Learning following exposure to a single dose of computer-based script training. International Journal of Speech-Language Pathology, 21(5), 448-458. 10.1080/17549507.2019.1661518

Cherney, L. R., Merbitz, C. T., & Grip, J. C. (1986). Efficacy of oral reading in aphasia treatment outcome. Rehabilitation Literature, 47(5-6), 112-118.

Cherney, L. R., Halper, A. S., Holland, A. L., & Cole, R. (2008). Computerized Script Training for Aphasia: Preliminary Results. American Journal of Speech-Language Pathology, 17(1), 19-34. 10.1044/1058-0360(2008/003)

Cherney, L. R., Halper, A. S., & Kaye, R. C. (2011). Computer-based script training for aphasia: Emerging themes from post-treatment interviews. Journal of Communication Disorders, 44(4), 493-501. 10.1016/j.jcomdis.2011.04.002

Cherney, L. R., & van Vuuren, S. (2012). Telerehabilitation, Virtual Therapists, and Acquired Neurologic Speech and Language Disorders. Seminars in Speech & Language, 33(3), 243-257. 10.1055/s-0032-1320044

Clegg, J., O'Flynn, P., & Just, P. (2021). Speech and language therapy during and beyond COVID-19: building back better with people who have communication and swallowing needs. Royal College of Speech and Language Therapy.

Clifford, C., Paulk, E., Lin, Q., Cadwallader, J., Lubbers, K., & Frazier, L. D. (2022). Relationships among adult playfulness, stress, and coping during the COVID-19 pandemic. Current Psychology (New Brunswick, N.J.), , 1-10. 10.1007/s12144-022-02870-0

Code, C. (2020). The Implications of Public Awareness and Knowledge of Aphasia around the World. Annals of Indian Academy of Neurology, 23(Suppl 2), S95-S101. 10.4103/aian.AIAN_460_20

Code, C., Hemsley, G., & Herrmann, M. (1999). The emotional impact of aphasia. Seminars in Speech and Language, 20(1), 19-31. 10.1055/s-2008-1064006

Code, C., & Petheram, B. (2011). Delivering for aphasia. Int J Speech Lang Pathol. 13(1), 3-10. 10.3109/17549507.2010.520090

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Routledge.

Cole, T., & Gillies, M. (2021). Thinking and Doing: Challenge, Agency, and the Eudaimonic Experience in Video Games. Games and Culture, 16(2), 187-207. 10.1177/1555412019881536

Collaboration of Aphasia Trialists (no date) *Aphasia Datasets.* Available at https://www.aphasiatrials.org/aphasia-dataset/ (Accessed 13.04.23)

Coltheart, M. (1981) The MRC psycholinguistic database. *The Quarterly journal of experimental psychology. A, Human experimental psychology,* vol. 33, no. 4, pp. 497-505.

Cornwell, E. Y., & Waite, L. J. (2009). Social Disconnectedness, Perceived Isolation, and Health among Older Adults. Journal of Health and Social Behavior, 50(1), 31-48. 10.1177/002214650905000103

Corsten, S., Schimpf, E. J., Konradi, J., Keilmann, A., & Hardering, F. (2015). The participants' perspective: how biographic–narrative intervention influences identity negotiation and quality of life in aphasia. International Journal of Language & Communication Disorders, 50(6), 788-800. 10.1111/1460-6984.12173

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. Bmj, 337, a1655. 10.1136/bmj.a1655

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2013). Developing and evaluating complex interventions: The new Medical Research Council guidance. International Journal of Nursing Studies, 50(5), 587-592. https://doi.org/10.1016/j.ijnurstu.2012.09.010

Creet, E., Morris, J., Howard, D., & Nickels, L. (2019). Name it again! investigating the effects of repeated naming attempts in aphasia. Aphasiology, 33(10), 1202-1226. 10.1080/02687038.2019.1622352

Cruice, M., Aujla, S., Bannister, J., Botting, N., Boyle, M., Charles, N., Dhaliwal, V., Grobler, S., Hersh, D., Marshall, J., Morris, S., Pritchard, M., Scarth, L., Talbot, R., & Dipper, L. (2021). Creating a novel approach to discourse treatment through coproduction with people with aphasia and speech and language therapists. Aphasiology, , 1-23. 10.1080/02687038.2021.1942775

Cruice, M., Hill, R., Worrall, L., & Hickson, L. (2010). Conceptualising quality of life for older people with aphasia. Aphasiology, 24(3), 327-347. 10.1080/02687030802565849

Cruice, M., Worrall, L., & Hickson, L. (2006). Perspectives of Quality of Life by People with Aphasia and Their Family: Suggestions for Successful Living. Topics in Stroke Rehabilitation, 13(1), 14-24. 10.1310/4JW5-7VG8-G6X3-1QVJ

Cullum, C. M., Hynan, L. S., Grosch, M., Parikh, M., & Weiner, M. F. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. Journal of the International Neuropsychological Society, 20(10), 1028-1033.

Cullum, C. M., Weiner, M. F., Gehrmann, H. R., & Hynan, L. S. (2006). Feasibility of telecognitive assessment in dementia. Assessment, 13(4), 385-390.

Cuypers, K., Krokstad, S., Holmen, T. L., Skjei Knudtsen, M., Bygren, L. O., & Holmen, J. (2012). Patterns of receptive and creative cultural activities and their association with perceived health, anxiety, depression and satisfaction with life among adults: the HUNT study, Norway. Journal of Epidemiology and Community Health, 66(8), 698-703. 10.1136/jech.2010.113571

D'Cruz, K., Douglas, J., & Serry, T. (2020). Narrative storytelling as both an advocacy tool and a therapeutic process: Perspectives of adult storytellers with acquired brain injury. Neuropsychological Rehabilitation, 30(8), 1409-1429. 10.1080/09602011.2019.1586733

Davidson, B., Worrall, L., & Hickson, L. (2003). Identifying the communication activities of older people with aphasia: Evidence from naturalistic observation. Aphasiology, 17(3), 243-264. 10.1080/729255457

De Luca, R., Leonardi, S., Maresca, G., Marilena, F. C., Latella, D., Impellizzeri, F., Maggio, M. G., Naro, A., & Calabrò, R. S. (2021). Virtual reality as a new tool for the rehabilitation of post-stroke patients with chronic aphasia: an exploratory study. Aphasiology, ahead-of-print(ahead-of-print), 1-11. 10.1080/02687038.2021.1998882

De Renzi, E., & Vignolo, L. A. (1962). The token test: A sensitive test to detect receptive disturbances in aphasics. Brain : A Journal of Neurology, 85, 665-678. 10.1093/brain/85.4.665

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3), 283-321. 10.1037/0033-295X.93.3.283

Des Roches, C. A., & Kiran, S. (2017). Technology-based rehabilitation to improve communication after acquired brain injury. Frontiers in Neuroscience, 1110.3389/fnins.2017.00382

Devane N., Behn N., Marshall J., Ramachandran A., Wilson S. & Hilari K. (2022) The use of virtual reality in the rehabilitation of aphasia: a systematic review, Disability and Rehabilitation, DOI: 10.1080/09638288.2022.2138573

Difrancesco, S., Pulvermüller, F., & Mohr, B. (2012). Intensive language-action therapy (ILAT): The methods. Aphasiology, 26(11), 1317-1351. 10.1080/02687038.2012.705815

Dipper, L. T., Franklin, S., De Aguiar, V., Baumgaertner, A., Brady, M., Best, W., Bruehl, S., Denes, G., Godecke, E., & Gil, M. (2022). An umbrella review of aphasia intervention description in research: The AsPIRE project. Aphasiology, 36(4), 467-492.

Dipper, L., Marshall, J., Boyle, M., Botting, N., Hersh, D., Pritchard, M., & Cruice, M. (2020). Treatment for improving discourse in aphasia: a systematic review and synthesis of the evidence base. Aphasiology, , 1-43. 10.1080/02687038.2020.1765305

Dodd, J. T., Sandell, R. E., & Scott, C. A. (2014). Cultural Value: User value of museums and galleries: a critical view of the literature. University of Leicester. Report. https://hdl.handle.net/2381/37043

Doedens, W. J., & Meteyard, L. (2022). What is Functional Communication? A Theoretical Framework for Real-World Communication Applied to Aphasia Rehabilitation. Neuropsychology Review, 10.1007/s11065-021-09531-2

Doedens, W. J., & Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: a critical review. *Aphasiology, 34*(4), 492-514. 10.1080/02687038.2019.1702848

Doedens, W., & Meteyard, L. (2018). The importance of situated language use for aphasia rehabilitation. PsyArXiv, https://doi.org/10.31234/osf.io/svwpf

Druks, J., & Masterson, J. (2000). An Object and Action Naming Battery. Psychology Press.

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on Lexical Retrieval of Content Words in Sentences in Persons with Aphasia. Aphasiology, 23(3), 402-424. 10.1080/02687030802291339

Efstratiadou, E. A., Papathanasiou, I., Holland, R., Archonti, A., & Hilari, K. (2018). A Systematic Review of Semantic Feature Analysis Therapy Studies for Aphasia. Journal of Speech, Language & Hearing Research, 61(5), 1261-1278. 10.1044/2018_JSLHR-L-16-0330

Efstratiadou, E. A., Papathanasiou, I., Holland, R., Varlokosta, S., & Hilari, K. (2019). Efficacy of Elaborated Semantic Features Analysis in Aphasia: a quasi-randomised controlled trial, Aphasiology, 33:12, 1482-1503, DOI: 10.1080/02687038.2019.1571558

Elkind. (2009). Outcomes After Stroke: Risk of Recurrent Ischemic Stroke and Other Events. *The American Journal of Medicine, 122*(4, Supplement 2), S7-S13. 10.1016/j.amjmed.2009.02.005

Evans, W. S., Cavanaugh, R., Gravier, M. L., Autenreith, A. M., Doyle, P. J., Hula, W. D., & Dickey, M. W. (2021). Effects of Semantic Feature Type, Diversity, and Quantity on Semantic Feature Analysis Treatment Outcomes in Aphasia. American Journal of Speech-Language Pathology, 30(1), 344-358. 10.1044/2020_AJSLP-19-00112

Evans, W. S., Hula, W. D., Quique, Y., & Starns, J. J. (2020). How Much Time Do People With Aphasia Need to Respond During Picture Naming? Estimating Optimal Response Time Cutoffs Using a Multinomial Ex-Gaussian Approach. Journal of Speech, Language, and Hearing Research : JSLHR, 63(2), 599-614. 10.1044/2019_JSLHR-19-00255

Falconer, C., & Antonucci, S. M. (2012). Use of semantic feature analysis in group discourse treatment for aphasia: Extension and expansion. Aphasiology, 26(1), 64-82. 10.1080/02687038.2011.602390

Farley, A., Kennedy-Behr, A., & Brown, T. (2021). An Investigation Into the Relationship Between Playfulness and Well-being in Australian Adults: An Exploratory Study. OTJR: Occupation, Participation and Health, 41(1), 56-64. 10.1177/1539449220945311

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2015). Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. Cerebral Cortex, 26(5), 2018-2034. 10.1093/cercor/bhv020

Forster, A., Dickerson, J., Young, J., Patel, A., Kalra, L., Nixon, J., Smithard, D., Knapp, M., Holloway, I., Anwar, S., Farrin, A., & TRACS Trial Collaboration. (2013). A cluster randomised controlled trial and economic evaluation of a structured training programme for caregivers of inpatients after stroke: the TRACS trial. Health Technology Assessment (Winchester, England), 17(46), 1-216. 10.3310/hta17460

Frank, A. W. (1995). The wounded storyteller: body, illness and ethics. University of Chicago Press.

Galliers, J. R., Wilson, S., Marshall, J., Talbot, R., Devane, N., Booth, T., Woolf, C., & Greenwood, H. (2017). Experiencing EVA Park, A Multi-User Virtual World For People With Aphasia. Association for Computing Machinery.

Gangwani, R., Cain, A., Collins, A., & Cassidy, J. M. (2022). Leveraging Factors of Self-Efficacy and Motivation to Optimize Stroke Recovery. Frontiers in Neurology, 13, 823202. 10.3389/fneur.2022.823202

Gauch, M., Leinweber, J., Plath, A., Spelter, B., & Corsten, S. (2022). Quality of life outcomes from aphasia telepractice: A scoping review. Aphasiology, , 1-25. 10.1080/02687038.2022.2079604

German, D.J. (1991*). Test of Word Finding in Discourse (TWFD).* Austin, TX: Pro-ed.

Ghai, S., Ghai, I., & Lamontagne, A. (2020). Virtual reality training enhances gait poststroke: a systematic review and meta-analysis. Annals of the New York Academy of Sciences, 1478(1), 18-42. 10.1111/nyas.14420

Giachero, A., Calati, M., Pia, L., La Vista, L., Molo, M., Rugiero, C., Fornaro, C., & Marangolo, P. (2020). Conversational Therapy through Semi-Immersive Virtual Reality Environments for Language Recovery and Psychological Well-Being in Post Stroke Aphasia. Behavioural Neurology, 202010.1155/2020/2846046

Gillespie, A., & Hald, J. (2017). The paradox of helping: Contradictory effects of scaffolding people with aphasia to communicate. Plos One, 12(8), e0180708. https://doi.org/10.1371/journal.pone.0180708

Gilmore, N., Dwyer, M., & Kiran, S. (2019). Benchmarks of Significant Change After Aphasia Rehabilitation. Archives of Physical Medicine and Rehabilitation, 100(6), 1131-1139.e87. S0003-9993(18)31190-0

Gilmore, N., Meier, E. L., Johnson, J. P., & Kiran, S. (2020). Typicality-based semantic treatment for anomia results in multiple levels of generalisation. *Neuropsychological Rehabilitation, 30*(5), 802-828. 10.1080/09602011.2018.1499533

Girvan, C. (2018). What is a virtual world? Definition and classification. Educational Technology Research and Development, 66(5), 1087-1100. 10.1007/s11423-018-9577-y

Godecke, E., Armstrong, E., Rai, T., Ciccone, N., Rose, M. L., Middleton, S., Whitworth, A., Holland, A., Ellery, F., Hankey, G. J., Cadilhac, D. A., Bernhardt, J., & VERSE Collaborative Group. (2021). A randomized control trial of intensive aphasia therapy after acute stroke: The Very Early Rehabilitation for SpEech (VERSE) study. International Journal of Stroke : Official Journal of the International Stroke Society, 16(5), 556-572. 10.1177/1747493020961926

Goldberg, D. P., Gater, R., Sartorious, N., Ustun, T. B., Piccinelli, M., Gurege, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychological Medicine, 27(1), 191-197. 10.1017/S0033291796004242

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders (2nd ed.). Lea & Febiger.

Graves, J. (2000). Vocabulary needs in augmentative and alternative communication: a sample of conversational topics between staff providing services to adults with learning difficulties and their service users. British Journal of Learning Disabilities, 28(3), 113-119. 10.1046/j.1468-3156.2000.00060.x

Graves R. E., Bezeau S. C., Fogarty J. & Blair R. (2004) Boston Naming Test Short Forms: A Comparison of Previous Forms with New Item Response Theory Based Forms, Journal of Clinical and Experimental Neuropsychology, 26:7, 891-902, DOI: 10.1080/13803390490510716

Gravier, M. L., Dickey, M. W., Hula, W. D., Evans, W. S., Owens, R. L., Winans-Mitrik, R., & 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(1), 438-453. 10.1044/2017_AJSLP-16-0196

Grechuta, K., Rubio Ballester, B., Espín Munné, R., Usabiaga Bernal, T., Molina Hervás, B., Mohr, B., Pulvermüller, F., San Segundo, R. M., & Verschure, P. F. M. J. (2020). Multisensory cueing facilitates naming in aphasia. Journal of NeuroEngineering and Rehabilitation, 17(1)10.1186/s12984-020-00751-w

Grechuta, K., Ballester, B., Munne, R., Bernal, T., Hervás, B., Mohr, B., Pulvermüller, F., San Segundo, R., & Verschure, P. (2019). Augmented Dyadic Therapy Boosts Recovery of Language Function in Patients With Nonfluent Aphasia. Stroke (00392499), 50(5), 1270-1274. 10.1161/STROKEAHA.118.023729

Grechuta, K., Bellaster, B. R., Munne, R. E., Bernal, T. U., Hervas, B. M., Segundo, R. S., & Verschure, P. (2017). The effects of silent visuomotor cueing on word retrieval in Broca's aphasies: A pilot study. IEEE ...International Conference on Rehabilitation Robotics: [Proceedings], 2017, 193-199. 10.1109/ICORR.2017.8009245

Grechuta, K., Rubio, B., Duff, A., Oller, E. D., Pulvermüller, F., & Verschure, P. (2016). Intensive language-action therapy in virtual reality for a rehabilitation gaming system. Journal of Pain Management, 9(3), 243-254.

Hall, N., Boisvert, M., & Steele, R. (2013). Telepractice in the assessment and treatment of individuals with aphasia: a systematic review. International Journal of Telerehabilitation, 5(1), 27-38. 10.5195/ijt.2013.6119

Harrison, M. (2019). Evaluating the Intervention Fidelity of Self-managed Computer Therapy for Aphasia Post-stroke. PhD thesis, University of Sheffield

Hartholt, A., Traum, D., Marsella, S., Shapiro, A., Stratou, G., Leuski, A., Morency, L. P., & Gratch, J. (2013). All Together Now, Introducing the Virtual Human Toolkit10.1007/978-3-642-40415-3_33

Hartholt, A., Traum, D., Marsella, S., Shapiro, A., Stratou, G., Leuski, A., Morency, L., & Gratch, J. (2013). All Together Now, Introducing the Virtual Human Toolkit10.1007/978-3-642-40415-3_33

Harvey, S. R., Carragher, M., Dickey, M. W., Pierce, J. E., & Rose, M. L. (2021). Treatment dose in post-stroke aphasia: A systematic scoping review. Neuropsychological Rehabilitation, 31(10), 1629-1660. 10.1080/09602011.2020.1786412

Harvey, S., Carragher, M., Dickey, M. W., Pierce, J. E., & Rose, M. L. (2022). Dose effects in behavioural treatment of post-stroke aphasia: a systematic review and meta-analysis. Disability and Rehabilitation, 44(12), 2548-2559. 10.1080/09638288.2020.1843079

Harvey, S., Carragher, M., Dickey, M. W., & Rose, M. L. (2022). Outcomes from a pilot dose comparison study of naming therapy in aphasia. Aphasiology, , 1-28. 10.1080/02687038.2022.2144112

Heilemann, C., Best, W., Johnson, F., Beckley, F., Edwards, S., Maxim, J., & Beeke, S. (2014). Investigating treatment fidelity in a conversation-based aphasia therapy. Aphasie Und Verwandte Gebiete/Aphasie Et Domaines Associés, 37(2), 14-26.

Hilari, K., Galante, L., Huck, A., Pritchard, M., Allen, L., & Dipper, L. (2018). Cultural adaptation and psychometric testing of The Scenario Test UK for people with aphasia. Taylor & Francis.

Hilari, K., Northcott, S., Roy, P., Marshall, J., Wiggins, R. D., Chataway, J., & Ames, D. (2010). Psychological distress after stroke and aphasia: the first six months. Clinical Rehabilitation. 2010 Feb;24(2):181-90. doi: 10.1177/0269215509346090..

Hilari, K., Lamping, D. L., Smith, S. C., Northcott, S., Lamb, A., & Marshall, J. (2009). Psychometric properties of the Stroke and Aphasia Quality of Life Scale (SAQOL-39) in a generic stroke population. Clinical Rehabilitation, 23(6), 544-557. 10.1177/0269215508101729

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

Hilari, K., Behn, N., James, K., Northcott, S., Marshall, J., Thomas, S., Simpson, A., Moss, B., Flood, C., McVicker, S., & Goldsmith, K. (2021). Supporting wellbeing through peer-befriending (SUPERB) for people with aphasia: A feasibility randomised controlled trial. Clin Rehabil, 35(8), 1151-1163. 10.1177/0269215521995671

Hilari, K., Byng, S., Lamping, D. L., & Smith, S. C. (2003). Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. Stroke, 34(8), 1944-1950. 10.1161/01.STR.0000081987.46660.ED

Hilari, K., Needle, J. J., & Harrison, K. L. (2012). What Are the Important Factors in Health-Related Quality of Life for People With Aphasia? A Systematic Review, Arch Phys Med Rehabil. 2012 Jan;93(1 Suppl):S86-95. https://doi.org/10.1016/j.apmr.2011.05.028

Hinckley, J. J., & Douglas, N. F. (2013). Treatment fidelity: Its importance and reported frequency in aphasia treatment studies. Am J Speech Lang Pathol. 2013 May;22(2):S279-84. doi: 10.1044/1058-0360(2012/12-0092).

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A., & Michie, S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ : British Medical Journal, 348(mar07 3), g1687. 10.1136/bmj.g1687

Holland, A. L., Halper, A. S., & Cherney, L. R. (2010). Tell me your story: analysis of script topics selected by persons with aphasia. American Journal of Speech-Language Pathology, 19(3), 198-203. 10.1044/1058-0360(2010/09-0095)

Holland, A., Frattali, C., & Fromm, D. (1999). Communication Activities of Daily Living-2. Pro-Ed.

Holland, A. (2007). Counseling/Coaching in Chronic Aphasia: Getting on With Life. Topics in Language Disorders, 27(4)

Holland, A. L., Nelson, R. L., & Goldberg, S. A. (2014). Counseling in communication disorders: a wellness perspective (Second ed.). Plural Publishing.

Hoover, E., DeDe, G., & Maas, E. (2021). A Randomized Controlled Trial of the Effects of Group Conversation Treatment on Monologic Discourse in Aphasia. Journal of Speech, Language, and Hearing Research : JSLHR, 64(12), 4861-4875. 10.1044/2021_JSLHR-21-00023

Horton, S., Lane, K., & Shiggins, C. (2016). Supporting communication for people with aphasia in stroke rehabilitation: transfer of training in a multidisciplinary stroke team. Aphasiology, 30(5), 629-656. 10.1080/02687038.2014.1000819

House, J. S. (2001). Social isolation kills, but how and why? Psychosomatic Medicine, 63(2), 273-274. 10.1097/00006842-200103000-00011

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

Hughes, J., & Orange, J. B. (2007). Mapping functional communication measurements for traumatic brain injury to the WHO-ICF. Revue Canadienne D'orthophonie Et D'audiologie-Vol, 31(3), 135.

International Conference on Primary Health Care. (1978). Declaration of Alma-Ata. WHO Chronicle, 32(11), 428-430.

James Lind Alliance (2021) *Stroke rehabilitation and long-term care top 10 priorities.* Available at https://www.jla.nihr.ac.uk/priority-setting-partnerships/stroke/stroke-rehabilitation-and-long-term-care-top-10-priorities.htm (Accessed 13.04.23)

Jefferies, E., & Lambon Ralf, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia : a case-series comparison. Brain (London, England : 1878), 129(Pt 8), 2132-2147. 10.1093/brain/awl153

Jones, F., & Riazi, A. (2011). Self-efficacy and self-management after stroke: a systematic review. Disability and Rehabilitation, 33(10), 797-810. 10.3109/09638288.2010.511415

Jüni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ (Clinical Research Ed.), 323(7303), 42-46. 10.1136/bmj.323.7303.42

Jutai, J., & Day, H. (2002). Psychosocial Impact of Assistive Devices Scale (PIADS). Technology and Disability, 14, 107-111. 10.3233/TAD-2002-14305

Kagan, A. (1998). Supported conversation for adults with aphasia: methods and resources for training conversation partners. Aphasiology, 12(9), 816-830. 10.1080/02687039808249575

Kalinyak-Fliszar, M., Martin, N., Keshner, E., Rudnicky, A., Shi, J., & Teodoro, G. (2015). Using Virtual Technology to Promote Functional Communication in Aphasia: Preliminary Evidence From Interactive Dialogues With Human and Virtual Clinicians. American Journal of Speech-Language Pathology, 24(4), S974-S989. 10.1044/2015_AJSLP-14-0160

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). The Boston Naming Test (1st Ed.). Philadelphia: Lea & Febiger.

Karamians, R., Proffitt, R., Kline, D., & Gauthier, L. V. (2020). Effectiveness of Virtual Reality- and Gaming-Based Interventions for Upper Extremity Rehabilitation Poststroke: A Meta-analysis. Archives of Physical Medicine and Rehabilitation, 101(5), 885-896. 10.1016/j.apmr.2019.10.195

Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? PharmacoEconomics, 34(7), 645-649. 10.1007/s40273-016-0389-9

Kauhanen, M., Korpelainen, J. T., Hiltunen, P., Nieminen, P., Sotaniemi, K. A., & Myllylä, V. V. (2000). Domains and determinants of quality of life after stroke caused by brain infarction. Archives of Physical Medicine and Rehabilitation, 81(12), 1541-1546. https://doi.org/10.1053/apmr.2000.9391

Keane, C., & Kiran, S. (2015). The nature of facilitation and interference in the multilingual language system: Insights from treatment in a case of trilingual aphasia. Cognitive Neuropsychology, 32(3-4), 169-194.

Kendall, D. L., Moldestad, M. O., Allen, W., Torrence, J., & Nadeau, S. E. (2019). Phonomotor Versus Semantic Feature Analysis Treatment for Anomia in 58 Persons With Aphasia: A Randomized Controlled Trial. Journal of Speech, Language, and Hearing Research : JSLHR, 62(12), 4464-4482. 10.1044/2019_JSLHR-L-18-0257

Kertesz, A. (2007). The Western Aphasia Battery–Revised (WAB-R) . Pearson.

Kessels, R. P. (2003). Patients' memory for medical information. Journal of the Royal Society of Medicine, 96(5), 219-222. 0960219

Keysers, C., & Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1644), 20130175. 10.1098/rstb.2013.0175

Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research : JSLHR, 51*(6), 1550-1568. 10.1044/1092-4388(2008/07-0038)

Kiran, S., & Thompson, C. K. (2019). Neuroplasticity of Language Networks in Aphasia: Advances, Updates, and Future Challenges. Frontiers in Neurology, 10, 295. 10.3389/fneur.2019.00295

Kirkevold, M., Martinsen, R., Bronken, B. A., & Kvigne, K. (2014). Promoting psychosocial wellbeing following stroke using narratives and guided self-determination: a feasibility study. BMC Psychology, 2(1), 4-4. eCollection 2014. 10.1186/2050-7283-2-4

Kladouchou, V., Papathanasiou, I., Efstratiadou, E. A., Christaki, V., & Hilari, K. (2017). Treatment integrity of elaborated semantic feature analysis aphasia therapy delivered in individual and group settings. International Journal of Language & Communication Disorders, 52(6), 733-749. 10.1111/1460-6984.12311

Kleim, J. A., & Jones, T. A. (2008). Principles of Experience-Dependent Neural Plasticity: Implications for Rehabilitation After Brain Damage. Journal of Speech, Language, and Hearing Research, 51(1), S225. 10.1044/1092-4388(2008/018)

Kok M (2018) Guidance Document: Evaluating public involvement in research. UWE Bristol. Available at: https://uwe-repository.worktribe.com/output/866884/guidance-document-evaluating-public-involvement-in-research (Accessed: 17.11.22)

Konnerup, U. (2015). Renegotiation of Self after a Brain Injury Using Immersive Virtual Environments: A Contribution to Technology-mediated Speech Therapy Ph.d.-serien for Det Humanistiske Fakultet, Aalborg Universitet, Aalborg Universitetsforlag. https://doi.org/10.5278/vbn.phd.hum.00011

Lam, J. M. C., & Wodchis, W. P. (2010). The Relationship of 60 Disease Diagnoses and 15 Conditions to Preference-Based Health-Related Quality of Life in Ontario Hospital-Based Long-Term Care Residents. Medical Care, 48(4), 380-387. 10.1097/MLR.0b013e3181ca2647

Lanyon, L. E., Rose, M. L., & Worrall, L. (2013). The efficacy of outpatient and community-based aphasia group interventions: a systematic review. International Journal of Speech-Language Pathology, 15(4), 359-374. 10.3109/17549507.2012.752865

Lin, Y., Lin, H., & Yang, Y. (2017). Players' Value Structure in Digital Games. Games and Culture, 12(1), 72-99. 10.1177/1555412015581710

Lo, S. H. S., Chau, J. P. C., Lam, S. K. Y., Saran, R., Choi, K. C., Zhao, J., & Thompson, D. R. (2022). Association between participation self-efficacy and participation in stroke survivors. BMC Neurology, 22(1), 361. 10.1186/s12883-022-02883-z

Lohse, K. R., Hilderman, C. G. E., Cheung, K. L., Tatla, S., & Van der Loos, H. F. M. (2014). Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PloS One, 9(3), e93318. 10.1371/journal.pone.0093318

Lorenz, A. & Ziegler, W. (2009) Semantic vs. word-form specific techniques in anomia treatment: A multiple single-case study. *Journal of neurolinguistics,* vol. 22, no. 6, pp. 515-537.

Magnuson, C. D., & Barnett, L. A. (2013). The Playful Advantage: How Playfulness Enhances Coping with Stress. Leisure Sciences, 35(2), 129-144. 10.1080/01490400.2013.761905

Manheim, L. M., Halper, A. S., & Cherney, L. (2009). Patient-Reported Changes in Communication After Computer-Based Script Training for Aphasia. Archives of Physical Medicine and Rehabilitation, 90(4), 623-627. https://doi.org/10.1016/j.apmr.2008.10.022

Manning, M., MacFarlane, A., Hickey, A., Galvin, R., & Franklin, S. (2022a). Regulating emotional responses to aphasia to re-engage in life: a qualitative interview study. International Journal of Language & Communication Disorders, 57(2), 352-365. 10.1111/1460-6984.12702

Manning, M., MacFarlane, A., Hickey, A., Galvin, R., & Franklin, S. (2022b). The relevance of stroke care for living well with post-stroke aphasia: a qualitative interview study with working-aged adults. Disability and Rehabilitation, 44(14), 3440-3452. 10.1080/09638288.2020.1863483

Manning, M., Cuskelly, C., Russ, E., & Franklin, S. (2020). Supporting people with post-stroke aphasia to live well: A cross-sectional survey of Speech & Language Therapists in Ireland. Health & Social Care in the Community, 28(6), 2105-2116. https://doi.org/10.1111/hsc.13021

Maples-Keller, J., Bunnell, B. E., Kim, S., & Rothbaum, B. O. (2017). The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. Harvard Review of Psychiatry, 25(3), 103-113. 10.1097/HRP.0000000000000138

Maresca, G., Maggio, M. G., Latella, D., Cannavò, A., De Cola, M. C., Portaro, S., Stagnitti, M. C., Silvestri, G., Torrisi, M., Bramanti, A., De Luca, R., & Calabrò, R. S. (2019). Toward Improving Poststroke Aphasia: A Pilot Study on the Growing Use of Telerehabilitation for the Continuity of Care. Journal of Stroke & Cerebrovascular Diseases, 28(10), N.PAG. 10.1016/j.jstrokecerebrovasdis.2019.104303

Marshall, J., Booth, T., Devane, N., Galliers, J., Greenwood, H., Hilari, K., Talbot, R., Wilson, S., & Woolf, C. (2016). Evaluating the Benefits of Aphasia Intervention Delivered in Virtual Reality: Results of a Quasi-Randomised Study. PLoS ONE, 11(8), 1-18. 10.1371/journal.pone.0160381

Marshall, J., Devane, N., Edmonds, L., Talbot, R., Wilson, S., Woolf, C., & Zwart, N. (2018). Delivering word retrieval therapies for people with aphasia in a virtual communication environment. Aphasiology, 32(9), 1054-1074. 10.1080/02687038.2018.1488237

Marshall, J., Devane, N., Talbot, R., Caute, A., Cruice, M., Hilari, K., MacKenzie, G., Maguire, K., Patel, A., Roper, A., & Wilson, S. (2020). A randomised trial of social support group intervention for people with aphasia: A Novel application of virtual reality.(Research Article). PloS One, 15(9), e0239715. 10.1371/journal.pone.0239715

Mayo Clinic. (2021). Stroke symptoms and causes. Available at: https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113#:~:text=Overview,and%20prompt%20treatment%20is%20crucial (Accessed: 04.03.21)

Mbuagbaw, L., Kosa, S. D., Lawson, D. O., Stalteri, R., Olaiya, O. R., Alotaibi, A., & Thabane, L. (2019). The reporting of progression criteria in protocols of pilot trials designed to assess the feasibility of main trials is insufficient: a meta-epidemiological study. Pilot and Feasibility Studies, 5(1), 120. 10.1186/s40814-019-0500-z

McClendon, M. J. (1991). Acquiescence and Recency Response-Order Effects in Interview Surveys. Sociological Methods & Research, 20(1), 60-103. 10.1177/0049124191020001003

McGill, K., Sackley, C. M., Godwin, J., McGarry, J., & Brady, M. C. (2020). A systematic review of the efficiency of recruitment to stroke rehabilitation randomised controlled trials. Trials, 21(1), 68. 10.1186/s13063-019-3991-2

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia Medica, 22(3), 276-282. biochem_med-22-3-276-4

McMenamin, R., Griffin, M., Grzybowska, B., & Pound, C. (2021). Working together: experiences of people with aphasia as co-researchers in participatory health research studies. Aphasiology, 1-22. 10.1080/02687038.2021.1923948

Mehri, A., & Jalaie, S. (2014). A Systematic Review on methods of evaluate sentence production deficits in agrammatic aphasia patients: Validity and Reliability issues. Journal of Research in Medical Sciences, 19(9), 885-898.

Mekbib, D. B., Han, J., Zhang, L., Fang, S., Jiang, H., Zhu, J., Roe, A. W., & Xu, D. (2020). Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. Brain Injury, 34(4), 456-465. 10.1080/02699052.2020.1725126

Menger, F., Morris, J., & Salis, C. (2020). The impact of aphasia on Internet and technology use. Disability and Rehabilitation, 42(21), 2986-2996. 10.1080/09638288.2019.1580320

Meteyard, L. & Bose, A. (2018) What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. Journal of Speech, Language, and Hearing Research, 61 (3). pp. 658-674. ISSN 1558-9102 doi:10.1044/2017_JSLHR-L-17-0214

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. Cortex, 48(7), 788-804. 10.1016/j.cortex.2010.11.002

Michalski, C. A., Diemert, L. M., Helliwell, J. F., Goel, V., & Rosella, L. C. (2020). Relationship between sense of community belonging and self-rated health across life stages. SSM - Population Health, 12, 100676. https://doi.org/10.1016/j.ssmph.2020.100676

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Science, 6(1), 42. 10.1186/1748-5908-6-42

Mongrain, M., & Anselmo Matthews, T. (2012). Do Positive Psychology Exercises Work? A Replication of Seligman et al. (). Journal of Clinical Psychology, 68(4), 382-n/a. 10.1002/jclp.21839

Moseley, A. M., Szikszay, T. M., Lin, C. -. C., Mathieson, S., Elkins, M. R., Herbert, R. D., Maher, C. G., & Sherrington, C. (2015). A systematic review of the measurement properties and usage of the Physiotherapy Evidence Database (PEDRO) scale. Physiotherapy, 101, e1043. 10.1016/j.physio.2015.03.1919

Napier, A. D., Ancarno, C., Butler, B., Calabrese, J., Chater, A., Chatterjee, H., Guesnet, F., Horne, R., Jacyna, S., & Jadhav, S. (2014). Culture and health. The Lancet, 384(9954), 1607-1639.

NICE. (2019). NICE's approach to public involvement in guidance and standards: a practical guide. Available at: https://www.nice.org.uk/about/nice-communities/nice-and-the-public/public-involvement/public-involvement-programme (Accessed: 14.12.22)

Nicholas, L. E., & Brookshire, R. H. (1993). A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults With Aphasia. Journal of Speech and Hearing Research, 36(2), 338-350. 10.1044/jshr.3602.338

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

Northcott, S., & Hilari, K. (2011). Why do people lose their friends after a stroke?: Friendship loss post stroke. International Journal of Language & Communication Disorders, 46(5), 524-534. 10.1111/j.1460-6984.2011.00079.x

Northcott, S., Marshall, J., & Hilari, K. (2016). What Factors Predict Who Will Have a Strong Social Network Following a Stroke? Journal of Speech, Language, and Hearing Research, 59(4), 772-783. 10.1044/2016_JSLHR-L-15-0201

Northcott, S., Simpson, A., Thomas, S., Hirani, S. P., Flood, C., & Hilari, K. (2019). SOlution Focused brief therapy In post-stroke Aphasia (SOFIA Trial): protocol for a feasibility randomised controlled trial. AMRC Open Research

O'Cathain, A., Croot, L., Sworn, K., Duncan, E., Rousseau, N., Turner, K., Yardley, L., & Hoddinott, P. (2019). Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot and Feasibility Studies, 5(1), 41. 10.1186/s40814-019-0425-6

O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. Academic Medicine : Journal of the Association of American Medical Colleges, 89(9), 1245-1251. 10.1097/ACM.0000000000000388

O'Cathain, A., Croot, L., Duncan, E., Rousseau, N., Sworn, K., Turner, K. M., Yardley, L., & Hoddinott, P. (2019). Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open, 9(8), e029954. 10.1136/bmjopen-2019-029954

Oliver, M. (2013). The social model of disability: thirty years on. Disability and Society, 28(7), 1024-1026. 10.1080/09687599.2013.818773

Olness, G. S., & Ulatowska, H. K. (2011). Personal narratives in aphasia: Coherence in the context of use. Aphasiology, 25(11), 1393-1413. 10.1080/02687038.2011.599365

O'Neill, B., Moran, K., & Gillespie, A. (2010). Scaffolding rehabilitation behaviour using a voice-mediated assistive technology for cognition. Neuropsychological Rehabilitation, 20(4), 509-527.

Palma, G., Freitas, T., Bonuzzi, G., Soares, M., Leite, P., Mazzini, N., Almeida, M., Pompeu, J., & Torriani-Pasin, C. (2017). Effects of virtual reality for stroke individuals based on the International Classification of Functioning and Health: a systematic review. Topics in Stroke Rehabilitation, 24(4), 269-278. 10.1080/10749357.2016.1250373

Palmer, R., Cooper, C., Enderby, P., Brady, M., Julious, S., Bowen, A., & Latimer, N. (2015). Clinical and cost effectiveness of computer treatment for aphasia post stroke (Big CACTUS): study protocol for a randomised controlled trial. Trials, 16(1), 18. 10.1186/s13063-014-0527-7

Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., Latimer, N., Julious, S., Cross, E., Alshreef, A., Harrison, M., Bradley, E., Witts, H., & Chater, T. (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. 10.1016/S1474-4422(19)30192-9

Palmer, R., Enderby, P., & Paterson, G. (2013). Using computers to enable self-management of aphasia therapy exercises for word finding: the patient and carer perspective. International Journal of Language & Communication Disorders, 48(5), 508-521. https://doi.org/10.1111/1460-6984.12024

Palmer, R., Hughes, H., & Chater, T. (2017). What do people with aphasia want to be able to say? A content analysis of words identified as personally relevant by people with aphasia. PloS One, 12(3), e0174065. 10.1371/journal.pone.0174065

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. Nature Reviews Neuroscience, 8(12), 976-987. 10.1038/nrn2277

Pedraza O., Sachs B. C., Ferman T. J., Rush B. K., Lucas J. A. (2011) Difficulty and discrimination parameters of Boston naming test items in a consecutive clinical series. Arch Clin Neuropsychol. Aug;26(5):434-44. doi: 10.1093/arclin/acr042.

Peñaloza, C., Martin, N., Laine, M., & Rodríguez-Fornells, A. (2022). Language learning in aphasia: A narrative review and critical analysis of the literature with implications for language therapy. *Neuroscience & Biobehavioral Reviews, 141*, 104825. 10.1016/j.neubiorev.2022.104825

Perdices, M., Tate, R. L., & Rosenkoetter, U. (2019). An Algorithm to Evaluate Methodological Rigor and Risk of Bias in Single-Case Studies. Behav Modif, , 0145445519863035. 10.1177/0145445519863035

Pettigrove, K., Lanyon, L. E., Attard, M. C., Vuong, G., & Rose, M. L. (2021). Characteristics and impacts of community aphasia group facilitation: a systematic scoping review. Disability and Rehabilitation, 1-15. 10.1080/09638288.2021.1971307

Picano, C., Quadrini, A., Pisano, F., & Marangolo, P. (2021). Adjunctive Approaches to Aphasia Rehabilitation: A Review on Efficacy and Safety, Brain Science, 2;11 (1):41. 10.3390/brainsci11010041

Pitt, R., Theodoros, D., Hill, A. J., & Russell, T. (2019). The development and feasibility of an online aphasia group intervention and networking program - TeleGAIN. International Journal of Speech-Language Pathology, 21(1), 23-36. 10.1080/17549507.2017.1369567

Pitts, L. L., Hurwitz, R., Lee, J. B., Carpenter, J., & Cherney, L. R. (2018). Validity, reliability and sensitivity of the NORLA-6: Naming and oral reading for language in aphasia 6-point scale. International Journal of Speech-Language Pathology, 20(2), 274-283. 10.1080/17549507.2016.1276962

Plaut, D. C. (1996). Relearning after damage in connectionist networks: toward a theory of rehabilitation. *Brain and Language, 52*(1), 25-82. 10.1006/brln.1996.0004

Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set"–a practical guideline. *Trials, 17*(1), 1-10.

Proyer, R. T. (2013). The well-being of playful adults: Adult playfulness, subjective well-being, physical well-being, and the pursuit of enjoyable activities. The European Journal of Humour Research, 1(1), 84-98.

Public Health England. (2018). Briefing document: First incidence of stroke. London: PHE Publications. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/678444/Stroke_incidence_briefing_document_2018.pdf

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. 10.1161/01.str.32.7.1621

Quique, Y. M., Evans, W. S., & Dickey, M. W. (2019). Acquisition and Generalization Responses in Aphasia Naming Treatment: A Meta-Analysis of Semantic Feature Analysis Outcomes. American Journal of Speech-Language Pathology, 28(1), 230-246. 10.1044/2018_AJSLP-17-0155

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. Nature Reviews Neuroscience, 18(1), 42-55. 10.1038/nrn.2016.150

REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators. (2022). Dosage, Intensity, and Frequency of Language Therapy for Aphasia: A Systematic Review-Based, Individual Participant Data Network Meta-Analysis. Stroke, 53(3), 956-967. 10.1161/STROKEAHA.121.035216

Renvall, K., Nickels, L., & Davidson, B. (2013a). Functionally relevant items in the treatment of aphasia (part I): Challenges for current practice. Aphasiology, 27(6), 636-650. 10.1080/02687038.2013.786804

Renvall, K., Nickels, L., & Davidson, B. (2013b). Functionally relevant items in the treatment of aphasia (part II): Further perspectives and specific tools. Aphasiology, 27(6), 651-677. 10.1080/02687038.2013.796507

Repetto, C., Paolillo, M. P., Tuena, C., Bellinzona, F., & Riva, G. (2020). Innovative technology-based interventions in aphasia rehabilitation: a systematic review. Aphasiology, 1-24. 10.1080/02687038.2020.1819957

Richardson, J. D., Hudspeth Dalton, S. G., Shafer, J., & Patterson, J. (2016). Assessment fidelity in aphasia research. American Journal of Speech-Language Pathology, 25(4S), S788-S797.

Ritchie, J., & Lewis, J. (2003). Qualitative research practice: a guide for social science students and researchers. SAGE.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology, 24,* 121–133.

Robson, J., Marshall, J., Pring, T., Montagu, A., & Chiat, S. (2004). Processing proper nouns in aphasia: Evidence from assessment and therapy. Aphasiology, 18(10), 917-935. 10.1080/02687030444000462

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. Evaluation (London, England.1995), 14(1), 29-48. 10.1177/1356389007084674

Rogers, T. T., & Lambon Ralph, M. A. (2022). Semantic tiles or hub-and-spokes? Trends in Cognitive Sciences, 26(3), 189-190. https://doi.org/10.1016/j.tics.2022.01.002

Roper, A., & Skeat, J. (2022). Innovation through participatory design: Collaborative qualitative methods in the development of speech-language pathology technology. Int J Speech Lang Pathol., 1-6. 10.1080/17549507.2022.2050943

Rose, D., & Pevalin, D. (2003). A Researcher's Guide to the National Statistics Socio-economic Classification. Sage: London. 10.4135/9780857024725

Rose, M. L., Nickels, L., Copland, D., Togher, L., Godecke, E., Meinzer, M., Rai, T., Cadilhac, D. A., Kim, J., Hurley, M., Foster, A., Carragher, M., Wilcox, C., Pierce, J. E., & Steel, G. (2022). Results of the COMPARE trial of Constraint-induced or Multimodality Aphasia

Therapy compared with usual care in chronic post-stroke aphasia. Journal of Neurology, Neurosurgery & Psychiatry, 93(6), 573-581. 10.1136/jnnp-2021-328422

Rose, T. A., Worrall, L. E., Hickson, L. M., & Hoffmann, T. C. (2012). Guiding principles for printed education materials: Design preferences of people with aphasia. International Journal of Speech-Language Pathology, 14(1), 11-23. 10.3109/17549507.2011.631583

Ross, A., Winslow, I., Marchant, P., & Brumfitt, S. (2006). Evaluation of communication, life participation and psychological well-being in chronic aphasia: The influence of group intervention. Aphasiology, 20(5), 427-448. 10.1080/02687030500532786

Royal College of Physicians. (2015). Sentinel Stroke National Audit Programme (SSNAP) post-acute organisational audit. Available at: https://www.strokeaudit.org/Documents/National/PostAcuteOrg/2015/2015-PAOrgGenericReportPhase2.aspx (Accessed: 14.12.22)

Royal College of Physicians Sentinel Stroke National Audit Programme. (2015). 'Is stroke care improving?' Second SSNAP Annual Report prepared on behalf of the Intercollegiate Stroke Working Party. Available at: https://www.hqip.org.uk/resource/is-stroke-care-improving-the-second-ssnap-annual-report/#.Y5maLuzP0_U Accessed: (14.12.22).

Sather, T. W., Howe, T., Nelson, N. W., & Lagerwey, M. (2017). Optimizing the Experience of Flow for Adults With Aphasia: A Focus on Environmental Factors. Topics in Language Disorders, 37(1)

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs (pp. 35- 37). Windsor, England: NFER-NELSON

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Services Research, 17(1), 88. 10.1186/s12913-017-2031-8

Sekhon, M., Cartwright, M., & Francis, J. J. (2022). Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. BMC Health Services Research, 22(1), 279. 10.1186/s12913-022-07577-3

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. American Psychologist, 55(1), 5-14. 10.1037/0003-066X.55.1.5

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive Psychology Progress: Empirical Validation of Interventions. The American Psychologist, 60(5), 410-421. 10.1037/0003-066X.60.5.410

Sella, F., Raz, G., & Cohen Kadosh, R. (2021). When randomisation is not good enough: Matching groups in intervention studies. Psychonomic Bulletin & Review, 28(6), 2085-2093. 10.3758/s13423-021-01970-5

Shadden, B. B., & Agan, J. P. (2004). Renegotiation of Identity: The Social Context of Aphasia Support Groups. Topics in Language Disorders, 24(3), 174-186. 10.1097/00011363-200407000-00005

Shiggins, C., Ryan, B., O'Halloran, R., Power, E., Bernhardt, J., Lindley, R. I., McGurk, G., Hankey, G. J., & Rose, M. L. (2022). Towards the consistent inclusion of people with aphasia in stroke research irrespective of discipline. Archives of Physical Medicine and Rehabilitation, https://doi.org/10.1016/j.apmr.2022.07.004

Shiggins, C., Soskolne, V., Olenik, D., Pearl, G., Haaland-Johansen, L., Isaksen, J., Jagoe, C., McMenamin, R., & Horton, S. (2020). Towards an asset-based approach to promoting and sustaining well-being for people with aphasia and their families: an international exploratory study. Aphasiology, 34(1), 70-101. 10.1080/02687038.2018.1548690

Shrubsole, K., Lin, T., Burton, C., Scott, J., & Finch, E. (2021). Delivering an iterative Communication Partner Training programme to multidisciplinary healthcare professionals: A pilot implementation study and process evaluation. International Journal of Language & Communication Disorders, 56(3), 620-636. https://doi.org/10.1111/1460-6984.12618

Sil, A., Kumar, P., Kumar, R., & Das, N. K. (2019). Selection of Control, Randomization, Blinding, and Allocation Concealment. Indian Dermatology Online Journal, 10(5), 601-605. 10.4103/idoj.IDOJ_149_19

Simmons-Mackie N, Damico J, S., & Damico H, L. (1999). A Qualitative Study of Feedback in Aphasia Treatment. American Journal of Speech-Language Pathology, 8(3), 218-230. 10.1044/1058-0360.0803.218

Simmons-Mackie, N., & Cherney, L. R. (2018). Aphasia in North America: Highlights of a White Paper. Archives of Physical Medicine and Rehabilitation, 99(10), e117. 10.1016/j.apmr.2018.07.417

Sirman, N., Beeke, S., & Cruice, M. (2017). Professionals’ perspectives on delivering conversation therapy in clinical practice. Aphasiology, 31(4), 465-494.

Snell, S., Martin, N., & Keshner, E. A. (2017). Engagement with a virtual clinician encourages gesture usage in speakers with aphasia. Paper presented at the , 2017-June10.1109/ICVR.2017.8007526

Spaccavento, S., Falcone, R., Cellamare, F., Picciola, E., & Glueckauf, R. L. (2021). Effects of computer-based therapy versus therapist-mediated therapy in stroke-related aphasia: Pilot non-inferiority study. Journal of Communication Disorders, 94, 106158. S0021-9924(21)00081-2

Spell, L. A., Richardson, J. D., Basilakos, A., Stark, B. C., Teklehaimanot, A., Hillis, A. E., & Fridriksson, J. (2020). Developing, Implementing, and Improving Assessment and Treatment Fidelity in Clinical Aphasia Research. American Journal of Speech-Language Pathology, 29(1), 286-298. 10.1044/2019_AJSLP-19-00126

Springer, L., Huber, W., Schlenck, K., & Schlenck, C. (2000). Agrammatism: Deficit or compensation? Consequences for aphasia therapy. *Neuropsychological Rehabilitation, 10*(3), 279-309.

Stanton, R., Ada, L., Dean, C. M., & Preston, E. (2015). Feedback Received While Practicing Everyday Activities During Rehabilitation After Stroke: An Observational Study. Physiotherapy Research International : The Journal for Researchers and Clinicians in Physical Therapy, 20(3), 166-173. 10.1002/pri.1612

Stone, C. A. (1998). The metaphor of scaffolding: its utility for the field of learning disabilities. Journal of Learning Disabilities, 31(4), 344-364. 10.1177/002221949803100404

Székely, A., D'Amico, S., Devescovi, A. *et al.* (2003) Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers* **35**, 621–633. 10.3758/BF03195542

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). Health measurement scales: a practical guide to their development and use. Oxford University Press, USA.

Stroke Association. (2018). State of the nation: stroke statistics. Stroke Association. Available at: https://www.stroke.org.uk/resources/state-nation-stroke-statistics (Accessed: 26.10.19)

Strong, K. A., Lagerwey, M. D., & Shadden, B. B. (2018). More than a story: My life came back to life. American Journal of Speech-Language Pathology, 27(1), 464-476. 10.1044/2017_AJSLP-16-0167

Strong, K. A., & Shadden, B. B. (2020). The Power of Story in Identity Renegotiation: Clinical Approaches to Supporting Persons Living With Aphasia. Perspectives of the ASHA Special Interest Groups, 5(2), 371-383. 10.1044/2019_PERSP-19-00145

Strong, K. A. & Shadden, B. B. (2020), Stories at the Heart of Life Participation: Both the Telling and Listening Matter. Chapter 5. In A. L. Holland & R. J. Elman (Eds.) Neurogenic communication disorders and the Life Participation Approach: The social imperative in supporting individuals and families (pp. 105-130) Plural Publishing.

Suresh, K. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. Journal of Human Reproductive Sciences, 4(1), 8-11. 10.4103/0974-1208.82352

Sutherland, L. M., Middleton, P. F., Anthony, A., Hamdorf, J., Cregan, P., Scott, D., & Maddern, G. J. (2006). Surgical simulation: a systematic review. Annals of Surgery, 243(3), 291-300. 00000658-200603000-00001

Szczepańska-Gieracha, J., & Mazurek, J. (2020). The Role of Self-Efficacy in the Recovery Process of Stroke Survivors. Psychology Research and Behavior Management, 13, 897-906. 10.2147/PRBM.S273009

Sze, W. P., Hameau, S., Warren, J., & Best, W. (2020). Identifying the components of a successful spoken naming therapy: a meta-analysis of word-finding interventions for adults with aphasia. Aphasiology, , 1-40. 10.1080/02687038.2020.1781419

Tarrant, M., Warmoth, K., Code, C., Dean, S., Goodwin, V. A., Stein, K., & Sugavanam, T. (2016). Creating psychological connections between intervention recipients: development and focus group evaluation of a group singing session for people with aphasia. BMJ Open, 6(2), e009652. 10.1136/bmjopen-2015-009652

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychological Rehabilitation, 23(5), 619-638. 10.1080/09602011.2013.824383 [doi]

Tate, R. L., Rosenkoetter, U., Wakim, D., Sigmundsdottir, L., Doubleday, J., Togher, L., McDonald, S., Perdices, M., & PsycBITE Group. (2015). Risk of bias in N-of-1 trials (RoBiNT) Scale: an expanded manual for the critical appraisal of single-case reports. PsycBITE.

Taubner, H., Hallén, M., & Wengelin, Å. (2020). Still the same? – Self-identity dilemmas when living with post-stroke aphasia in a digitalised society. Aphasiology, 34(3), 300-318. 10.1080/02687038.2019.1594151

Tausch, A. P., & Menold, N. (2016). Methodological Aspects of Focus Groups in Health Research: Results of Qualitative Interviews With Focus Group Moderators. Global Qualitative Nursing Research, 3, 2333393616630466. 10.1177/2333393616630466

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health and Quality of Life Outcomes, 5(1), 63. 10.1186/1477-7525-5-63

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, Daniëlle A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60(1), 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., & Bond, M. (2020). EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis [computer software]. EPPI-Centre Software. London: UCL Social Research Institute:

Thomas, M., Cinganotto, L., & Heike, P. (2018). Digital Game-Based Language Learning in 3D Immersive Environments: The GUINEVERE Project

Thomas, S. A., Walker, M. F., Macniven, J. A., Haworth, H., & Lincoln, N. B. (2013). Communication and Low Mood (CALM): a randomized controlled trial of behavioural therapy for stroke patients with aphasia. Clinical Rehabilitation, 27(5), 398-408. 10.1177/0269215512462227

Thompson, C. K. (2007). Complexity in language learning and treatment. American Journal of Speech-Language Pathology, 16(1), 3-5. 16/1/3

Thompson, C. K. (1989). Generalization research in aphasia: A review of the literature. Clinical Aphasiology, 18, 195-222.

Thompson, C., Choy, J. J., Holland, A., & Cole, R. (2010). Sentactics®: Computer-automated treatment of underlying forms. Taylor & Francis Ltd. 10.1080/02687030903474255

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The Role of Syntactic Complexity in Treatment of Sentence Deficits in Agrammatic Aphasia: The Complexity Account of Treatment Efficacy (CATE). *Journal of Speech, Language, and Hearing Research, 46*(3), 591-607. 10.1044/1092-4388(2003/047)

Thorndike, E. L. (1927). The law of effect. The American Journal of Psychology, 39, 212-222. 10.2307/1415413

Togher, L., Wiseman-Hakes, C., Douglas, J., Stergiou-Kita, M., Ponsford, J., Teasell, R., Bayley, M., & Turkstra, L. S. (2014). INCOG recommendations for management of cognition following traumatic brain injury, part IV: Cognitive communication. The Journal of Head Trauma Rehabilitation, 29(4), 353-368.

Townend, E., Brady, M., & McLaughlan, K. (2007). Exclusion and inclusion criteria for people with aphasia in studies of depression after stroke: a systematic review and future recommendations. Neuroepidemiology, 29(1-2), 1-17. 000108913

Trebilcock, M., Shrubsole, K., Worrall, L., & Ryan, B. (2021). Development of an online implementation intervention for aphasia clinicians to increase the intensity and comprehensiveness of their service. Disability and Rehabilitation, 1-10. 10.1080/09638288.2021.1910867

Tucker-Drob, E. M. (2011). Individual differences methods for randomized experiments. Psychological Methods, 16(3), 298-318. 10.1037/a0023349

van der Meulen, I., van de Sandt-Koenderman, W. M., Duivenvoorden, H. J., & Ribbers, G. M. (2010). Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. International Journal of Language & Communication Disorders, 45(4), 424-435. 10.3109/13682820903111952

van der Meulen, I., van de Sandt-Koenderman, W. M. E., Duivenvoorden, H. J., & Ribbers, G. M. (2010). Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. International Journal of Language & Communication Disorders, 45(4), 424-435. 10.3109/13682820903111952

Varley, R. (2011). Rethinking aphasia therapy: A neuroscience perspective. International Journal of Speech-Language Pathology, 13(1), 11-20. 10.3109/17549507.2010.497561

Vestal, L., Smith-Olinde, L., Hicks, G., Hutton, T., & Hart Jr, J. (2006). Efficacy of language assessment in Alzheimer's disease: comparing in-person examination and telemedicine. Clinical Interventions in Aging, 1(4), 467.

Vickers, C. P. (2010). Social networks after the onset of aphasia: The impact of aphasia group attendance. Aphasiology, 24(6-8), 902-913. 10.1080/02687030903438532

Volkmer, A., Spector, A., Swinburn, K., Warren, J. D., & Beeke, S. (2021). Using the Medical Research Council framework and public involvement in the development of a communication partner training intervention for people with primary progressive aphasia (PPA): Better Conversations with PPA. BMC Geriatrics, 21(1), 642. 10.1186/s12877-021-02561-8

Voorberg, W. H., Bekkers, V. J. J. M., & Tummers, L. G. (2015). A Systematic Review of Co-Creation and Co-Production: Embarking on the social innovation journey. Public Management Review, 17(9), 1333-1357. 10.1080/14719037.2014.930505

Wadams, A., Suting, L., Lindsey, A., & Mozeiko, J. (2022). Metacognitive Treatment in Acquired Brain Injury and Its Applicability to Aphasia: A Systematic Review. Frontiers in Rehabilitation Sciences, 3

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2016). Remote Neuropsychological Assessment in Rural American Indians with and without Cognitive Impairment. Archives of Clinical Neuropsychology, 31(5), 420-425. 10.1093/arclin/acw030

Wallace S, J., Bridget, S., Rose, T. A., Worrall, L., Le Dorze, G., & Shrubsole, K. (2021). Core Outcome Set Use in Poststroke Aphasia Treatment Research: Examining Barriers and Facilitators to Implementation Using the Theoretical Domains Framework. Journal of Speech, Language, and Hearing Research, 64(10), 3969-3982. 10.1044/2021_JSLHR-20-00683

Wallace, S. J., Worrall, L., Rose, T. A., Alyahya, R. S. W., Babbitt, E., Beeke, S., de Beer, C., Bose, A., Bowen, A., Brady, M. C., Breitenstein, C., Bruehl, S., Bryant, L., Cheng, B. B.

Y., Cherney, L. R., Conroy, P., Copland, D. A., Croteau, C., Cruice, M., . . . Dorze, G. L. (2022). Measuring communication as a core outcome in aphasia trials: Results of the ROMA-2 international core outcome set development meeting. *International Journal of Language & Communication Disorders, n/a*10.1111/1460-6984.12840

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A. C., Marshall, J., . . . Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. International Journal of Stroke : Official Journal of the International Stroke Society, 14(2), 180-185. 10.1177/1747493018806200

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A., Marshall, J., . . . Webster, J. (2018). A core outcome set for aphasia treatment research: the ROMA consensus statement. SAGE Publications.

Wallace, S. J., Worrall, L., Le Dorze, G., Brandenburg, C., Foulkes, J., & Rose, T. A. (2022). Many ways of measuring: a scoping review of measurement instruments for use with people with aphasia. Aphasiology, 36(4), 401-466. 10.1080/02687038.2020.1836318

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Kirke, E., & Kolomeitz, D. (2018). Report from ROMA: an update on the development of a core outcome set for aphasia research. Aphasiology, 32(sup1), 241-242. 10.1080/02687038.2018.1487020

Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2017). Which treatment outcomes are most important to aphasia clinicians and managers? An international e-Delphi consensus study. Aphasiology, 31(6), 643-673. 10.1080/02687038.2016.1186265

Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop "fishing"? A review of generalisation following aphasia intervention. Aphasiology, 29(11), 1240-1264. 10.1080/02687038.2015.1027169

Whitworth, A., & Perkins, L. (1997). Conversation analysis profile for people with aphasia (CAPPA). Whurr Publishers Ltd.

Whitworth, A., Webster, J., & Howard, D. (2013). A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia: A clinician's guide (2nd ed.). Psychology Press.

WHO. (2012). The World Health Organization Quality of Life (WHOQOL). World Health Organisation. Available at: https://www.who.int/publications/i/item/WHO-HIS-HSI-Rev.2012.03 (Accessed: 14.12.22)

Wight, D., Wimbush, E., Jepson, R., & Doi, L. (2016). Six steps in quality intervention development (6SQuID). Journal of Epidemiology and Community Health, 70(5), 520-525. 10.1136/jech-2015-205952

Wilkinson, R., & Wielaert, S. (2012). Rehabilitation targeted at everyday communication: can we change the talk of people with aphasia and their significant others within conversation? Archives of Physical Medicine and Rehabilitation, 93(1 Suppl), 70. 10.1016/j.apmr.2011.07.206

Williams, K., Moffatt, K., McCall, D., & Findlater, L. (2015). Designing Conversation Cues on a Head-Worn Display to Support Persons with Aphasia. Paper presented at the 231-240. 10.1145/2702123.2702484

Wilson, S., Roper, A., Marshall, J., Galliers, J., Devane, N., Booth, T., & Woolf, C. (2015). Codesign for people with aphasia through tangible design languages. CoDesign, 11(1), 21-34. 10.1080/15710882.2014.997744

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

Woolf, C., Caute, A., Haigh, Z., Galliers, J., Wilson, S., Kessie, A., Hirani, S., Hegarty, B., & 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.

Woollams, A. M. (2012). Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe. Frontiers in Human Neuroscience, 6, 85. 10.3389/fnhum.2012.00085

World Health Organization. (2001). International classification of functioning, disability and health : ICF. World Health Organization; World Health Organization.

Worrall, L., Ryan, B., Hudson, K., Kneebone, I., Simmons-Mackie, N., Khan, A., Hoffmann, T., Power, E., Togher, L., & Rose, M. (2016). Reducing the psychosocial impact of aphasia on mood and quality of life in people with aphasia and the impact of caregiving in family members through the Aphasia Action Success Knowledge (Aphasia ASK) program: study protocol for a randomized controlled trial. Trials, 17, 153. 10.1186/s13063-016-1257-9

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. Aphasiology, 25(3), 309-322. 10.1080/02687038.2010.508530

Worrall, L. E., PhD, Hudson, K., PhD, Khan, A., PhD, Ryan, B., PhD, & Simmons-Mackie, N., PhD. (2016). Determinants of living well with aphasia in the first year post stroke: a prospective cohort study. Archives of Physical Medicine and Rehabilitation, 10.1016/j.apmr.2016.06.020

Wray, F., Clarke, D., Cruice, M., & Forster, A. (2021). Development of a self-management intervention for stroke survivors with aphasia using co-production and behaviour change theory: An outline of methods and processes. PloS One, 16(11), e0259103. 10.1371/journal.pone.0259103

Wray, F., Clarke, D., & Forster, A. (2018). Post-stroke self-management interventions: a systematic review of effectiveness and investigation of the inclusion of stroke survivors with aphasia. Disability and Rehabilitation 40(11), 1237-1251. 10.1080/09638288.2017.1294206

Wray, F., Clarke, D., & Forster, A. (2020). "Guiding them to take responsibility": exploring UK speech and language therapists' views of supporting self-management of aphasia. Disability and Rehabilitation, 34(4), 411-430. 10.1080/02687038.2019.1697865

Yarwood, J. (2015). The Most Beautiful Video Games Inspired by Famous Artists. Vice. Available at: https://www.vice.com/en/article/qbxm9d/the-most-beautiful-video-games-inspired-by-famous-artists-211 (Accessed: 17.12.21)

Yasuda, K., Beckmann, B., & Nakamura, T. (2000). Brain processing of proper names. Aphasiology, 14(11), 1067-1089. 10.1080/02687030050174638

You Sung, H., Jang, S. H., Kim Yun-Hee, Mark, H., Ahn, S. H., Kwon Yong-Hyun, Kim, J. H., & Young, L. M. (2005). Virtual Reality–Induced Cortical Reorganization and Associated Locomotor Recovery in Chronic Stroke. Stroke, 36(6), 1166-1171. 10.1161/01.STR.0000162715.43417.91

Zhang, Y., Chen, P., Li, X., Wan, G., Xie, C., & Yu, X. (2017). Clinical research on therapeutic effect of virtual reality technology on Broca Aphasia patients. Paper presented at the 2nd International Conference on Information Technology (INCIT), 1-5. 10.1109/INCIT.2017.8257880