Computer Gesture Therapy for Adults with Severe Aphasia

Abigail Heidi Roper

Thesis submitted for the degree of Doctor of Philosophy in Language and Communication Science

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

March 2017

Dedication

For my Nan and Granddad, Rose and Phil Roper

Table of Contents

List of Tables

List of Figures

Acknowledgements

Thank you to everyone who supported me to complete this piece of work.

The following is a non-exhaustive list of some of those who have come with me along the way:

- My wonderful supervisors Jane Marshall and Stephanie Wilson.
- The endlessly supportive Jo Piper and Alasdair Eaglestone who provided me with both food and emotional nutrition.
- My long-suffering and ever-patient housemates Simon Dyson and Ed Cole.
- My family, who provided continuous encouragement throughout the journey, and who hosted, fed, watered and funded me throughout the data collection process in the South West.
- Stroke Association communication support workers who helped me to recruit participants for this project.
- The excellent technicians Robert Davey and Colin Day who supported the logistical aspects of video presentation during the blinded gesture scoring sessions. And, the 40 or so speech and language therapy students who donated their time to score said videos.
- My very supportive PhD colleagues including Dr. Anneline Huck, Judith Kistner, Dr. Madeline Pritchard, Ioanna Georgiadou, Becky Moss, Helen Cain, Dr. Nick Kingston, Livia Bernadi and Lisa –Maria Muller.
- My crack team of last-minute proof readers: Charlotte Bennett, Jen Ryder, Lucy Dyson, Shabana Jiwaji, Fergus Campbell, Vicki Flood, Helen Burgess and Ryc Aquino.
- Each and every participant, their friends and family members who welcomed me into their homes and donated their time and energy to take part this project.
- City University London for funding me to undertake this PhD studentship.

Declaration

Abstract

Aphasia intervention has made increasing use of technology in recent years. The evidence base, which is largely limited to the investigation of spoken language outcomes, indicates positive treatment effects for people with mild to moderate levels of aphasia. Outcomes for those with severe aphasia however, are less well documented and - where reported - present less consistent gains for measures of spoken output. In light of this issue for existing approaches, and due to the fact that non-speech focussed interventions might therefore be more suitable, the current thesis explores the use of computer gesture therapy for people with severe aphasia. An initial review of gesture therapy is presented, followed by a systematic review of current computer therapy literature. A pseudo-randomised, wait-list control study of twenty participants with severe aphasia forms the experimental body of the thesis. The study investigates the effects of two purpose-built gesture therapy technologies: GeST and PowerGeST. The latter of these was developed for the purposes of the thesis. Following completion of a range of candidacy measures examining gesture comprehension, language, cognition and praxis, participants undertook a five-week intervention period comprising practice with GeST and PowerGeST. Primary outcomes were assessed using a measure of gesture production in isolation. Secondary outcome measures included an assessment of naming production, a novel assessment of interactive gesture abilities and an accessible computer use and confidence measure. These two latter measures were developed for the purposes of the thesis. Study outcomes show significant improvement in gesture production abilities for adults with severe aphasia following computer intervention. They indicate no transfer of effects into naming gains or interactive gesture. Findings reveal comparatively low levels of access to everyday technologies for this group. Outcomes therefore, indicate the positive effects of a purpose-built computer-delivered therapy for a population who commonly experience challenges with access to everyday technology. Insights gained within this thesis offer encouraging results for computer therapy methods within this hitherto under-researched population and propose a case for further development of the evidence base in this field.

Chapter 1. Introduction

Aphasia limits the ability to express and/or understand spoken and/or written language. It occurs as a result of brain damage, most commonly stroke. Estimates reported at the time of writing suggest that there are more than 1.2 million stroke survivors living in the UK alone (The Stroke Association, 2016) and that around 45% will experience aphasia as a result (Ali, Lyden, & Brady, 2015). Individuals with aphasia find that they are no longer able to make use of language as they did before their brain damage occurred. Those with severe aphasia may find themselves almost entirely unable to speak.

Speech and language therapists aim to help people with aphasia to re-establish or compensate for those linguistic abilities affected. Methods for compensation might include writing, drawing or gesture. For those with severe aphasia, therapy focusing on gesture offers an opportunity to communicate a concept that they are unable to express verbally. There are also indications that therapies involving gesture may help to stimulate speech (M. L. Rose, Attard, Mok, Lanyon, & Foster, 2013). Therapy however, requires time and repetition to be most effective (Kelly, Brady, & Enderby, 2010), although therapist resources are often limited (Enderby & Petheram, 2002).   Service providers are therefore increasingly looking to extend and maximse the effectiveness of the treatments available using non-traditional means.

It is now widely accepted that computer-based exercises offer an opportunity to raise therapy dose, increase autonomy and address a range of language functions impaired by aphasia (van de Sandt-Koenderman, 2004; R. Varley, 2011). There is growing evidence regarding the use and effectiveness of computer-delivered therapy approaches for spoken and written language modalities (Zheng, Lynch, & Taylor, 2016). However, there is currently limited work which investigates computer therapy in relation to severe aphasia - particularly in relation to non-linguistic, compensatory strategies such as gesture.

Within the above context, this thesis examines computer therapy for aphasia. Specifically, it examines computer gesture therapy for adults with severe aphasia. It explores the established evidence base surrounding offline gesture therapy and online computer therapy and presents a new study of a previously established computer gesture intervention (GeST - Marshall et al., 2013), examined in depth to explore its effects for naming, gesture and technology use for a larger group of users with severe aphasia.

The objective of this research is to develop the evidence base regarding the effects of computer intervention for adults with severe aphasia. This thesis first introduces the background to those topics central to the research (aphasia, gesture, intervention, computer intervention - chapter 2). Chapter 3 presents a systematic literature review of evidence regarding computer-delivered therapy for gesture and spoken language in aphasia. Chapter 4 introduces details of a previous research project which piloted a computer gesture therapy for aphasia (GeST) and which forms the background to the present study. Next, chapter 5 presents details of the particular computer gesture therapy whose effects will subsequently be examined (GeST + PowerGeST). Chapter 6 introduces specific therapeutic and outcome measure components developed for the purpose of the study reported here and describes the process behind their development. These components comprise a novel computer therapy adjunct - created to support extended levels of therapy practice, and two novel assessment tools – developed to examine the effects of the described therapy upon both interactive gesture and also technology use and confidence. Chapter 7 reports the conduct of the pseudo-randomised waitlist control group study that forms the central contribution of this thesis. This study aimed to assess the effects of the described computer gesture therapy in terms of both gestural and spoken performance at a single item level as well as gesture at an interactive communication level. Further investigations also examined the effects of computer gesture therapy upon levels of use and confidence-in-use of a series of items of everyday technology. Results for the above study are presented in chapter 8. The thesis concludes with a discussion of the study's outcomes and limitations and their implications in relation to existing evidence in the field of computer therapy for severe aphasia (chapter 9). Findings are further discussed within the wider context of aphasia, technology and clinical practice more generally.

Specific contributions of this thesis are as follows:

- A systematic review of literature pertaining to computer-delivered therapy for gesture and spoken language deficits in aphasia
- Further evaluation of an existing computer gesture therapy tool for adults with severe aphasia (GeST)
- The development of a novel computer therapy adjunct for gesture practice by adults with severe aphasia (PowerGeST)
- The development of an assessment of gesture in interaction with a participant's regular communication partner
- The development of an assessment of technology use and confidence – specifically created to be accessible to adults with severe aphasia
- An examination of the use of technology by adults with severe aphasia in relation to other adults with and without stroke
- An examination of the effects of the described computer gesture therapy on:
    - Gesture production in isolation
    - Spoken picture naming
    - Gesture production in interaction with a regular communication partner
    - Wider technology use and confidence
- An exploration of prognostic factors relating to the scale of therapeutic outcomes

Crucially, this thesis reports methods for the delivery of a successful and largely autonomous computer therapy for adults who have severe aphasia. Furthermore – with a focus on those with severe aphasia, it presents evidence about the effects of computer therapy for a section of the aphasia population whose technology needs are presently relatively poorly understood. In a context where there is growing emphasis on the provision of healthcare support through digital means (Lane Fox, 2015), the evidence presented herein provides a valuable contribution to the discussion for a largely under-represented and ‘voiceless’ population.

Chapter 2. Background literature review

2.1 Introduction

To introduce the context for this work, the main themes pertaining to this research will now be presented. This chapter will first describe the population of people with aphasia before examining evidence regarding executive function in aphasia and specifically, severe aphasia. Following the discussion of severe aphasia, the field of gesture will be introduced – with a subsequent section on aphasia intervention, gesture intervention and the associations between aphasia, gesture and limb apraxia. A summary of Rose's 2013 systematic literature review of gesture therapy is presented, followed by a review of subsequent relevant gesture literature and then an introduction to computer intervention for aphasia alongside selected descriptions of existing computer interventions.

2.2 Aphasia

Aphasia is a language impairment caused by brain injury. Early descriptions of the condition reported impairments of spoken language production (Broca, 1861) and also of spoken language comprehension and written comprehension and production (Wernicke, 1874). Aphasia limits the ability to express and/or understand spoken and/or written language. People affected are no longer able to make use of language in the same way they did before the brain damage occurred. Aphasia most commonly occurs as a consequence of brain injury sustained through cerebral vascular accident (more commonly referred to as stroke), although it can also arise as a result of traumatic brain injury or degenerative neurological conditions. Estimates suggest that around 45% of those who experience a stroke will also experience aphasia as a result, with 24% going on to experience symptoms which persist for at least 3 months (Ali et al., 2015).

Aphasia exists in many forms and those affected may experience difficulties with one modality in isolation (for example verbal expression or reading comprehension) or multiple or all modalities being affected. Wernicke identified the notion of distinct and isolated deficits in his 1874 report "The Aphasic Symptom-Complex". Individual facets may be impaired to differing extents; with the ability to name spoken objects seriously impaired for example, and yet the ability to understand spoken words and sentences relatively spared. Alternatively, individuals may be affected in the opposite fashion with severe difficulties in understanding in contrast to a relatively mild impairment of word production. Naming deficits however, are perhaps the most common defining feature of the condition and are evident to a greater or lesser extent within most presentations.

A key identifying feature of aphasia is the relative preservation of cognitive skill, with those affected typically retaining abilities of reasoning, object and real world knowledge and understanding and sensory perception. However, in spite of this retained knowledge, people with aphasia are commonly left significantly impaired in their abilities to share and express ideas, thoughts and needs with others using the linguistic means to which they are accustomed (Mayo Foundation for Medical Education and Research, 2015). Moreover, whilst aphasia itself is not a direct impairment of cognition, as a result of stroke comorbidities, some individuals affected may demonstrate some associated cognitive difficulties. Details of this relationship between cognition (specifically executive function) and aphasia will be discussed further in section 2.2.2 Executive Function and Aphasia.

2.2.1 Severe aphasia

As previously identified, aphasia can affect language components to differing extents. Symptoms can be classified along a spectrum from mild to severe.

The level of severity may be classified clinically by a qualified speech and language therapist (e.g in Parr, 2007), or through the use of either a full battery of assessments which comprise a standardised assessment measure (for example the Western Aphasia Battery Aphasia Quotient (Kertesz, 1982) or the Aachen Aphasia Test (Huber, Poeck, & Willmes, 1984)) or a given subtest from a standard measure, which is taken to represent a key feature of the condition such as a naming deficit. Caute et al (2013) and Marshall et al (2013 identified severe participants as those who scored below 20% on the spoken naming subtest of the Comprehensive Aphasia Test (Swinburn, Porter, & Howard, 2004). This final classification approach has been adopted for the purposes of the experimental research presented within the current thesis – due to the close relevance of the two afore-mentioned studies to the topics being investigated here.

To the observer, individuals with severe aphasia typically demonstrate very limited spoken ability and are likely to “experience major difficulties with communication as a result of aphasia” (Parr, 2007). The range of classification methods presented demonstrates that there is presently no single agreed standard by which to identify severe aphasia. What we can state with certainty however, is that it will have a profound effect on an individual’s ability to express and understand language.

2.2.2 Executive Function and Aphasia

The topic of associations and distinctions between performance on cognitive assessment and performance on language assessment has been extensively investigated in the field of aphasia. Of particular relevance to this thesis is the study of a range of skills referred to under the umbrella term “executive function” (EF).   Purdy (2002) described EF as a range of skills that allow us to flexibly manage goal-directed activities through the use of planning, sequencing, organisation and monitoring. Purdy compared performance on a number of recognised EF assessments (specifically the Porteus Maze Test – Porteus, 1959, the Wisconsin Card Sorting Test – Grant & Berg, 1993 and the Tower of London task – Shallice, 1982) for a group of 12 people with aphasia to a group of 15 people without. She reported a significant reduction in the speed and efficiency of task performance for the aphasic group and attributed this reduction to decreased levels of EF in those individuals with aphasia - ascribing these performance deficits to difficulties with cognitive flexibility and planning. Of note in this study is that the range of "nonverbal" measures chosen reflects those that have typically been used with a non-aphasic population (i.e. those with a typical competence in language use). Purdy reported adaptations to the existing tasks to make the instructions accessible to the aphasic participants.

Using an alternative range of measures of EF with 13 aphasic participants – developed to particularly cater to those with aphasia - Helm-Estabrooks (2002) reported age-typical EF performance in two of the 13 participants tested and a below typical performance for the remaining 11. Assessments implemented comprised tasks of symbol cancellation, symbol trails, design memory and mazes. These tasks aim to assess a number of facets including visual attention and perception, working memory and mental flexibility and the executive functions involved in planning a course of action, rejecting/inhibiting incorrect choices, and correcting mistakes when made. When comparing the outcomes of these measures to performance on a range of language measures (memory of personal facts, confrontation naming, generative naming, story telling and paragraph comprehension), Helm-Estabrooks additionally found no significant association between the scale of linguistic deficit observed and the scale of EF deficit observed. Both studies (Helm-Estabrooks; and Purdy) point towards some reduced level of performance on EF tasks for a number of individuals with aphasia but do not indicate a clear profile of specific aphasic characteristics that might be correlated with these difficulties.

Further investigations with adults with severe aphasia additionally support Helm-Estabrooks's finding that the scale of linguistic deficit (or aphasia severity) is not the sole contributor to poor performance in tests of EF. These are reported in the following section.

2.3 Severe Aphasia and Executive Function

Examples of three different EF skills investigated in relation to severe aphasia come from R. A. Varley, Klessinger, Romanowski, & Siegal (2005), Bek, Blades, Siegal, & Varley (2010) and Marshall et al.(2013). R. A. Varley et al. (2005) reported an ability to carry out “basic computational procedures” of mathematical calculation in three individuals with severely impaired grammatical ability due to aphasia when tested on a range of algebraic functions. Bek et al. (2010) found that a similarly impaired group of aphasic adults demonstrated an intact ability to re-orientate themselves to a physical object following systematic disorientation (namely being blindfolded and turned around a number of times). Marshall et al. (2013) found that, during observation, four of nine participants with severe expressive aphasia demonstrated difficulties in navigating between levels of a computer delivered gesture therapy. The above suggests that whilst computational procedures and physical orientation may remain intact, some individuals with aphasia may experience difficulties with ‘navigation’ through a digital landscape which are not present for navigation within a physical landscape. Further investigation is warranted to establish how the EF skills described might impact upon users’ ability to carry out independent intervention activities using a computer-delivered therapy. Moreover, investigations may reveal if measures of EF can be used to effectively predict who will benefit from an intervention of this format.

2.4 Severe Aphasia and Gesture

2.4 Gesture

In cases of severe aphasia, individuals may find that they are no longer able to produce any spoken or written output, or that their utterances are limited to a very small set of words or non-words. In such cases, speech and language therapists may advocate the use of gesture to compensate for the spoken language that is no longer available (Marshall, 2006).

2.4.1 Gesture

As a context for discussions of gesture production, it is necessary to understand more about gesture itself and the way in which it is understood in the context of this thesis. It is worth noting that the field of human-computer interaction and the field of human communication science use of the term gesture to represent two different concepts. Within human computer interaction, the term gesture is now very commonly used to refer to a physical interaction with a computer interface. An example of such an interaction might be the use of a finger pinch movement on a touch screen to allow a user to zoom in and out of an image. Within communicative science however, the term gesture is most commonly applied in reference to a physical action used to convey some form of expression (Kendon, 2000). It is this second definition that I shall employ within the subsequent discussion. The following sets out a brief description and framework for understanding gesture. This sets the scene for subsequent exploration of the specific variant of gesture being treated within the current study.

Communicative gesture literature makes use of a widely adopted framework of gesture classification first introduced by Kendon (1980) and subsequently updated by McNeill (1992). Kendon and McNeill classify gesture types according to a continuum – described in Figure 1.

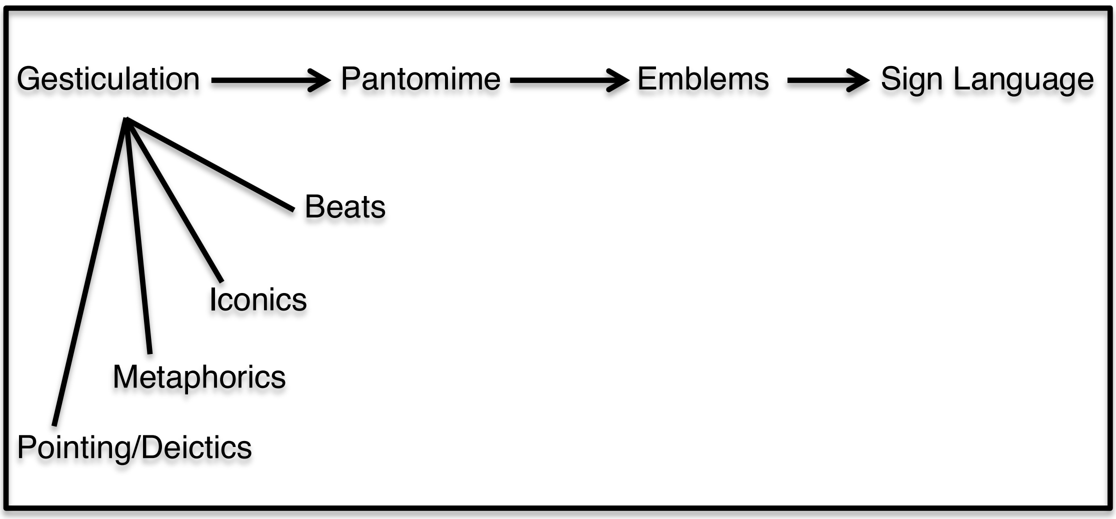


Figure 1. Continuum of gesture classification as proposed by Kendon (1980) and updated by McNeill (1992)

Using this framework, McNeill (2006) classifies gestures as follows:

Gesticulation

In broad terms, this classification refers to a movement that accompanies speech.

McNeill further delineates gesticulation into four sub-categories:

1. Beats – where a hand appears to beat out time according to the cadence of an utterance.
2. Iconics – a depiction or physical manipulation of an imagined concrete entity or action being referred to within the accompanying speech.
3. Metaphorics – a description or manipulation of an imagined metaphoric entity being referred to within the accompanying speech.
4. Pointing or Deictics – the act of pointing (with a finger or any other physical component) towards a real or metaphoric entity being referred to within the accompanying speech.

Pantomime

In contrast to gesticulation, pantomimes are often produced in the absence of speech. They depict objects or actions and can be presented in sequence to depict a narrative. Pantomime gestures may be accompanied by vocal (but nonverbal) gestures – for example the production of an engine noise when producing a car pantomime. They may additionally stand alone to represent one entity or action.

Emblems

Where pantomimes may be idiosyncratic in style, emblems are conventionalized within a culture. Emblems represent accepted gestural symbols of a given concept, for example – the thumbs up gesture to indicate that something is good, or the touching of the tip of thumb and forefinger on one hand to represent “OK”. These gestures can occur with or without speech.

Sign Language

Signs are lexical items in their own right. They exist in a linguistic construct and hold a specific meaning in the same way as a spoken word. They conform to grammatical constraints and can be produced in syntactically governed sequences. They can convey both concrete and abstract ideas. Sign languages are largely used in the absence of speech, typically within deaf communities, and require both the person delivering the message and the person receiving the message to have knowledge of that given sign language. Further evidence of the separate linguistic status of sign languages comes from Marshall, Atkinson, Smulovitch, Thacker, & Woll (2004) who demonstrate a dissociation between gesture and sign abilities in a sign language user with aphasia.

The above descriptions give a sense of the types of gestures that might be targeted within the sphere of gesture therapy. The gestures that will be utilised within the computer therapy subsequently examined within this thesis (listed in Appendix A) can be classified as pantomime gestures. Such gestures can be produced in the absence of spoken language and do not necessarily pre-suppose prior cultural or linguistic knowledge on the part of the individual observing and interpreting the gesture.

2.5 Intervention

2.5.1 Speech and Language Therapy and Aphasia

As described within section 2.2, the implications of aphasia are many and varied. Speech and language therapy aims to address the challenges of aphasia through a number of methods. The Royal College of Speech and Language Therapy Resource Manual (Enderby et al., 2009) identifies four therapy strategies:

1. **Impairment** – aiming to rehabilitate language skills lost through aphasia - for example the ability to produce spoken words and sentences.

2. **Activity** - aiming to support those with aphasia to employ compensatory and alternative means to convey or interpret language - for example through the use of drawing, gesture or other alternative and augmented means of communication.

3. **Participation** - aiming to support individuals to access aspects of their life, which may have changed as a result of their aphasia – for example, engagement with a social circle, access to employment or education.

4. **Well being** - aiming to help people with aphasia and their families understand the condition and the ways in which communication can be adapted to accommodate language impairments.

The topics are couched against the World Health Organisation's International Classification of Functioning, Disability and Health (World Health Organization, 2001) – a framework commonly used to conceptualise medical and social approaches to healthcare.

2.5.2 Principles of aphasia intervention

The field of aphasia intervention has benefitted in recent years from a growing evidence base. An increasing number of approaches now exist which seek to address many of the myriad aspects of language that can be variously affected by the condition.

Kelly et al, (2010) conducted a Cochrane review of "speech and language therapy for aphasia following stroke". From this, the authors found that the effects of therapy seem most beneficial for functional communication (including spontaneous speech and daily communicative activities) and expressive language (including skills of picture naming, writing and spoken repetition) with less available evidence to support therapy for receptive language skills (including comprehension of auditory and written language and also gesture). Furthermore, they argue that the strongest evidence for the effectiveness of therapy is provided by studies of intensive regimes, as opposed to regimes that deliver the same dose over an extended period. Aside from a general factor of intensity however, authors note insufficient evidence amongst the current literature to effectively advocate one specific therapy approach over another.

Kelly et al.'s review comprises a very strictly selected set of literature as the Cochrane review process includes only those articles with a randomised control trial (RCT) design. Current aphasia intervention literature comprises numerous studies that do not conform to this design however. For this reason, there is a large body of evidence available for review that will not have been included in Kelly et al.'s report.

The relative infancy of the speech and language therapy evidence base does not lend itself to the extensive production of a large body of RCT work. Indeed, Garret and Thomas (2006) argue in support of systematic reviews of literature that take into account both RCT designs as well as quasi-experimental and single-subject designs. Bhogal et al (2003) conducted such a review of the intensity of aphasia therapy amongst a broader range of clinical trials – including those with a quasi-experimental design. Having established a base of 10 studies reporting on a total of 864 participants with aphasia, authors found that studies reporting positive treatment effects demonstrated a significantly higher number of hours of total therapy (*M*=98.4 hours, *S.D.*=28.2) and a significantly higher number of hours per week (*M*=8.8 hours, *S.D.*=2.0) than those studies reporting no treatment effect (Total *M*=43.6, *S.D.*=8.3; Weekly *M*=2.0, *S.D* Not reported). Bhogal et al.'s research provides evidence for the fact that aphasia intervention works and that high dose, intensive therapy over a restricted period of time results in significant improvements in therapy outcomes.

2.5.3 Gesture Intervention

We shall now consider gesture intervention. Using the framework outlined by Enderby et al. in 2009, gesture intervention can be seen to accord with the *Activity* principle alongside drawing and other compensatory methods such as augmentative communication aids. Gesture offers certain advantages above other compensatory approaches. It does not require the use of any additional equipment such as pen and paper or hi-tech or lo-tech communication aids. Another advantage of this lack of equipment is that it aligns the user with other verbal communication partners who will also make use of gesture within their typical conversational exchanges (De Ruiter, 2000). While some individuals with aphasia may employ gestures spontaneously as a communication strategy, others may need therapeutic support to do so (Marshall, 2006). Gesture may be inhibited by a generalised impairment in the employment of symbols (as described in Rose, 2006) or by motor difficulties such a limb weakness (hemiparesis) or paralysis (hemiplegia) on one side – a common sequela of stroke. Due to the location of the neurological injury, individuals with aphasia may experience weakness or paralysis in the right arm and hand – very often the limb that has previously been dominant for writing and other fine motor use. This consequence leads to an additional complication that any compensatory activity, be it writing, drawing or gesture must now be carried out by the non-dominant left hand. Rose (2006) further highlights an additional factor that can impair gesture production in aphasia, namely, the presence of limb apraxia.

2.6 Gesture, Aphasia and Limb Apraxia

This section reports two studies of severe aphasia and gesture. Both studies additionally address a topic not yet discussed in detail – that of limb apraxia. Limb apraxia is a disorder of purposeful movement of the limb (Cubelli, Marchetti, Boscolo, & Della Sala, 2000). It is a relatively common sequela of left hemisphere stroke, occurring in an estimated 28-37% of cases

(Donkervoort, Dekker, van den Ende, Stehmann-Saris, & Deelman, 2000). As mentioned earlier, Rose (2006) acknowledges this as a potential inhibiting factor for gesture performance in aphasia. The first study reported here (Caute et al., 2013; Marshall et al., 2012) describes details of a gesture intervention study and its findings in relation to apraxia. The second (Hogrefe, Ziegler, Weidinger, & Goldenberg, 2012), examines the topic of limb apraxia in aphasia more directly.

Marshall et al. (2012) and Caute et al. (2013) report details of an intervention study where 14 individuals with severe expressive aphasia received a programme of pantomime gesture and naming therapy delivered by a speech and language therapist. Participants were assessed both before and after intervention to identify the presence of any change in gesture and naming abilities. Significant improvement in both modalities was observed with changes being most marked in the sphere of spoken naming. Gestures were assessed both in isolation - in response to a confrontation gesturing task and also within a more embedded context – in response to an interactive message and story delivery task. Participants were explicitly advised to make use of gesture in their responses and outcome assessments followed a period of fairly extensive and explicit gesture therapy. Following analysis of outcomes, Marshall et al found that limb apraxia scores on a customized version of the Apraxia Battery for Adults (Dabul, 2000) were *not* predictive of the size of gains observed in participants' gesture assessments. However, authors acknowledge that this may be attributed to the relatively small number of participants and recommend that the issue should be further addressed in larger scale studies of gesture intervention.

Hogrefe et al. (2012) examined the impact of limb apraxia on the spontaneous gesture abilities of people with severe aphasia. Using a series of assessments, 24 people with aphasia and "highly limited output" were instructed to re-tell a series of short video clips commonly used in gesture analysis research (namely animated cartoon clips featuring the Looney Tunes character “Tweety Pie” and video clips from the British comedy television series “Mr Bean”). Upon viewing the clips, participants were advised to: “Retell the story in a vivid manner, so that someone who had not seen the video could infer what the story was about”. These story re-tells were video recorded and analysed for the diversity of hand gestures used and the comprehensibility of the participants’ gestural communication more generally. The authors found that performance on measures of gestural diversity and comprehensibility did not correlate with scores on standard aphasia assessments (The Aachen Aphasia Test, Huber et al., 1984). However, scores of limb apraxia (assessed through a pantomime-to-command task: Goldenberg, Hartmann, & Schlott, 2003; Goldenberg, Hermsdorfer, Glindemann, Rorden, & Karnath, 2007) could be used to predict comprehensibility. Participants in this study varied in the length of time post-stroke. Whilst a number could be classified as within the chronic stages of aphasia, 11 out of 24 were less than 6 months post-onset. Treatment studies within this field typically identify participants as chronic only when they have been living with a diagnosis of aphasia for 6 months or more. For this reason, the cohort may not be directly comparable to other groups reported. The authors here assert that a breakdown of language alone will not be paralleled by a breakdown of equal measure in the ability to produce gesture. It should be noted that the gestures assessed here were produced within a discourse and hence reflect an embedded production. Additionally – and in contrast to the prior study - the instructions given to participants did not request that they made specific use of gesture, rather that they re-tell the video clip “in a vivid manner”. For these reasons, the experiment reflects a relatively naturalistic employment of gesture production.

The above two studies provide a mixed picture regarding the involvement of limb apraxia in severe aphasia - indicating a need for more evidence exploring the presence and effects of limb apraxia for this specific population.

2.7 Systematic Review of Gesture Treatments for Aphasia

Returning now to the wider topic of gesture therapy for individuals with aphasia, we can explore what is known about the existing evidence base for this topic by examining a systematic review of the literature, conducted by Rose et al. (2013). Information from individual articles included in this review has been summarised in Table 1. A further summary of findings will also be provided in the following text before specific studies of interest to the present thesis are examined in more detail.

Authors found that from a total of 177 citations relating to the use of gesture therapy in aphasia intervention, 23 adhered to the specified inclusion criteria they had identified (briefly, that they were peer-reviewed; contained original data regarding only aphasia arising from stroke; included gesture and incorporated a group or single case experimental design). Of these 23 studies, four were group designs and 19 single case experimental designs. Final data included reported upon outcomes from a total of 134 participants with aphasia - most (but not all) of whom had a diagnosis of non-fluent, moderate to severe chronic aphasia. Rose and colleagues report that the typical gesture treatment reviewed was administered on average 2-3 times per week for an average of 11.2 hours (range 6.5 – 32 hours). The reviewed studies employed two rather different approaches to therapy. In 19 of the studies, gesture therapy was used in order to cue speech (typically alongside other tasks designed to stimulate spoken words). Here, a key outcome measure was an assessment of spoken naming. In the four remaining studies, gestures were taught as a compensatory strategy, with the expectation that these would be used in place of speech (Code & Gaunt, 1986; Coelho, 1991; Daumuller & Goldenberg, 2010; Marshall et al., 2012). Here, the key outcome measure was an assessment of gesture production. Within cueing studies, significant naming gains were typically seen within the groups of items that had been trained within an intervention protocol but generalisation to untrained items was limited. As we can see from the final

Key Findings column of Table 1, gesture outcomes across the 23 reviewed studies were variable.

Rose et al (2013) conclude their review by calling for a number of investigations to further elucidate our understanding of the field of gesture treatments for aphasia. Amongst other features, they advocate further studies to investigate:

1. Gesture therapy in general
2. The outcomes of a higher dose and greater intensity of gesture therapy
3. Group studies with between subject designs - allowing for random allocation across treatment conditions.

The study reported in this thesis addressed each of the above three topics by exploring the effects of a large-dose, high intensity gesture intervention for adults with severe aphasia within a quasi-randomisd waitlist control study. Further examination of studies most pertinent to the current thesis, e.g. those which focus on the use of gesture as a compensatory strategy for absent spoken language (as opposed to those which focus upon its use to facilitate spoken naming) will enable a better understanding of how this specific topic may be best explored. These will be examined next.

Table 1. Articles included in Rose et al.'s 2013 systematic review of gesture therapy for aphasia – summarised for the purposes of the present chapter (continued overleaf)

Study Authors and Year	Group study or Single Case Experimental Design	Number of Pts	Aphasia Severity	Quality Scores (out of 10) on PEDRO-p / SCED scale	Amount of practice reported (hours)	Key Findings
Attard, Rose, & Lanyon, 2013	Single case	2	2 Severe	9	64	2 of 2 pts: improved naming following multimodal aphasia therapy. 1 of 2: improved naming after CIAT (no gestures)
Boo & Rose, 2011	Single case	2	1 Severe; 1 Moderate	9	40	2 of 2 participants: improved in verb naming for semantic and semantic + gesture conditions. 0 of 2 for gesture only condition
Code & Gaunt, 1986	Single case	1	1 Severe	4	NR	Improvement in trained words/pantomimes reported. No improvement in untrained words/pantomimes.
Coelho, 1991	Single case	2	1 Moderate-Severe; 1 Severe	8	NR	2 of 2 pts: improved number of trained signs. 2 of 2: improved on untreated signs
Ferguson, Evans, & Raymer, 2012	Single case	4	2 Mild; 2 Moderate-Severe	9	13	2 of 4 pts: improved naming for nouns trained using intentional gesture. 0 of 4 for untrained & trained with pantomime gesture. 1 of 4: improved gesture production for nouns trained using intentional gesture. 3 of 4: improved gesture production for nouns trained using pantomime gesture. 1 of 4: improved gesture production for untrained nouns in both the intentional and pantomime gesture condition.
Hoodin & Thompson, 1983	Single case	2	2 Moderate-Severe	5	NR	2 of 2 pts: improved naming of trained nouns. 0 of 2: improved naming of untrained nouns, 2 of 2: improved gesture production for trained gestures
G. V. Pashek, 1997	Single case	1	1 Severe	7	NR	1 of 1 pts: large naming improvements for trained nouns. 0 of 1: improvements for untrained nouns. 'Minimal' gesture improvements reported.
G. Pashek, 1998	Single case	1	1 Severe	4	NR	1 of 1 pts: greater naming improvements in trained nouns for gesture + verbal over gesture only. Gesture measures not implemented.

Study Authors and Year	Group study or Single Case Experimental Design	Number of Pts	Aphasia Severity	Quality Scores (out of 10) on PEDRO-p / SCED scale	Amount of practice reported (hours)	Key Findings
Purdy & van Dyke, 2011	Single case	2	1 Moderate; 1 Severe	6	10-12	2 of 2 pts: small increase in pointing + increased switching of communication modality (usually to gesture). 1 of 2: small increase in gestures.
A. M. Raymer & Thompson, 1991	Single case	1	NR	6	NR	No gains in picture naming. Gains in repetition for trained phonemes and some untrained phonemes. Small increase in reported in gesture production. (No data provided)
A. M. Raymer, Rowland, Haley, & Crosson, 2002	Single case	1	1 Mild	7	20	1 of 1 pts: Increase in trained sentence production. Some increase in untrained sentence production. Some improvement in time to produce sentences. Gesture measures not implemented.
A. Raymer et al., 2006	Single case	9	5 Moderate; 3 Severe; 1 Profound	9	20	5 of 9 pts: improved naming for trained nouns. 6 of 9: improved naming for trained verbs. 0 of 9: improved naming for untrained nouns or verbs. 8 of 9: improved gesture production for trained nouns. 6 of 9: improved gesture for trained verbs. 3 of 9: improved gesture for untrained nouns. 4 of 9: improved gesture for untrained verbs.
A. M. Raymer et al., 2007	Single case	4	2 Mild; 1 Moderate; 1 Severe;	8	20 sessions	Outcomes for gesture + verbal training: 3 of 4 pts: improved naming for trained nouns. 1 of 4: improved naming for untrained nouns. 2 of 4: improved gesture production for trained nouns. 1 of 4: improved gesture production for untrained nouns
Raymer S. et al., 2011	Single case	8	7 Moderate; 1 Severe	9	<40	Outcomes for gesture + verbal training: 3 of 8 pts: improved naming for trained nouns. 2 of 8: improved naming for untrained nouns. 6 of 8: improved gesture production for trained nouns. 3 of 8: improved gesture production for untrained nouns
Richards, Singletary, Gonzalez-Rothi, Koehler, & Crosson, 2002	Single case	3	2 Moderate	6	22.5	3 of 3: improved naming for trained nouns. 3 of 4: improved naming for untrained nouns. Gesture measures not implemented.

Study Authors and Year	Group study or Single Case Experimental Design	Number of Pts	Aphasia Severity	Quality Scores (out of 10) on PEDRO-p / SCED scale	Amount of practice reported (hours)	Key Findings
Rodriguez, Raymer, & Gonzalez Rothi, 2006	Single case	4	2 Moderate; 2 Severe	9	20	Outcomes for gesture + verbal training: 1 of 4: improved naming for trained nouns. 0 of 4: improved naming for untrained nouns. 3 of 4: improved gesture production for trained nouns. 0 of 4: improved gesture production for untrained nouns.
M. Rose & Douglas, 2008	Single case	1	1 Mild	8	14	Improvements in picture naming for gesture training and verbal training. Gesture measures not implemented.
M. Rose & Sussmilch, 2008	Single case	3	3 Moderate	9	19	Outcomes for gesture + verbal training: 2 of 3: improved naming for trained verbs. Gesture measures not implemented.
Rose M., Douglas, & Matyas, 2002	Single case	1	1 Mild	8	13.5	Increases in naming trained nouns for gesture training, verbal training and verbal/gesture naming. No increases in naming untrained nouns. Gesture measures not implemented.
Caute et al., 2012	Group	14	14 Severe	4	15	Significant improvement in message delivery skills after gesture therapy. No significant difference in narrative assessment. (Gesture outcomes for the same group reported separately by Marshall et al., 2012)
Crosson et al., 2007	Group	34	12 Moderate; 11 Severe; 11 Profound	4	7.5	Significant improvement on trained words for moderate-severe group and profound group. Significant improvement on untrained words for moderate-severe group only. Gesture measures not implemented.
Daumuller & Goldenberg, 2010	Group	34	34 Severe	3	7.5	Significant increase on measure of gesture intelligibility for both trained and untrained items.
Marshall et al., 2012	Group	14	14 Severe	2	15	Significant improvement in gesture production after gesture therapy. However, naming effects of naming therapy significantly greater than gesture effects of gesture therapy.

CIAT=constraint induced aphasia therapy; NR=Not Reported; pt=participant

2.8 The use of gesture as a compensatory strategy

Four of 23 studies reported in the review by Rose et al (2013) explore the use of aphasia therapy for compensatory gesture. In order to ensure the inclusion of all relevant data in the current literature review, the search described by Rose et al (2013) was repeated to look for evidence published subsequent to the September 2013 (the end date of the search reported by Rose et al, 2013) and up until February 2017. 42 articles were found. After removal of duplicates, 27 remained. Following subsequent abstract review, 25 articles were excluded as they did not address the topic of gesture treatments in aphasia. One further article was excluded as it did not address gesture as a compensatory strategy but instead as a means for cuing speech. The remaining article, by Marshall et al. (2013) will be considered alongside the four previously identified articles (Code & Gaunt, 1986; Coelho, 1991; Daumuller & Goldenberg, 2010; Marshall et al., 2012) to identify factors specifically relevant to the study of gesture as an alternative to speech. From this comparative reduction in numbers across both searches, it is immediately apparent that the majority of research published around gesture therapy in aphasia has had a focus on its effects for cuing speech. This alone creates a case for further research in this area. A summary of the information presented in the compensatory gesture literature is presented in table 2. This summary enables us to identify some key features as they relate to the current thesis

2.8.1 Summary of findings from articles exploring the use of gesture as a compensatory tool

The five studies identified in Table 2 report data for 60 participants in total (Code & Gaunt, 1986; Coelho, 1991; Daumuller & Goldenberg, 2010; Marshall et al., 2012; Marshall et al., 2013 reporting data for 1; 2; 34;14 and 9 participants respectively). All participants are reported to demonstrate severe aphasia with the exception of one participant in the Coelho (1991) study.

Table 2. Articles exploring the use of gesture as a compensatory tool within aphasia therapy

Study authors and Year	Description of gesture therapy	Reported treatment frequency	Total duration of therapy	Number of Items assessed	Gesture measures	Results
Code & Gaunt, 1986	Group sessions: improving confidence, initiating alternative strategies Articulation programme: reducing apraxia of speech Intensive group: gestural practice	45 min per wk	8 months (with 7 week break) i.e. approx 28 weeks = approx. 21 hours	20 (10 treated and 10 untreated)	Pantomime cued sequentially by spoken label for target: *Trained items* *Untrained items*	Gain of 3 gestures No gains
Coelho, 1991	Clinical setting: sign repetition and modelling "Easy street" setting: context elicits sign production	5 practices per session per item, 1 time per week.	Pt 1 = 8 weeks Pt 2 = 10 weeks	12 (12 treated)	# correct signs: Trained Generalised: Easy Street Natural Setting	2 of 2 Improved 2 of 2 Improved 1 of 2 Improved in clinic phase 1 of 2 improved in Easy street phase
Daumuller & Goldenberg, 2010	Gesture: object/picture mime training; role plays, group and individual	50 mins × 3/wk × 3 wks per set (3 sets)	7.5 hours	24 items treated in three batches of 8	Intelligibility of gesture production Trained gestures Untreated gestures	No significant increase for controls Significant increase(p = .0005 on 3 sets); d = 4.86 Significant increase(p = .02–.001 on 3 sets); d = *3.07*
Marshall et al., 2012	Gesture-only treatment: gesture recognition and production training	1 hr × 2/wk × 7–8 wks	15 hours	60 (20 treated for gesture, 20 treated for naming, 20 untreated)	Gesture to picture production: Blind scoring – assessors asked to identify target from video of gesture.	Significant increase gesture production. Main effect of treatment (p < .01,ηp2 = .29) – i.e. treated gesture items scored more highly than items receiving no treatment or naming treatment.
Marshall et al., 2013	GeST – computer gesture therapy employing gesture recognition. Three levels of practice: repetition of gesture demonstrated in video; repetition of gesture within virtual, responsive environment; repetition of gesture in response to everyday scenario.	64.4 sessions over 6 wks – average 10 sessions / wk of around 13 min / session (logged)	13.9 hrs average across 6 weeks (3 supported by a therapist & 3 unsupported)	30 gestures trained across two conditions (15 therapist supported; 15 unsupported)	Pantomime gesture production of the same set of 60 imagable words – (15 treated by SLT + computer; 15 treated by computer only; 15 treated by SLT only; 15 untreated) were scored as correct or incorrect and rated 1-5 for accuracy. Untrained items and identical picture naming task used to assess for generalisation.	Significant improvement on gesture production of items treated by SLT + computer, but no others. Gains were maintained 6 weeks post-intervention. No changes observed for accuracy rating over time. Untrained items and naming items did not change.

All five treatments reported indicate improvements for treated gesture items, with
data from 36 participants indicating generalisation to untreated items. (Coelho, 1991; Daumuller & Goldenberg, 2010). Four studies reported face-to-face therapy approaches (Code & Gaunt, 1986; Coelho, 1991; Daumuller & Goldenberg, 2010; Marshall et al., 2012) and one reported the effects of computer-delivered therapy both with and without therapist support (Marshall et al., 2013).

2.8.1.1 Dosage, intensity and number of items trained

Reported therapy dose varied between 7.5 hours (Daumuller & Goldenberg, 2010) to an estimated 21 hours (Code & Gaunt, 1986), across a period of three weeks (Marshall et al., 2013 – supported therapy period) to 28 weeks (Code & Gaunt, 1986). Therapy duration in minutes / hours was not reported for Coelho(1990). Of the remaining four studies, therapy was administered for 45 minutes per week for Code & Gaunt(1986) and between two and two and a half hours per week for the remaining three studies (Daumuller & Goldenberg, 2010; Marshall et al., 2012; Marshall et al., 2013). The number of gesture items treated ranged from 10 (Code & Gaunt, 1986) to 24 (Daumuller & Goldenberg, 2010). [Marshall et al., 2013 reported significant improvements for the group of 15 items trained with a computer PLUS therapist support only - and not the group of 15 trained using computer practice alone.].

2.8.2 Implications

The above summary indicates a small but positive evidence base for the successful use of compensatory gesture intervention for adults with severe aphasia. Data suggest the benefits of around two hours of practice per week for a set of between 10 and 24 items over a period of three to 28 weeks - acknowledging that – with the exception of the participant reported by Code & Gaunt, 1986– these gains were all observed within a period of three to 10 weeks. Due to the limited number of studies and a lack or evidence directly comparing different therapy methodologies against one another, there is presently no conclusive evidence about which therapy approach is optimal. It can be observed however, that each of the explored studies employed some level of gesture imitation and subsequent practice of gesture production – indicating that the inclusion of such methods might be associated with positive outcomes. An important observation to make, regards the relatively low number of acquired gesture items reported. Gains appear smaller for the gesture literature reported above, than for the scale of gains reported in typical naming studies (Wisenburn & Mahoney, 2009). This comparison might be attributed to the comparatively high level of aphasia severity for candidates engaged in compensatory gesture therapy (e.g. it is the very severity of their impairment which necessitates that a non-speech intervention be explored). Whilst this is one possible explanation, Marshall et al (2012) found the learning of words to substantially outstrip the learning of gestures even within the same severely aphasic candidates. An alternative explanation for the discrepancy therefore, is that in naming, theoretical accounts suggest that participants are regaining access to a previously established phonological form (Schwartz, 2013). In gesture no such form exists. So learning is more strategic and possibly dependent on cognitive skills such as executive function which enable an individual to switch out of their typical verbal linguistic resource and into a multimodal communication mode in order to solve the problem of how to convey their message (Purdy & Koch, 2006). Whilst the group findings from Marshall et al's (2012) paper may at first seem to make a case against the use of gesture and in favour of naming therapy, it should be acknowledged that there remain a number of participants reported in that study (three out of 14) who responded more favourably to the gesture intervention than the naming intervention. Taken in conjunction with the limited evidence reporting systematically on the use of gesture intervention for this group, we find a case for further, well-designed research of the area and the investigation of specific capacities such as executive function in relation to any outcomes observed.

2.9. Computer-based intervention

Having now explored the evidence base regarding the use of gesture therapy for aphasia, the issue of computer-based intervention will next be examined to establish an understanding of mechanisms for achieving an appropriate therapy dose. An introduction to computer intervention techniques in aphasia therapy is provided first. This is followed by a systematic literature review regarding the effects of computer-based therapy for expressive deficits in aphasia (chapter 3).

The following section introduces the reader to the general context for computer-based therapies in aphasia, providing examples of some key methodologies adopted in the application of computing within the therapy domain.

The use of computers in aphasia intervention has been increasingly advocated over time (as described by Petheram, 2004). Computer-based technologies have been developed to address a number of facets of aphasia. These technologies range from assistive communication devices through to technologies that provide structured exercises for the rehabilitation of impaired language skills (van de Sandt-Koenderman, 2004; van de Sandt-Koenderman, 2011). The technologies available can be broadly categorized into two of the four aphasia therapy strategies introduced earlier in section 2.5.1: namely *Impairment* and *Activity.* Those technologies addressing impairment aim to rehabilitate impaired communication skills – such as spoken word retrieval, whereas those addressing activity aim to compensate for impairments or augment communication by another means. In a 2004 review of the field of computer technology in aphasia therapy, van de Sandt-Koenderman further comments that the remaining two therapy strategies earlier introduced - *Participation* and *Wellbeing* - may also be addressed to some extent by computer technology. She acknowledges that there is a lot of general information on aphasia available on the Internet, which might serve this aim. She comments, however, that this information may not always be presented in a format accessible to users with aphasia. Hence the computer technology available at the time of writing the article (2004) was used far less commonly for goals of *Participation* and *Wellbeing* than for *Impairment* and *Activity* focused outcomes. With the advent of pervasive computing and advances in Internet connectivity, this is beginning to change. However – for the purposes of this literature review, the remaining focus will address just the topics of *Impairment* and *Activity* based computer interventions.

The following section introduces a number of examples of impairment and activity-focused computer-based interventions for aphasia. It aims to report a representative selection of computer-based interventions in some depth. Examples of the impairment-focused technologies AphasiaScripts and Step-by-step are presented first, followed by descriptions of two activity-focused technologies - SentenceShaper and C-Speak.

It is acknowledged that such an approach does not provide a comprehensive overview of the available evidence however. To more fully examine the breadth of research conducted in this field therefore, chapter 3 presents a systematic review of literature regarding computer intervention in aphasia. Chapter 4 then reports details of GeST, a computer based gesture therapy that straddles both the impairment and activity focussed categories – aiming to help severely aphasic users to compensate for their language limitations (activity) by providing intensive training of the specific skill of pantomime gesture (impairment).

2.9.1 Impairment-focused technologies

van de Sandt-Koenderman (2004) defines technologies falling within this category of the International Classification of Functioning (World Health Organization, 2001) as those that aim to restore an impaired linguistic ability. Technologies may, for example, deliver therapeutic activities aiming to remediate semantic, phonological or syntactic deficits.

In a technology developed to address the syntactic deficits experienced by some individuals with aphasia, several authors (Bilda, 2011; Cherney, Halper, Holland, & Cole, 2008; Cherney, Kaye, & van Vuuren, 2014; Manheim, Halper, & Cherney, 2009; Youmans, Holland, Muñoz, & Bourgeois, 2005, 2014) have reported research into a therapeutic system named AphasiaScripts™ (Rehabilitation Institute of Chicago, 2007). This system comprised a software program used alongside a microphone and audio playback on a laptop or desktop personal computer. The program aimed to support users in practising conversational scripts that have been separately scripted and recorded into the system. Through employment of a virtual agent whose animated face appears on screen alongside written prompts, users were given the opportunity to first listen to, then repeat their lines within the script conversation. Users were provided with differing levels of written and spoken support. Within the final stage of practice, users aimed to (independently and unprompted) speak-aloud scripted lines in response to allotted turns in conversation.

Cherney et al. (2014) report on eight individuals with chronic aphasia (mean time post onset = 26 months, range = 8 – 59 months) who received 6 weeks of independent computer practice with AphasiaScripts™. Two participants within this group were classified as having severe aphasia, five moderate and one mild according to scores on the Western Aphasia Battery Aphasia Quotient (WAB AQ, Kertesz, 1982). Authors report significant improvement on a measure of phonological, semantic and grammatical accuracy (NORLA-6, Gingrich, Hurwitz, Lee, Carpenter, & Cherney, 2013) and on the rate of production when comparing performance on scripts before treatment and after treatment.

A second example of activity-focused computer intervention is that of StepByStep - described in a further range of studies (Mortley, Wade, Davies, & Enderby, 2003; Mortley, Wade, Enderby, & Hughes, 2004; Palmer et al., 2012) this program was similarly delivered via a laptop or PC with a microphone and audio playback. Intended to target aphasic difficulties of varying degrees, the software comprised a choice of 13,000 activities allowing users to practice activities of word recognition and word production – both in isolation and in sentences. As with other programs reported, users were presented with varying levels of support to complete the therapy tasks.

Palmer et al. report outcomes for a group of 16 participants with aphasia who carried out 5 months of independent practice (on average 60 minutes per week) with one initial support visit from a therapist and subsequent follow up support from a trained volunteer (weekly for the first month and monthly for subsequent months). Authors found that participants demonstrated a 19.8% increase in performance on naming assessment when compared to a control group of 17 participants tested over the same period who did not receive the computer intervention. Participants demonstrated a positive mastery of the therapy, achieving 75% of the recorded practise hours independently, without additional volunteer support or supervision. Authors do identify however, that of those who received the intervention, the two participants "with more severe aphasia showed little benefit" (Palmer et al., 2012, p. 1904).

2.9.2 Activity-focused technologies

We can refer again to the 2004 review of computer technology in aphasia therapy for a definition of the activity-focused area of intervention. van de Sandt-Koenderman (2004) categorises these technologies as those that aim to enable users to partake in effective day-to-day communication in spite of the presence of aphasia. The aim of such technologies is not to restore the impaired ability but to allow an individual to communicate in spite of the impaired ability. Common examples of this type of technology include augmentative communication devices, which produce a synthetic voice to supplement or extend the users own verbal output.

Considering now an example of such an augmentative and alternative communication (AAC) device, several authors (Bartlett, Fink, Schwartz, &

Linebarger, 2007; M. Linebarger, Schwartz, & Kohn, 2001; M. C. Linebarger, Schwartz, Romania, Kohn, & Stephens, 2000; McCall, Virata, Linebarger, & Berndt, 2009) report studies regarding a “processing prosthesis” named SentenceShaper^{TM}. This technology comprises a software program used alongside a microphone and audio playback on a laptop or desktop personal computer. It aims to facilitate users with aphasia who are no longer able to produce rapid and fluent sentence constructions spontaneously to record individual speech samples (perhaps words or key phrases) that they can then concatenate to produce more complex sentences to be played back.

Linebarger et al. (2001) and Bartlett et al. (2007) found that supported use of the system enabled some users with moderate levels of aphasia to extend their existing spoken productions and to facilitate the expression of both make-believe and also hypothetical everyday narratives. Due to the requisite recording and manipulation of speech samples taken of the user’s own voice however, authors concede that those “who are unable to produce even single words or short phrases are unlikely to use the program effectively.” (Bartlett et al., 2007, p. 492).

Another example of an AAC device used to support expression in users with aphasia is C-Speak Aphasia (Reported in Nicholas, Sinotte, & Helm-Estabrooks, 2011, 2011). This technology comprises a software program used alongside audio playback on a laptop or desktop personal computer. The program aims to facilitate users with aphasia who are no longer able to produce rapid and fluent sentence constructions spontaneously to compose phrases and sentences from a selection of icons – organized by semantic category. These phrases and sentences can then be spoken aloud using the system’s electronic speech synthesizer.

Nicholas et al. (2011) found that six months supported use of the system enabled four out of ten users with severe aphasia to communicate more information with C-Speak Aphasia than without on a series of communication tasks such as describing pictures and videos and making telephone calls. We can infer, however, that six of ten users did not demonstrate such communicative gains. Authors found levels of non-verbal executive function to be linked to success in using the system to communicate more content after training, a finding that suggests that measures of such skills may prove to offer useful insights into therapy candidacy.

2.9.3 Impairment + activity-focused technology

In an intervention which could be seen to combine both the strategy of impairment and that of activity, Galliers et al. (2012) and Marshall et al. (2013) report a computer gesture therapy tool named GeST. Aiming to provide a structured and intensive means of practising and learning a vocabulary of pantomime gestures, GeST supports practice of up to 30 items by using a system of vision-based gesture recognition. The tool comprises a software program used on a laptop with audio playback. In addition, it utilises a simplified external keyboard and an external webcam. To facilitate the vision-based recognition, users must wear a yellow cotton glove on the hand they are using to gesture. (Full details of the system are provided in section 5.3). Training items are presented via video and then practiced using immediate and then delayed repetition via one of three training levels. Levels make use of both video instruction and also a 3D virtual environment where participants' gestures allow them to interact – in a limited way - with a novel gaming world. Users receive feedback in the form of applause for each gesture correctly produced and identified by GeST.

In a study of 9 participants with severe aphasia, Marshall et al. (2013) found that users who received 3 weeks of independent practice with GeST plus three hours of therapist support made significant gains in the numbers of identifiable gestures they produced for those items trained. The same participants, receiving a further 3 weeks of independent practice with no therapist support, made no measurable gains on items practiced within this additional unsupported phase.

Further details of the research project behind the development of GeST and the pilot trial of its effectiveness are described in section 4.2.

The above section has introduced the reader to some examples of different computer-based interventions in relation to aphasia therapy. A systematic review of published literature pertaining to computer-based therapy for gesture and expressive language is presented in the next chapter.

Chapter 3. Computer-based intervention – systematic literature review

3.1 Introduction

The structure for this chapter adheres to PRISMA guidelines for reporting systematic review (Moher, Liberati, Tetzlaff, & Altman, 2009). As such, items covered herein reflect those suggested within Liberati et al.'s (2009) guideline document. Briefly, a review abstract is first presented, followed by introduction, methods, results, discussion and funding sections.

3.2 Title

Self-administered computer therapy for gesture and speech deficits in aphasia – a systematic review of evidence.

3.3 Abstract

Background: Zheng, Lynch, & Taylor (2016) provide a review of the effects of computer-based therapy for aphasia - comprising evidence drawn from controlled group studies only. Much additional evidence exists to supplement our understanding however, in the form of single case experimental designs (SCEDs) – more typical in clinical fields such as speech and language therapy, which are relatively infant in comparison to areas such as pharmaceutical research.

Objectives: To examine the effects of computer-delivered aphasia therapy for gesture and spoken language – reviewing both group and single case study experimental designs. Specifically, this review examines such effects in comparison to no therapy/usual care/sham therapy, alterative clinical treatment, or alternative computer treatment. It additionally addresses comparisons between treated and non-treated stimuli, the maintenance of

Table 4. Participant characteristics of group studies (continued overleaf)

Study Authors and Year	Group	Number of Participants	Age	Education	Gender	Time Post Onset (months)	Handedness	Aetiology	Aphasia type	Aphasia severity	Coexisting communication Impairments
1. Cherney, Kaye, & van Vuuren, 2014	High cue–low cue treatment order	4	Mean=53.9; Range 42.9-66.4	Mean=14.5; Range 11-18 yrs	M=3; F=1	Mean=25.8; Range 8-59	R=2; L=2	CVA=4	Fluent=1; Non-fluent=3;	Severe=1; Moderate=3; (established using WAB AQ)	NR
	Low cue–high cue treatment order	4	Mean=50.2; Range 25-64.5	Mean=14.8; Range 13-16 yrs	M=3; F=1	Mean=27; Range 10-48	R=2; L=2	CVA=4	Fluent=1; Non-fluent=4;	Severe=1; Moderate=2; Mild=1 (established using WAB AQ)	NR
2. Doesborgh et al., 2004	Treated	8	Mean=66.7; Range NR	NR	M=4; F=4	Mean=13; Range 11-16	R=7; L=1	CVA=8	NR	Moderate-severe (established via BNT)	NR
	Control – no treatment	10	Mean=65; Range NR	NR	M=5; F=5	Mean=13; Range 11-17	R=10; L=0	CVA=10	NR	Moderate-severe (established via BNT)	NR
3. Fridriksson et al., 2009	Treated (within-subject design)	10	Mean=59.4; Range 33-74	NR	M=8; F=2	Mean=85.3; Range 17-216	R=8; L=0	CVA=10	Broca=8	Very severe=1; Severe=4; Moderate=5 (established using WAB AQ)	AoS=10
	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4. Manheim, Halper, & Cherney, 2009	Treated (within-subject design	20	Mean=54.8; Range 26-78	Mean=15.1; Range 10-22	M=13; F=7	Mean=53; Range 10.6-273.7	R=20; L=0	CVA=20	NR	M=64.6=Moderate; R=30.5-85.3=Severe-mild (Based on WAB AQ)	NR
	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5. Marshall et al., 2013	Treated (within-subject design	9	Mean=68.8; Range 31-90	NR	M=6; F=3	Mean=77.9; Range 24-276	R=7; L=2	CVA=9	NR	Severe=9 (established through naming performance of <20% on CAT naming test)	Limb apraxia = 3
	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Study Authors and Year	Group	Number of Participants	Age	Education	Gender	Time Post Onset (months)	Handedness	Aetiology	Aphasia type	Aphasia severity	Coexisting communication Impairments
6. Palmer et al., 2012	Treated	16	Mean=69.5; Range 37.8-82.6	NR	M=9; F=7	Mean=74.4; Range 12-348	NR	CVA=16	Fluent=3; Non-fluent=12;	Severe=2; Moderate=5; Mild=9 (established via CAT)	AoS=3
	Control – usual care (no impairment focused treatment)	17	Mean=66.2; Range 48.2-83.7	NR	M=12; F=5	Mean=79.2; Range 13-348	NR	CVA=17	Fluent=3; Non-fluent=13; Global=1	Severe=2; Moderate=4; Mild=11 (established via CAT)	AoS=3
7. Thompson, Choy, Holland, & Cole, 2010	Treated	6	Mean=50.5; Range 30-68	Mean=16.7; Range 12-20 yrs	NR	Mean=73.8; Range 13-196	R=5; L=1	CVA=6	non-fluent	Severe=1; Moderate=1; Mild=4 (established using WAB AQ)	nil
	Control (no treatment)	6	Mean=48.5; Range 37-59	Mean=15.6; Range 12-22 yrs	NR	Mean=45.8; Range 14-106	R=6	CVA=6	non-fluent	Severe=1; Moderate=1; Mild=4 (established using WAB AQ)	nil
8. Wenke et al., 2014	Treated computer therapy	10	Mean=63.5; S.D.=12.6	Mean=10.4; S.D.=2.8	M=8; F=2	Mean=9.1; S.D.=8.3	NR	NR	NR	NR	NR
	Treated – group therapy	8	Mean=61.6; S.D.=15.5	Mean=11.3; S.D.=2.9	M=7; F=1	Mean=27.5; S.D.=28.4	NR	NR	NR	NR	NR
	Treated – SPTA therapy	6	Mean=59.8; S.D.=4.9	Mean=15.5; S.D.=0.8	M=4; F=2	Mean=10; S.D.=15.4	NR	NR	NR	NR	NR
	Control – standard service	22	Mean=66.7; S.D.=13.1	Mean=12; S.D.=2.9	M=14; F=8	Mean=6.1; S.D.=9.9	NR	NR	NR	NR	NR

Notes: AoS=Apraxia of Speech; BNT=Boston Naming Test (Kaplan, Goodglass, Weintraub, Segal, & van Loon-Vervoorn, 2001); CAT=Comprehensive Aphasia Test (Swinburn, Porter, & Howard, 2004); CVA=Cerebrovascular Accident; F=Female; L=Left; M=Male; NR=Not Reported; R=Right; S.D=Standard Deviation; SPTA=Speech Pathology Therapy Assistant; WAB AQ=Western Aphasia Battery Aphasia Quotient (Kertesz, 1982).

Table 5. Participant characteristics of single case experimental design studies (continued overleaf)

Study Authors and Year	Number of Participants	Age	Education	Gender	Time Post Onset (months)	Handedness	Aetiology	Aphasia type	Aphasia severity	Coexisting communication Impairments
9. Choe, Azuma, & Mathy, 2010	3	Mean=47.3; Range 39-52	NR	M=1; F=2	Mean=35; Range=16-67	R=3; L=0	CVA=3	Nonfluent=3	NR	AoS=3
10. Choe & Stanton, 2011	2	Mean=55.5; Range 55-56	Mean=15; Range 14-16	M=1; F=1	Mean=111; Range 49-173	R=2; L= 0	CVA=2	Anomia=1; Broca=1;	Moderate=1; WNL=1	nil
11. Choe, Azuma, Mathy, Liss, & Edgar, 2007	4	Mean=57; Range=48-76	NR	M=2; F=2	Mean=78.75; Range=17-156	R=4; L=0	CVA=4	Nonfluent=4	NR	AoS=4
12. De Luca et al., 2014	1	56	12	M=0; F=1	9	R=1	CVA=1	NR	AAT results; Token=23/0 (sic.); comprehension 83/120; written 40/90; naming 63/120; repetition 125/150 (Based on AAT scores)	Psychological impairments reported
13. Fink, Brecher, Schwartz, & Robey, 2002	6	Mean=60.5; Range 54-64	Mean=14; Range 12-16 yrs	M=5; F=1	Mean=49.2; Range 28-92	R=4; L=2	CVA=6	Anomia=2; Broca=1; Conduction=3	Moderate difficulties (BDAE 2)=4; Some moderate to minor difficulties (BDAE 3)=1; Some minor difficulties (BDAE 4)=1 (Based on BDAE Severity Scores)	NR
14. Kurland, Wilkins, & Stokes, 2014	8	Mean=66.8; Range 55-80	NR	M=4; F=4	Mean=31.9; Range 8-84	R=8; L=0	CVA=8	Anomia=3; Conduction=1; Crossed Wernicke=1; Mixed Transcortical=1; Transcortical Sensory=1; Wernicke=1	Severe difficulties (BDAE 1to2)=2; Moderate difficulties (BDAE 2to3)=5; Some minor difficulties (BDAE 4)=1 (Based on BDAE Severity Scores)	Nil

Study Authors and Year	Number of Participants	Age	Education	Gender	Time Post Onset (months)	Handedness	Aetiology	Aphasia type	Aphasia severity	Coexisting communication Impairments
15. Laganaro, Di Pietro, & Schnider, 2003	11	Mean=53.3; Range 32-80	NR	M=4; F=7	Mean=14.3; Range 2-120	NR	CVA=10; TBI=1	Anomia=2; Broca=1; Conduction=3; Mixed=1; Mixed-Broca=2; Transcortical motor=1; Wernicke=1	Moderate to Severe=11 (based on unspecified clinical reports)	NR
16. Linebarger, Schwartz, & Kohn, 2001	5	Mean=44.4; Range 20-64	Mean=15; Range 12-16 (Education NR for 1 pt)	M=3; F=2	Mean=71.4; Range 15-135	R=4; L=1	CVA=3; TBI=2	Agrammatism=5	NR	NR
17. Pedersen, Vinter, & Olsen, 2001	3	Mean=65.3	NR	M=3; F=0	Mean=11.7; Range 6-21	NR	CVA=3	Anomia=3	Moderate=1; Mild=2 (established using WAB AQ)	NR
18. Ramsberger & Marie, 2007	4	Mean=67.5; Range 63-74	Mean=15; Range 12 - 16	M=3; F=1	Mean=31.5; Range 6-72	NR	CVA=4	Anomia=1; Broca=1; Conduction=1; Wernicke=1	Severe=1; Moderate=3 (established using WAB AQ for 2 moderate pts and the ADP for one moderate and one severe pt)	nil
19. Routhier, Bier, & Macoir, 2016	2	Mean=56; Range= 51-61	Mean=11.5; Range 11-12	M=1; F=1	Mean=42; Range 12-72	R=2; L= 0	CVA=2	Nonfluent=1; fluent=1	Severe=2 (based on performance on the DVL-38; Hammelrath, 2001)	nil
20. Wieczorek, Huber, & Darkow, 2011	4	Mean=54.4; Range 35.2-65.8	NR	M=3; F=1	Mean=47; Range 17-73	R=4; L=0	CVA=4	Broca=2; Wernicke=2	Moderate=4 (Based on AAT Severity Scores)	NR

Notes: AAT= Aachen Aphasia Test (Huber, Poeck, & Willmes, 1984); ADP=Aphasia Diagnostic Profiles (Helm-Estabrooks, 1992); AoS=Apraxia of Speech; BDAE=Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2001); BNT=Boston Naming Test (Kaplan et al., 2001); CAT=Comprehensive Aphasia Test (Swinburn et al., 2004); CVA=Cerebrovascular Accident; DVL-38= Test de dénomination des verbes lexicaux. (Hammelrath, 2001); F=Female; L=Left; M=Male; NR=Not Reported; Pt=Participant; R=Right; TBI=Traumatic Brain Injury; WAB AQ=Western Aphasia Battery Aphasia Quotient (Kertesz, 1982)

therapy effects over time, the duration and intensity of effects and the benefits in specific relation to people with severe aphasia.

Data sources: A systematic review of English articles using the EBSCHOST platform was conducted. Search terms comprised three key words: aphasia, computer and therapy as well as variations on each of these terms.

Study eligibility criteria, participants, and interventions: Only those studies with an experimental design were included. Both group and single case methodologies were reviewed. Participants all demonstrated a primary communication diagnosis of aphasia following acute brain injury (CVA or TBI). Interventions included were restricted to those that targeted gesture or expressive language symptoms. Each required a period of self-administered practice on the part of the participant.

Study appraisal and synthesis methods: Study quality was appraised using two rating measures (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003; Tate et al., 2008). Outcomes were synthesised to document key features of studies including the severity of participants' aphasia, the nature of the examined therapy, the choice of measure used to explore its effect, the duration and intensity of therapy input and the nature and duration of any observed therapy effects.

Results: 20 studies were included – eight group studies and 12 SCEDs. 17 studies reported statistical outcomes. Outcomes largely supported the clinical effect of computer therapy for the improvement of naming treated items. There was little evidence to support the generalisation of therapy effects to untreated items. Limited evidence is available regarding the effect of therapies beyond the single word level - just one study each to suggest an effect of computer training for communicative gesture production, conversational scripts, complex sentence production and spoken grammar terms. Evidence largely supports the maintenance of therapy effects in the short term (up to two months post-therapy-cessation). The duration and intensity of therapies reported varied although there is good evidence to suggest that participants with aphasia could use the majority of programs

reported autonomously. The effects of computer therapy appear more variable for participants with severe aphasia than for other participants.

Limitations: A variety of techniques are reported within the reviewed studies. These include variety in both therapeutic methods and outcome measures. Inconsistency of methodologies across the literature means that outcomes should be interpreted with caution.

Conclusions and implications of key findings: Computer therapies can provide effective outcomes for participants with aphasia – particularly in relation to naming. Less evidence is available in relation to computer gesture therapy and other computer therapies targeting speech beyond the single word level. Single word and gesture outcomes are typically limited to treated items with some evidence of generalisation for sentence level and grammatical interventions. Additionally, whilst effects are largely maintained in the short term, less is known about their persistence after two months or more. Participants with aphasia are able to use therapy technologies to carry out independent practice and the amount of practice reported varies from individual to individual. Finally, those with severe aphasia can benefit from computer therapies, although performance can be variable. Little is currently known about the factors that contribute to the relative success or failure of computer practice for people with severe aphasia.

3.4 Introduction

3.4.1 Rationale

There is an increasing move towards the provision of digital methods of healthcare (Lane Fox, 2015) - creating a need for the full understanding of existing evidence regarding the use of technology within different health domains. Within aphasia treatment, increased therapy intensity is associated with increased clinical gains (Bhogal et al, 2003) and computer therapy offers a cost-effective means of increasing therapy intensity without additional therapist demand (Varley, 2011). Whilst the above issues are acknowledged generally within the aphasia research community, there is currently limited documentation of the overall validity of computer-delivered therapy as a means of achieving the levels of therapy practice required to effect clinical change. Zheng et al. (2016) conducted a systematic review of controlled group trials - examining the effect of computer therapy targeting verbal and/or orthographic deficits within aphasia. They reported evidence from seven trials to provide tentative support for the effect of computer therapy versus no therapy. The choice to exclude single case experimental designs from Zheng et al.'s review however, resulted in a substantial weight of existing evidence being omitted from consideration. The relative infancy of the speech and language therapy evidence base does not lend itself to the extensive production of a large body of randomised control trials (RCTs) and Garrett & Thomas (2006) argue in support of systematic reviews of literature that take into account both RCT designs as well as quasi-experimental and single-subject designs. It should be acknowledged that RCTs represent the gold standard of research evidence in healthcare as the rigorous methods employed diminish the risk of bias in the conduct of the study. Procedures such as registration also minimise the risk of publication bias (where only studies with significant outcomes are published), a risk which is much greater for small group and case studies. Trials, critically, also allow for the pooling of cumulative data to enable meta-analyses – allowing us to draw wider conclusions about a body of evidence (as forms the basis of the Cochrane Library of reviews). In spite of their primacy however, there currently exist comparatively few RCTs within the field of speech and language intervention. Indeed, Howard, Best, & Nickels (2015) argue that single-case studies and single-case series can provide a better window of the merits of specific therapy approaches than large scale RCTs which are better placed to comment on the effectiveness of aphasia therapy in general. The aim of the current literature review therefore, is to document and examine the existing evidence base surrounding the use of remedial self-administered computer therapy for gesture and speech deficits in aphasia – taking into account both group and single case experimental designs. As the overarching thesis which subsumes this review is concerned with the effects of a computer delivered therapy upon gesture and spoken output, therapies which target orthographic deficits will be excluded in favour of a focus upon gesture and spoken output therapies only. The resulting review is intended to add to our understanding of computer therapy for gesture and/or spoken language in aphasia by exploring more fully a broad range of the evidence that is presently available in the field of speech and language therapy research.

3.4.2 Objectives

The questions addressed within this review are:

1. What is known about the effect of remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes when compared to a clinical non-computer therapy, an alternative computer therapy or no therapy/usual care?

2. What is known about the durability/maintenance of effects achieved using remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes?

3. What is the range in the duration and intensity of practice reported for computer therapy?

4. What is the effect of computer based therapy for adults with severe aphasia?

These questions are examined within the existing literature using the Participant; Intervention; Comparison; Outcome; Study design (PICOS) components identified within the eligibility criteria in Table 3.

3.5 Methods

3.5.1 Eligibility criteria

Table 3. Inclusion/Exclusion criteria identified using the PICOS format

	Inclusion Criteria	Exclusion Criteria
Participants	Adults with aphasia. No exclusion regarding severity of, comorbidities, age, sex or setting.	Absence of a reported aphasia diagnosis. Communication partners of adults with aphasia. Primary Progressive Aphasia or other progressive diagnoses (such as semantic dementia).
Interventions	Primary delivery method - Remedial computer-delivered therapy for aphasia. Period of self-administered practice included. Interventions address deficits of gesture or spoken language production.	Therapy that does not include technology. Therapy whose primary focus is not aphasia, for example apraxia of speech or dysarthria. Technological aphasia interventions not expected to remediate gesture or spoken language e.g. augmented or alternative communicative devices. Technological applications where intervention is not self-administered via computer – e.g. tele-rehabiltation or no period of autonomous practice without therapist mediation. Drug intervention. Interventions that entail transcranial magnetic stimulation (TMS) or transcranial direct current stimulation (tDCS). Primary focus of intervention to address deficits of reading or writing.

	Inclusion Criteria	**Exclusion Criteria**
Comparisons	Control group comprising no treatment/sham treatment/usual care or a non-computer treatment. Control group comprising an alternative computer treatment. Untreated control items. Untreated period for within participant comparisons.	No control items or comparison group. No comparative untreated period for within participant comparisons.
Outcomes	Change in performance on language measures or measures of the given skill being targeted for remediation.	No quantitative measures used.
Study Design	Group studies comparing a treatment and control group. Group studies comparing participant performance over time for treated and untreated items. Group studies comparing participant performance over time for treated and untreated periods. Single case studies comparing participant performance over time for treated and untreated items. Single case studies comparing participant performance over time for treated and untreated periods.	Non-experimental methods.
Publication Details	Peer-reviewed publications English language articles	Non-peer reviewed publications, for example conferences or unpublished theses Non-English articles

3.5.2 Information sources

Electronic searches of the following databases were conducted in July 2014 and again in February 2016 using the EBSCHOST platform: Academic search complete; CINAHL plus full text; E-Journals; Health and Psychosocial Instruments; Health Economic Evaluations Database; Library, Information Science and Technology abstracts; MEDLINE with full text; PsychARTICLES; PsychINFO; SocINDEX.

3.5.3 Search

Search terms were:

Aphasia	Computer	Therapy
Aphas* or anomia or dysphas* or agrammatism	Comput* or Tech* or AAC or digital or mobile	Treat* or intervention or therap*

All searches were at abstract level. All stages of the search strategy were limited to the English language. Sources were limited to academic journals.

Output from the second of the above searches (February 2016) was restricted to include only articles published in the time since the first was conducted (i.e between July 2014 and February 2016).

3.5.4 Data collection process

Customised data extraction tables were developed using the models set out in Rose, Raymer, Lanyon, & Attard 's 2013 systematic review of gesture therapy for aphasia and Zheng et al's 2016 systematic review of computer intervention for aphasia.

3.5.5 Data items

The following data was extracted for all studies: study objective, design, inclusion criteria, exclusion criteria, recruitment procedures, participant details, comparator details (for studies with a control group), aphasic symptom targeted, description of therapy, reported treatment frequency and total number of hours duration, comparisons, outcome measures and summary of key findings.

3.5.6 Quality assessment

Following the convention adopted by Rose et al. (2013) and Zheng et al. (2016), the methodological quality of group studies was assessed using the Physiotherapy Evidence Database (PEDro) scale (Maher et al., 2003). This scale is suitable for assessment of controlled trials and comprises a series of 11 items, which receive a score of one point if the quality measure is met and zero points if it is not. Scores for 10 items are summed to give a total score that reflects the study's measure of internal validity. The outstanding item represents a measure of the study's external validity. The methodological quality of non-group studies was assessed using the Single-case Experimental Design (SCED) scale (Tate et al., 2008). This follows the precedent set by Rose et al. (2013). Again, the scale comprises 11 measures of quality – with a one-point score being awarded if the criterion is met and zero if it is not. As for the PEDro scale, one item represents the study's external validity and the remaining 10 are totalled to provide a measure of internal validity.

3.5.7 Summary measures

Behavioural measures of gesture or spoken language production – either standardised or unique to the tested intervention, were used as the primary outcome measures to examine change.

3.5.8 Synthesis of results

Data were first synthesised descriptively within summary tables (as described in 0). Subsequent summaries of data reported within tables was completed descriptively within the text.

3.5.9 Risk of bias across studies

As previously stated in section 0, two scales were used to assess risk of bias within studies. These are the PEDro scale (Maher et al., 2003) and the SCED scale (Tate et al., 2008). Items within these scales allow for an assessment for potential bias from a range of fields including: allocation to treatment group, experimenter bias, bias arising from a failure to analyse participants who are not followed up and a failure to report comparative treatment effect sizes.

3.5.10 Additional analyses

Meta-analyses were not undertaken due to the heterogeneity of study types, outcome measures and intervention techniques.

3.6 Results

3.6.1 Study selection

Searches were carried out on two separate dates. An initial search conducted in 2014 was later repeated to capture any additional evidence published in the intervening 19 months.

Search 1: conducted July 2014

1,251 references were found. After automatic deduplication 808 remained and after manual deduplication a total of 702 references remained. The abstracts of the remaining 702 references were screened against the exclusion criteria. Where sufficient detail was unavailable from the abstract, a full-text version of the article was consulted. 684 were excluded for reasons described in Figure . The remaining 18 papers were systematically reviewed.

Search 2: conducted February 2016

The search was repeated in Feb 2016 to include all articles published within the date range July 2014 – Feb 2016. 124 references were found. After automatic deduplication, 101 remained and after manual deduplication 96 remained. The abstracts of the remaining 96 references were screened against the exclusion criteria. Where sufficient detail was unavailable from the abstract, a full-text version of the article was consulted. Following application of exclusion criteria, 2 additional papers were added to the systematic review.

Figure 2 shows the study flow incorporating both of the above searches. Reasons for exclusion are stated in the order in which decisions were made and exclusions applied.

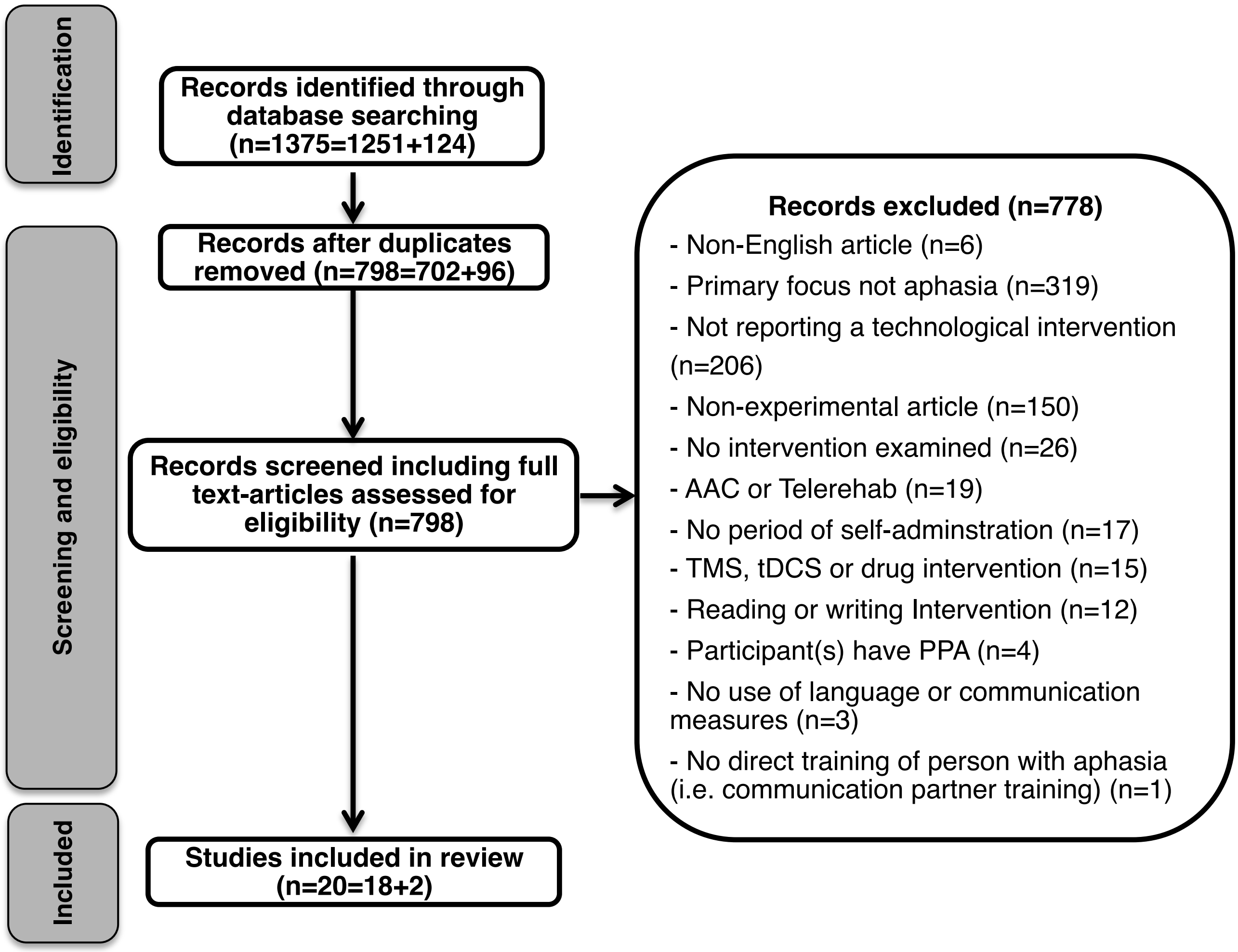


Figure 2. Flow diagram showing systematic selection of articles

3.6.2 Study characteristics

For the purposes of this review, studies where data were analysed as a group were classified as group studies. Those studies where data were examined on a case-by-case basis were categorised as single case/case series designs. Within these parameters, a total of eight group studies and 12 single case/case-series designs were included in the review.

Participant characteristics are presented in Table 4 and Table 5.

A total of 195 participants are reported (115 male, 68 female, gender not reported for 12). 142 participants are reported within group studies (85 male, 45 female, gender not reported for 12). 53 participants were reported within single case experimental designs (30 male, 23 female). The reported age of participants ranges from 20 to 90. The mean ages of participants range from 44.4 years to 69.5 years.

The total number of years in education was reported for 10 of 20 studies. The reported number of years in education range from 10-22 and the mean number of years of education range from 10.4 years to 16.7 years.

Nineteen of 20 studies reported the length of time post onset. The reported range of time post onset was 2 – 348 months and the mean time post onset ranges from 11.7 to 111 months. Pre-morbid handedness was reported for 15 of 20 studies covering 110 participants (101 right-handed, 9 left handed).

Aetiology of aphasia was reported for 19 of 20 studies (163 participants). The most typical cause of aphasia was stroke (cerebrovascular accident - CVA) listed as the aetiology for 160 participants. Aetiology for remaining three participants was traumatic brain injury (TBI).

Aphasia type was reported for 15 of 20 studies. Reporting of aphasia type was varied, with labels ranging from categorisation solely by fluency through to categorisation using clinical aphasia subtypes. A summary of the types of aphasia reported is provided in Table 6. Where subtypes have been used, these have been additionally categorised within fluent, nonfluent or mixed groupings.

Table 6. Types of aphasia reported within studies

Type of aphasia reported	Number of participants
Fluent (no clinical subtype listed)	8
Anomia (fluent)	12
Conduction (fluent)	8
Wernicke (fluent)	5
Crossed Wernicke (fluent)	1
Transcortical sensory (fluent)	1
Total number of fluent participants	**35**
Nonfluent (no clinical subtype listed)	52
Broca (nonfluent)	14
Mixed Broca (nonfluent)	2
Transcortical Motor (nonfluent)	1
Global (nonfluent)	1
Agrammatism (nonfluent)	5
Total number of nonfluent participants	**75**
Mixed (no clinical subtype listed)	1
Mixed Transcortical (mixed)	1
Total number of mixed participants	**2**
Total number of participants (all types)	**112**

Severity of aphasia is reported for 16 of 20 studies (130 participants). As was the case for the aphasia type, reporting of aphasia severity varied, with participants classified against a range of standardised measures such as the Western Aphasia Battery Aphasia Quotient (Kertesz, 1982), the Aachen Aphasia Test (Huber et al., 1984), the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983) or the Comprehensive Aphasia Test (Swinburn et al., 2004) and others whose severity level was identified clinically by a speech and language therapist. A summary of the severity of aphasia reported is provided in Table 7. Studies where diagnoses for individual participants were not reported (Manheim et al., 2009), or where no definitive categorisation was provided (De Luca et al., 2014), have been excluded from the summary statistics presented for reasons of clarity.

Table 7. Severity of aphasia reported

Severity of aphasia reported	Number of participants
Very severe to severe	27
Moderate to severe	29
Moderate	39
Mild to moderate	1
Mild	33
Within normal limits	1
Total number of participants (all severities)	**130**

Apraxia of speech is the most commonly reported co-existing communication impairment (23 participants). Limb apraxia is also reported for three participants (Marshall et al., 2013) and psychological impairments are reported for one participant (De Luca et al., 2014).

3.6.3 Risk of bias within studies

Quality ratings for group and SCED studies are reported in Table 8 and Table 9 respectively. Scores for the quality rating of group studies range from three to eight out of a maximum of 10 - with two studies scoring eight or above (Doesborgh et al., 2004; Palmer et al., 2012). All group studies specified their eligibility criteria, reported measures of one key outcome and provided point measures and measures of variability. No studies demonstrated blinding of participants or therapists to the treatment conditions. Four demonstrated blinding of assessors (Doesborgh et al., 2004; Manheim et al., 2009; Marshall et al., 2013; Palmer et al., 2012).

Scores for the quality rating of SCED studies range from three to nine out of a maximum of 10 - with five studies scoring eight or above (Choe et al., 2010; Choe & Stanton, 2011; Fink et al., 2002; Ramsberger & Marie, 2007; Routhier et al., 2016). All studies demonstrated evidence of clinical history, had a study design that allowed for the examination of cause and effect and all reported raw data points. No studies reported assessor independence.

3.6.4 Results of individual studies

Therapy characteristics and outcomes for group and SCED studies are reported in Table 10 and Table 11 respectively.

Table 8. Study design and quality scores for group studies

Study Authors and Year	Eligibility criteria specified	Random allocation	Concealed allocation	Groups similar at baseline	Blinding of participants	Blinding of therapists	Blinding of assessors	Measure of at least one key outcome from more than 85% of Pts.	Intention to treat	Between group statistical comparison	Point measures/measures of variability	Total PEDro rating Score
1. Doesborgh et al., 2004	1	1	1	1	0	0	1	1	1	1	1	**8**
2. Cherney et al., 2014	1	1	0	1	0	0	0	1	1	1	1	**6**
3. Fridriksson et al., 2009	1	N/A	0	N/A	0	0	0	1	1	1	1	**4**
4. Manheim et al., 2009	1	N/A	0	N/A	0	0	1	1	0	0	1	**3**
5. Marshall et al., 2013	1	N/A	0	N/A	0	0	1	1	1	1	1	**5**
6. Palmer et al., 2012	1	1	1	1	0	0	1	1	1	1	1	**8**
7. Thompson et al., 2010	1	0	0	1	0	0	0	1	1	1	1	**5**
8. Wenke et al., 2014	1	0	0	0	0	0	0	1	1	1	1	**4**
TOTALS	**8**	**3**	**2**	**4**	**0**	**0**	**4**	**8**	**7**	**7**	**8**	**N/A**

Table 9. Study design and rating score for single case experimental studies

Study Authors and Year	Clinical history	Target behaviours	Design	Baseline	Sampling during treatment	Raw data	Inter-rater reliability	Assessor independence	Statistical analysis	Replication	Generalisation	Totals
9. Choe & Mathy, 2010	1	1	1	0	1	1	1	0	1	1	1	**8**
10. Choe & Stanton, 2011	1	1	1	0	1	1	1	0	1	1	1	**8**
11. Choe et al., 2007	1	1	1	0	0	1	1	0	1	1	0	**6**
12. De Luca et al., 2014	1	0	1	0	0	1	0	0	0	0	1	**3**
13. Fink et al., 2002	1	1	1	1	1	1	1	0	1	1	1	**9**
14. Kurland et al., 2014	1	1	1	1	1	1	0	0	0	1	1	**7**
15. Laganaro et al., 2003	1	1	1	0	0	1	0	0	1	1	1	**6**
16. Linebarger et al., 2001*	1	1	1	0	0	1	0	0	1	0	1	**5**
17. Pedersen et al., 2001	1	1	1	1	0	1	0	0	1	1	1	**7**
18. Ramsberger & Marie, 2007	1	1	1	1	1	1	0	0	1	1	1	**8**
19. Routhier et al., 2016	1	1	1	1	1	1	1	0	1	1	1	**9**
20. Wieczorek et al., 2011	1	1	1	1	0	1	0	0	1	1	1	**7**
TOTALS	**12**	**11**	**12**	**6**	**6**	**12**	**5**	**0**	**10**	**10**	**11**	**N/A**

Table 10. Therapy characteristics and outcomes for group studies (continued overleaf)

Study authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy	Comparisons	Outcome measures	Summary of key findings
1. Cherney et al., 2014	Script acquisition	AphasiaScripts – Dialogue scripts trained and delivered by a 'digital therapist' who could model speech and interactively guide treatment. Treatment comprises three parts: listen silently to whole conversation as it appears – in text form – on the screen; repeatedly practise each turn of the script – first in unison with 'therapist' then independently; entire conversation rehearsed whilst taking turns with the 'therapist' Six personalised dialogue scripts of equal length and grammatical complexity developed per person.	90 min / day for 6 days a wk, over 3 wks for each training condition (high-cue and low-cue)	Estimated 27 hrs / condition (two conditions). i.e. total = 54 hrs	High-cue computer therapy vs. low-cue computer therapy.	Accuracy of script acquisition measured using NORLA-6 (Gingrich, Hurwitz, Lee, Carpenter, & Cherney, 2013). Rate of script production established through automatic capture of production duration using computer.	Significant improvement in script acquisition and rate of production for both conditions (high-cue and low-cue). No significant difference between conditions. Maintained at 3 and 6 wks. Larger effect sizes demonstrated for more severe group.
2. Doesborgh et al., 2004	Spoken naming	Multicue – a choice of 80 pictures (high and low frequency words) presented randomly. Choice to utilise up to four prompts to cue naming: semantic, orthographic, sentence completion, distraction (take a break).	30-45 minute sessions two – three times / wk for 8 wks	10 to 11 hrs	Computer treatment vs. no treatment.	Boston Naming Test– spoken naming performance for 60 items. Each response scored on a 4 point rating scale (0, 1, 2 or 3 points). Secondary measure of verbal communicative ability: Amsterdam Nijmegen Everyday Language Test, scale A (understandability). BNT; ANELT-A (Kaplan et al., 1983; Blomert, Kean, Koster, & Schokker, 1994)	Treated group improved significantly on the BNT. Untreated (control) group did not. Mean improvement was not significantly different between groups however. Neither group improved on the ANELT-A
3. Fridriksson et al., 2009	Spoken naming	Computer program presenting 18 nouns delivered in a 3-level hierarchy. Level 1 – six items randomly presented 12 times; level 2 – six previous plus an additional six new items presented randomly eight times; level 3 – 12 previous items plus additional six items presented randomly five times. 2 conditions treated with this procedure – For Audio Visual (AV) condition, picture shown then video with audio of word	30 min/day, 5 days/wk logged. Minimum 5 sessions / level (3 levels). Maximum 15 sessions / level	Minimum 7.5 hrs; maximum 22.5 hrs / condition. 2 conditions i.e. totals = min 15 hrs and max 45 hrs	Audio-visual computer treatment vs. audio only computer treatment.	Spoken naming of the same set of 36 imagable words – (18 treated in audio visual condition AV; 18 treated in audio only condition AO) were scored as correct or incorrect. Additionally, the 175-item Philadelphia Naming Test, PNT (Roach, Schwartz, Martin, Grewal, & Brecher, 1996) was used to assess generalisation.	Significant improvement observed for AV items but not for AO items. AV condition improved significantly more than AO condition. Significant improvements of PNT after AV phase but not AO phase, indicating generalisation of AV treatment effect.

Study authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy	Comparisons	Outcome measures	Summary of key findings
		presented directly afterwards, for Audio Only (AO) condition picture shown then audio of word presented directly afterward.					
4. Manheim et al., 2009	Conversational skills	AphasiaScripts – Dialogue scripts trained and delivered by a 'digital therapist' who could model speech and interactively guide treatment. Treatment comprises three parts as described above within Cherney et al study. Three personalised dialogue scripts developed per person.	30 min/day over 9 wks (63 days) suggested. Logged times not reported.	Estimated 31.5 hrs	No treatment period vs. computer treatment period.	No operationally defined target behaviour identified. Patient-reported communication difficulty assessed using sub-scales of the Burden of Stroke Scale, BOSS (Doyle et al., 2004)	Significant improvement (decrease) in BOSS scores indicating reduced levels of communication difficulty.
5. Marshall et al., 2013	Communicative gesture	GeST – 30 gestures trained across two conditions (15 therapist supported; 15 unsupported). Three levels of practice: repetition of gesture demonstrated in video; repetition of gesture within virtual, responsive environment; repetition of gesture in response to everyday scenario. Gesture recognition technology employed to provide feedback on performance.	64.4 sessions over 6 wks – average 10 sessions / wk of around 13 min / session (logged)	13.9 hrs average	Items treated by computer + therapist vs.: Items treated by computer only, Items treated by therapist only, untreated items.	Pantomime gesture production of the same set of 60 imagable words – (15 treated by SLT + computer; 15 treated by computer only; 15 treated by SLT only; 15 untreated) were scored as correct or incorrect and rated 1-5 for accuracy. Untrained items and identical picture naming task used to assess for generalisation.	Significant improvement on gesture production of items treated by SLT + computer, but no others. Gains were maintained 6 weeks post-intervention. No changes observed for accuracy rating over time. Untrained items and naming items did not change.
6. Palmer et al., 2012	Generalised language therapy – measured though spoken naming output	StepbyStep exercises as configured by an SLT and supported by a volunteer. Selection of over 13,000 exercises available following steps from listening to target words, producing words using cues through to saying words in sentences.	20 min / day, 3 days / wk suggested for five months (approx. 20 wks)	Estimated 20 hrs	Computer treatment vs usual care (i.e. no impairment-focused treatment)	Spoken naming performance on 48 items from the Object and Action Naming Battery (Druks, 2000). All 48 items were treated in treatment group and untreated in control group	Treated group improved significantly more on the naming measure than untreated group at 5 months follow-up. This difference was not maintained at 8 months. Excluding pts with <10% naming performance at baseline, increases scale of increase at 5 months.
7. Thompson et al., 2010	Complex sentence production	Sentactics ® - Virtual clinician (Sabrina) presents a series of tasks. 14 sentences trained. First; sentence-to-picture matching task, second; sentence production 'priming task' showing two related pictures and presenting a model phrase to describe the first (e.g. "It was the woman who the man saved") and prompting, in relation to the	2 days X 2-hr sessions / wk for 4-5 wks	16 to 20 hrs	Computer treatment vs. no treatment.	Elicitation of 30 sentences using sentence production priming procedure test within "Sentactics" program. (10 trained 'object relatives' items, 10 untrained 'object clefts' and 10 untrained 'object wh-questions'.) Secondary measure of sentence	Treated group improved significantly on sentence production measure. Untreated (control) group did not. Mean improvement was significantly different between groups. Improvement was significant for treated items and untreated

Study authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy	Comparisons	Outcome measures	Summary of key findings
		second picture, "For this picture you could say…"				comprehension was the Northwestern Assessment of Verbs and Sentences. (Cho-Reyes & Thompson, 2012)	'object wh-questions' but not for untreated 'object clefts', indicating some generalisation
8. Wenke et al., 2014	Generalised spoken language output	Computer-based therapy employed using "a range of software programs including REACT-2, Aphasia Tutor, Language Links and Synonyms, Homonyms and Antonyms to target individual goals."	Intense therapy condition (computer therapy group; group therapy group; SLP therapy assistant group): 9 hrs / wk for 11 wks (4 to 5 1-hr sessions with a therapist +1-1.5 hrs computer/group/assistant practice session) logged by researchers Standard therapy condition: 3 1-hr sessions / wk for ~ 8 wks	Intense therapy group (computer therapy group; group therapy group; SLP therapy assistant group): estimated 99 hrs total comprising 40-50 hrs therapist work + 49-59 hrs computer work. Standard therapy condition: estimated 24 hrs	Intense therapy (9hrs/wk) vs. non-intense therapy (3hrs/wk). Intense therapy condition utilised either computer therapy, group therapy or SLT assistant group therapy. Non-intense comprised clinical, non-computer therapy.	Spoken language subtests of the Comprehensive Aphasia Test (CAT). Additional outcome measure – the Disability Questionnaire of the CAT. (Swinburn et al., 2004)	Pts in all conditions (computer therapy group; group therapy group; SLP therapy assistant group; non-intense, standard therapy group) improved significantly on spoken language sub-tests of the CAT. No significant differences were observed between groups. Intense therapy condition (computer therapy group; group therapy group; SLP therapy assistant group) improved (decreased) significantly on the Disability Questionnaire, regular therapy group did not.

Note: primary outcome measures for Palmer et al., 2012 were measures of feasibility – recruitment rate, completion rates and statistical variability. Clinical measures were reported as secondary outcomes. Wenke et al., 2014 similarly assessed costs and consumer satisfaction in addition to clinical outcomes. For the purposes of this review, key clinical outcomes are reported only.

Table 11. Therapy characteristics and outcomes for single case experimental studies (continued overleaf)

Study Authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy hours	Comparisons	Outcome measures	Summary of key findings
9. Choe & Mathy, 2010	Spoken naming	PowerPoint (ppt) show offering 5 picture items with increasing levels of cue. Cues were provided as both audio clips and accompanying written text. Five cues of increasing level of support were provided for each item in order: "What is this?"; initial phoneme; full word; full word X 5; full word X five + instruction to "say it with me".	Pt 1. 8.5 hrs ppt + 7.7 hrs AAC / wk Pt 2. 19.2 hrs ppt +14.6 hrs AAC /wk Pt 3. 13.5 hrs + 3 hrs AAC / wk. As logged by pt for 15 wks of treatment	Estimated figures: Pt 1. 127½ hrs ppt; 115½ hrs AAC Pt 2. 288 hrs ppt; 219 hrs AAC Pt 3. 202½ hrs ppt; 45 hrs AAC	Items treated by computer vs. items practised using AAC device vs. untreated control items.	30 imagable words – tailored for each pt were scored for naming success using Porch Index of Communicative Ability (Porch, 1981) (10 trained on ppt, 10 trained on AAC, 10 untrained).	Items practiced on computer (ppt) improved significantly. AAC items and control items did not.
10. Choe & Stanton, 2011	Spoken naming	PowerPoint (ppt) show offering 10 picture items with increasing levels of cue. Cues were provided as either audio visual (AV) video clips of a person speaking or audio only (AO) clips – depending on the condition. Six cues of increasing level of support were provided for each item in order: "What is this?"; semantic clue; initial phoneme; full word; full word X 5; full word X five + instruction to "say it with me".	One pt (TV) had 20 (30 min) sessions over 4 wks. One pt (ML) had 10 sessions (30 min) over 4 wks.	10 hrs for one pt. 5 hrs for second pt	Audio-visual computer treatment vs. audio only computer treatment.	30 imagable words – tailored for each pt were scored for naming success using PICA (Porch, 1981) (10 trained in auditory only (AO) ppt condition; 10 trained in auditory visual (AV) condition; 10 untrained)	TV made significant improvement in AV condition immediately post therapy, and maintained at 4-wk maintenance and 4-mnth follow-up. AO condition significant improvement at 4-wk maintenance only and changes in control condition not significant. ML demonstrated significant improvement in both AO and AV conditions immediately post therapy and all three conditions (AO, AV, control) at 4-wk maintenance and 4-mnth follow-up
11. Choe et al., 2007	Spoken naming	PowerPoint (ppt) show offering 10 picture items with increasing levels of cue. Cues were provided as both audio clips and accompanying written text. Seven cues of increasing level of support were provided for each item in order: "What is this?"; semantic clue; initial phoneme; full word; full word x 5; full word x five + instruction to "say it with me": full word x 5 + drawing of place of articulation for initial sound.	Pt 1. 39 min / day Pt 2. 24 min / day Pt 3. 34 min / day Pt 4. 20 min / day As logged by pt for 14 wks of treatment	Estimated figures: Pt 1. 63.7 hrs Pt 2. 39.2 hrs Pt 3. 55.5 hrs Pt 4. 32.7 hrs	Computer practice vs. weekly therapist practice (non-computer) vs. untreated control items.	30 imagable words – tailored for each pt were scored for naming success using PICA (Porch, 1981) (10 ppt home practice items; 10 once-weekly therapist practice items; 10 untrained)	2 of 4 pts improved significantly on ppt items immediately post-therapy rising to 3 after 5-wk maintenance. 1 of the above 3 pts improved significantly on weekly therapist items immediately post-therapy but this was not maintained. No improvements were observed for control items.
12. De Luca et al., 2014	Generalised language therapy	Power AFA - PC based tool comprising multiple tasks: "sound exercises, image recognition exercises, word and letter	6 sessions / wk for 12 wks	72 sessions. Duration of sessions not	Computer treatment + therapist	No target measure identified. Full battery administered pre and post therapy including Mini Mental	No statistical analysis was provided. Post-therapy improvements were reported on measures of written

Study Authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy hours	Comparisons	Outcome measures	Summary of key findings
		exercises, verbs and action exercises.”		stated.	treatment vs. therapist treatment (no computer) only	State Exam. Attentive Matrices, Trial Making Test A & B and Reversal Motor Learning, Digit Span Test, Aachen Aphasia Test, Hamilton Rating Scale for Depression and Coping Orientation to Problems Experienced – New Italian Version. MMSE; AM; TMT & RML; DST; AAT; HRSD; COPE; (Folstein, Folstein, & McHugh, 1975; Spinnler & Tognoni, 1987; Giovagnoli et al., 1996; Orsini et al., 1987; Huber et al., 1984; Hedlund & Vieweg, 1979; Sica et al., 2008).	language, naming, comprehension, repetition and the token test from the AAT. Additional improvements were observed in MMSE, AM, DST and COPE.
13. Fink et al., 2002	Spoken naming	MossTalk Words – cued naming module. Choice to utilise up to six prompts to cue naming: three spoken (initial phoneme, sentence completion, whole word repetition), and three written (as above but in print). Two groups of 3 pts; One clinician guided (CG) and one partially self-guided (PSG). 40 items treated over two phases of 20 each.	Sessions 30-45 min. Criterion performance. Pts had to get 17/20 correct on 3 of 4 consecutive sessions or stop at 4 wks. 3 sessions / wk for up to 4 wks across 2 conditions: three full clinician-guided (CG) sessions / wk, 1 or 1 clinician guided and 2 independent sessions /wk (PSG).	Estimated figures: Minimum possible duration = 4 X 30 min = 2hrs (1 hr clinician guided; 1 hr independent). Maximum possible duration = 12 X 45 min = 9hrs (3 hrs clinician guided; 6 hrs independent).	Computer practice under full therapist guidance vs. computer practice with some therapist guidance and some independent practice.	Spoken naming of 40 imagable words – tailored for each pt (20 treated; 20 untreated after phase 1; all 40 treated by end of phase 2) were scored as correct if target or acceptable alternative was produced. Non-target responses scored as errors. Additionally, a 339-item picture-naming test from MossTalk program assessed for generalisation.	5 of 6 pts demonstrated medium to large naming gains for treated items after phase 1, maintained at post phase 2 follow-up. Untrained items did not improve after phase 1. Acquisition and maintenance effects were greater in the CG group than the PSG group.
14. Kurland et al., 2014	Spoken naming	Following 2 weeks of offline intensive language action treatment (ILAT) with a therapist, 2 iBooks files created for each pt for delivery via iPad. First contained 20 object items, second contained 20 action	20 min / day, 5 to 6 days / wk for 6 months suggested (approx. 26 wks)	Estimated 43 1/3 hrs (min) to 52 hrs (max)	Items treated by a therapist then practised on computer vs. items treated by a	Spoken naming of 80 imagable words – tailored for each pt (20 treated by SLT and practised with iPad TR-PR; 20 not treated by SLT and practised with iPad UNTR-PR; 20 treated by SLT and	No statistical analysis was provided. 3 of 8 pts not included in analysis. 5 pts demonstrate ‘strong evidence of treatment effect’ for UNTR-PR items, “Clinical significance was achieved on all practice words.”

Study Authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy hours	Comparisons	Outcome measures	Summary of key findings
		items. 10 items from each iBook were previously trained with ILAT, 10 were untrained. Treatment comprises four stages per item: naming from repetition and cue; word-to-picture matching; rhyming/spelling task; semantic odd-one-out.			therapist then not practised on computer vs. items not treated by a therapist but practised on computer vs. items both untreated by therapist and unpractised on computer.	not practised with iPad TR-UNPR; 20 not treated by SLT and not practised with iPad UNTR-UNPR) Untrained items assessed to test for generalisation.	
15. Laganaro et al., 2003	Spoken naming	Three different types of computer aided treatment employed; Receptive metaphonological tasks – syllable count judgement and initial phoneme identification; picture naming with graphemic cues available; written naming with next-letter prompts available.	4 chronic pts: 2-3 Computer Aided Therapy (CATh) sessions / wk for 2 wks 7 acute pts: daily CATh therapy in addition to daily clinical therapy (Session duration not stated)	Estimated figures: Chronic pts: minimum 4 sessions, maximum 6 sessions. Acute pts: minimum 10 sessions, maximum 14 sessions. (Session duration not stated)	Clinical (non-computer) therapy vs computer therapy.	Spoken naming of 144 imagable words (48 treated by computer; 48 treated by SLT; 48 untreated). Untrained items and verbal fluency task used to assess for generalisation.	Chronic pts: significant effect for items treated by computer for 3 of 4 pts. 2 of 4 also had SLT treatment - which also had significant effects for trained items. Acute pts: significant effect for items treated by computer for 3 of 7 pts. Significant effect for items treated clinically for 1 of 7 pts. No improvements observed for untrained items or lexical fluency.
16. Linebarger et al., 2001	Spoken use of locatives and preposition s	Natural Language Processing software + augmentative communication system. 5 separate experiments reported. Expt 1: Computer task involving picture description to generate a sentence with a locative component. Computer fed back to user on accuracy of spoken production. Followed by sentence verification task (i.e. "is the following sentence correct for the above picture?").	Expt 1: 3-5 sessions / wk for 10 wks. Sessions lasting 15-40 min each. (Family reports and limited logging data) Expt 2: 30 hrs total. Time to complete 30 hrs not reported. Expt 3: 15 hrs total.	Estimated figures: Expt 1: Minimum 7.5 hrs, maximum 33.3 hrs. Reported Figures: Expt 2: 30 hrs Expt 3: 15 hrs Expt 4:not	Experiments 1-4: items treated within computer therapy vs. untreated items. Experiment 5. No clear comparison condition.	For experiments 1-4: Sentence production test from pictures. Items scored as correct or incorrect. Untrained items used to assess for generalisation. Expt 1. 50 pictures intended to elicit single locative assertions (12 trained; 10 untrained but based on trained prepositions; 18 simple active transitives; 10 dative	Expt 1. 2 of 3 pts showed significant increase in performance for the 22 trained + untrained locative assertions. Expt 2. Significant increase in correct picture descriptions and correct prepositions. Expt 3. Both pts showed significant increase in performance for 42 trained + untrained pictures. Expt 4. Both pts showed significant

Study Authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy hours	Comparisons	Outcome measures	Summary of key findings
		Expt 2: as above with additional preposition-to-picture matching task included prior to sentence generation task. Prepositions trained instead of locatives. Expt 3: as for expt 2, expanded to include further locative prepositions. Expt 4: as for expt 3 but training locative and directional modifiers in transitive structures. Expt 5: assessment of the communication system (a pre-cursor to sentence shaper). Pts encouraged to compose narratives on topics of personal interest using sentence construction prosthesis which allowed them to record short utterances and configure them into longer productions for playback. Encouraged to include trained prepositions in productions.	Time to complete 30 hrs not reported. Expt 4: not reported Expt 5: Pt 1: Condition 1: 3 sessions / wk of 45 min for 7 wks. Condition 2: 2.5 sessions / wk of 64.8 min for 8 wks. Pt 1: Condition 1: 5 sessions / wk of 43 min for 5 wks. Condition 2: 6.3 sessions / wk of 48 min for 2.5 wks.	reported Expt 5: Pt 1: Condition 1: 17.2 [15.8] hrs Condition 2: 22.5 [21.6] hrs Pt 2: Condition 1: 16.4 [17.9] hrs Condition 2: 13.5 [12.6] hrs [Note: totals reached by multiplying reported intensity for expt. 5 disagree with reported total duration of therapy hrs. Calculated estimates are provided in square brackets after reported figures]		structures). Expt 2. 36 pictures (18 trained; 18 untrained). Expt 3. 52 pictures (21 trained; 21 untrained but based on trained prepositions; 10 simple active sentences). Expt 4. 36 pictures (9 trained locatives; 9 untrained locatives; 9 trained directionals; 9 untrained directionals). Expt 5. Boston Naming Test (BNT), Noun/Verb naming test (NVNT). Narrative production test from silent film. Narratives analysed using Quantitative Production Analysis (QPA). (BNT, NVNT, QPA Kaplan et al., 2001; Zingeser & Berndt, 1990; Saffran, Berndt, & Schwartz, 1989)	increase in performance for trained + untrained locatives. One patient went on to receive further treatment for directional items and showed significant increase in trained + untrained directional items. Expt 5. Statistical analysis not provided. BNT: pt 1 showed strong gains, pt 2 showed small gains. NVNT: both pts showed increase in verbs but not nouns. QPA: pt 1 made 'striking' improvements, pt 2 made 'solid gains' on narrative without the augmentative system.
17. Pedersen et al., 2001	Spoken naming	Two custom-built programs employed offering three different types of task in succession: semantic, phonological and written naming. Semantic tasks involved word-to-picture matching. Phonological tasks involved word-to-picture matching and written letter word completion. Written naming tasks involved typed copying; re-ordering from an anagram and written picture naming.	P1. 39+ hrs over 24 wks (not all were logged). Pt 2. 201 hrs over 28 wks. P3. 59 hrs over 9 wks.	P1. 39+ hrs (not all were logged) Pt 2. 201 hrs P3. 59 hrs	Items treated within computer therapy vs. untreated items.	Spoken naming of 260 imagable words (S&V, Pt 1. 101 trained; 159 untrained; Pt 2. 184 trained; 76 untrained; pt 3. 68 trained; 192 untrained). Items were scored as correct or incorrect. Untrained items and Western Aphasia Battery Aphasia Quotient (WAB AQ) used to assess for generalisation. S&V, WAB AQ (Snodgrass & Vanderwart, 1980; Kertesz, 1982)	Pt 1: Significant improvement on treated words and significantly more improvement in treated words vs untreated words. Effect maintained after 3 mnths. No change for WAB AQ. Pt 2: Significant improvement on treated words. Effect maintained after 2 mnths. Improvement on WAB AQ observed but significance level not reported. Pt 3: Significant improvement on treated words. Effect maintained after 1.5 mnths. No change for WAB AQ.
18. Ramsberger &	Spoken naming	MossTalk Words – cued naming module. Choice to utilise up to eight prompts to cue	2 pts had 15 sessions over 8 wks and	First 2 pts: 11.25 hrs (min)	Intense computer	80 imagable words – (40 treated in phase 1; 40 treated in phase 2)	Significant increase in naming for 3 of 4 pts after phase 1 and significant

Study Authors and Year	Aphasic symptom targeted	Description of computer therapy	Reported treatment frequency	Total duration of therapy hours	Comparisons	Outcome measures	Summary of key findings
Marie, 2007		naming: four spoken (initial phoneme, sentence completion, whole word repetition, spoken description of item), and four written (as above but in print). Two conditions: One intense (5 sessions / wk) and one nonintense (2 sessions / wk). 80 items treated over two phases of 40 each.	another 15 sessions over 3 wks. A further two pts had 20 sessions over 10 wks and another 20 sessions over 4 wks. (Sessions lasted 45-60 min)	to 15 hrs (max) / condition (2 conditions). 22.5 -30 hrs total. Second 2 pts: 15 hrs (min) to 20 hrs (max) / condition (2 conditions). 30-40 hrs total.	therapy (5 sessions/wk) vs. non-intense therapy (2 sessions/wk). Also, items treated within computer therapy vs. untreated items.	tailored for each pt were scored as either fully correct (1point) fully correct after a delay or self-correction (½ point) or incorrect (0 points). Untrained items used to assess for generalisation.	increase again for 1 of these 3 pts after phase 2. No evidence of increase for untrained items in any of the pts.
19. Routhier et al., 2016	Spoken verb naming	PowerPoint (ppt) show offering 25/31 video items with increasing levels of cue to prompt the verb. 5 cues of increasing level of support were provided for each item in order: Watch the action; listen to an 'inducing sentence'; initial syllable/initial phoneme; written word; spoken word. Two conditions, one was cued as above, the other – named repetition trained a separate 25/31 video items employing just the first cue alone (watch the action).	20 sessions over 5 wks (approx. 60 min / session) suggested. Estimated 4 sessions / wk.	20 hrs	Cued computer practice items vs repeated computer practice items vs untreated items.	5-second videos used to elicit single spoken verbs (all were treated). Items were scored as correct or incorrect. Pt was tested on 75 items; Pt 2 tested on 93 items. (Each item list was split evenly into three groups: trained; repeated; control) A verb-to-verb production task was used using 113 untrained verbs to elicit semantically matched or unmatched verbs.	Pt 1. Significant increase in verb naming immediately post-therapy for treated items. Effect maintained. No gains for repeated or control items and no generalisation to verb-to-verb task. Pt 2. Significant increase in verb naming between baseline and maintenance for treated items. No gains for repeated or control items and no generalisation to verb-to-verb task.
20. Wieczorek et al., 2011	Use of aspect – tense production	Computerised version of tasks undertaken in live therapy. Picture shown illustrating a prospective, on-going or past tense action. Written sentence frame provided with choice of 6 possible real words available to complete the sentence. Once an option was selected, the whole sentence was read aloud and pt had to judge whether or not the created sentence was grammatical and correctly matched the picture.	10 X clinician sessions of 30 min over 2 wks + computer 'homework' to be completed between sessions. Duration of homework not reported	5 hrs clinician therapy + unreported amount of computer-based 'homework'. Duration of home practice not reported.	Items treated within computer therapy vs. untreated items.	20 triplets of imagable verbs (prospective, on-going and past tense) – tailored for each pt, were used to elicit sentences (10 trained; 10 untrained) Untrained items used to assess for generalisation.	For 2 pts: Significant improvement on trained and untrained items. For 1 pt: Significant improvement on trained items only. For 1 Pt: no improvement.

Note. For reasons of brevity, outcome measures reported are limited to only the key outcome measure stated plus one selected generalisation measure. Generalisation measure selected firstly on the basis of it being the closest to the key outcome and secondly as indicated by clinical interest. Where no key outcome measure is reported (De Luca et al., 2014) all measures employed are reported.

3.6.5 Synthesis of results

3.6.5.1 Dates of research

Study publication dates range from 2001 to 2016.

3.6.5.2 Focus of intervention

Eleven of 20 studies identify spoken naming ability as the focus of their therapy. Of the remaining nine, just one study targets communicative gesture (Marshall et al., 2013); two target spoken grammar (Linebarger et al., 2001; Wieczorek et al., 2011); three report generalised language therapy/spoken language therapy (De Luca et al., 2014; Palmer et al., 2012; Wenke et al., 2014), one targets complex sentence production (Thompson et al., 2010); one targets script acquisition (Cherney et al., 2014) and one conversational skills (Manheim et al., 2009).

3.6.5.2 Programs used for therapy

Computer therapies reported include a range of programs – some of which are commercially available and some of which were custom built for the study reported.

AphasiaScripts (Rehabilitation Institute of Chicago, 2007) and MossTalk Words (Fink, Brecher, Montgomery, & Schwartz, 2001) were tested in two studies each (AphasiaScripts: Cherney et al., 2014; Manheim et al., 2009; MossTalk Words: Fink et al., 2002; Ramsberger & Marie, 2007). Custom shows created using Microsoft Office PowerPoint software were used to deliver therapy practice in four studies (Choe et al., 2007; Choe et al., 2010; Choe & Stanton, 2011; Routhier et al., 2016). Evidence for remaining therapies is limited to one study per technique. Stimuli within nine programs were tailored to each individual participant (typically by a speech and language therapist). Six studies delivered the same programme and stimuli to all participants involved in their given project. Five did not specify whether training stimuli were uniform for each participant or tailored to the individual.

3.6.5.4 Outcome measures

For the majority of studies reported (14 of 20), the primary clinical outcome measures were created for the purposes of the study and do not represent clinically standardised measures. Four studies use standardised measures to assess for clinical change (Cherney et al., 2014; Doesborgh et al., 2004; Palmer et al., 2012; Wenke et al., 2014) and two did not identify primary clinical outcome measures – reporting outcomes from a variety of standardised measures instead (De Luca et al., 2014; Manheim et al., 2009).

3.6.5.5 Key Findings

Seventeen of 20 studies used statistical methods to report findings for primary outcome measures. Two of 20 did not report statistics (De Luca et al., 2014; Kurland et al., 2014). Additionally, one of the aforementioned articles (De Luca et al., 2014) and one other remaining study did not identify a primary outcome measure (Manheim et al., 2009).

Literature Review Question 1. What is known about the effect of remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes when compared to a clinical non-computer therapy, an alternative computer therapy or no therapy/usual care?

Evidence for treatment effect of computer-delivered therapy for non-orthographic features of aphasia upon expressive language outcomes

Comparisons

A number of contrasting conditions were used as comparators within studies. Seven studies report comparisons of computer therapy versus either a no treatment/usual care control condition or set of untreated items (Doesborgh et al., 2004; Linebarger et al., 2001; Manheim et al., 2009; Palmer et al., 2012; Pedersen et al., 2001; Thompson et al., 2010; Wieczorek et al., 2011). Six studies report comparisons of computer therapy versus an alternative treatment - two studies report comparisons with a clinical non-computer treatment delivered by a speech and language therapist (De Luca et al., 2014; Laganaro, Di Pietro, & Schnider, 2006); four studies report comparisons with an alternative computer treatment (Cherney et al., 2014; Choe & Stanton, 2011; Fink et al., 2002; Fridriksson et al., 2009). The remaining seven studies report multiple comparisons – typically including an untreated condition plus one or more of the other afore mentioned comparisons (Choe et al., 2007; Choe et al., 2010; Kurland et al., 2014; Marshall et al., 2013; Ramsberger & Marie, 2007; Routhier et al., 2016; Wenke et al., 2014).

Statistical Analyses

Three studies did not report statistical analyses (29 participants). One reports outcomes for treatment in relation to no therapy (Manheim et al., 2009) and two report therapy in relation to alternative treatments (De Luca et al., 2014; Kurland et al., 2014). These studies describe three different therapy approaches: generalised language therapy (Power AFA - De Luca et al., 2014), spoken naming for objects and actions (iBooks - Kurland et al., 2014) and conversation skills (AphasiaScripts - Manheim et al., 2009). Due to the lack of statistical analysis reported, evidence from these studies has not been used here to examine treatment effects for the stated approaches. Additionally, Linebarger et al. (2001) analyse outcomes by combining both treated and untreated items and so no clear conclusions about treatment effects can be observed.

Of the 16 studies that reported statistical analyses (161 participants), 10 report outcomes for computer treatment in relation to no therapy (five single comparisons + five mixed comparisons), 10 report outcomes in relation to alternative treatments (five single comparisons + five mixed comparisons).

Results for no treatment comparisons

We shall first look at outcomes immediately post-therapy for the 10 studies comparing computer therapy outcomes to a no therapy/usual therapy condition or to items that received no input (92 participants). Six report a significant improvement for items treated by computer therapy. A further four single case designs report improvements for some but not all participants (nine of 14 participants - Choe et al., 2007; Ramsberger & Marie, 2007; Routhier et al., 2016; Wieczorek et al., 2011)

Improvements reported for seven of the above 10 studies relate to single word spoken naming for spoken noun or verb production. Of the remaining three studies, Marshall et al. (2013) report a significant effect of computer treatment for gestures, Thompson et al.(2010) a significant effect of computer treatment for complex sentence production and Wieczorek et al. (2011) a significant effect of computer-based therapy targeting spoken grammar.

Results for alternative treatment comparisons

Ten of the reviewed studies compared computer therapy outcomes to an alternative therapy condition – either clinical or computer-based (87 participants).

Seven studies report a significant improvement for items treated by computer therapy (Cherney et al., 2014; Choe et al., 2010; Choe & Stanton, 2011; Fink et al., 2002; Fridriksson et al., 2009; Marshall et al., 2013; Wenke et al., 2014). A further three single case designs report improvements for some but not all participants (6 of 10 participants - Choe et al., 2007; Laganaro et al., 2003; Routhier et al., 2016).

Investigating individual comparisons, Choe et al.(2007), Laganaro et al. (2003) and Marshall et al.(2013) reported outcomes comparing computer-delivered therapy and clinician-delivered therapy. Choe et al. (2007) found that one of four participants who received therapy for spoken naming made significant gains in the clinical therapy (no computer) condition and two (including the one who had made gains through clinical therapy) made gains through computer therapy. Laganaro et al. (2003) found that two chronic participants who received therapy for spoken naming made significant gains in the clinical therapy (without computer) condition as well as in the computer therapy condition. Seven participants with acute aphasia were also tested. Of these, one made gains with clinical therapy (no computer) and three (including the one who had made gains through clinical therapy) made gains through computer therapy. Marshall et al.(2013) found that participants made significant gesture gains for items trained using computer therapy in combination with therapist support, but no gains for items familiarised briefly by the therapist alone or items trained using the computer but no therapist support.

Turning now to comparisons against other forms of computer therapy, Choe et al. (2010) found a naming effect for computer therapy but not AAC practice and Choe & Stanton (2011) & Fridriksson et al. (2009) found a naming effect for items practised with audio visual training stimuli but only a limited (1 of 2 participants - Choe & Stanton, 2011) or no effect of audio only computer training stimuli (Fridriksson et al., 2009). Cherney et al. (2014) found that script acquisition improved for both a high-cue computer therapy and a low cue computer therapy, whereas Routhier et al.(2016) found that verb naming improved for a high cue computer therapy but not for a repetition only therapy. Finally, Fink et al. (2002) found significant naming improvements for both fully-supported computer therapy and partially self-guided computer therapy and Wenke et al. (2014) found significant spoken language improvements for participants treated in four therapy conditions: computer therapy, group therapy, speech and language therapy assistant group therapy and non-intensive clinical speech and language therapy (no computer)

Again, improvements reported for seven of the above 10 studies relate to single word spoken naming for spoken noun or verb production. Of the remaining three studies, Marshall et al. (2013) report a significant effect of computer treatment for gestures, Cherney et al.(2014) a significant effect of computer treatment for script acquisition and Wenke et al. (2014) a significant effect of computer-based therapy targeting generalised spoken output.

Evidence for generalisation to untreated items

Twelve studies report analysis of untreated control items (76 participants). Of these, 10 report no significant improvement in untreated items after computer therapy (66 participants). For the remaining two studies, Thompson et al. (2010) report generalisation from treated to untreated 'object wh-questions' following complex sentence training with Sentactics® and Wieczorek et al. (2011) report significant improvement for untreated aspect and tense items for two of their four participants.

Literature Review Question 2. What is known about the durability/maintenance of effects achieved using remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes?

Evidence for maintenance of treatment effects

Maintenance effects are reported for 11 studies (68 participants). Of these, eight report a significant improvement when comparing performance for treated items after a wash out / maintenance period to performance at baseline / pre-therapy (45 participants). Two studies report maintenance of therapy effects for some but not all participants (two out of six participants whose analyses were reported - Choe et al., 2010; Ramsberger & Marie, 2007 . One additional participant, reported by Choe et al.(2010), did not complete maintenance assessment. Maintenance outcomes for this participant were therefore not reported or analysed). One study reports evidence that gains are not maintained (16 participants - Palmer et al., 2012). Maintenance periods reported vary in length from two weeks to eight months (mean reported length = 7.3 weeks). It is worth noting that the study that reported no maintenance of effects (Palmer et al., 2012) was that with the longest maintenance period (eight months). Choe et al.(2010), who showed persisting effects for just one of their two participants, had a maintenance period of four to eight weeks and Ramsberger & Marie (2007), showed maintained improvement on naming for one of their four participants although they did not report the duration of their 'treatment withdrawal' maintenance phase.

Literature Review Question 3. What is the range in the duration and intensity of practice reported for computer therapy?

Reporting of therapy frequency and duration.

Using all of the reported information, we observe that therapy was conducted over a variety of periods ranging from 2 to 26 weeks, with the reported number of sessions undertaken ranging from 4 to 66. Total duration of computer practice reported ranges greatly from 2 to 288 hours. Therapy duration was recorded and reported in a variety of ways, however. For four studies, computer logs were used to report the amount of practice undertaken at the computer (Fridriksson et al., 2009; Linebarger et al., 2001; Marshall et al., 2013; Pedersen et al., 2001). Average duration of computer practice as logged by computers is approximately 40.3 hours on average and the range of logged duration goes from approximately 14 hours over 6 weeks (Marshall et al., 2013) to 100 hours over 9 to 28 weeks (Pedersen et al., 2001).

For two studies, (Choe et al., 2007; Choe et al., 2010), records were based on paper logs kept by the participant or their family during practice. The mean duration of total logged practice here was 206 hours over 15 weeks and 48 hours over 14 weeks respectively (mean=127 hours). Three studies report only the suggested practice schedule and do not state observed amount of therapy undertaken (Kurland et al., 2014; Manheim et al., 2009; Routhier et al., 2016). Recommended duration of practice for this group ranges from 20 hours over 20 weeks (Palmer et al., 2012) or 5 weeks (Routhier et al., 2016) to 47.7 hours over 26 weeks (mean=~30 hours total). One study did not report the duration of computer practice (Wieczorek et al., 2011). The remaining nine studies report therapy duration as logged by clinicians or researchers. Two studies did not report duration of sessions but only frequency (De Luca et al., 2014; Laganaro et al., 2003) Logged duration of practice for the remaining seven studies ranges from 4-54 hours (mean=25.5 hours) over a period ranging from 3 to 11 weeks, (mean=5.25 weeks).

Literature Review Question 4. What is the effect of computer based therapy for adults with severe aphasia?

As previously reported in Table 7, 27 participants were described as demonstrating either severe or very severe aphasia. These participants were drawn from eight of the 20 studies reported in this review.

Two group studies report evidence from a participant group that included individuals with severe aphasia (Palmer et al., 2012; Thompson et al., 2010). Palmer et al. (2012) report data on four participants from this category within their total participant group of 33. Whilst the overall group trend for participants who received computer therapy within this study is for significant improvement in naming performance, authors state that, “it is clear that this trend is not applicable to participants who were able to name <10% of words at baseline” (Palmer et al., 2012, p.1908). This indicates a limited effect of generalised language therapy upon naming for those individuals identified as severe. Thompson et al.(2010) report data on two participants from this category within their total participant group of 12. Authors again observe a significant effect on sentence production for the group of six participants who received Sentactics® computer-delivered therapy, however no comparison of performance by severity was reported in this study and no figures regarding individual performance were presented for inspection.

Six SCED studies report evidence from a participant group that included individuals with severe aphasia (Cherney et al., 2014; Fridriksson et al., 2009; Kurland et al., 2014; Marshall et al., 2013; Ramsberger & Marie, 2007; Routhier et al., 2016).

Demographic data from Cherney et al., 2014 reports two participants who can be categorised as severe within the WAB Aphasia Quotient (WAB AQ of 26 to 50 - Kertesz, 1982). Within the analysis however, the authors identify four of their total of eight participants as "more severe" (achieving <60 on the WAB AQ) and compare their outcomes to the remaining four whom they classify as "less severe" (achieving >60 on the WAB AQ). Using this delineation, no significant differences are found between the key measures of production rate or accuracy of script acquisition when contrasting therapy change according to severity. Authors identify a relatively greater effect size for items treated in the high-cue condition versus the low-cue condition for "more severe" participants compared to "less severe" participants although no statistical comparisons are provided to endorse this finding.

Fridriksson et al.(2009) report data on four participants from the severe category and one participant within the very severe category of the WAB AQ from amongst their total participant group of 10. Authors comment "more severe aphasia tended to be associated with less or no improvement in naming" (Fridriksson et al., 2009, p. 857). Whilst this is true for the participant whose aphasia was classified as "very severe" and for three of the "severe" participants whose individual scores indicate little or no improvement on the primary outcome naming measure, one participant from this category demonstrated an improvement of around seven naming items in the audio visual training condition and one in the audio only training condition (WAB AQ = 44.3). This suggests a good response to computer-delivered naming treatment in an audio-visual condition for one of five participants identified as being either severe or very severe.

Kurland et al., 2014 report data on two participants from the "severe difficulties" category of the Boston Diagnostic Aphasia Examination (BDAE - Goodglass et al., 2001) within their total participant group of eight (Scoring 1-2 out of 5 on the BDAE aphasia severity rating scale). Two participants with a diagnosis of moderate difficulties (BDAE Severity scores of 2 out of 5) did not complete the therapy programme and a further participant with severe difficulties (BDAE = 1.5) did complete the programme but did not contribute data due to a technical error. For this reason, one of five participants whose data was analysed can be classified as demonstrating severe difficulties (BDAE = 1.5). This participant and the other three participants in the group achieved a “clinically significant” improvement (on at least 20% of items) for the naming of words which had been treated by computer practice alone, and those which had been treated by computer practice plus prior intensive language treatment with a speech and language therapist.

Marshall et al., 2013) report data on 9 of 9 participants with severe aphasia - scoring <20% on the spoken naming subtest of the Comprehensive Aphasia Test (CAT - Swinburn et al., 2004). Group data here demonstrate a significant improvement in gesture production for items practiced with a computer and supported by therapist input. No evidence of improvement in naming performance was observed.

Ramsberger & Marie, 2007 report data on one participant from this category within their total participant group of four. One participant was classified as severe on the Aphasia Diagnostic Profiles (ADP - Helm-Estabrooks, 1992), a second participant was classified as moderate on the ADP and the two remaining participants were reported with a moderate classification according to the WAB Aphasia Quotient (Kertesz, 1982). Three of these four participants (including the participant with severe aphasia) made significant improvements following both intense (5 sessions/week) and non-intense (2 sessions/week) MossTalk computer therapy practise for cued naming. The participant with severe aphasia demonstrated the largest post-treatment gain at the maintenance assessment (166%).

Routhier et al., 2016) report data on two participants from the severe aphasia category within their total participant group of two. Both participants were classified as exhibiting severe verb anomia on the naming of lexical verbs test (DVL-38; Hammelrath, 2001), reporting scores of 9/114 and 32/102. The participant with more severe impairment demonstrated significant improvement immediately post-therapy and at maintenance for items treated using high cueing computer verb practice. The remaining participant demonstrated significant improvement at maintenance but not immediately post-therapy for these items. Items which were untrained and those in the repetition only condition did not significantly improve for either participant.

3.6.6 Additional analyses

No further additional analyses were conducted on the articles reviewed.

3.7 Discussion

3.7.1 Summary of evidence

1. What is known about the effect of remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes when compared to a clinical non-computer therapy, an alternative computer therapy or no therapy/usual care?

The evidence examined largely supports the positive and statistically significant effect of remedial, self-administered computer therapy for single word spoken naming outcomes (typically nouns with the exception of the study by Routhier et al., 2016 which focused on verb naming). The evidence reviewed reports success for treated items within 12 studies - comprising a group of 128 participants with aphasia. Counter evidence for this naming effect is seen in just one of three individual participants reported within a study by Choe et al. (2007) however this one participant represents a very small constituent of a largely positive trend. Obviously, there may be individuals within group studies whose naming effects did not reach significance. However, the group analyses within the two group studies investigating this particular aphasic symptom (Doesborgh et al., 2004; Palmer et al., 2012) provided overall evidence in support of this effect across their combined participant group of 61. It should be noted that the most common form of evidence for the success of computer therapy is observed when it is compared to a no-treatment control. This enables us to consider its effects in comparison to no input but tells us less about the relative strength of the approach in comparison to other offline therapies.

Additional studies included in this review examine computer intervention for aspects of expressive output other than spoken word naming for nouns and verbs. Some limited evidence is available to support the effectiveness of script acquisition and rate of production (one study, 8 participants - Cherney et al., 2014), complex sentence production (one study, 12 participants - Thompson et al., 2010), spoken grammar (two studies, 9 participants - Linebarger et al., 2001; Wieczorek et al., 2011) and communicative gesture (one study, 9 participants - Marshall et al., 2013). Study quality for the group study by Thompson et al however is limited to 5 out of 10 on the PEDRo rating scale, suggesting that outcomes should be interpreted with caution. Whilst quality ratings for the remaining studies are relatively favourable (each scoring 7 or above on the SCED scale), participant numbers still remain relatively low for these investigations, creating a case for further investigation before any more definitive conclusions can be drawn.

Additional research reporting on generalised language therapy (Power AFA - De Luca et al., 2014), spoken naming for objects and actions (iBooks - Kurland et al., 2014) and conversation skills (AphasiaScripts - Manheim et al., 2009) was not included in the above evidence due to the lack of statistical analysis.

No evidence was found to support the generalisation of treatment to untreated items for therapies targeting naming, script acquisition or gesture. Thompson et al.'s 2010 study provides some evidence of generalisation for complex sentence production and Wieczorek et al. (2011) show some generalisation on aspects of spoken grammar. This pattern mirrors the findings for non-computer therapy where better generalisation is observed for therapies targeting grammatical components than for those aimed at improving skills of single word retrieval (Webster, Whitworth, & Morris, 2015) suggesting the mechanisms employed within computer therapy create similar effects to those utilised within offline therapy. Taking into consideration that transfer effects should to be considered unlikely for the majority of naming treatments (with similar tentative evidence for script acquisition and gesture), it is crucially important that target items be selected wisely - to promote maximum benefit to participants.

2. What is known about the durability/maintenance of effects achieved using remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes?

Slightly less evidence is available to explore the maintenance of treatment effects over time when compared to outcomes available immediately after therapy. Nonetheless, the majority of studies that explored this topic provided evidence to support the persistence of therapy gains – outcomes for 52 participants after periods of 2 weeks to 2 months were largely in favour of preservation of gains (statistics for 46 of 51 participants demonstrated maintained gains and one of 51 was not reported in authors final analysis). One study explored the maintenance of gains over a longer-term period of 5 to 8 months (Palmer et al., 2012). This study found that naming gains established immediately after therapy were not maintained.

3. What is the range in the duration and intensity of practice reported for computer therapy?

Evidence of the duration of therapy comes from a variety of sources. Computer-logged practice reflects a total duration of 40.3 hours on average, participant-logged practice reports an average of 127 hours although this figure is somewhat inflated by one study (Choe et al., 2010), therapist/researcher-recommended practice levels are reported at approximately 30 hours on average and therapist/researcher-logged practice levels are reported at approximately 25.5 hours. This reflects an interesting variation in totals across the different methods of reporting duration and therapy figures. The variability in methods used also indicates that data summarised from this information should be interpreted with caution. The figures described suggest that participants with aphasia are able to use computer-delivered tasks to practice therapy independently. Variation in reported figures may represent acceptability or usability of individual programs. It may additionally represent reliability of reporting methods for capturing this data as well as individual variation in therapy uptake.

4. What is the effect of computer based therapy for adults with severe aphasia?

Eight of 20 studies report evidence regarding the effects of computer-based therapy for participants with severe or very severe aphasia. A total of 27 people fitting this classification are reported. Eight of these participants are members of group studies that report significant therapy effects across the groups for measures of naming (Palmer et al., 2012) and complex sentence production (Thompson et al., 2010). Whilst we cannot be certain of their individual performance from the reported outcomes, Palmer et al (2012) do indicate that exclusion of participants scoring <10% on tests of naming, increases the improvement observed across the group. This indicates a more limited response to generalised language therapy for more severe participants when compared to those with relatively superior naming abilities.

Moving on to consider SCED studies, Kurland et al. (2014) demonstrated significant improvement in spoken naming for a participant with severe aphasia as did Ramsberger & Marie (2007). In contrast Fridriksson et al. (2009) found that three of four severe participants in their study demonstrated little or no effect of naming therapy. One however, demonstrated significant and meaningful gains in an audio-visual naming therapy condition. Marshall et al. (2013) provide evidence of significant improvement in gesture production in an analysis of the nine participants with severe aphasia who took part in computer gesture therapy – although it should be noted that no comparisons were made against other users with less severe aphasia – who may have achieved larger gains. Cherney et al. et al (2014) did compare participants with "more severe" aphasia against those with "less severe" aphasia. Here, authors report relatively smaller effect sizes for rate of spoken script production in the more severe participant group when contrasted to the less severe participant group. Authors did however, find no statistically significant differences between the groups' performance on outcome measures of script acquisition.

The data reviewed above suggests that whilst participants with severe aphasia may be subject to smaller therapy gains than their less severe counterparts, there is evidence, nonetheless, that they do respond positively to self-administered computer therapy. There is mixed evidence regarding the effectiveness of naming therapy, with some indication that training stimuli with both audio-visual components may be of greater benefit to some individuals than audio stimuli alone. There is also evidence in favour of the positive effect of gesture therapy. Further investigation of the impact of computer-based practice for individuals with severe aphasia is indicated in order to better understand which symptoms respond favourably to therapy of this form and which components of therapy may be most effective with the group.

3.7.2 Limitations

3.7.2.1 Risk of publication bias

As elucidated by (Liberati et al., 2009), clinical trials are more likely to be published if they report statistically significant outcomes (i.e. $p<0.05$). This risk is particularly redolent for single case or small group trials that do not require registration on a clinical trials database before commencement. Single case or small group studies where an intervention is not successful may not be reported for a number reasons – including a reluctance on the part of the study's authors and also the observed publication bias in acceptance rate for studies reporting positive outcomes (Robey, 1998).

Registered clinical trials are subject to an obligation to publish outcomes regardless of their direction (i.e. even if the intervention was unsuccessful) and this obligation extends to the key journals where outcomes could be published (Laine et al., 2007). Whilst this obligation is not always fully adhered to (Jones et al., 2013), the mitigation against publication bias in this form of research is felt to be greater. Taking into account the preceding arguments, it should be noted that any review of published, clinical evidence (particularly in relation to single case and small group study designs) might not fully represent the full range of knowledge established through existing clinical research. Within such boundaries, we can nonetheless seek to summarise existing published evidence whilst acknowledging the need for full reporting of both positive and negative outcomes in order to both establishing which interventions are effective and meaningfully rule out those which are ineffective.

3.7.2.2 Limitations of Quality Indicators

The choice to use the Physiotherapy Evidence Database (PEDro) scale (Maher et al., 2003) as an indicator of quality for the speech and language therapy research presented here was based on its previous use in related reviews (e.g. Rose et al., 2013; Zheng et al., 2016). It should be acknowledged however, that this scale was originally developed for the purposes of appraising randomised control trials within physiotherapy evidence. In spite of its physiotherapy origins however, the scale has been extensively used as a tool for appraising neuro-rehabilitation evidence more widely - for example for the appraisal of evidence on the database "PsychBITE" (Psychological Database for Brain Impairment Treatment Efficacy). Here, the tool is used to examine the quality of both randomised and non-randomised controlled group trials. Its attributes reflect many of the quality indicators used by Brady, Kelly, Godwin, Enderby, & Campbell (2016) in their Cochrane review of RCTs reporting on the general effectiveness of speech and language therapy for problems after stroke. For example, both Brady et al. and Maher et al. (2003) examine the methodological quality of research by reporting on the presence (or absence) of factors such as concealment of group allocation, blinding of assessors and 'intention to treat' analyses. As identified by Layfield, Ballard, & Robin (2013) however, whilst the PEDro scale constructively allows us to address pre-treatment, performance and analysis biases in the appraisal of aphasia intervention studies, it is not uncommon for important work – reporting "invaluable clinical information" (pg. 6) to score relatively poorly on the scale. Layfield et al. (2013) attribute this challenge to difficulties regarding the challenges in blinding therapists and participants to treatment conditions and requirements around the specific reporting of statistical analyses at different stages in the trial. The ongoing discussions about what constitutes methodological vs clinical quality (e.g. in Howard et al., 2015 should therefore be taken into account when interpreting the outcomes of quality rating scales.

1. What is known about the effect of remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes when compared to a clinical non-computer therapy, an alternative computer therapy or no therapy/usual care?

The variety and application of computer therapy techniques explored within the examined studies provide limited specification regarding the 'active ingredients' of computer therapy – namely, which aspects or approaches are most beneficial. The majority of evidence discussed supports the effect of naming therapy (forming the focus for 12 of 20 studies examined), however evidence remains limited for our understanding of the effects of other types of therapy. More research regarding the effects of computer therapy upon other aspects of expressive language and communication is warranted.

2. What is known about the durability/maintenance of effects achieved using remedial, self-administered computer therapy targeting gesture and/or spoken language outcomes?

There are inconsistencies in the amount of time between post-treatment and maintenance assessments. Of the eleven studies that reported exploration of maintenance effects, the majority of these reflected a period of two to eight weeks with just one study reporting outcomes extending beyond this period. For this reason, whilst there is positive evidence of the preservation of therapy gains over a period of up to two months post-intervention, the long-term maintenance of improvements is far less well documented.

3. What is the range in the duration and intensity of practice reported for computer therapy?

There are inconsistencies in the reporting of intensity and duration of self-administered practice. Figures provided by authors range from suggested levels/intensity of practice through to self-reported/family-reported levels of practice and finally computer logged levels/intensity of practice with varying levels of reliability.

4. What is the effect of computer based therapy for adults with severe aphasia?

The scales used to classify severity of aphasia for participants vary greatly. Taking this into account, it is difficult to precisely state the overlapping features of participants identified as presenting with severe aphasia. It should be acknowledged that firm claims regarding the specific effects of computer therapy for individuals with severe aphasia cannot be made without additional evidence provided using consistent categorisation approaches.

3.7.3 Conclusions

Computer therapy for expressive aphasia appears to work for spoken naming deficits. Participants tested within the studies reviewed responded positively to self-administered therapy, demonstrating significant improvements in expressive output measures and revealing a good uptake of independent computer practice. Additional limited evidence supports the use of computer practice for a variety of other aspects of language including script acquisition, sentence production and gesture. Outcomes for participants with severe aphasia can be varied although there is evidence to support positive and significant therapy effects in this group – potentially assisted by the presence of multimodal training stimuli, which utilise both sound and vision. Studies suggest improvements can be achieved in both naming and gesture production for this population. Generalisation to untreated items is observed rarely, with effects observed for grammar or sentence level interventions but not for single item word or gesture treatments. The maintenance of gains appears to be quite high in the period of up to two months following cessation of therapy. Little is known about the longer-term persistence of improvements.

Findings from the above review indicate a need for further investigation of computer therapy outcomes for gesture and other aspects of language beyond the single spoken word, exploration of the longer-term maintenance of therapy outcomes and the extension of research pertaining to those with severe aphasia.

3.8 Funding

This review was undertaken within the bounds of the present PhD thesis. Funding for the project was provided in the form of a three-year City University PhD scholarship award with a £1,000 stipend for project expenses.

Chapter 4. Previous computer gesture therapy research – the GReAT project

4.1 Introduction

This chapter provides a detailed overview of previous research regarding GeST – one of the two computer therapy tools investigated in this thesis. The first section provides a brief introduction to the GeST tool and details about an initial pilot study into its effectiveness. The chapter concludes by considering all of the literature review so far (in chapters 2, 3 and 4) and then presenting the resultant research questions which form the basis of the subsequent experimental study.

4.2 The GReAT project

The GReAT project set out to develop and pilot a novel computer-based tool to allow users with severe aphasia to practice gestures in their own home. The project was funded by the Engineering and Physical Sciences Research Council and comprised a tool development phase and a therapy pilot study. Details of various aspects of the tool's development phase have been documented elsewhere (Galliers et al., 2011; Galliers et al., 2012; Wilson et al., 2015). A brief description of the tool is provided below, followed by further information regarding the therapy pilot study.

4.2.1 GeST – computer gesture therapy

The tool developed within the GReAT project was named GeST. The GeST software program ran on a Windows laptop computer, with an adapted external keypad and a separate high quality webcam (Figure 3). GeST presented users with a video demonstration of a gesture and invited them to copy it. The system then used vision-based gesture recognition to identify whether a user had produced the targeted gesture correctly. Feedback was provided to indicate successful gesture production. Practice activities were presented in three different formats, including straightforward video repetition of the gesture in isolation, repetition activities presented within a virtual world, game-like context and repetition of gestures presented within videos of brief, real-world scenarios.

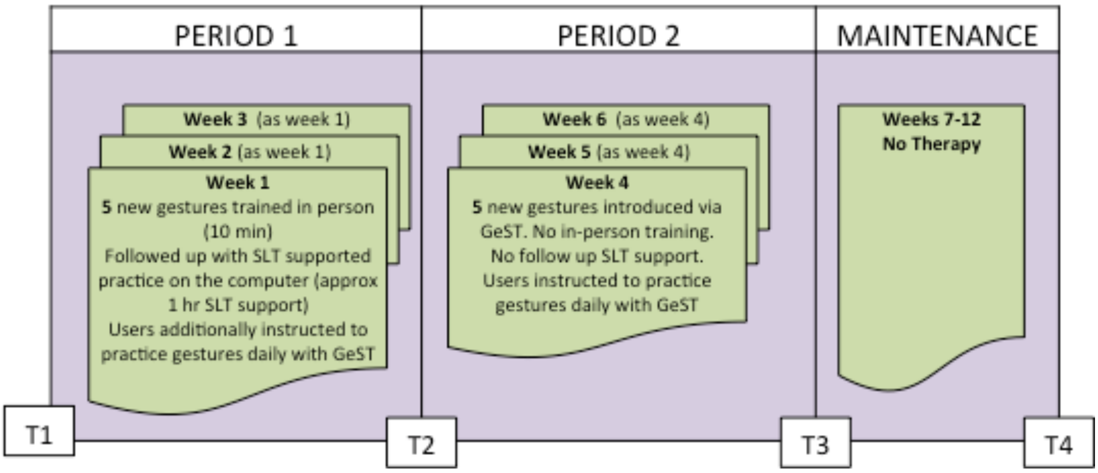
Figure 3. Image of the GeST therapy tool

4.2.2 Pilot testing the GeST prototype

Following development of the GeST tool, a pilot study was set up to ascertain its effectiveness in training gestures, to establish a therapy protocol for delivery of the treatment, to review the tool's usability in a real-world setting and to garner opinions on the tool's acceptability from users and their families.

Pilot Design

Reported in Marshall et al (2013), the study employed a two-phase, repeated measures design. 10 participants were recruited, and 9 participants (three female) completed the protocol. Each followed the protocol set out in Figure 4 below. Following an initial series of background assessments, participants undertook a first three-week practice period with GeST – rehearsing a different set of five gestures each week. A speech and language therapist researcher introduced the five new gestures - in person – at the start of each week and then participants were requested to practise these independently, every day, with GeST. Also during period one, in addition to those gestures trained via therapist + GeST, a further set of 15 gestures (five per week) were introduced by the therapist but not subsequently practised with GeST. During period two - the second three-week practice phase, participants rehearsed five new gestures per week using GeST alone. These 15 gestures received no in-person introduction. Repeated measure assessments of gesture and naming were conducted at four time points: T1 - before therapy; T2 - directly after period one; T3 - directly after period two; T4 – following a maintenance period of six weeks. At each time point, participant competence on a total of 60 gestures was assessed (using methods reported in Marshall, 2012 and described in chapter 7 of the current report). Items tested comprised the 15 gestures trained in period one using therapist + GeST, 15 items trained in period 1 using therapist only, 15 items trained in period 2 using GeST only and 15 matched control items which received no training. Naming accuracy for the same 60 items was also assessed at each time point. Measures aimed to examine the effects of the different therapy components over time.



Assessment 1 (T1)	Assessment 2 (T2)	Assessment 3 (T3)	Assessment 4 (T4)
The imitation element of the Limb Apraxia subtest of the Dabul (Dabul, 2000)			
Object picture matching and gesture picture matching test			
Comprehensive Aphasia Test – Spoken picture naming subtest. (Swinburn et al., 2004)			
Gesture assessment	Gesture assessment	Gesture assessment	Gesture assessment
Naming Assessment	Naming Assessment	Naming Assessment	Naming Assessment

Figure 4. Design of the GReAT project pilot study

Outcomes of the project

Analysis of outcomes revealed that the 9 participants produced significantly more gestures that could be identified by strangers after having received the therapy than at the outset of therapy (by an average of two items). Closer analysis revealed that trained gestures accounted for these gains, specifically those trained within the first 3 weeks of therapy which were supported by weekly visits by a speech and language therapist. Gestures trained in the second, unsupported phase did not demonstrate a significant improvement across the group. Untrained gestures matched for linguistic frequency to the spoken name of the trained items did not demonstrate a significant improvement as a result of the therapy intervention. When tested, spoken naming demonstrated no significant improvement as a result of the pilot gesture intervention. Aside from the quantitative measures, carer reports suggested mixed outcomes regarding how effectively gains achieved were translated from the practice schedule into everyday communication. Four of nine thought that participants were using more words and gestures since completing the practice but two did not, e.g. “While she works on it here [points to computer], it doesn’t necessarily translate.” (Marshall et al., 2013, p. 1140).

In addition to gains made in gesture production, some carers reported that participants with aphasia had expressed an increased interest and use of additional household technologies as a consequence of their involvement in the research project. This finding however arose in carer interviews only and was not measured formally.

Outcomes from this study, whilst promising, indicated a number of further areas for exploration. The relatively small size of the gesture gain observed (two items) raised the question of whether it is possible to increase the dosage (amount of exposure to each target gesture) and active ingredient (exposure to therapist supported practice) of the intervention provided in the GReAT study in order to improve therapy gains. Additionally, further investigation was indicated to establish whether gains in isolated gesture production translate into interactive communication, away from the computer as was observed for the offline gesture therapy conducted by Caute et al (2013). Finally, was there way to capture the increase in participants’ interest in technology reported by carers using a more systematic measure?

4.3 Implications of background literature for the current project

Taken together, the range of studies presented within chapters two, three and four show evidence that technologies can be meaningfully applied to remediate and compensate for aphasic language impairments. Technological treatments are well received by participants with aphasia and compliance appears good (Marshall et al., 2013; Palmer et al., 2012), however the range of tasks implemented is commonly reminiscent of the approaches applied in offline therapy and typically makes limited use of the additional modalities available within a computer delivery. The impact of a more novel game-like approach to intervention – such as that utilised in GeST - is yet to be explored fully.

Findings from the systematic literature review in chapter three suggest that further investigations are warranted to further elucidate:

1. The effects of computer therapy for adults with severe aphasia
2. The maintenance of effects over time
3. The effects of therapy duration on outcomes
4. The factors that contribute to successful outcomes for participants with severe aphasia

Effects of computer therapy for adults with severe aphasia

Findings from previous research into computer interventions suggest that outcomes for those with severe aphasia are less consistently positive than for those with less severe aphasia. For example, outcomes for naming therapies for this group are mixed (Fridriksson et al., 2009; Palmer et al., 2012; Ramsberger & Marie, 2007). Initial research into the use of computer gesture therapy with this group appears promising (Marshall et al., 2013) however the evidence base here is much smaller (one study) and requires further expansion to be able allow us to speak with more certainty on the topic. Similarly, whilst Caute et al (2013) found evidence for translation from the face-to-face training of individual gestures into a more interactive, message conveyance task, might we observe similar translation for a computer-delivered therapy? More comprehensive investigation of the field, with a larger number of participants and a greater level of experimental rigour is warranted.

Maintenance of effects over time

Evidence from chapter 3 suggests that once therapy has been withdrawn, naming or gesture gains commonly persist in the relative short term (e.g. Marshall et al., 2013; Pedersen et al., 2001; Routhier et al., 2016). However, evidence for longer-term retention of naming gains is limited and somewhat less emphatic (Palmer et al., 2012). Furthermore, evidence pertaining to the longer-term retention of gesture gains is yet to be established. Further research regarding the persistence of gains would enhance our understanding of this issue.

Effects of therapy duration on outcomes

Computer-delivered therapy offers us the opportunity to explore duration and intensity issues as practice is not dependent on the constant presence of a therapist and yet the delivery method enables a remote or retrospective measure of therapy dosage to be obtained. The 'dose' of therapy achieved by Marshall et al. (2013) using GeST, will be extended in the current study through the use of a novel computer therapy practice adjunct (PowerGeST) which has been developed for this research (full details of PowerGeST's development and implementation are provided in chapters six and seven respectively).

Factors that contribute to successful outcomes for participants with aphasia

Chapter 2 highlighted some key areas that might be investigated to address this question – namely, limb apraxia (section 2.6) and aspects of executive function (section 2.3). Work by Hogrefe, Ziegler, Weidinger, & Goldenberg (2012) suggests that the presence of limb apraxia may have an effect on an individual with aphasia's ability to produce comprehensible gesture, yet Marshall et al. (2012) and Caute et al. (2013) found that individuals with severe aphasia did demonstrate improvements in pantomime gesture production abilities in response to intervention delivered by a speech and language therapist. Marshall et al. and Caute et al. called for a larger scale study of gesture intervention to address the role of limb praxis on response to therapy.

Beyond the issue of limb apraxia, a further question remains as to whether non-linguistic factors might inhibit a participant's response to autonomously-operated computer-delivered therapy. Evidence from Helm-Estabrooks (2002) suggests that severe aphasia may be accompanied by variable performance on non-linguistic measures of cognitive performance and Marshall et al. (2013) found four of nine participants with severe expressive aphasia demonstrated difficulties in navigating between levels of a computer delivered gesture therapy. Additionally, several of the studies presented in section 2.8 acknowledge the limits of computer-interventions for some users with severe aphasia (Bartlett, Fink, Schwartz, & Linebarger, 2007; Nicholas, Sinotte, & Helm-Estabrooks, 2011; Palmer et al., 2012). An inclusion of measures of limb praxis, executive function and other language abilities will enable us to investigate issues of candidacy in further detail – with an aim of predicting who might derive the most value from the tool. A measure of the impact of aphasia on levels of technology use is also prompted by such acknowledged limitations within existing research and preliminary findings from Marshall et al. (2013).

The study reported in this thesis aims to shed light upon each of the above issues by examining in more detail the implications and effects of a computer gesture therapy protocol which employs GeST + PowerGeST upon measures of gesture and spoken naming. Issues of the translation of gains into interactive communication are explored, as are the effects upon general technology use. Furthermore, a range of candidacy measures is explored to establish whether clinically relevant prognostic indicators can be established. Employing an approach indicated by Rose et al. (2013) the current study aims to provide not only a further study of gesture intervention, but one that employs therapy at a high dose and intensity and using a randomised, between-group design.

4.4 Research Questions

Based on the themes discussed above, the following research questions are addressed within the current study:

1. **Primary outcome measure**
    a. Does practice with GeST + PowerGeST improve gesture production in isolation?
        i. Are gains confined to items trained in therapy or do they generalise to untrained items?

2. **Secondary outcome measures**
    a. Does practice with GeST + PowerGeST improve spoken naming?
        i. Are gains confined to items trained in therapy or do they generalise to untrained items?
    b. Can people with severe aphasia use learned gestures within interactive communication?
    c. Does access to GeST + PowerGeST affect participants' use of technology and confidence in its use?

3. **Additional Questions**
    a. How much computer practice do participants undertake and at what intensity?
    b. Are therapeutic gains maintained after therapy has been withdrawn for 10 weeks?

c. Can we identify clinically relevant prognostic indicators for those who might be good candidates for GeST + PowerGeST therapy? Specifically in relation to executive function, praxis and language.
d. Is there a relationship between therapy intensity and the size of the therapy effect?
e. Is there a relationship between gains on assessment of gesture in isolation and gains on assessment of other skills? Specifically, spoken naming and gesture in interactive communication.

Chapter 5. Description of GeST and PowerGest – computer-delivered gesture therapy tools for aphasia

5.1 Introduction

Two computer-delivered gesture therapy tools were used to deliver the intervention examined in the current project. These were GeST – as trialled within the GReAT project (described in section 4.2.2) and PowerGeST, a novel adjunct created for the purposes of this project to develop and extend the gesture practice dosage achieved in the GReAT pilot project. Full details of each tool and their uses are outlined below. Additional details regarding the rationale and development process behind PowerGeST are provided within the chapter Development of Tools (section 6.4). The choice to use GeST within the present therapy study was based upon two factors. The first of these factors was my involvement in the GReAT project. In my role as therapist researcher, I helped to develop the GeST tool and to create, deliver and analyse outcomes from the tool's original pilot study. The second factor in determining to use GeST for this research was that, at the time the study was carried out, no other computer-delivered gesture therapy for aphasia existed. Therefore, in order to assess the effectiveness of computer delivered gesture therapy for severe aphasia, GeST was the only tool available offer such practice. The pilot study for this tool additionally contributes towards the background evidence for the current project.

5.2 Accessibility features of the computer therapy

Key to the usability of the computer therapy in the present study, were a series of principles utilised to ensure that both GeST and PowerGeST were accessible to people with severe aphasia. Features to enable this accessibility included:

- The use of an external, simplified keypad to interact with the tools; no QWERTY keyboard, mouse or touchscreen was required.
- Each key press on the keypad was mapped to a single action in the given tool.
- No text instructions were displayed on the computer screen; e.g. video was used to demonstrate the gestures.
- All screens were clear, uncluttered and consistent in their layout.
- Therapy tool software was set to run from start-up; pressing the green ON button on the keypad started both the computer and launched the application.
- A progress bar at the bottom of the screen showed, within GeST, which gesture of five was being practised, and, within PowerGeST, the proportion of the entire gesture set which had so far been completed (of 20).
- Navigation was minimal and consistent. In PowerGeST, pressing the OK button started the next gesture video. In GeST, pressing the OK button moved to the next gesture in a level. Arrow keys could be used to navigate between levels and between gesture items.

5.3 Description of GeST

5.3.1. What was it?

GeST was a computer-delivered gesture therapy tool designed to help people with severe expressive aphasia to learn and practice communicative gestures. Users were presented with a video demonstration of a gesture and instructed to copy it. The system monitored the user's gesture production using a vision-based gesture recognition technique. When the target gesture was identified, the user heard a round of applause and the computer moved on to present the next target gesture (item). GeST offered three levels of practice (further discussed in section 5.3.5). Overall, people practised 20 gesture items. These were practised as a different set of five items per week over a four-week period.

Figure 5. GeST Computer Therapy Tool

5.3.2. Hardware

The GeST therapy tool trialled within this research project ran on a 14" Samsung laptop computer with additional Logitech webcam and a separate customised Korg "Nanopad" input device as a simplified keyboard (Figure 5). Users wore a yellow cotton glove on the hand they were using to gesture. This enabled the tool's vision-based recognition system to identify their hand against other elements of the background.

5.3.3. Software

Originally developed and trialled within the GReAT pilot project (Marshall et al, 2013), GeST was a computer-delivered gesture therapy tool developed with the Unity development platform and operating on the Windows 7 operating system. It offered three practice levels and a main menu. To access the levels, users had to operate the system as described below.

5.3.4. Navigation

Operation of the GeST system was entirely carried out using the external keyboard and the computer's power switch. No mouse or QWERTY keyboard use was required. Navigation within GeST consisted of a two-tier structure, with one Main Menu screen (tier one) and three subordinate Gesture Activity levels (at tier two). Tier one – the main menu (as shown in Figure 6) – showed three level icons. These comprised an image taken from each level. Selecting the given image (using the [←] and [→] keyboard buttons) and then pressing [OK] allowed the user to navigate to either level 1, 2 or 3 of tier two, below. Each of these subordinate levels used a cyclical progression beginning at item 1, continuing sequentially to item 5 and then returning to item 1 again. Users were required to press a separate menu button ([Menu]) in order to return to the menu screen on tier one.

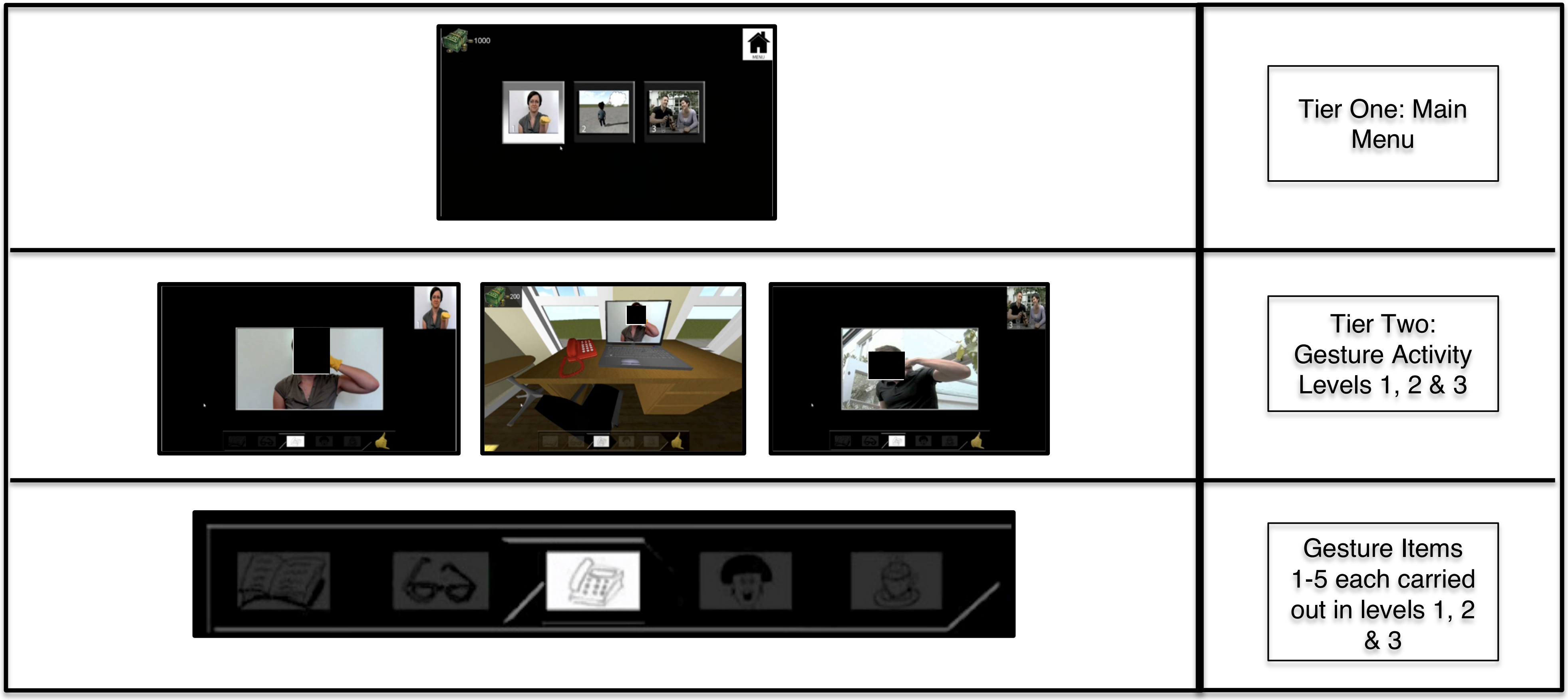


Figure 6. Structure of activities in GeST

5.3.5. Getting Started

Users first switched on the laptop using the computers power button

(labelled **ON**)

GeST was initiated automatically with no additional navigation through the operating system.

5.3.5.1 Menu Screen

Following startup, the menu screen was the first component encountered by users (Figure 7).

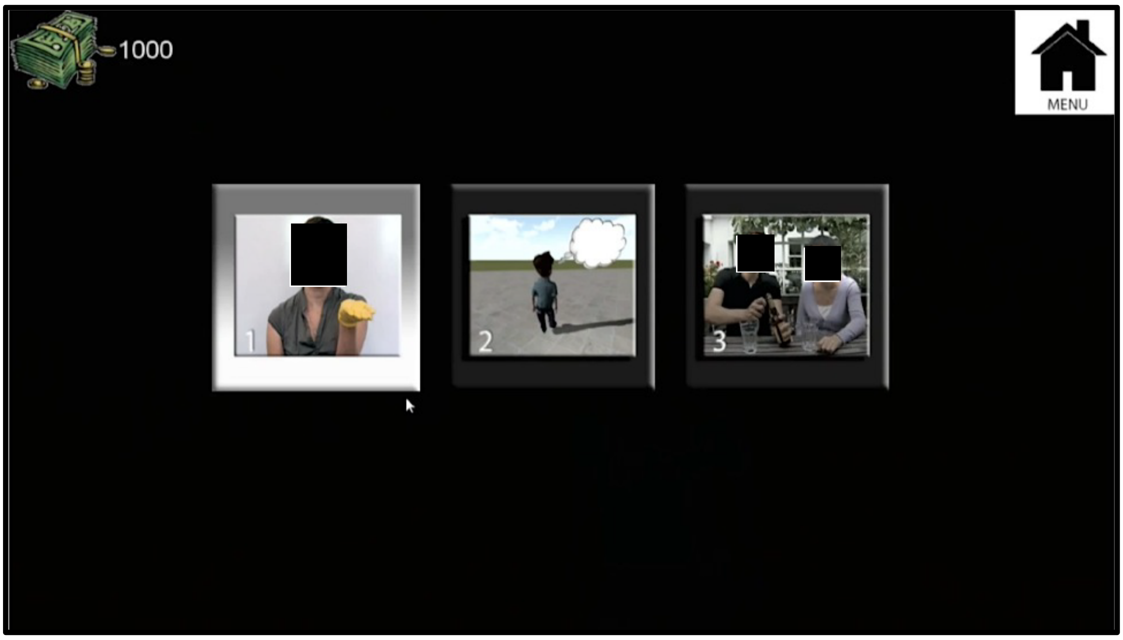


Figure 7. GeST Menu screen

As previously noted, users' could select a level icon from the menu screen by pressing the left or right arrow keys ( ) on the external keyboard.

Once the selected level icon was highlighted, pressing the **OK** button would navigate the user to the corresponding level. Each level will be further described below.

5.3.5.2 Level 1

Level 1 consisted of a series of five gesture video prompts. Each video showed the speech and language researcher presenting a model target gesture and then encouraging the user to repeat the action demonstrated. Videos were displayed against a blank, black background with a navigation panel situated along the lower edge of the screen (Figure 8).

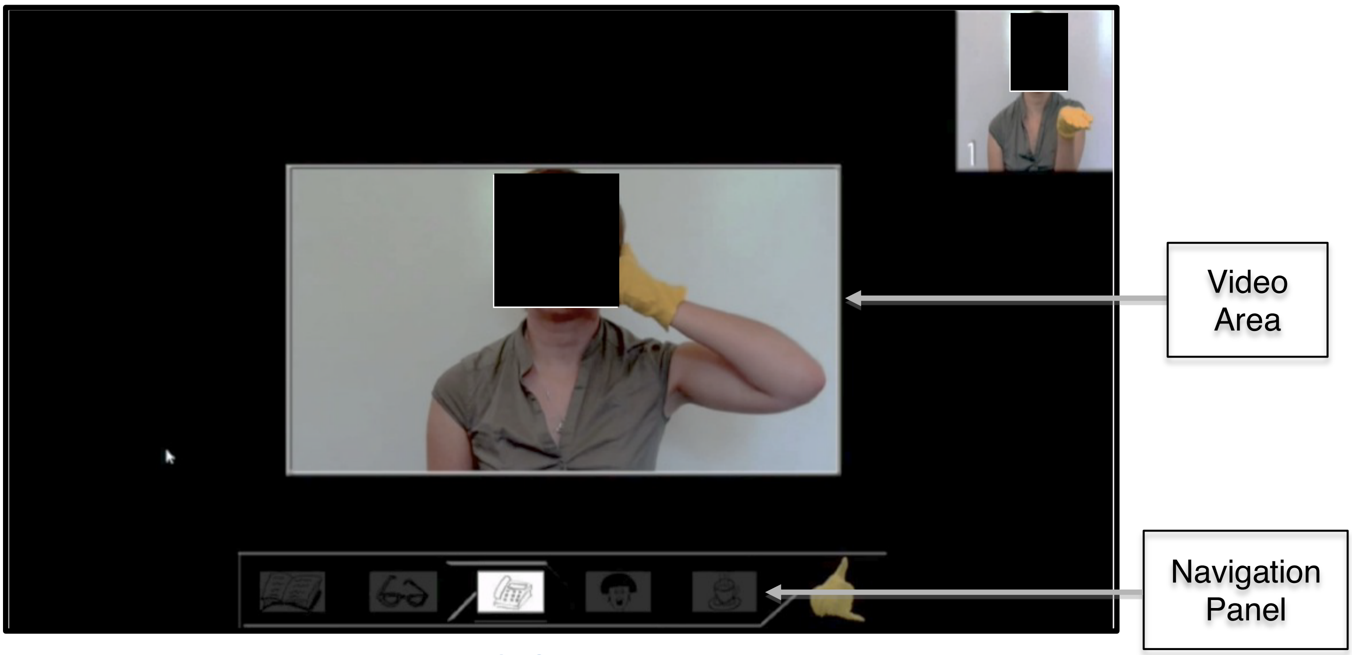
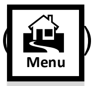


Figure 8. GeST Level 1 screen

When the user was ready to begin, they could press **OK** on the keyboard.

A video of a speech and language therapist was presented, instructing: "Here is the gesture for *X*" (where *X* is the spoken label for the target gesture e.g. "telephone"). The model gesture was demonstrated twice. Users were next instructed, "Now it's your turn".
A numerical countdown was presented (3, 2, 1), concluding with the sound of a bell. At this point computer gesture recognition was initiated and the video image on the screen was replaced by a real-time outline view of the user's gloved hand as it moved in front of the camera. Users were able to see only

the yellow glove presented against a black background (as illustrated in Figure 9). Here, the user had to attempt to produce the target gesture they had just seen demonstrated. If they produced the target gesture successfully, they were rewarded with the sound of a round of applause and the computer continuing on to present the next item. If they failed to correctly produce the gesture, they received no feedback from the computer. Users were given an unlimited length of time to attempt gesture production. If they failed to produce the gesture but wished to continue practising, they could move to the next item manually by pressing the right arrow button.

All five gesture items were targeted in this way at level 1, with the program progressing through the items either as each gesture was recognised or if the right arrow button was depressed. Once they had completed practice on this level, users could step up to tier one by pressing the menu button (

)

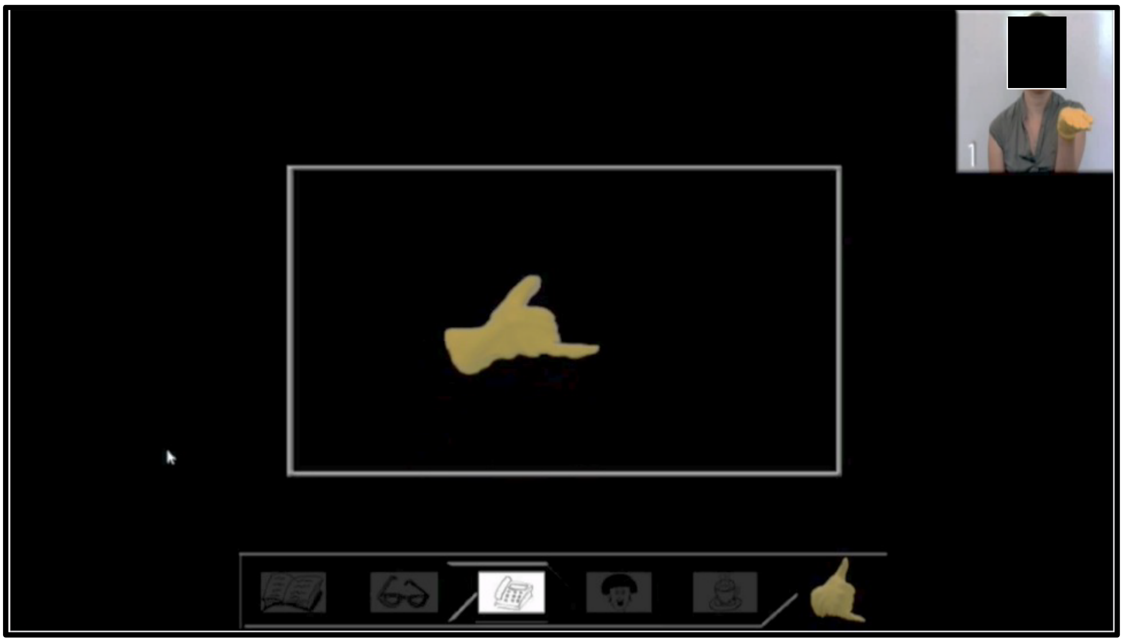


Figure 9. Gesture recognition screen within Level 1

5.3.5.3 Level 2

Level 2 consisted of the same series of five gesture video prompts from level 1 – with each video showing the speech and language researcher presenting a model target gesture and then encouraging the user to repeat the action demonstrated. This time however, instead of being shown against a plain, black background, videos were presented within a 'virtual' world and viewed from the perspective of an avatar, "Gerry", as he navigated his way from one location to another (Figure 10).

Figure 10. GeST Level 2 screen

As before, users pressed the OK button on the keypad when they were ready to begin. At each destination, Gerry encountered a virtual instance of the target item (in this example, the red telephone pictured in Figure 11). The user was then presented with a video demonstration – as for level 1, and again instructed to copy the target (Figure 11). Gesture recognition was again initiated following a countdown ending with the sound of a bell, and the user again saw their gloved hand appear against a black background in the location previously occupied by the video. As for level 1, correctly identified gestures were met with a round of applause and progression to the next item. Once again, incorrect gesture production was met with no response and the user was required to depress the right arrow key in order to move on to the subsequent item. The same five target gesture activities continued to cycle in series until the user pressed the MENU button. In contrast to level 1, level 2 additionally featured a score tally (as shown in the top, left-hand corner of Figure 10). Here, correctly produced gestures were rewarded with the addition of 100 points to a user's points total. This tally accumulated points with each correctly identified gesture produced within level 2. Scores were added to each time a user visited level 2, meaning that at the end of a week's practice the total score reflected a number that was 100 X the number of correct gestures produced within level 2 across the whole week.

Figure 11. Gesture recognition screen within Level 2

5.3.5.4 Level 3

Level 3 consisted of the same five target gestures as the two preceding levels, this time presented within a short video vignette of the target gesture being used within a real-life exchange. Each video was of the format:

situation, gesture, spoken acknowledgment of gesture, resolution of situation. (Figures 12, 13, 14 and 15)

Users began this level by depressing the OK button. This initiated the video in its entirety. After viewing the video, users were shown a numerical countdown concluding with the sound of a bell. Gesture recognition was initiated here and the user again saw their gloved hand appear against a black background in the location previously occupied by the video. In contrast to levels 1 and 2, level 3 provided users with no verbal prompt to copy the gesture. They were instead encouraged to respond at the sound of the bell signifying the onset of gesture recognition. Level 3 also differed from levels 1 and 2 in the fact that the target gesture was presented within a short video narrative and not explicitly demonstrated by the established therapist instructor immediately prior to gesture recognition being initiated. Here, users were required to recall and reproduce the gesture from a video where its production was embedded within a brief, two-way interaction. This inclusion aimed to introduce more naturalistic examples of gesture use for practice than those available in levels 1 and 2. In doing so, level 3 required users to retain the target gesture production over an extended period, however this increase on memory demands was deemed an appropriate progression to attempt to promote carryover of acquired gesture skills into everyday communication.

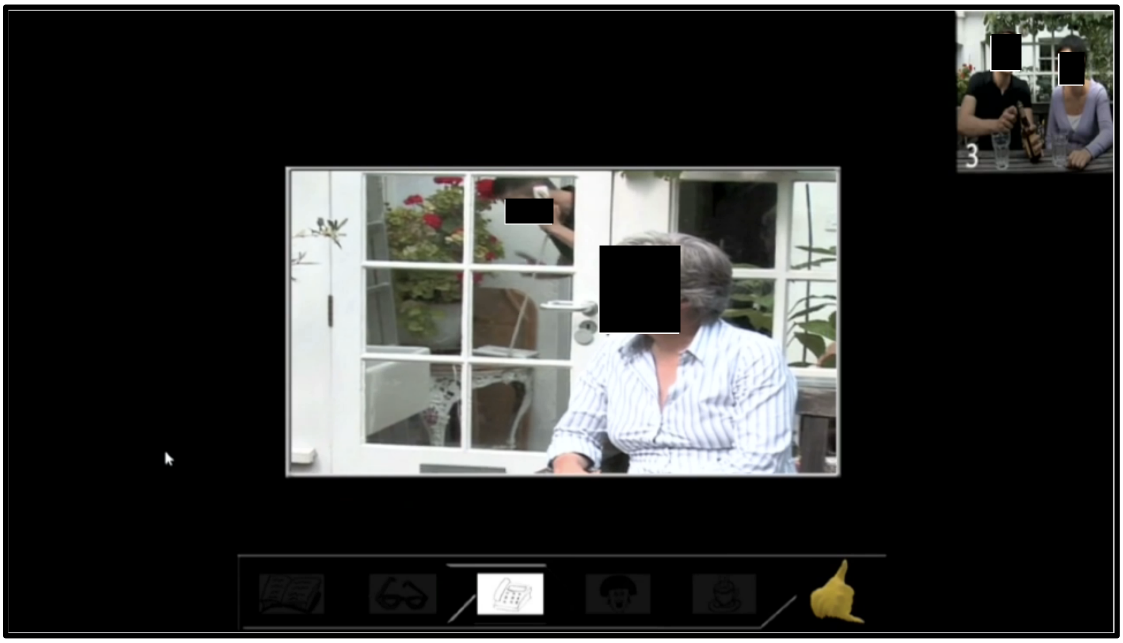


Figure 12. Situation: Telephone rings inside the house

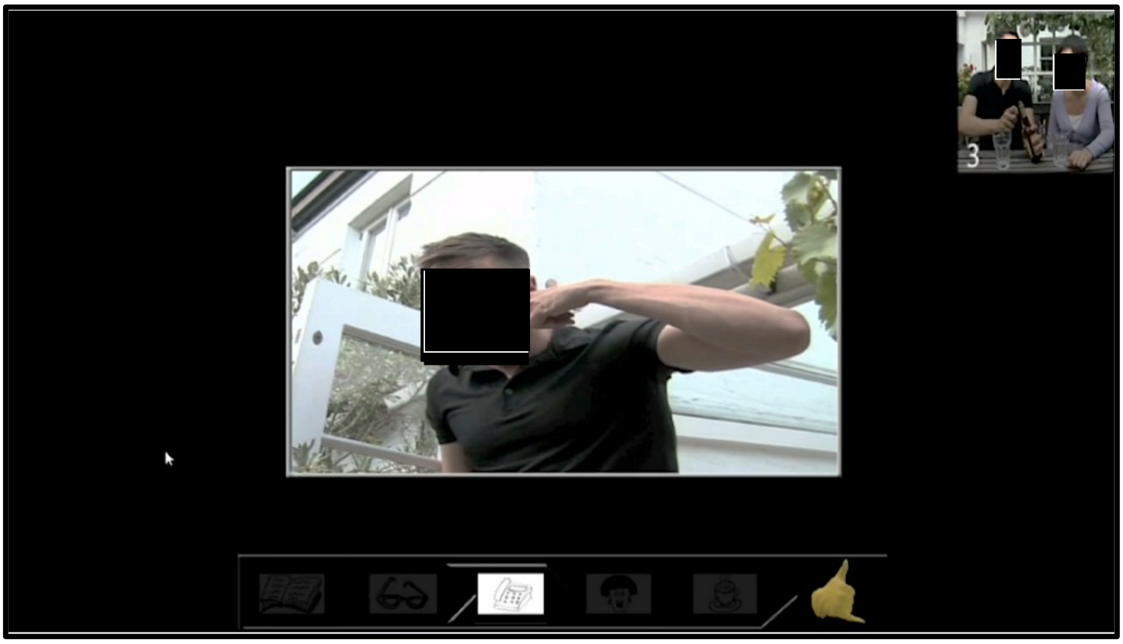


Figure 13. Character A gestures 'Telephone'

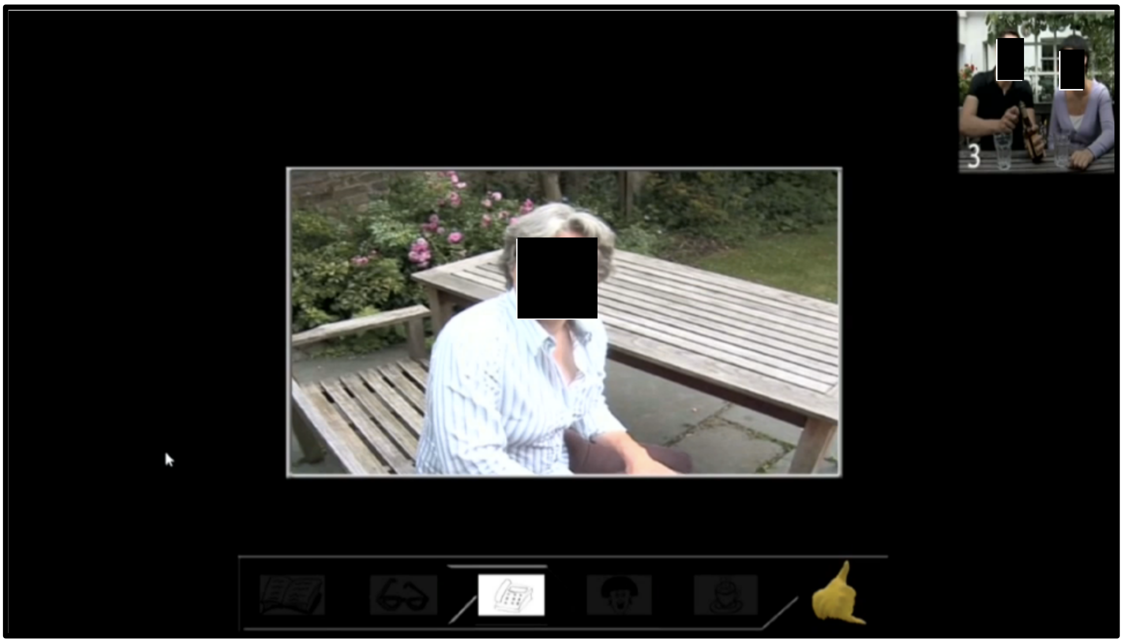


Figure 14. Character B provides spoken acknowledgement of the gesture: "Telephone?"

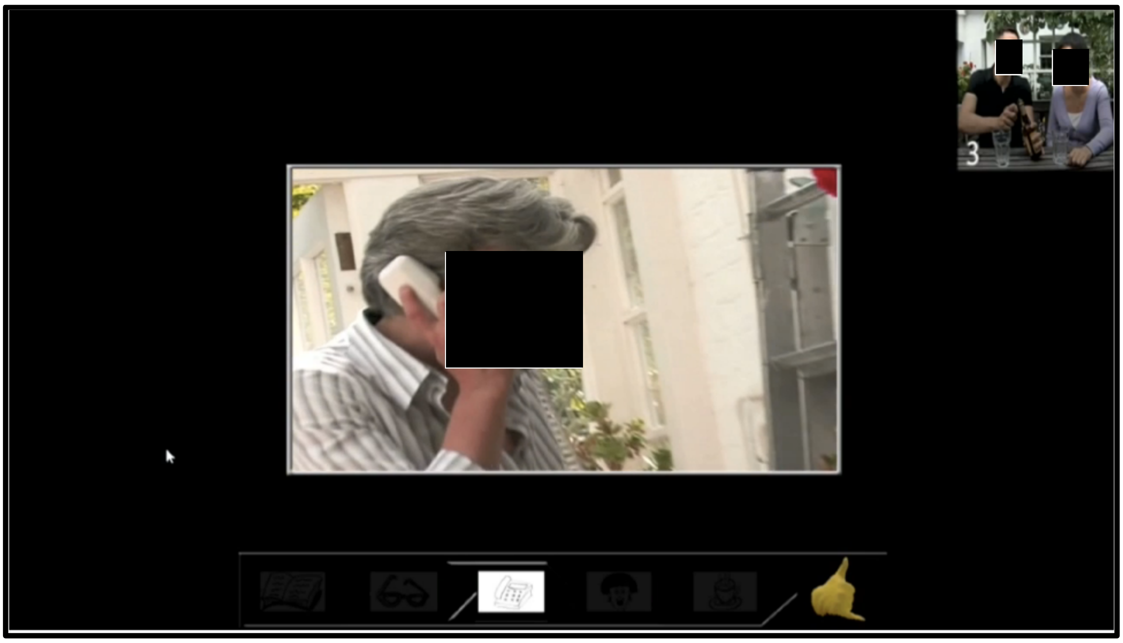


Figure 15. Resolution of situation: Character B answers the telephone

Following completion of activities, users could switch the entire system off (including shutting down the computer) by depressing the off button (OFF).

Within the current project, gestures were trained in sets of five. A total of 20 gestures were trained by the system, split across four sets. Participants practiced one set of five gestures per week for the first four weeks. In the fifth week, they practiced all 20 previously practised gestures using PowerGeST.

5.4 PowerGeST

5.4.1 What was it?

PowerGeST was a simple computer tool designed to help people with severe expressive aphasia to practice communicative gestures previously introduced to them during their use of GeST. Users were presented with a video demonstration of a gesture to copy. Having copied the gesture, they had to depress the OK button to move forward to view the next item. The PowerGeST system employed no gesture recognition and aimed to encourage increased autonomy in gesture practice by requiring the user to decide for themselves when to proceed to the next item. This contrasts with GeST where the program would progress forwards automatically when the user had correctly produced the target gesture item.

5.4.2 Hardware

The PowerGeST therapy tool ran on a 14" Samsung laptop computer and used the same separate customised Korg "Nanopad" input device as was employed for GeST. No vision-based recognition was used within PowerGeST and hence the additional webcam and cotton glove used within GeST were no longer required.

5.4.3 Software

Developed and piloted as an extension to GeST, PowerGeST presented a series of twenty demonstration videos using a Microsoft PowerPoint Show (.ppsx) run through PowerPoint 2010 software.

5.4.3.1 Activity Format

PowerGeST comprised a series of 20 gesture video prompts. These videos were excerpts from the short video vignettes used within level 3 of GeST. Each video was of the format: situation, gesture. This time, the spoken acknowledgement of the target gesture and the resolution of the situation were not shown. Videos were presented against a blank, pale blue background with a progress bar situated along the lower edge of the screen. (Figure).

5.4.3.2 Navigation

Navigation within PowerGeST was limited to a simple forward progression from start to finish, with no implementation of a hierarchical structure. This meant that, unlike GeST, there was no Main Menu, instead only one 'level' that allows practice of 20 items in the same fashion.

As for GeST, operation of the PowerGeST system was entirely carried out using the external keyboard and the computer's power switch. Again, no mouse or QWERTY keybaord use was required.

5.4.3.3 Getting Started

As for GeST, users first switched on the laptop using the computers power button (labelled ON).

PowerGeST was initiated automatically with no additional navigation through the operating system.

Following startup, users were presented with a green start screen and a picture prompt to press the OK button on their keyboard. Upon depression of the OK button, users were shown a blue screen with a video presentation of a situation. This was immediately followed by a demonstration of the target gesture (of the nature described in GeST figures Figure 12 and Figure 13.)

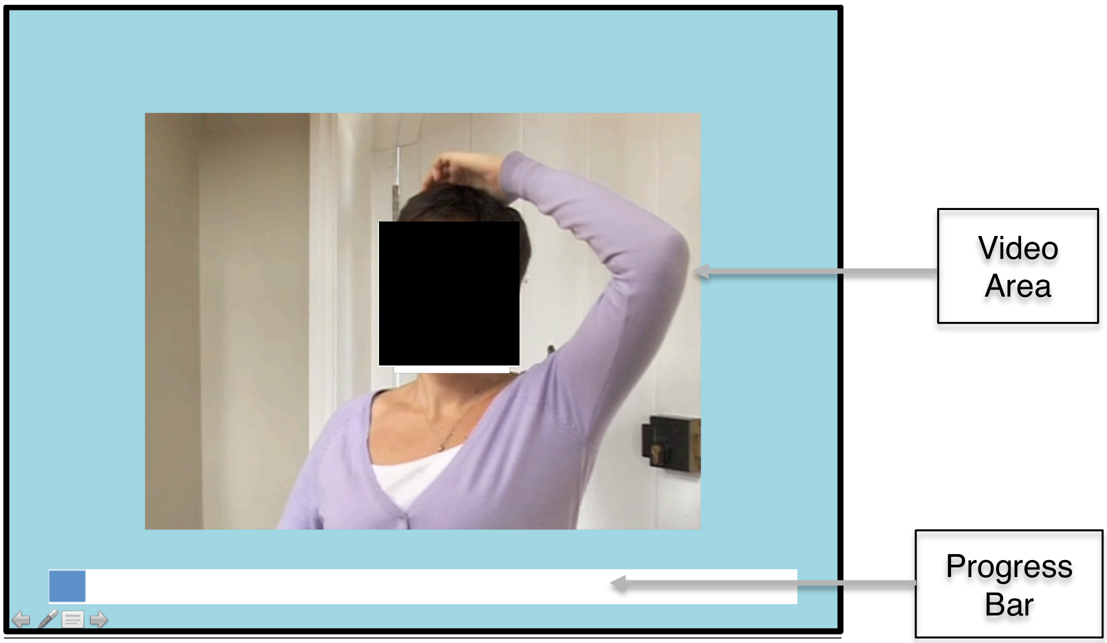


Figure 16. PowerGeST video scenario paused at the production of the gesture 'Hat'

Following demonstration of the gesture, the image of the action in its concluding location was frozen on the screen and users heard a bell to signify that they should copy the target. At this point, users were expected to copy the gesture demonstrated. No gesture recognition was employed within PowerGeST and users did not receive any visual or auditory feedback as to the accuracy of their attempt. Two seconds after the bell had sounded, a picture of the OK button appeared (Figure 17). This acted as a prompt to encourage users to proceed (by pressing the OK button on the keypad).

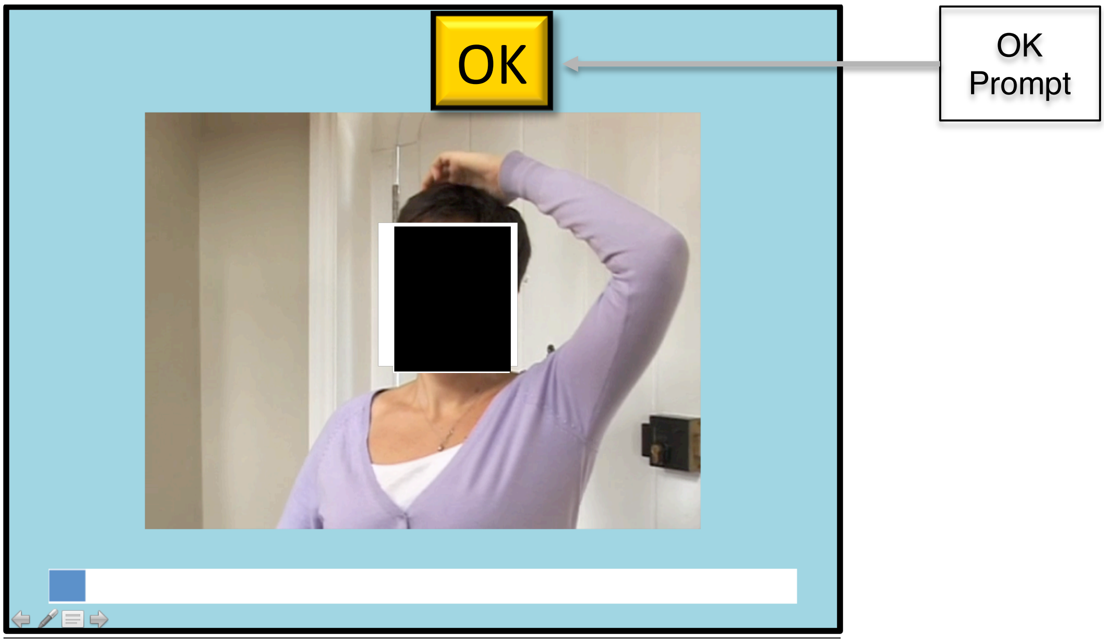


Figure 17. PowerGeST screen plus OK prompt

Once more, following completion of activities, users could switch the entire system off (including shutting down the computer) by depressing the OFF button.

PowerGeST employed a start-to-finish progression from item 1 to item 20 and then concluding. Users were not required to navigate to a separate menu screen at any point.

A total of 20 gestures were trained by the system and all were presented in one set. Participants used PowerGeST to practice all 20 gestures within the fifth and final week of intervention.

Chapter 6. Development of new tools for the current project

6.1 Introduction

Five novel components were utilised in the present study. Three were entirely novel and two were adapted from existing measures. The development of each will be described within this chapter. To begin, a report will be presented explaining how two measures used within a previous pilot study of GeST - the GReAT project - were adapted to ensure their relevance to the present study. The following section will then provide an account of two novel assessment measures which were developed. These were created to document the impact of computer gesture intervention on the wider communication, technology use and technology confidence, and also to capture general levels of everyday technology use for participants. Finally, a novel computer therapy adjunct will be described. This was created in order to refine and extend the therapy protocol previously investigated by Marshall et al (2013).

6.2 Revision of Existing Outcome Measures

6.2.1 Naming and Gesture Assessments

Both the gesture assessment and the naming assessment used as primary and secondary outcome measures (respectively) within this project were adapted from related measures used within the pilot study of GEST (Marshall et al., 2013). In brief, stimuli comprised a set of coloured pictures of objects printed on white paper and shown to participants one at a time. For each picture shown, participants were asked to demonstrate either a gesture to represent that picture (for the gesture assessment) or (for the naming assessment) to name the item in the picture. (Full details of the exact nature and administration of these measures can be found in Methods sections 7.9 and 7.10.) For the purposes of the previous study, these two measures comprised the same 60 items - 30 treated items which had been trained within the therapy intervention, and 30 untreated items matched for lexical frequency. The present study intended to train 20 items and to test both these and a matched group of a further 20 untreated items. For this reason, it was necessary to reduce the list used previously from 60 down to 40.

6.2.1.1 Item selection

Within the proposed therapy intervention, treated items could be practised in batches of five. Item selection was constrained in this way due to the development of the GeST tool as a set of six individual therapy programs – each programmed to train a specific set of five items. The implication of this constraint is that the choice to train 20 items required a selection of four of the six available batches of five items. The choice of batches was decided by assigning a randomly generated number to each batch, sorting the list of random numbers into ascending order and opting to use the first four batches assigned in the list. Undertreated items were selected from the remaining list of 40 items. In order to achieve an untreated set matched for its 'gesturability', the following procedure was adopted. An item-by-item analysis of all 60 gestures assessed within the GeST pilot project was conducted. This created a tally of the number of times each gesture was produced and correctly identified at T1 in the pilot study. As there were 9 participants in that study, this tally created a total gesturability score out of 9 for each gesture. An untreated item was selected from the remaining list to match the gesturability score of each treated item. The full list of selected items and matched untreated items can be found in appendix A. This method of matching was felt to be preferable to referencing lexical databases of spoken or written words on the basis that although such databases may be drawn from a larger corpus, their relevance is limited by the fact of their modality being one of non-gesture.

6.3 Novel Assessment Tools

Two novel therapy assessments have been developed:
The first aims to capture the effect of GeST practice upon gesture interaction between a participant and their everyday communication partner.
The second aims to capture people's use of and confidence with items of everyday technology both before and after Computer therapy practice.
Both assessments were developed through discussion with supervisors and have been trailed with a participant from the previous GReAT project pilot study.

6.3.1 Interactive Gesture Assessment (IGA)

The aim in developing this assessment was to enable collection of data regarding the use of gesture within a participant's interactive communication. Whilst the primary gesture outcome measure (6.2.1) allows for measurement of change of performance in gesture in isolation, for purposes of clinical relevance it was felt important to establish the carryover of any gesture gains into interactive communication with a participant's regular communication partner. Due to the relative paucity of established gesture assessments suitable for use with this population (i.e. those with severe aphasia), it was necessary to create a novel assessment to meet this need for this study. Full details of the final administration of this assessment are reported in methods section 7.10.2.

6.3.1.1 Development of the tool

In contrast to the primary outcome gesture measure, where recognition would be carried out using video review at a later date, this assessment was developed to capture live recognition by a participant's regular communication partner at the time of testing. Using a 'barrier' technique to obscure the target image from the communication partner, the participant would be asked to produce a gesture from a picture for their partner to identify. To make this interaction more reflective of live communication, the partner could speak to the participant to clarify or check their understanding of the gesture – as might be expected in a real life communicative exchange. Once identified, the partner would be asked to record their answer for each gesture on a score sheet. The score sheet would then be marked against the assessor's list of targets to establish an accuracy score.

6.3.1.2 Item selection

As for the primary outcome gesture measure, it was determined that the items examined with this tool should comprise equal numbers of treated and untreated items. This decision would enable examination for the presence of a therapy effect by looking at change for treated versus untreated items over time. The assessment was to be carried out at four time points across the study. It was decided to test six treated and six untreated items at each time point, creating a total of 12 items to be tested within each measurement. The motivation behind the length of this item list was to present a relatively small item sample which would not create excessive burden on the participant but which would reflect a meaningful sample of the trained and untrained items. This 12-item measure created the requirement for a total of 4X6 items of each category to be tested across all time points. Unlike the primary gesture assessment, stimuli items here were not coloured pictures printed on white paper but instead brief video clips showing a situation relevant to the target object followed by the presentation of a digital colour image of the specific item (Figure 18). The aim in using video stimuli was to more effectively simulate a real-life situation where a gesture might be called for. The completion of the video with a static picture prompt was to ensure the specific target to be gestured was clear to the participant in a way that might not be as obvious as if the stimuli comprised a video scenario alone.

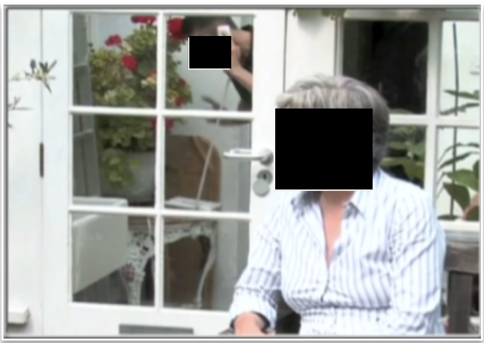
	
Part one: video, showing a telephone ringing	Part two: a static picture of the specific target item, telephone

Figure 18. Example Test Stimuli from the IGA

Video clips were drawn from the materials utilised in level 3 of GeST (described in section 5.3.3.6). This provided 20 video stimuli for use as the treated items, and an additional 10 video stimuli that could be used as untreated items. These comprised the additional 10 videos developed for the original set of 30 items trained within GeST but not selected for use within the current study. As noted above, the design of this measure necessitated that a total of 24 (4X6) items from each category (treated and untreated) be tested across the four time points. Due to the limited number of video items available for use as stimuli, it was necessary to present some items more than once across the four time points assessed. No item would be presented more than once within an individual assessment, however, some items would need to be presented up to three times across the different time points to ensure that 6 items from each category were represented at each time point. The distribution of these items across assessments is reported in appendix B.

6.3.1.3 Psychometric Properties

In order to investigate the face and content validity of the construct measured by this tool (i.e. whether the IGA might meaningfully capture the production of a gesture within an interactive communication context), assessments of feasibility were carried out. These aimed to provide evidence to expert judges (the author – a clinician with experience of gesture assessment in aphasia; a professor of aphasiology and a senior technology researcher) that the measure captured the skill of interactive gesture production. Tests of reliability (i.e. the ability of the IGA to measure interactive gesture production in a reproducible fashion), and validity (i.e. the actual success of the IGA in measuring the skill of interactive gesture production) were not undertaken prior to the main experimental study due to resource constraints. However, outcomes from the main study do provide opportunity for further examination of these properties and were examined retrospectively, subsequent to the main study and are reported in section 8.13.

6.3.1.3.1 Feasibility testing

An early version of the assessment was tested with two language researchers – one assuming the role of test participant and a second assuming the role of communication partner. Following this trial, task instruction wording was refined and two practice items were added to the start of the assessment. These items would be unscored but participants would receive feedback about the accuracy of their written response and the acceptable and unacceptable components of their interaction within the remit of the assessment. The updated version of this assessment was then trialled with a test user. This user was a former participant on the GReAT project and had severe aphasia and a familiarity with alternative methods of gesture assessment used within that project. His wife also took part in the trial – in her role as regular communication partner. Both participants gave feedback regarding their comprehension of the test materials and instruction format. Both reported that they had found the task acceptable. Both identified that they had found the presence of practice items helpful but that the communication partner would like better guidance that they were expected to identify a noun from the gesture and not a verb. In response to this, an additional reminder was added to the task instructions and to the partner's response sheet to highlight that the participant would be gesturing either a picture or an object. The words 'picture' and 'object' were additionally emboldened in the written instruction to make this more salient. As previously noted, full details of the final administration of this assessment are reported in section 7.10.2.

6.3.1.3.2 Retrospective analysis of test-retest reliability for the IGA

Following completion of the main study, an analysis of test–retest reliability was conducted for a small number of participants. This analysis found good agreement between total IGA scores across two time points separated by a five-week period comprising no intervention. Full results are reported in section 8.13.1.1.

6.3.1.4 Strengths

This measure has been tested for face and content validity. It represents an early-stage test, which offers a novel contribution in terms of capturing a more real-world measure of gesture success within an individual's everyday communication environment. Furthermore, good test-retest reliability was observed within retrospective assessment undertaken following completion of the main study. This is reported in section 8.13.1.1 and lends further support to the credibility of the measure.

6.3.1.5 Limitations

Prior to its use in the main experiment, this measure had not been tested in regards to a rigorous complement of psychometric properties. Whilst test-re-test reliability was explored retrospectively, other properties such as construct validity and sensitivity to change were not examined. For this reason, as for any novel, early-stage measure, its results should be interpreted with a measure of caution.

6.3.2 Technology Use and Confidence Measure

Anecdotal reports from the GeST pilot study indicated that the partners of some participants reported an improvement in the participant's levels of technology use and confidence in technology use throughout their involvement in the research project. Additional demographic information gathered at the outset of this same project indicated that many participants experienced a reduction in their use of everyday technology such as mobile phones and computers subsequent to the onset of their stroke. These two features were felt to be important factors to investigate to reveal the wider context within which computer delivered therapy should be considered for people with severe aphasia. The aim in developing the Technology Use and Confidence measure therefore, was to enable systematic collection of data regarding participant's use of and confidence in their use of technology. Due to the relative paucity of established technology assessments accessible for use with this population (i.e. those with severe aphasia), it was necessary to create a novel assessment to meet this need for this study. Full details of the final administration of this assessment are reported in methods section 7.10.3.

6.3.2.1 Development of the tool

To investigate this topic a simple questionnaire listing 17 items of everyday technology was developed. Participants were asked to indicate whether they had recently used the stated technology and also express their level of confidence in using that technology. This questionnaire was piloted with an individual with severe aphasia who had taken part in the GReAT project as a participant. Subsequent to its finished design, the questionnaire was also assessed for test retest reliability and used to capture information about the test items with related populations (i.e. adults with stroke but no aphasia and older adults with no stroke and no aphasia).

6.3.2.2 Item selection

Technology items for this questionnaire were selected on the basis of their likelihood to occur in a participant's everyday environment. Item selection aimed to represent a variety of digital and non-digital technologies and aimed to present a relatively short list that would not create excessive burden on the participant. A full list of items selected is included in appendix C.

6.3.2.3 Accessibility

Existing measures of technology use for older adults (Czaja, 2006; Office of National Statistics 2014) and even those with aphasia (Finch & Hill, 2014) rely heavily on the use of complex written material in order to assess levels of technology use. The current questionnaire presented items in pictorial format with supporting written text comprising only the target name of the item. The decision to present items in this way made to eliminate performance issues which might arise a consequence of not being able to understand the content of individual assessment items due to impaired reading abilities as a result of stroke. Furthermore, this prevents the reliance on carer support to complete the assessment, enabling participants to exercise autonomy over their responses. For similar reasons related to difficulties with expressive language, participants were not required to write or speak their response but instead to point to a "thumbs up" picture with the word **yes** beneath it if their response was yes, and a "thumbs down" picture with the word **no** beneath it if their response was no.

6.3.2.4 Confidence Scale

In order to gauge a participant's confidence in the use of a given item of technology, an existing scale measure was adapted for this assessment. The chosen scale was adapted (with permission) from an item used within the Visual Analogue Scale of Self-Esteem (VASES, S. Brumfitt & Sheeran, 1999a). The VASES is a measure of self esteem standardised for use with adults with severe aphasia. The choice to adapt it for use in this measure was taken to enable a scalable measure of participants' confidence in using technology across the duration of the study. Its previous validation with adults with severe aphasia (S. M. Brumfitt & Sheeran, 1999 b) gave an indication that it would be accessible to the population being assessed within the present study.

6.3.2.5 Psychometric assessment of the tool

Figure 19. Initial prototype of technology questionnaire

In order to investigate the psychometric properties of this tool, face and content validity was examined through assessments of feasibility and explorations of discriminative validity and test-retest reliability for an age-matched population were carried out. Further tests of of reliability (i.e. the ability of the measure to capture technology use and confidence in a reproducible fashion), and validity (i.e. the measure's in capturing the constructs of technology use and confidence) were not undertaken prior to the main experimental study due to resource constraints. However, outcomes from the main study do provide opportunity for further examination of these properties in relation to participants with severe aphasia and were examined retrospectively, subsequent to the main study. These are reported in section 8.13

6.3.2.5.1 Feasibility testing

An early version of this task was discussed with a speech and language therapist researcher. It comprised a one-page sheet displaying each item and asking participants to rate confidence on a numerical scale (Figure 19). Following review it was felt that the clarity of this questionnaire could be improved if items were presented one at a time, with a yes / no referent for participants to point to in response to each item. Furthermore, the numerical scale was adapted to include a visual metaphor (6.3.2.4) and to utilise a – to ++ scale in place of the number scale – utilising an item developed for the VASES assessment (Brumfitt and Sheeran, 1999). An example similar to the confidence scale used is presented in Figure20. For the measure used, 'Image A' showed a line drawing of a figure walking precariously along a tightrope. 'Image B' showed the same figure walking confidently along a tightrope. It was decided to ask participants their confidence level for each item they had reported using instead of the group as a whole. Finally, it was agreed to further specify the question 'which of these do you use now?' to include a timeframe. As a result, the question was adapted to become "Have you used this in the last month?"

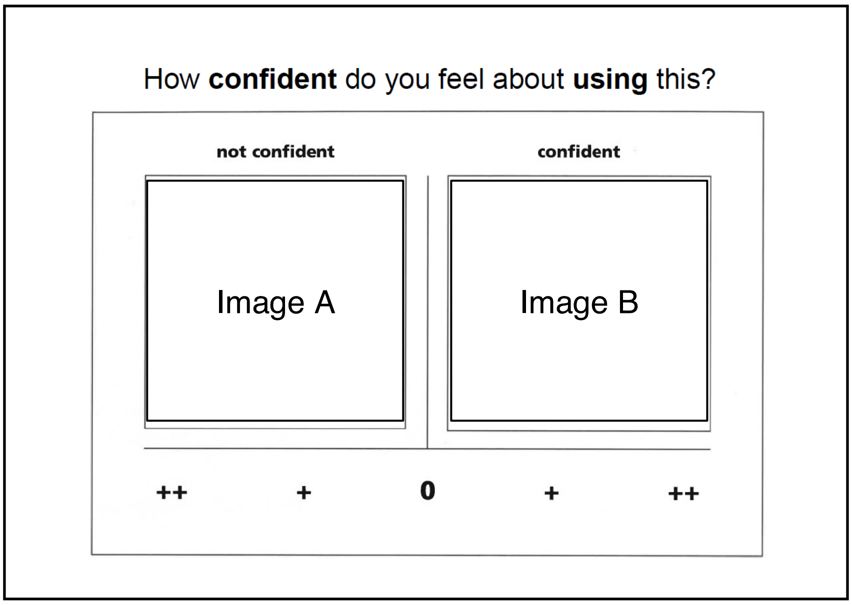


Figure 20. Example format of confidence scale used alongside the technology measure

Following the above refinements, the adapted version of this assessment was trialled with a test user. As for the development of the Interactive Gesture Assessment (IGA), this user was a former participant on the GReAT project and had severe aphasia and a familiarity with alternative methods of gesture assessment used within that project. The assessment was filmed and feedback was received regarding the comprehension of the test materials and instruction format. The test user identified that he had felt he understood what was expected of him and that the images and format of the items was acceptable. He made use of the yes / no text and image referent to indicate his recent use of each item and used pointing to identify his confidence levels on the picture scale. During interview, he indicated that he had experienced difficulty switching from the yes/no response to the confidence rating response for each item and also expressed a preference for being able to report confidence levels for each item and not solely those which he had reported using in the previous month. In response to this feedback, administration of the task was adjusted so that participants should provide yes or no responses to all items consecutively and then re-visit each item to rate it on the confidence scale consecutively – regardless of whether or not they had reported making recent use of it.

6.3.2.5.2 Test-retest reliability

The test-retest reliability of the Technology Use and Confidence Measure was examined by Shika Patel – as a research dissertation contributing to the qualification of MSc in Speech and Language Therapy at City University London. 50 participants without aphasia or stroke and between the ages of 45 and 95 were asked to complete the measure on two separate occasions. The test / retest sessions were separated by at least one month. A two-way, mixed method intraclass correlation (ICC) was conducted to compare outcomes between test totals on both occasions. An excellent degree of reliability (>0.9) was found between the two score sets for both the technology use score and the technology confidence score. For the technology use scores, the average measure ICC was .994 with a 95% confidence interval from .975 to .995 ($F(24,24) = 181.11$, $p<.001$). For the technology confidence scores, the average measure ICC was .998 with a 95% confidence interval from .996 to .999 ($F(24,24) = 1204.75$, $p<.001$). This indicates a high level of reliability for the use of this test over time, when measuring technology use and confidence in an age-matched population with no aphasia.

6.3.2.6 Use of the test with participants without aphasia – discriminative validity

To examine its use with the wider population of adults, the Technology Use and Confidence Measure was administered to a number of participants without aphasia. This data was collected by Emma Gould and Shauna Brown – as research dissertations contributing to the qualification of BSc in Speech and Language Therapy at City University London.
41 participants representing two different user groups undertook the measure: healthy adults (n=23); and adults with a stroke but no aphasia (n=18).

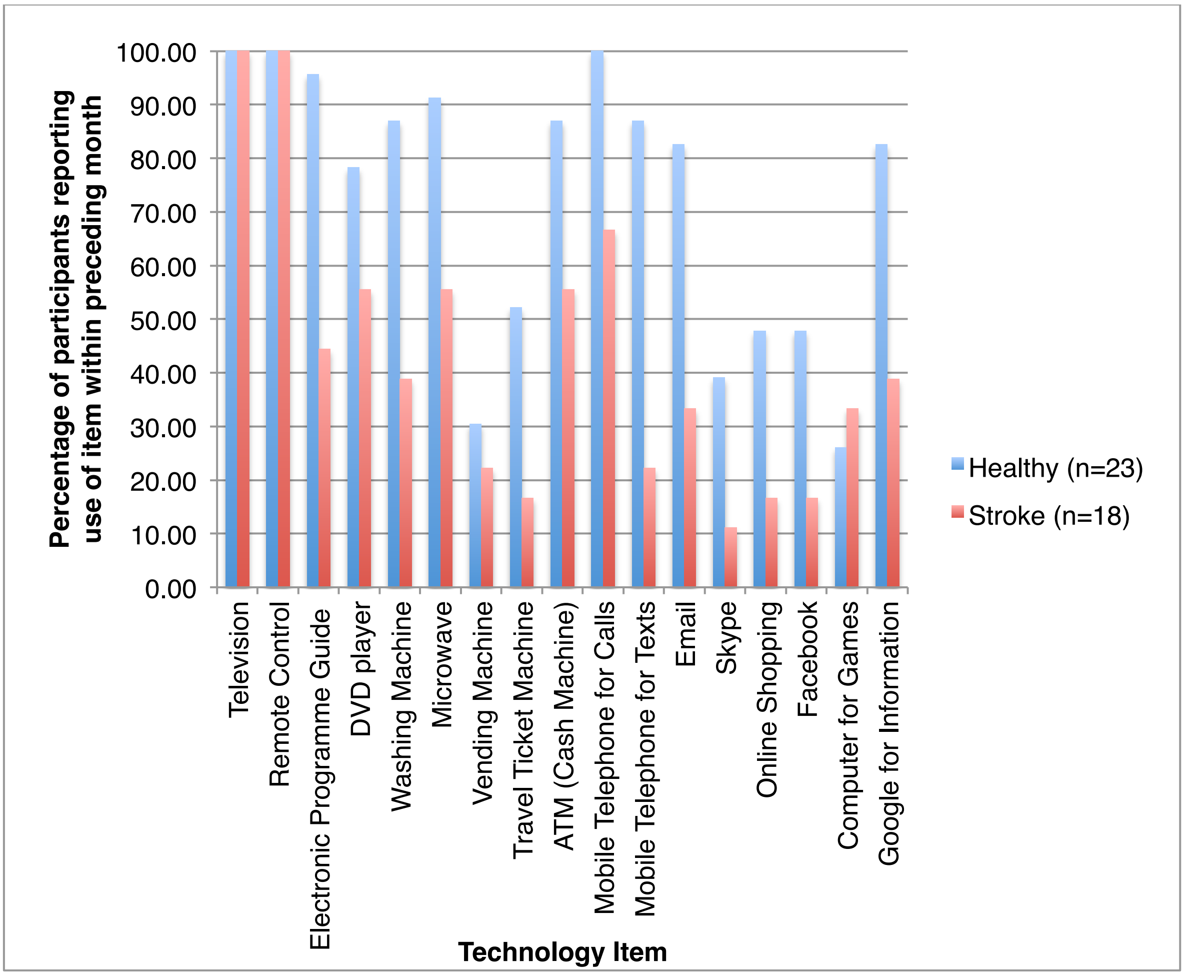


Figure 21 Percentage of participants from the two groups reporting use of specific technology items

Participants in both groups were recruited in urban and rural locations, via community groups. They were of mixed ethnicities and socio economic status. Groups did not differ with respect to age, or gender distribution. The healthy adults had no other diagnosed neurological or visual impairment. Participants in the stroke but no aphasia group demonstrated competent language abilities and reported no aphasia subsequent to stroke. Additionally, they reported no other diagnosed neurological or visual impairment. Further details of participant demographics are provided in appendix D

Outcomes from this deployment (Figure 21) indicate a substantial difference in the use of all items of technology examined with the exception of the Television and Remote Control items. Aside from this, all items with the exception of “computer use for games” showed greater levels of reported use in the healthy group. Additional ANOVA analysis revealed that participants in the stroke group reported significantly less use of technology than those in the healthy group ($F=21.05$ (1, 39), $p<0.05$). There were no significant differences in levels of confidence reported. These findings suggest that all strokes impair access to technology. Moreover, the fact that the stroke group could be differentiated from the healthy group indicates the discriminative validity of the technology use measure – in its ability to distinguish between groups. Outcomes for participants with aphasia, who took part in the main study, are presented in section 8.9.1.

6.3.2.7 Retrospective analysis of test-retest reliability for participants with severe aphasia

Following completion of the main study, an analysis of test–retest reliability was conducted for a small number of participants. This analysis found excellent agreement between total technology scores across two time points separated by a five-week period comprising no intervention. This mirrors the excellent agreement observed for age-matched participants with no aphasia in section 6.3.2.5.2. In contrast to the findings for participants with no aphasia however, poor agreement was found for the confidence totals across the two time points. Full results are reported in section 8.13.1.2.

6.3.2.8 Strengths

Using the data collected to standardise this measure, we can see that it offers good test-re-test reliability and discriminant validity for adults with no aphasia and that it is accessible for use with populations without aphasia and with a participant with aphasia. Due to the consultative nature of the development of the assessment – through discussion with expert peers – we can also assume face and content validity. Furthermore, good test-retest reliability of technology use (although not confidence) scores was observed within retrospective assessment undertaken following completion of the main study. This is reported in section (8.13.1.2) and lends further support to the credibility of the measure in relation to its use as a measure of technology use. In light of theses findings, this assessment can be seen as a novel tool with potential to capture levels of technology use in a group unable to access existing measures of this construct.

6.3.2.9 Limitations

For the technology confidence measure, in spite of good test-retest reliability for adults without aphasia, retrospective analysis of this property for adults with severe aphasia showed poor levels of agreement 8.13.1.2. In addition, other properties such as construct validity and sensitivity to change were not examined. For this reason, as for any novel, early-stage measure, its results should be interpreted with a measure of caution – particularly in relation to findings around confidence levels.

6.4 Software

6.4.1 PowerGeST

PowerGeST was developed as a means of extending the practice opportunities available within GeST. The present study sought to refine the GeST therapy protocol adopted within GReAT study (Marshall et al., 2013) to try and increase the magnitude of gains (i.e. the number of gestures learned). To illustrate precisely how this was achieved, a brief summary of the protocol adopted for the GReAT study will first be provided before an overview of the intervention protocol for the present study.

6.4.1.1 GReAT study protocol

Therapy delivery within the GReAT study comprised three weeks of therapist-supported intervention (i.e. autonomous individual practice supported by a weekly session of therapist facilitation) followed by three weeks of autonomous individual practice. Within each week, participants practised a different set of five gestures. This meant that the spread of 30 gestures was distributed across a six-week period. Findings revealed that participants using GeST demonstrated significant gains in gesture production directly after completion of the therapy protocol. These were additionally maintained after a maintenance period. The gains were, however, restricted to those 15 items treated within the three weeks of therapist-supported intervention. Additionally, the margin of gains was small (an average of around 2 items per participant).

6.4.1.2 Present study protocol

The current study adopted only the therapist-supported period of intervention within its protocol - in order to capitalize on the most fruitful intervention components established within the GReAT project. This, previously three-week, period was extended to a five-week period within the current study. During weeks one to four, participants practised a different set of five gestures weekly (20 in total) using GeST. Week five aimed to provide an additional novel 'revision' period, which allowed participants equal exposure to all 20 previously trained gestures within the same practice period. (Full description of therapy protocol is reported in Methods chapter - section 1.7) The inclusion of this final 'revision' period, offered participants an opportunity to receive an increased 'dose' of therapy for each item. The addition of this adjunct also sought to reinforce gestures learnt in the first one or two weeks of therapy that had not been practiced within GeST as recently as those introduced in weeks three and four. The aim here was to diminish the disadvantage these gestures may have faced in testing from not having been practiced so recently.

In order to achieve this final week's revision period, a supplemental practice intervention – PowerGeST – was developed by the researcher. Using principles established during the development of GeST, the researcher compiled the tool not in the 'Unity' programming environment used to create GeST but instead in the more widely available Microsoft PowerPoint software. The choice to utilise PowerPoint to deliver the supplemental practice material was based on the researcher's previous experience with creating therapy tools using this software and the absence of additional external resources to develop and extend the development of GeST within Unity. The created tool was dubbed PowerGeST – to reflect its origins in the GeST tool and its utilization of PowerPoint software.

PowerGeST was developed by the author of this thesis using insights gained through her involvement in the original development of GeST (Galliers et al, 2012). Many design choices implemented within PowerGeST arose from principles established during the development of GeST, which was developed in consultation with five users with aphasia to provide a system and interface accessible to users with severely impaired language and concomitant hemiplegia. The author of this thesis worked as a research speech and language therapist on the development of GeST - alongside fellow speech and language researchers, human computer interaction design researchers and a software developer. The following describes the development process for design choices made within PowerGeST.

A number of iterations of PowerGeST were created and refined before the final design was decided. An early prototype of PowerGeST was trailed over a four-day period at home with a test user. This user was a former participant on the GReAT project and hence had a familiarity with the process of computer gesture therapy and the GeST software. He shall, henceforth, be referred to as test-user. Following a brief demonstration and introduction to the prototype, the test-user was requested to practise with the system once a day. Following this period, the researcher filmed the use of the PowerGeST prototype and interviewed the test-user in situ about the use of the system and any barriers he had encountered during its trial. Outcomes from this trial were later reviewed in consultation with a human computer interaction design (HCID) researcher in order to address any outstanding issues of usability. The following section will report each design component of PowerGeST, identifying any changes that were implemented during the design process.

6.4.1.3 PowerGeST file format

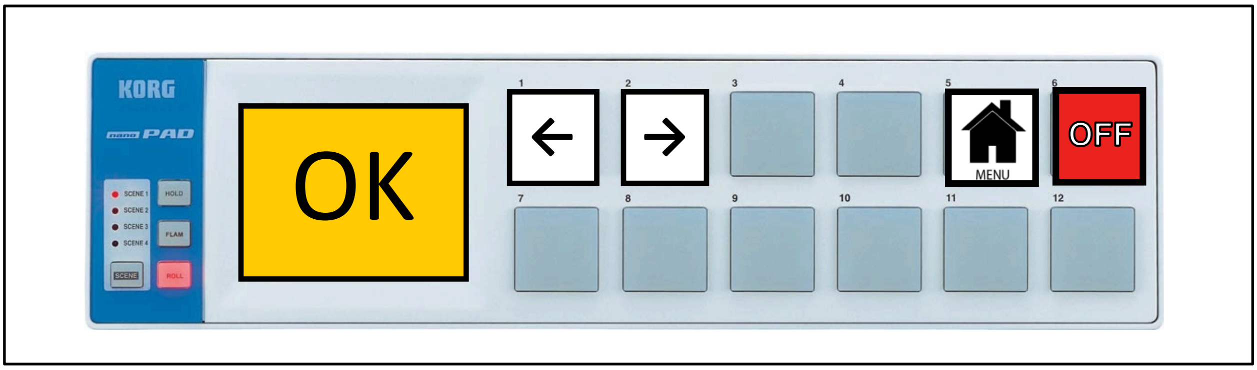
As previously stated, PowerGeST is operated through PowerPoint software. It runs as a modified version of a typical PowerPoint Presentation, in which a sequence of slides is presented in order. Some 'slides' may contain videos, animations and sound effects. Participants can progress through a predefined set of content using key presses or mouse clicks. For PowerGeST, the created resource was saved as a PowerPoint Show (.ppsx) instead of the more typical PowerPoint Presentation file (.pptx). The key difference between these two file formats is that the former enables the presentation to run directly from being opened without the need to navigate through the menu system within PowerPoint to launch or 'view' the slide show. For example, selecting and opening the file from its icon launched the presentation in full screen, instead of opening up the PowerPoint editing window.

6.4.1.4 Run from Startup

PowerGeST was operated within a laptop running the Windows 7 operating system and employing Microsoft PowerPoint 2010. The PowerGeST.pps file was saved within the Windows startup facility in order to launch directly as soon as the computer was switched on. This startup method employs the same technique used to ensure the startup of GeST from switch on. The aim for this method was to avoid placing multiple navigational demands on participants when accessing the therapy intervention. Utilizing the PowerPoint Show format to save the PowerGeST tool, further eschewed the navigational demands required to operate a typical presentation file within PowerPoint.

6.4.1.5 Interacting with the PowerGeST slideshow

In order to interact with the PowerGeST slideshow, navigational buttons were mapped to a KORG "Nanopad" keyboard (Figure 22). This simplified, customized keyboard was initially created to match the demands of the tasks within the GeST tool, meaning that a given button press would create a predictable interaction with the GeST software (e.g. the ‘OK’ button would move to the next screen). The principles applied for GeST were modified here to allow similar keyboard navigation within the PowerGeST show. Using the understanding that a regular PowerPoint show can be navigated using keyboard buttons, the external keyboard was linked to PowerPoint keyboard shortcuts and Windows operating system commands as described in Figure 12.



Keyboard Button	OK	←	→	MENU	OFF
Mapped to Windows Command	Spacebar	Left arrow	Right arrow	-	Windows Shutdown
Rationale	Within PowerPoint, spacebar enables progression to the next slide or action identified in the slideshow.	Within PowerPoint, the left and right cursor keys enable progression backwards and forwards (respectively) through the allocated sequence of slides and actions within the slideshow.		Not mapped -PowerGeST did not employ a menu-based navigation system.	This enables Windows to shutdown entirely, eliminating need for additional system navigation after practice.

Figure 12. Initial keyboard and operating system commands for PowerGeST

6.4.1.5.1 Refinement

During the user trial with the former participant from the GReAT project, it became apparent that the above mapping did not allow users to consistently navigate forward through PowerGeST using the OK button. The test-user demonstrated that the slideshow sometimes froze and that depression of the OK button (or the <- or -> buttons) failed to resolve this. Upon investigation, it was established that this occurred due to the Window's system losing 'focus' from the PowerPoint application and switching it to a different background application running within Windows. Whilst this created no obvious visual impact upon the presentation screen in terms of notification windows etc., it meant that that keyboard was no longer instructing the PowerPoint application and hence the button presses could have no effect. Through trial and error, it was established that 'focus' for the PowerPoint application could be re-established by clicking anywhere on the screen with the mouse. This re-enabled the keyboard mappings to interact with the PowerGeST slideshow and operate it as intended. A limitation of this resolution however, is that it required additional task-switching demands from the user/participant (to resolve an unpredicted breakdown in anticipated use of therapy software), plus the use of an additional input device – the mouse – which had not been necessary for the previous software tool, GeST. To resolve this, the **OK button was mapped instead to the 'right-mouse-click'** command within Windows. This solution meant that participants could re-gain command of PowerPoint using the OK button they were used to using to continue through the program. Conveniently, the right-mouse-click also allows progression forward through a PowerPoint presentation in the same way the space bar had done in the previous iteration. This proved to solve the difficulty experienced by the test-user.

6.4.1.6 Choice of stimuli and proposed interaction

PowerGeST was developed for use after a four-week period of intervention with GeST. It used video stimuli developed for GeST in order to stimulate gesture practice for participants. Aiming to build on the increased demands of unprompted gesture elicitation required by level 3 of GeST (5.3.3.6), gesture stimuli used were the videos employed in this 3rd level of GeST. The intention for this was to promote increased autonomy in participant gesture production by excluding an explicit verbal prompt to repeat the gesture modelled in the stimulus video. Videos for all 20 items (previously trained in batches of five using GeST) were presented in sequence using PowerPoint slides, with the intention that participants mirror each gesture in turn. Within level 3 of GeST, the video for each stimuli progresses as a series of four steps:

1. Situation
2. Gesture
3. Spoken acknowledgement of gesture label
4. Resolution of situation

(This is illustrated fully in section 5.3.3.6)

All four steps of the video are presented in sequence level 3 of GeST. However, the decision was made to segment the videos within PowerGeST so that participants saw sections 1 and 2 followed only. At this point the sound of a bell was played to encourage participants to practice the demonstrated gesture. Then, upon pressing OK to continue, the 3rd and 4th sections of the clip were presented as a resolution (Figure 23). Having watched these final sections of the video, participants could press OK again to begin watching the next video.

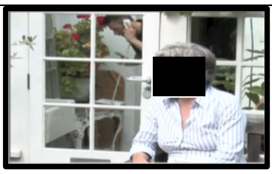
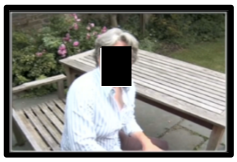
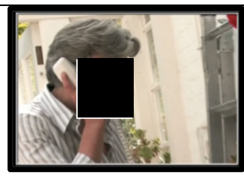
		Participant copies gesture, then presses OK		
1. Situation	2. Gesture		3. Spoken Label	4. Resolution

Figure 23. Initial sequence of interactions with each gesture item in PowerGeST

6.4.1.6.1 Refinements

During the user trial and feedback the test-user indicated that he felt sections 3 and 4 of the videos were not helpful for his practice, demonstrating that he would prefer instead to continue directly on to the next gesture without watching them. This sentiment was echoed by the interaction design researcher who advised that the inclusion of sections 3 and 4 of the video interrupted the user's goal in using the system – namely to imitate gestures. She advised that PowerGeST could be made more usable by eliminating these components. As a result, the final version of PowerGeST contained only sections 1 & 2 of each video (i.e. the situation and the gesture) and the subsequent OK-press progressed the participant on to viewing sections 1 & 2 of the next gesture item instead of playing back the 3rd and 4th components of the video. All 20 trained items were presented in sequence using this format.

6.4.1.7 Screen presentation

Content developed for PowerGeST was created with the intention of being relevant and engaging for target users. All screens utilised were intended to present limited visual clutter and to maximise the focus of the attention towards the required target of interaction. During presentation, participants were presented with one video at a time, which appeared in the centre of the screen and began playing upon depression of the OK button. Once completed a bell sounded to indicate that the user should imitate the gesture. After a 2 second pause, the OK image appeared in the centre of the screen indicating that the participant should press the OK button to proceed.

6.4.1.7.1 Refinements

The user trial revealed a tendency for the test-user to depress the OK button multiple times in a row. This had the effect of skipping the PowerGeST sequence forward through a number of subsequent actions with the undesired consequence that the OK button appeared over the top of the video as it was playing - hence masking the video model of the target gesture (Figure 34)

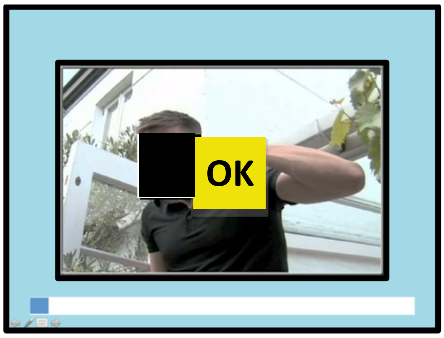


Figure 34. Gesture demonstration obscured by OK button prompt

This issue was resolved by relocating the OK button to the top centre of the screen so that even if it did appear accidentally – as the result of over-zealous button clicking – it did not obscure the video model of the target gesture (Figure 45).

Figure 45. Visible gesture demonstration and relocated OK button prompt

6.4.1.8 Progress Indicator

It was felt that as PowerGeST comprised a series of twenty gestures in a row – in contrast to batches of five for GeST – it would be constructive to indicate to users how far they had progressed through the sequence of 20 and hence how much practice they had left to complete. This was initially illustrated by the inclusion of a progress slide after each batch of five gestures. This slide presented a circle cut into four segments, with each segment being filled in after five consecutive items. After all 20 items had been demonstrated, the circle was again presented and shown to be complete – indicating the completion of the exercise.

6.4.1.8.1 Refinements

This initial progress indicator proved unclear to test-user, who expressed confusion as to the purpose of the device. Subsequent discussion with the HCID researcher supported this critique on the same grounds as she had challenged the inclusion of the 3rd and 4th sections of the video stimuli clips – namely that it interrupted the user's main goal in using the system. With this in mind, progress through the sequence of videos was next explored by showing a number at the bottom of the screen to indicate which gesture of 20 was being presented (e.g. 12/20). This eliminated the issue of interrupting the flow of use, however it was felt to be too reliant on number comprehension – a feature which can be a challenge for individuals with severe aphasia. The final design made use of a white progress bar, divided into 20 segments and situated along the bottom of the screen. Segments turned from white to blue with each incremental item filling the bar from left to right as the practice progressed (Figure 4).

6.4.1.9 Logging use of PowerGeST

PowerPoint software does not automatically create an output log of the time, date and manner in which its presentations have been used. This feature would be a desirable component of PowerGeST in order to monitor dosage and patterns of use in a similar manner to GeST. In order to try and implement this feature for PowerGeST, the researcher contacted an open web forum of developers working on the use of Microsoft Office[1]. A member of the Microsoft "Most Valuable Professional" community created an Add-in which could be installed to individual versions of PowerPoint in order to document which slides have been visited, and the times and dates of each use. The data was captured as a .csv file (as is the case for GeST) and saved to a designated output folder.

[1] http://answers.microsoft.com/en-us/office/forum/office_2010-powerpoint/is-there-a-way-to-log-usage-of-a-powerpoint-show/fedec0af-9c35-45e1-a9cd-e98e3ad2465d

6.4.1.10 Use of Gesture Recognition Software

Gesture Recognition software was not utilised within PowerGeST. The choice not to include a gesture recognition component within PowerGeST was based on the developing researcher's lack of experience in this field of programming. Building on the premise that participants would have received four weeks of gesture practice with the use of gesture recognition prior to embarking upon their use of GeST, it was felt that the ability to practice gestures with this subsequent reduced level of feedback, would be beneficial in moving towards autonomous, unprompted use of gesture away from the computer and in real life. The absence of computer gesture recognition further dispensed of the need for the participant to wear a yellow glove during practice, again moving towards a more real-life environment for using gesture.

The above design process created the final version of PowerGeST implemented within the current study. The use of this tool is described fully in chapter 4 of this thesis - Description of Therapy.

6.4.1.11 Usability of PowerGeST Design in Relation to GeST Design

The features described aimed to not only capitalise on the design lessons learnt through the development of GeST, but also to address some of the shortcomings identified during it's pilot study. Overall, outcomes from the pilot study of GeST were very favourable in terms of usability – with family members and carers reporting that participants had largely used the tool independently. Marshall and colleagues however, do report the following challenges:

> "Two partners noted that the gesture recognition was variable, and two recalled instances when the computer failed to shut down. One partner indicated that navigation between levels was challenging."
>
> Pg. 1140 Marshall et al. 2013

PowerGeST circumvents issues of gesture recognition variability – eliminating this component altogether. It utilises a shutdown key mapped to the external "nanopad" keyboard instead of on the "Enter" button located on the laptop's busy, main keyboard. Furthermore, navigation is limited to just one tier – with no menu hierarchy - avoiding the challenges introduced by this additional demand.

6.4.1.12 Strengths

PowerGeST offers a computer gesture practice tool developed to be accessible to adults with severe aphasia. It is operated using an external keyboard with demonstrated acceptability and usability and provides a means of extending intervention opportunities established through the use of GeST

6.4.1.13 Limitations

It should be noted that the early testing of this tool in consultation with only one user with aphasia is a limitation of the design. Further reflection on the usability and effectiveness of PowerGeST can be established through its subsequent use by a larger participant group within the current study.

6.5 Summary

The preceding chapter documents the novel contributions developed and refined within the current study. These include contributions to the assessment of gesture and technology and also to the intervention of impaired communication using computer-based gesture practice. Critically, these tools have been created / adapted to ensure their suitability for a population with comparatively little established research in these areas. Each novel component demonstrates a measure of feasibility and acceptability. Use within the current study will reveal further information about the practical limitations and benefits of these early-stage tools.

Chapter 7. Methods

7.1 Introduction

The structure for this chapter adheres to CONSORT group guidelines for reporting randomised group trials (Schulz, Altman, & Moher, 2010). As such, items covered herein reflect those suggested within Moher et al.'s (2010) guideline document. Briefly, the trial design is first introduced, followed by participant inclusion criteria; details of screening measures, candidacy measures, repeated measures, scoring methods and concluding with proposed analysis techniques.

7.2 Funding and Ethical Approval

7.2.1 Funding

Funding for this project was provided in the form of a three-year City University PhD scholarship award with a £1,000 stipend for project expenses.

7.2.2 Ethical Approval

Ethical approval to conduct the study reported herein was granted by the City University London School of Community and Health Sciences Research Ethics Committee (Appendix E).

7.3 Trial Design

7.3.1 Research Design

This was a waitlist controlled feasibility study with balanced [1:1] randomisation.

7.3.2 Summary of Research Protocol

This design develops previous research into computer-delivered gesture therapy (Marshall et al, 2013). Computer gesture therapy is not only subject to better scrutiny with a larger number of participants, the more complex study design utilised here allows an examination of the feasibility of both the therapeutic intervention being employed (described in 7.8) and the research methods utilised to examine it. An outline of the design is provided in Figure 26. The research design enables two types of analysis – between group comparison and within group comparison. The between group design comprises parallel groups: an **immediate** treatment group and a **delayed** waitlist control group. Using pseudo-randomisation, the group allocation ratio is intended as 1:1. This design enables us to establish the required numbers for a randomised control trial (RCT) of the described therapy computer gesture therapy model and to establish the feasibility of a larger scale study of the model. Within group features of the research design allow for comparison of individual participants' performance against themselves over a series of time points. This allows us to establish the longer-term effects of intervention. Video recordings of individual therapy sessions and log data provided by the therapy software enable an examination of if and how the therapy technology is used. Furthermore, they facilitate a discussion as to the usability of such computer therapy by participants with severe aphasia.

Following screening, 20 participants completed a range of candidacy assessments examining gesture comprehension, language, cognition and praxis (7.7). Participants were then randomly allocated to one of two groups: immediate therapy or delayed therapy via the randomisation process described in sections 7.16, 7.17 and 7.18. Both groups underwent the first of four repeated measure assessments of gesture, naming, interactive gesture and technology use and confidence (time point 1). As indicated in26, those in the **immediate** therapy group next received a 5-week period of computer gesture intervention (four weeks supported therapy with GeST and a further week with PowerGeST). Those in the **delayed** therapy group received no intervention for this five-week period. Repeated measure assessments were carried out again for both groups (time point 2). Those in the **delayed** therapy group then received a 5-week period of gesture intervention whilst those in the **immediate** group receive no intervention. Repeated measures were carried out a third time for both groups (time point 3). Following a final five-week period with no intervention, participants from both groups underwent a fourth and final set of repeated measure assessments (time point 4). In addition to the assessments noted above, video recordings of the participants using PowerGeST were taken at the outset of its use and after one week of practice. Assessments included within the research design were chosen for their usefulness and feasibility within the home setting where assessment would take place and with the population being tested in mind – namely adults with severe aphasia and associated physical and cognitive limitations.

This waitlist control design was chosen in order to enable both within subject and between subject comparisons. If we look only at the data gathered up to and including time point two, the design mirrors that of a randomised control trial. We can compare immediate therapy (treatment group) outcomes with delayed therapy (control group) outcomes and assume that differences between outcomes can be attributed to the therapy intervention. Looking at the data within groups, we are able to compare individuals against themselves over time to establish whether changes occur at all and whether these can be attributed to a specific time point of intervention.

In addition to investigating computer gesture therapy with a larger group of participants with severe aphasia, the above design extends previous research in a number of ways: comprising a novel computer therapy component; novel assessments of interactive gesture; technology use and confidence as well as the inclusion of a range of candidacy measures not previously investigated.

	Time Point 1		← 5 weeks →	Time Point 2	← 5 weeks →	Time Point 3	←5 weeks→	Time Point 4
Immediate Therapy Group	**Candidacy assessments**: • CAT single word naming • CAT spoken single word and sentence comprehension • CAT written single word comprehension • CLQT non-linguistic executive function • BUPS limb praxis	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence	**Gest + PowerGeST Intervention** 2 x video recordings of PowerGeST being used	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence	No input	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence	No input	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence
Delayed Therapy Group			No input		**Gest + PowerGeST Intervention** 2 x video recordings of PowerGeST being used		No input	

Figure 26. Research Design Protocol

7.4 Changes to trial design

No significant changes were made to the methods after trial commencement.

7.5 Study settings

All assessments and therapy were carried out in the participants' homes. These comprised locations around the South East and South West of England.

7.6 Participants

7.6.1 Recruitment

The initial cohort of participants was recruited via community stroke support groups around London and the South East of England or via referral from either a fellow researcher or a family member or friend contacting the university. A small number, additionally based in the South East of England were recruited via a neurological rehabilitation charity. A subsequent cohort of participants was additionally recruited from the South West of England, again via community stroke support groups and also referral from a family member.

7.6.2 Eligibility Criteria

Eligibility criteria for participants were as follows:

1. Severe expressive aphasia following one or more Cerebral Vascular Accident(s) (CVAs)
2. 6 months or more post CVA
3. Fluent pre-morbid use of English
4. Willingness to take part in a computer-delivered gesture therapy in their home
5. Ability to relate real world objects to photographic and line drawing representations of the type used within assessment and intervention (scoring 60% or more on a novel assessment of object to picture matching)

6. Limited ability to name pictures aloud (scoring less than 30% on a standardised spoken picture naming assessment – CAT spoken naming subtest [Swinburn, Porter, & Howard, 2004])

Criteria 1 – 4 were judged through discussion / background interview and criteria 5 and 6 were judged using novel and standardised assessments.

7.6.3 Screening

Three measures were used to **screen** participants, and ensure that they met the study recruitment criteria. Measures were administered once consent had been gained. These comprised a case history interview, a novel object-picture matching assessment and a standardised assessment of picture naming. The following section describes the measures in more detail. Further rationale for specific assessment choice is provided in sections 7.6.4, 7.6.5 and 7.6.6.

7.6.4 Case History/Demographic Information

7.6.4.1 Rationale

For the purposes of judging eligibility according to criteria 1-4 and further to describe participant and group characteristics, a brief initial questionnaire of basic demographic and health information was conducted (included in Appendix F).

7.6.4.2 Administration

The questionnaire was typically completed with support from the participant's family member / carer / friend. It comprised an informal interview read aloud from a question sheet by the researcher. Additional written and gesture support was provided as appropriate[1]. Participants were asked their age, the date of their stroke and whether they experienced any visual, hearing or mobility issues. For instance, whether they required glasses to view a

[1] For example, if a participant indicated that they had not understood a spoken question, the written text on the question sheet would be shown to

computer screen or had difficulties viewing any aspect in their range of vision. They were also asked to report pre-morbid levels of computer and mobile telephone use, primary and additional languages and their work history.

7.6.5 Object picture matching

7.6.5.1 Rationale

Picture recognition skills form a sizable component of experimental measures used within the current project and are additionally necessary for successful use of GeST. This object picture-matching test examined participants' ability to relate objects and gestures to both photographs and line drawings.

This assessment examined participants' ability to relate objects and gestures to both photographs and line drawings, as these were required for completion of the experimental measures, and for successful use of GeST. Based on a novel assessment developed by Caute et al.(2013) and used to screen participants in two previous gesture studies (Marshall et al., 2012; Marshall et al., 2013Marshall et al, 2012), it provided a basic guide to an individual's ability to match a physical object to one of a choice of four photographs and later to a choice of four line drawings.

7.6.5.2 Administration

Participants were presented with four photographs of objects (Figure 27). The researcher then presented a single, physical object and the participant was asked to point to the corresponding picture from the four available options. The given array included the target picture and three distractor items. Distractors were chosen from the range of target items tested throughout the assessment. All were therefore items that could be manipulated and therefore represented by a manual pantomime.

Figure 27. Photographs from the object-picture matching test

The above process was repeated for five items, after which, participants were next asked to identify a further five objects in turn from a set of four line drawings at a time (Figure 28).

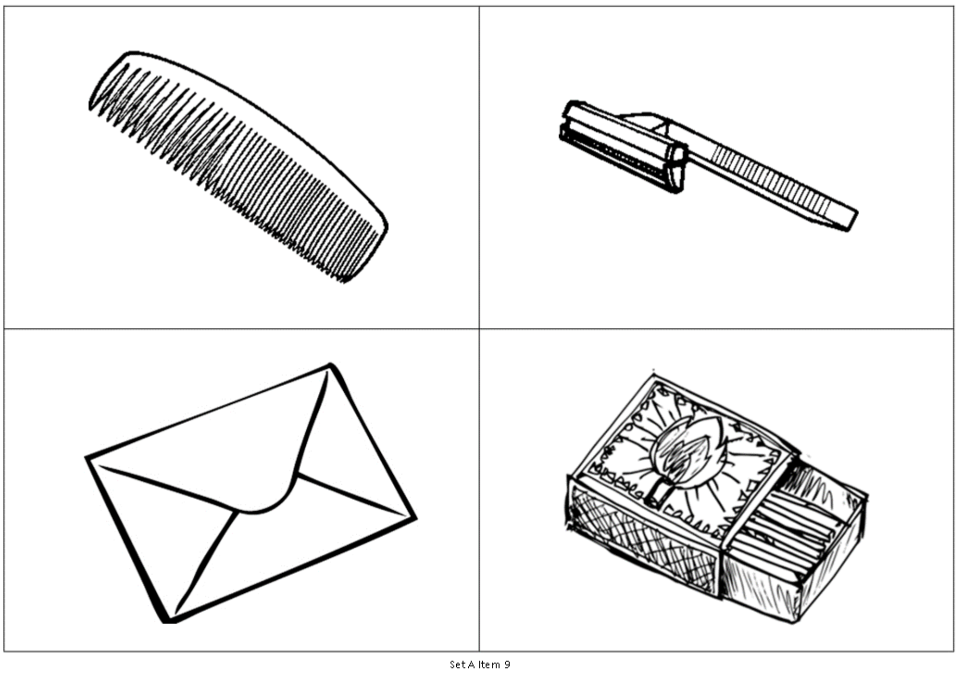


Figure 28. Line drawings from the object-picture matching test

Participants scoring 60% or less (≤6/10) for this assessment were excluded from progression to the main study on the grounds of a limited ability to relate real world objects to photographic and line drawing representations.

7.6.6 CAT single word naming assessment

7.6.6.1 Rationale

Severity of expressive aphasia can be measured through performance on a standardised assessment of picture naming. The *single word spo*ken naming assessment within the Comprehensive Aphasia Test (CAT, Swinburn et al., 2004) was used in this instance. This measure is a standardised naming subtest used widely in clinical practice. It has additionally been employed in previous gesture therapy trials (Marshall et al., 2012; Marshall et al., 2013).

7.6.6.2 Administration

Participants were presented with a line drawing of an individual object (Figure 29) and asked to state the name of that object. 24 items were presented in total. Responses were transcribed/recorded by the researcher and scored for accuracy.

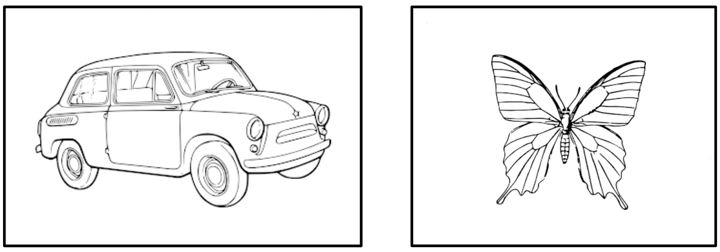


Figure 29. Example images similar to picture stimuli for the CAT single word naming assessment (Swinburn et al., 2004)

7.6.6.1 Scoring

Participant responses were recorded as a total score out of 48 – with two points being awarded for each item identified immediately and one point for each item identified after a delay or a self-correction. As GeST+PowerGeST was designed as a tool for people with severe expressive aphasia, participants scoring 30% or higher (≥16/48) on this measure were excluded from proceeding any further on the grounds that a compensatory gesture therapy at this level may not provide a substantial level of benefit to their existing communicative capacity. Those scoring less than 30% (<16/48) were invited to proceed to the full research protocol. The choice was made to assess candidate on all 24 items of this assessment (not applying the 'discontinue after eight consecutive failures' rule). This decision was made to ensure a comparative picture of all candidates across the full set of 24 items.

7.7 Candidacy Measures

Participants meeting the specified screening criteria would next undertake the following range of candidacy measures. A total of six measures were used to gather background information about the participants and to explore **candidacy** for GeST + PowerGeST through later correlation with demographic information and experimental measures. These comprised five standardised assessments and one informal assessment. The standardised assessments consisted of three measures of aphasia, one of executive function and one of praxis. The informal assessment administered examined gesture comprehension. Further rationale for individual assessment choice is provided in sections 7.7.1 to 7.7.13.

7.7.1 Gesture picture matching

7.7.1.1 Rationale

The ability to comprehend gesture may provide useful guidance as to candidacy – indicating a participant's overall ability to make use of gesture both receptively and expressively. Whilst there are no formalised assessments of receptive gesture comprehension, this study makes use of a measure that has previously been implemented in existing research of gesture intervention for aphasia (Caute et al., 2013). It should be noted that this measure has not been scrutinised for psychometric properties; however, this may create a case for investigation of this property in future related work.

7.7.1.2 Administration

Using photograph stimuli from the object-picture matching test (Figure 27), participants were asked to identify a gesture produced by the researcher from a selection of four photographs. This was repeated for 10 items. Gestures comprised a one-handed pantomime production of the use of each target item pictured. The vocabulary of gestures used within the assessment was consistent for each participant. Gestures produced were separate from those items trained or tested within the primary gesture outcome measure used within the study.

7.7.2 CAT Subtests - spoken single word and sentence comprehension, written single word comprehension

The first three spoken and written language comprehension measures described below have been taken from the Comprehensive Aphasia Test (CAT, Swinburn, 2004). As is the case for the CAT single word naming assessment used during screening, these assessments are used widely in clinical practice and have additionally been employed within previous gesture therapy trials (Marshall et al., 2012; Marshall et al., 2013).

7.7.2.1 Rationale

Basic spoken language recognition skills are required to complete experimental measures used within the current project and are additionally necessary for navigation through GeST. The severity of receptive aphasia for spoken language can be measured using standardised assessments of spoken word-to-picture matching and spoken sentence-to-picture matching. The *spoken single word and sentence comprehension assessment*s within the CAT (Swinburn et al., 2004) were used in this instance. Basic single-word written prompts are additionally employed within GeST. Comprehension of single written words can be measured through performance on a standardised assessment of written word-to-picture matching. The *written single word naming assessment* within the CAT (Swinburn et al, 2004) was used in this instance.

7.7.2.2 Administration - CAT spoken single word comprehension assessment

This is a standardised comprehension subtest from the CAT (Swinburn et al, 2004). Participants were presented with an array of four line drawings of objects (Figure 30). The researcher then read aloud a single word and the participant was asked to point to the corresponding picture from the four available options. The given array included the target picture, e.g. mouse; a semantically related distractor e.g. rabbit; a phonologically related distractor, e.g. house and a semantic distractor for the phonological distractor, e.g. tent. Following the introduction of the task with one practice item, 15 target items were then tested and scored.

7.7.2.3 Scoring

Participant responses were recorded as a total score out of 30 – with two points being awarded for each item identified immediately and one point for each item identified after a delay, as a self-correction or following the request of a repetition of the from the test administrator.

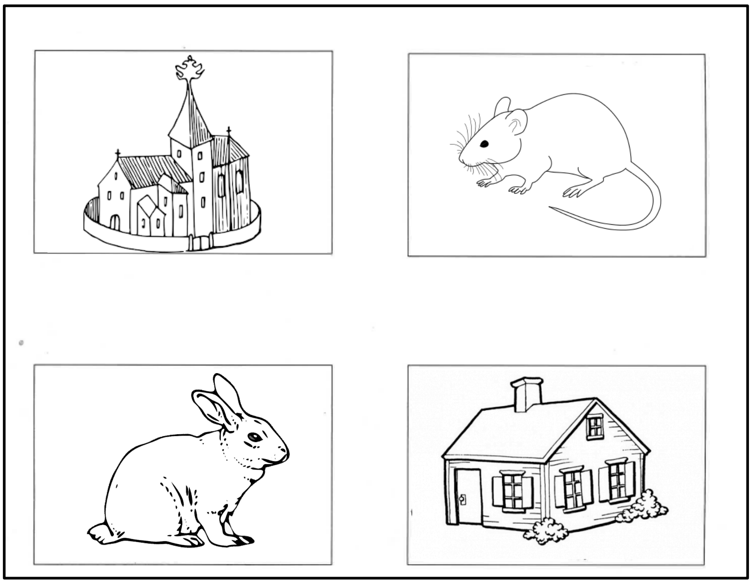


Figure 30. Example picture stimuli similar to the CAT spoken single word comprehension assessment (Swinburn et al., 2004)

7.7.2.4 Administration - CAT spoken sentence comprehension assessment

This is a standardised sentence comprehension subtest from the CAT (Swinburn et al, 2004). As for the single spoken word comprehension test above, participants were presented with an array of four line drawings. The researcher read aloud a single sentence and the participant was asked to point to the matching picture. For example, for the target sentence “She is sitting” the array included a target picture of a woman sitting on a stool. Distractor items varied from the target by either one or two key words. Distractors here included a picture of a man sitting, a picture of a woman standing and a picture of a man leaning. Here, keywords differing from the target are underlined. Following the introduction of the task with one practice item, 16 target items were then tested and scored.

7.7.2.5 Scoring

Participant responses were recorded as a total score out of 32 – with two points being awarded for each item identified immediately and one point for each item identified after a delay, as a self-correction or following the request of a repetition from the test administrator.

7.7.2.6 Administration – CAT written single word comprehension assessment

The CAT written single word comprehension assessment is a standardised written comprehension subtest from the Comprehensive Aphasia Test (Swinburn et al, 2004). As for the spoken comprehension tests above, participants were presented with an array of four line drawings (Figure 31). A single target word was written in printed text at the centre of the page. The participant was asked to point to the matching picture. As above, the array includes the target picture, a semantically related distractor, a phonologically related distractor (which is also visually similar to the target when the two words are written down), and a semantic distractor for the phonological distractor. Following the introduction of the task with one practice item, 15 target items were then tested and scored.

7.7.2.7 Scoring

Participant responses were recorded as a total score out of 30 – with two points being awarded for each item identified immediately and one point for each item identified after a delay, as a self-correction or following the request of a repetition from the test administrator.

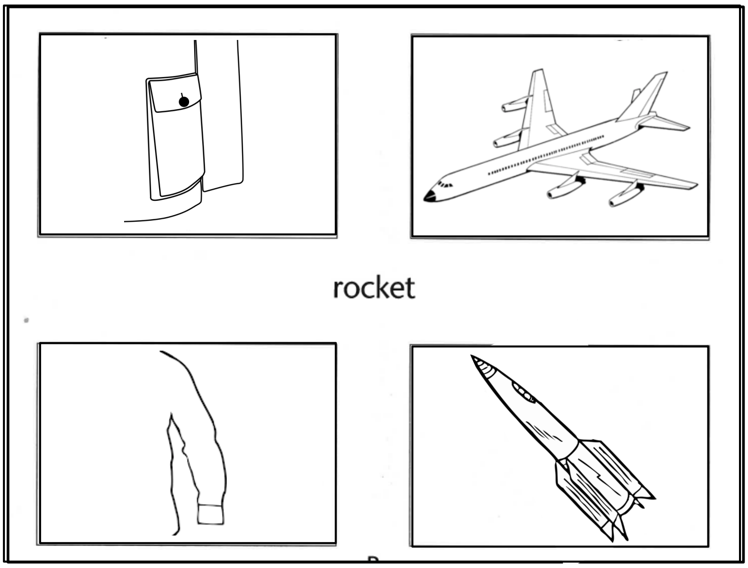


Figure 31. Example picture stimuli similar to the CAT written single word comprehension assessment (Swinburn et al., 2004)

7.7.3 Non-linguistic executive function subtests from the Cognitive Linguistic Quick Test

Following a review of cognitive therapy assessments a standardised assessment developed and tested with people with severe aphasia was selected to assess whether non-linguistic cognitive skills could be used to predict therapy gains for participants.

7.7.3.1 Rationale

Findings from a pilot trial of GeST (Marshall et al., 2013) indicated that the severity of expressive aphasia alone was not a predictor of the size of any therapeutic gain. Moreover, participants demonstrated varying levels of success in relation to their mastery of the system. For example, although all

9 participants demonstrated an ability to navigate forwards within a level (during observation, after three weeks use of GeST), 4 participants were less able to effectively navigate between levels using the menu screen (only 5 out of 9 participants demonstrated this after three weeks use of GeST).
Executive functions outside of language are commonly credited as being responsible for our ability to switch between tasks or to navigate through a series of steps or sequence (Purdy, 2002). These tasks however, are commonly tested using activities that draw upon some degree of linguistic competence in order to expose such skills. Common measures of executive function include the Wisconsin Card Sorting Test (Grant & Berg, 1948), the Trail Making Test (U.S. Army, 1944), the Towers of Hanoi (Welsh, Pennington, Ozonoff, Rouse, & McCabe, 1990), Ravens Progressive Matrices (Raven, 1977) and the Stroop Test (Stroop, 1935). Each of these assessments makes use of spoken or written instructions - with some additionally including the use of written stimuli and/or spoken output within test materials. This dependence upon a baseline level of linguistic competence means that those tests most commonly employed to measure skills of executive function would not be appropriate measures of such capacities in those people with severe expressive and often severe receptive aphasia as they depend upon input or output modes not accessible to this population.
During an initial review, several standard assessments of cognitive skills (e.g. Cognitive and Semantic subtests within the Comprehensive Aphasia Test, Swinburn et al, 2004; Wisconsin Card Sorting Task, Heaton et al, 1993) were considered. However, the CAT subtests, were rejected in favour of a more detailed assessment of executive function and the Wisconsin Card Sorting Task on the basis of its language complexity. The Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975) was similarly presumed not to be accessible for this level of aphasia severity.

The assessment chosen for this project was the Cognitive Linguistic Quick Test (CLQT), developed by Helm-Estabrooks et al (2001). The full version of this assessment examines a range of linguistic and cognitive features. However, those sections of the test pertaining to language skills were omitted for this project as for this project, as a more detailed assessment of language was carried out elsewhere. Further justification of this decision comes from Nicholas et al (2011) who employed only those 'non-linguistic' measures of the CLQT to successfully predict outcomes of use for C-Speak, a computer based system augmentative and alternative communication device in users with severe expressive aphasia (described in section 2.8.2)

7.7.4 Symbol Deletion

7.7.4.1 Administration

This task examines participants' "visual attention, scanning, discrimination, inhibition, and response shifting" (Helm-Estabrooks, 2001).

Participants were shown a target symbol (e.g ) in isolation. They were then shown a page with several examples of the target symbol amongst numerous other shapes with greater or lesser degrees of visual similarity (Figure 32). Following modelling by the examiner, participants were asked to identify examples of the target symbol and to use a pen to mark through them with a cross. Participants were given two minutes to complete the test.

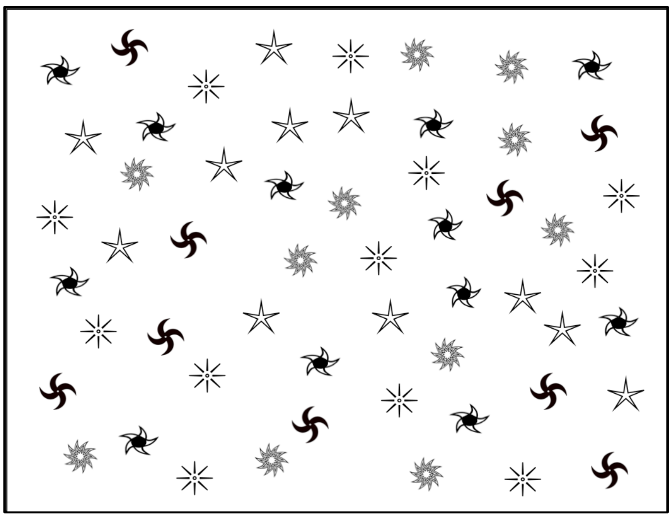


Figure 32. Example symbol page similar to the CLQT Symbol Deletion subtest (Helm-Estabrooks, 2001)

7.7.4.2 Scoring

Upon completion, crossed through target symbols from each quadrant were totalled. Non-target symbols were also totalled and the figure was subtracted from the total. The maximum total score was 12. The scoring of the results by quadrant provided an additional record of the participants' attention to each of the four quadrants of their visual field when attempting the task.

7.7.5 Symbol Trails

7.7.5.1 Administration

This task examines participants' "working memory, […] planning and mental flexibility" (Helm-Estabrooks, 2001). Participants were required to draw lines connecting a sequence of shapes in ascending size order (Figure 33). Critically, they had to alternate between shape types, e.g. by moving from a circle to a triangle.

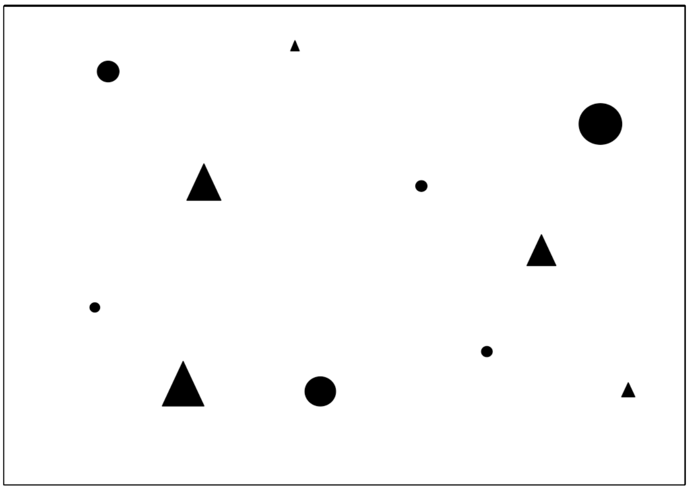


Figure 33. Example trails page similar to the CLQT Symbol Trails subtest (Helm-Estabrooks, 2001)

The task was made more demanding by increasing the number of shape types and exemplars.

7.7.5.2 Scoring

Upon completion, lines correctly linking any of the shapes were totalled and participants were awarded a score out of 10.

7.7.6 Design Memory

7.7.6.1 Administration

This task aims to examine participants' "immediate/working visual memory" (Helm-Estabrooks, 2001). Participants were asked to look at and remember a pair of shapes on a page (Figure 34). After 20 seconds, the first page was obscured and a new page was presented showing both the original shapes plus four visually related distractors. Participants were asked to point to the two shapes from this list that matched the originals.

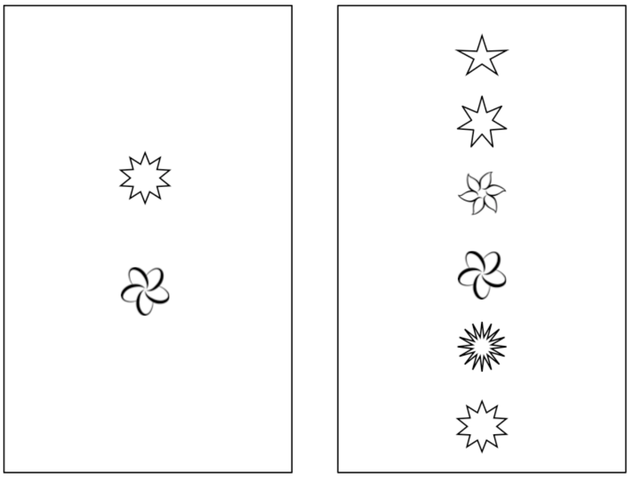


Figure 34. Example 'Target' and 'Target plus distractor' stimuli similar to the CLQT Design Memory subtest (Helm-Estabrooks, 2001)

The task was made more demanding by increasing the visual complexity and similarity of the designs over three trials.

7.7.6.2 Scoring

Participants were awarded one point for each correctly identified design and no points for incorrectly identified distractors - up to the maximum total score of 6.

7.7.7 Mazes

7.7.7.1 Administration

This task aims to examine participants' "attention and visuospatial skills" and also the ability to "plan a course of action, reject/inhibit incorrect choices, and correct any mistakes he or she makes" (Helm-Estabrooks, 2001). Participants were asked to complete two mazes by drawing a line from the start point through to the pile of money. They were instructed to start at the arrow and stay within the walls of the maze. A 60 second time limit was allotted for the first maze and a 2-minute time limit allotted for the second, more complex maze.

7.7.7.2 Scoring

Participants were awarded 4 points for correctly completing each maze with points deducted for errors where the examinee's line travelled up an incorrect path but was then self-corrected.

7.7.8 Design Generation

7.7.8.1 Administration

This task aims to examine participants' "productivity and creativity, the ability to vary responses rapidly, to self-monitor, to remember and follow rules, and to develop and use effective strategies." (Helm-Estabrooks, 2001). Participants were shown a square comprising four dots (one in each corner). Following modelling by the examiner, participants were asked to join up the four dots using four lines to create a design. They were asked to do this for a further series of duplicate blank squares, creating a new design each time (Figure 35). Participants were given three minutes to complete the task.

7.7.8.1 Scoring

Participants were awarded one point for each novel design produced. Repeated designs, incomplete designs or designs using more than four lines to join the dots were not awarded any points. A maximum score of 13 was available for this task.

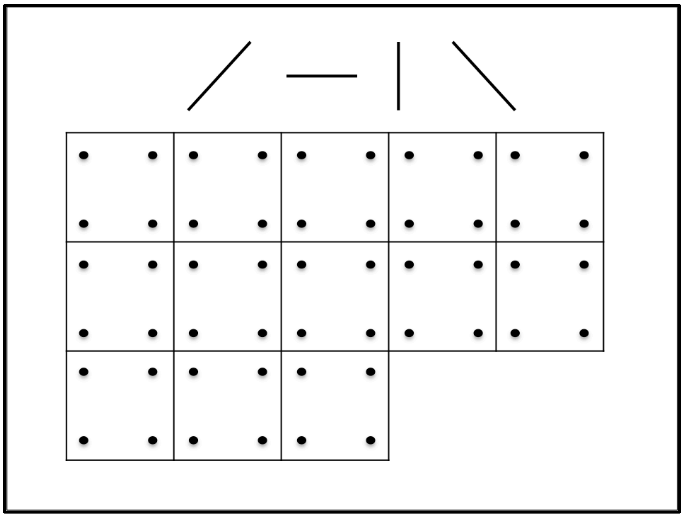


Figure 35. Example Design Generation test page similar to the CLQT Design Generation subtest (Helm-Estbrooks, 2001)

7.7.8.2 CLQT Scoring

Scores from the above five subtests were totalled to give a composite 'visuospatial skill domain' score for each participant.

7.7.9 Birmingham University Praxis Scale (BUPS)

7.7.9.1 Rationale

As for assessments of cognitive function, tests of praxis commonly depend upon an ability to accurately follow spoken instructions of varying complexity - for example, the Action Research Arm Test (Lyle, 1981) and the Test for Motor Apraxia (Poek, 1986). Many such standardised assessments of praxis however, rely upon complex verbal instruction as a means of assessing gesture competence as an independent variable. The high co-morbidity of severe language *comprehension* deficits alongside severe *expressive* deficits in aphasia poses a conflict when implementing such measures to assess praxis in those with severe levels of aphasia. For example, the Apraxia Battery for Adults (ABA, Dabul, 2000) provides a praxis measure widely used to assess such skills in people with aphasia. This subtest however is limited both by its reliance, once again, on verbal comprehension and also its limited scope of gesture competency explored – being dependent on gesture production following a short phrase. A more diverse assessment with additional language support is desirable to investigate the capacity more thoroughly amongst the target population being investigated here – namely, those with severe aphasia. Such breadth of assessment should allow for more finely grained examination of the relationship between initial gestural capacity and subsequent performance on a gesture training intervention.

The Birmingham University Praxis Scale (BUPS, Samson - as cited in Bickerton et al, 2012) covers a broader range of praxis assessments than the ABA and includes a "tool use" subtest (putting together and using a torch), which may enable comparisons between successful tool use for a torch versus successful tool use of a computer to receive therapy.

The BUPS comprises four subtests (versus the ABA's one), each supplying written prompts alongside spoken instruction and thus reducing linguistic and memory demands for participants. The abilities targeted by each subtest are described below alongside an outline of its administration.

As a whole, this assessment examined participants' fine and gross motor skills and core gesture abilities. It is included in the study design to examine whether successful response to gesture training using GeST/PowerGeST was linked to any aspect of limb praxis or the ability to produce, understand or copy gestures.

The BUPS comprises four subtests. These are outlined below.

7.7.10 Multi-step Object Use

7.7.10.1 Administration

This task aims to assess participants' ability to carry out a series of goal directed actions requiring fine and gross motor coordination. Participants were presented with a series of objects and instructed to "Make the torch work". The objects were arranged as presented below in Figure 36.

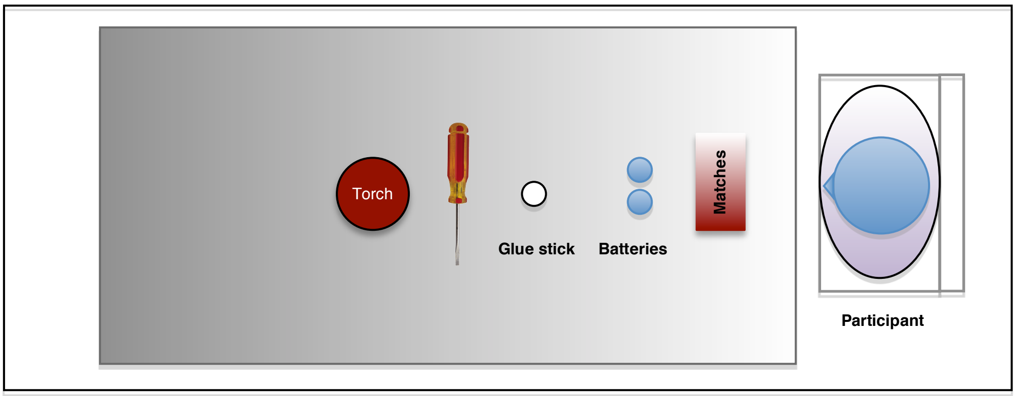


Figure 36. Order of objects presented in front of participant. L-R: Torch, screwdriver, glue stick, batteries and matches

To operate the torch, the participant was required to insert the two batteries. All other objects presented were distractors and served no purpose in making the torch operate.

7.7.10.2 Scoring

Participants were awarded a point for each correct action carried out towards completing the task with additional points for not using any irrelevant objects and not perseverating any action. A total of 12 points were available for this task.

7.7.11 Gesture Production

7.7.11.1 Administration

This task aims to assess participants' ability to produce recognisable transitive and pantomime gestures from a set of 6 single word or phrase prompts presented in both written and spoken form. For intransitive gestures, following a demonstration item produced by the examiner, participants were instructed "show me the gesture for X" where X is an iconic gesture such as *Military Salute*. For pantomime gestures, following a demonstration item produced by the examiner, participants were asked: "how would you use X?" where X is an imagined object such as *Hammer*.

7.7.11.2 Scoring

Participants were awarded two points for a fast and accurate gesture production and 1 point for "recognisable but inaccurate" productions. A total of 12 points was available for this task.

7.7.12 Gesture Recognition

7.7.12.1 Administration

This task aims to assess participants' ability to identify a target intransitive or pantomime gesture from a set of 5 single word or phrase prompts presented in both written and spoken form. For intransitive gestures, participants were shown a common gesture (e.g. *Goodbye*) and asked: "If I show you this gesture, what does it mean?" They were then presented with a set of five written prompts that were read aloud by the examiner and asked to point to the word which matched the gesture. For pantomime gestures, participants were shown the pantomime use of an imagined object (e.g. *Key*) and asked: "If I show you this gesture, which object do I pretend to use?" They were then presented with a set of five written prompts, which were read aloud by the examiner and asked to point to the word which matched the gesture.

7.7.12.2 Scoring

Participants were awarded one point for each gesture identified correctly. A total of 6 points was available for this task.

7.7.13 Meaningless Gesture Imitation

7.7.13.1 Administration

This task aims to assess participants' ability to accurately imitate novel gestures and hand shapes immediately after their presentation. For the first two items and following demonstration by the examiner, the participant was asked to copy a sequence of two actions (see Bickerton et al, 2012 for an example).

For the second two items, following demonstration by the examiner, the participant was asked to copy: “how I position my fingers” (see Bickerton et al, 2012 for an example).

7.7.13.2 Scoring

Participants were awarded three points for a fast and accurate gesture production, two for correct and accurate production after a second demonstration, and one point for a broadly correct gesture containing only one error after a second demonstration. A total of 12 points was available for this task.

7.7.13.3 BUPS Scoring

Scores from each of the above four subtests were totalled to give a composite gesturing score for each participant. This score was used to examine the relationship between gesturing ability at the outset of treatment with gains made over the course of treatment.

7.8 Intervention

Participants in both groups undertook a five-week therapy intervention period (Table 12) and a five-week non-therapy period between time points 1, 2 and 3. Those in the immediate therapy group completed the intervention between time points 1 and 2 and those in the delayed therapy group did so between time points 2 and 3. Details of the intervention undertaken within the therapy period follow.

7.8.2 Intervention Overview

Table 12. Figure showing week-by-week administration of therapy intervention

Week 1	Week 2	Week 3	Week 4	Week 5
Day 1 10-minute familiarisation task for **five** gestures. Researcher Demonstration and participant rehearsal of **GeST** use for above gestures **Days 2-7** Autonomous use of GeST.	**Day 1** 10-minute familiarisation task for **five** new gestures. Participant practice use of **GeST** for above, new gestures **Days 2-7** Autonomous use of GeST.	**Day 1** 10-minute familiarisation task for **five** new gestures. Participant practice use of **GeST** for above, new gestures **Days 2-7** Autonomous use of GeST.	**Day 1** 10-minute familiarisation task for **five** new gestures. Participant practice use of **GeST** for above, new gestures **Days 2-7** Autonomous use of GeST.	**Day 1** 20-minute re-familiarisation task for **twenty** previously trained gestures. Demonstration and rehearsal of **PowerGeST** use for above gestures. **Days 2-7** Autonomous use of PowerGeST.

Participants were allocated to practice blocks of five gestures in one of two different orders across weeks 1-4 (see Appendix G for allocation). This aimed to counter-balance the risk of practice order effects across participants.

Procedure for administration of the above intervention is described below:

7.8.3 Familiarisation Exercise – for use prior to working with GeST

This exercise introduced five gestures with associated pictures to a participant.

The exercise comprised four stages:

1. **Identifying a gesture from a pair of pictures**
   Researcher demonstrated a gesture. Participant was asked to identify the related object from a choice of two pictures. If an incorrect response was selected, the researcher repeated the gesture and pointed to the correct target. This was repeated for five gestures and their associated set of picture pairs.
2. **Immediate repetition of a gesture**
   Researcher demonstrated a gesture. Participant was asked to immediately repeat the gesture. If an incorrect/crude gesture was produced, the researcher modelled the gesture again until the participant produced it correctly – moulding the participant's production if necessary. The level of researcher support required was noted on the exercise record form. This was repeated for all five gestures and their associated individual picture.
3. **Production of a gesture from a picture**
   Participant was shown the above set of five individual pictures in turn and asked to produce the gesture they had previously been shown in association with that image. If an incorrect/crude gesture was produced, the researcher modelled the gesture again until the participant produced it correctly - moulding participant's production if necessary. The level of researcher support required was again, noted on the record form. This was repeated for all five gestures and their associated individual picture.
4. **Barrier task**
   Participant turned a picture over from a pile of five individual pictures. They were asked to produce the associated gesture for the

researcher. The researcher was unable to see the chosen picture due to a cardboard barrier placed between them and the participant (Figure 37). The researcher attempted to identify the relevant target from the participant's gesture production. If the participant was unable to produce the gesture, they were asked if they would like to come back to it or for researcher to show them the target. If requested, the researcher modelled the target gesture until the participant was able to produce it correctly – moulding the participant's production if necessary. This was repeated for all five pictures. The researcher recorded correct/incorrect identifications on the record form.

The above task was undertaken for ten minutes. If ten minutes finished before the end of the exercise then the most recently started stage (i.e. 1, 2, 3 or 4) was completed before stopping the task. If all four stages were completed within ten minutes, the task was repeated as necessary until ten minutes was complete.

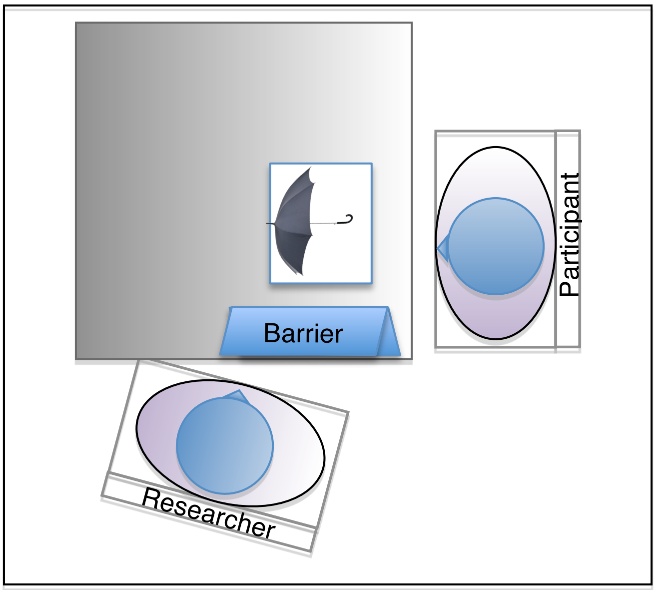


Figure 37. Diagram from above showing the seating and barrier configuration for stage 4 of the familiarisation task

7.8.4 Re-familiarisation Exercise - for use prior to working with PowerGeST

This exercise re-introduced all twenty previously practised gestures with associated pictures to a participant.

As above, the exercise comprised four stages:

1. **Identifying a gesture from a set of four pictures**
   Researcher demonstrated a gesture. Participant was asked to identify the related object from a choice of four pictures. If an incorrect response was selected, the researcher repeated the gesture and pointed to the correct target. This was repeated for all 20 gestures and an associated set of four pictures.
2. **Immediate repetition of a gesture**
   Researcher demonstrated a gesture. Participant was asked to immediately repeat the gesture. If an incorrect/crude gesture was produced, the researcher modelled the gesture again until the participant produced it correctly – moulding the participant's production if necessary. The level of researcher support required was noted on the exercise record form. This was repeated for all twenty gestures and their associated individual picture.
3. **Production of a gesture from a picture**
   Participant was shown the above set of five individual pictures in turn and asked to produce the gesture they had previously been shown in association with that image. If an incorrect/crude gesture was produced, the researcher modelled the gesture again until the participant produced it correctly - moulding participant's production if necessary. The level of researcher support required was again, noted on the record form. This was repeated for all twenty gestures and their associated individual picture.
4. **Barrier Task**
   Participant turned a picture over from a pile of five individual pictures. They were asked to produce the associated gesture for the

researcher. The researcher was unable to see the chosen picture due to a cardboard barrier placed between them and the participant (Figure 37). The researcher attempted to identify the relevant target from the participant's gesture production. If the participant was unable to produce the gesture, they were asked if they would like to come back to it or for researcher to show them the target. If requested, the researcher modelled the target gesture until the participant was able to produce it correctly – moulding the participant's production if necessary. This was repeated for all twenty pictures. The researcher recorded correct/incorrect identifications on the record form.

The above task was undertaken for twenty minutes. If twenty minutes finished before the end of the exercise then the most recently started stage was completed before stopping the task. If all four stages were completed within twenty minutes, the task was repeated as necessary until the twenty minutes was completed.

7.8.5 Demonstration and Rehearsal of GeST

As summarised in Table 12, weeks 1-4 of therapy not only included familiarisation but also demonstration and practice of a new block of five gestures within GeST each week. This was facilitated by the researcher as detailed below.

Week 1:

The aim of this session was to provide an initial introduction to use of GeST. Following the above familiarisation exercise (7.8.3), the researcher provided a full demonstration of GeST use (assuming the role of participant for modelling purposes). During this demonstration, use of the keyboard to navigate the software was increasingly deferred away from the researcher to the real participant. Researcher demonstration was immediately followed by one full use by the participant with prompts as required. The participant was advised to use GeST to practice for 1 hour per day – to be carried out across one or more sessions as desired. The participant was advised that computer would log their use of GeST and that the speech and language therapist would review this log with them the following week. They were additionally advised that they would not break the system by using it, and that the researcher would fix any problems that occurred. Participants were encouraged to use the first week to "really get to know" the tool.

Weeks 2-4:

The researcher asked the participant to demonstrate use of GeST. She addressed any challenges with navigation/use during the review, or noted them for discussion at subsequent use of the next story. A note was taken of any buttons pressed by the participant and any responses exhibited by the computer system during this period. Next, that participant's use of GeST was reviewed through consultation with GeST log. The researcher reported to the participant the number of day's-use recorded. She then set up laptop with new 5-gesture therapy block. This was followed by the familiarisation exercise above. Without providing a demonstration, the speech and language therapist then asked the participant to use GeST to practise new gestures. Any issues with navigation/use were addressed and practice was repeated until successful independent use was observed.

7.8.6 Demonstration and Rehearsal of PowerGeST

Week 5:

To continue established pattern from above and preceding familiarisation – the researcher asked the participant to demonstrate their use of GeST. As for weeks 2-4, independent use of GeST was additionally reviewed through consultation with GeST log. The number of day's-use recorded was reported to the participant.

The researcher next set up the laptop with PowerGeST – removing both the camera and the yellow gloves. The re-familiarisation exercise was then completed (7.8.4). Without providing a demonstration, the researcher then asked the participant to use PowerGeST to practise all twenty gestures. Any issues with navigation/use were addressed and practice was repeated until successful use was observed. NB PowerGeST operates differently to GeST in that it requires only flat navigation and includes no system gesture recognition component. The researcher filmed the participant's initial use of PowerGeST using a portable video camera. She additionally gave the participant a sheet, marked with days, on which to record sessions (with a tick). Participants were advised to keep a physical record of each session they complete by marking the sheet with a tick.

End of week 5:
The researcher asked the participant to demonstrate their use of PowerGeST. This was filmed. The use of PowerGeST was reviewed through consultation with participant's paper log. If the log had not been used or participant indicated that it was incomplete, the researcher read aloud the sheet day-by-day, asking the participant to recall whether they practised on this occasion and if so then how many times did they go through the sequence from start to finish. The paper log was updated to reflect any additional sessions reported. This completed the therapy block.

7.9 Outcomes

In addition to the candidacy measures previously introduced (7.7), four assessments were utilised as outcome measures. These were each administered at time points 1, 2, 3 and 4 (

Figure 26). The primary outcome measure was a gesture production assessment. The three secondary outcome measures were a spoken naming assessment, an interactive gesture assessment and a technology assessment. Each assessment is described, in turn, below. Descriptions will include instructions for administration and scoring. Administration instructions include any adjustments made to accommodate additional challenges to completion either due to the participant's aphasia or additional difficulties. A further measure of computer usage was available through a data logging facility within the GeST software. This was collected across weeks 1 – 4 of the 5-week therapy period. This will be analysed to examine the extent to which GeST was used to practice during the intervention period as well as the ways in which it was used. Data recorded will be described as will interim analysis methods used to summarise the data. Finally, video recordings of PowerGeST use for each participant will be examined to assess the manner and success of use for this additional therapy component.

7.10 Primary Outcome Measure

7.10.1 Gesture assessment

A 40-item gesture assessment was chosen as the primary outcome measure for this study. The reason for this selection was that the intervention being examined has been developed primarily as a tool for the training of single communicative gestures. It is hence in the assessment of individual gestures that we might most anticipate change as a result of intervention. Remaining outcome measures aim to capture any change in associated skills that might arise as secondary results of the computer-delivered gesture intervention.

7.10.1.1 Administration

The gesture assessment (primary outcome measure) and the naming assessment (secondary outcome measure) each had two alternate versions: A and B. Both A and B contained the same set of test items, however the order of presentation was different for each – having been separately randomly assigned. Participants undertook alternating test versions across the four testing time points: i.e. A B A B or B A B A. This alternating presentation aimed to counteract any order presentation effects within participants and across groups.

Derived from a measure employed by Marshall et al. (2013) to pilot test GeST, this assessment comprised 40 photograph images of objects (Figure 38). Twenty of the objects presented were items trained within the intervention protocol being tested. A further 20 were items matched for lexical frequency. This structure allowed assessment of change in items that have been exposed to treatment in comparison to assessment of change in items of equal lexical frequency but which have not been trained. The presence of items from both trained and untrained categories allows for the examination of generalisation effects from treated to untreated stimuli.

Within the gesture assessment, participants were presented with a photograph of an object and instructed: “Show me a gesture for this. Use your hands and your face”. The assessment was video recorded and not scored by the examiner. Using a method described in Marshall et al (2012 and 2013), videos were later edited into a series of four new videos, each comprising the forty gestures elicited but distributed across a range of time points. Independent video scorers were then asked to identify and write down the target being gestured. These transcripts were then compared to the target item and scored for accuracy - with one point being awarded for each accurate identification or acceptable synonym (appendix H). A maximum score of 40 was available for this assessment.

Figure 38. Example '*telephone*' and '*glasses*' picture stimuli from the gesture production assessment

7.10.1.2 Scoring

7.10.1.2.1 Procedure for Scoring Gesture Assessments

Gesture production assessments were recorded on video. Videos were filmed in the participant's home using a Toshiba 'Camileo' x400 camcorder on a Joby 'Gorillapod' flexible camera tripod. Recordings were captured at a resolution of 1080 X 30 in full HD and audio was captured using the device's internal microphone.

For blinded scoring purposes, videos of each gesture assessment were edited together to form new composite videos. Editing was carried out upon a 2008 iMac computer running OS X Yosemite version 10.10.2 and using iMovie software version 10.0.7. Videos were exported at resolution of 1280 X 720 pixels at a 'medium' quality setting and stored in .mp4 format.

7.10.1.2.2 Structure of Composite Videos shown to blind scorers

40 test items from each of the 4 time points were shuffled across four new videos per participant. Each video comprised ¼ of the items from time points 1, 2, 3 and 4. This means that one ¼ of the data from each time point was presented in any given video. (In cases where data was collected at only 2 or 3 time points (e.g. participants F and N respectively) items were distributed evenly across either 2 new videos comprising ½ of the items from each time point or 3 new videos comprising 1/3 of the items from each time point). An independent scorer, blinded to the time of assessment for each item, viewed each video.

The aim of shuffling videos across time points was two-fold:

1. To ensure scorers were blinded to the time of assessment for each item.
2. To account for variation in scorer ability e.g. contributions from a blind scorer who was less able to identify any gestures account for ¼ of the score awarded at each given time point. Equally, contributions from a blind scorer who is more capable of identifying gestures account for ¼ of the score awarded at each time point. By dividing the videos in

equal proportions across four scorers, the negative or positive impact of a less or more abled interpreter of gesture were weighted equally across all four assessment time points.

Within each scoring video, items were presented in sequence as follows:

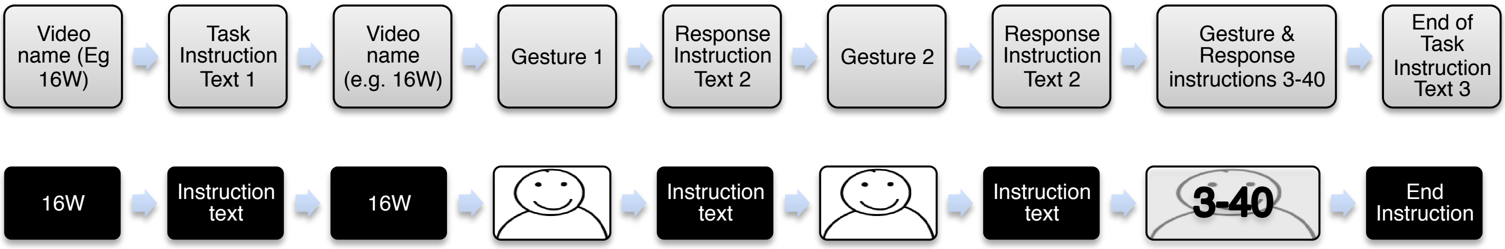


Figure 39. Example format of videos shown to blind markers

7.10.1.2.3 Editing of videos

The video clip of each gesture was segmented as follows:

- Clip begins at the moment a given item picture is presented to the participant. The researcher is not visible in the clip.
- Clip ends when participant indicates that they have completed their attempt for that gesture or as the researcher introduces the next picture.

On occasions where participants enlisted props to support their gesture, these sections of the video were cut out.

7.10.1.2.4 Audio

Audio recording from the video was included in the clip presentation (to capture vocal gestures such as the revving of an engine etc.). However, video sound was muted where participants produced real spoken words during assessment (whether on target or off target). The aim of excluding spoken words was to allow an interpretation of the gesture alone and hence to ensure scoring of this assessment reflected changes in gesture production in isolation from changes in spoken word production. (This contrasts to the interactive gesture assessment where scorers were the participant's real-life communication partners – see 7.11.2). Audio of the video clips was also muted where the researcher is heard to use the participant's name – to enable preservation of anonymity of the participant's name.

7.10.1.2.5 Video Scoring Sessions

Scoring of video gesture data was carried in a university computer laboratory.

Scorers attended a 90-minute session within which they were presented with two to four edited videos via PC media player and asked to complete a web form to state what they believed the gestures produced to show. The web form as generated using Google Docs for viewing within a web browser. The use of an online form allowed multiple scorers to enter responses simultaneously. These responses were collated automatically into a spreadsheet, which was later downloaded and analysed separately. An abbreviated URL address (www.tinyurl.com/GeST15) was generated to make the web form easy to find. All scorers were student speech and language therapists. At the beginning of a scoring session each student was seated at a computer displaying a web form and a media player. Headphones were provided and scorers were asked to wear these whilst completing the scoring exercise. Instructions regarding how to complete the activity were provided via the web form.

Scorers were required to reply a response for each clip. Having completed answers for the 40 items shown within the first video and pressing the submit button, the following message was presented:

> **GeST Gesture Scoring Form**
> Thank you for completing part one of this exercise.
> Your response has been recorded.
> Please close the video that you have completed. If you have been asked to score another video, please click "Submit another response" below and open the next video file.
>
> Submit another response

Scorers were asked to watch and score up to four videos of different participants within the 90-minute scoring session. No scorer saw more than one video from the same participant. This ensured that up to four scorers equally contributed responses for any one assessment. [Each participant's data was shared amongst 4 videos (with the exception of two participants who completed fewer than four assessments in total. Data for these participants was split equally across two and three videos to reflect the number of assessments they had completed in total)]

Once collected, scorer responses were re-assigned to their original assessment time point and awarded points according to the below scoring criteria.

7.10.1.2.6 Scoring Criteria

Those responses that matched the target item or a synonym of the target item exactly were awarded 2 points. [Synonym items were generated using WorldNet 3.1 (Princeton University, 2010). A full list of synonyms is provided in appendix H.]

Those responses where a scorer or participant had supplied no response or indicated that they did not know the answer, were awarded 0 points. All other responses were deemed ambiguous and were subject to a further scoring judgement by a second scorer to examine their semantic acceptability. This process was conducted as follows for each target word – ambiguous-response pair:

1. A list of the target words was created (list 1). This list was duplicated to give a list double in size and containing each of the target words twice.
2. A list of the ambiguous responses to those targets was created (list 2). This list was duplicated to give a list double in size and containing each of the ambiguous response words twice.

Responses from list 2 were linked once to their intended target and once to a randomly selected word from list 1 (i.e. a foil). (The second scorer was unaware of this allocation or this design.) The second scorer was asked to identify whether the given response was an acceptable response for the target. For a response to score a point, it must have been deemed acceptable in the target condition and unacceptable in the foil condition.[2] An example of the score award system is provided in Table 13.

Table 13. Example of ambiguous responses judged by a second scorer and the corresponding system for awarding points

ITEM NUMBER	EXAMPLE RESPONSE	TARGET (List 1)	Judged acceptable	FOIL (List 2)	Judged acceptable?	Score Awarded
1	Sleeping	Bed	YES	Spider	NO	1
2	Diving	Swimming	YES	Football	YES	0
3	Newspaper	Piano	NO	Book	YES	0

7.10.1.3 Rationale for scoring process

The decision to extend scoring from a 0 or 1 judgement of incorrect/correct to a 0, 1, 2 measure for degree of accuracy was taken to account for the clinical observation that although participants may not always be able to accurately convey the precise target when gesturing, their actions often convey related semantic meaning to the communication partner. This additional information may give rise to more accurate funnelling of information to support both parties in moving towards the intended shared understanding. Progress towards this improved mutual understanding achieved within this fashion is not captured within a scoring system that accounts for target or synonymous answers alone (i.e. a 0 or 1 bimodal response). For this reason, the decision to include a further level of scoring as a measure of communicative success was taken. Items judged acceptable within this framework were awarded one point and items which matched precisely with the target or a synonym,

[2] A response was deemed acceptable where the communication partner had gathered some relevant, useful information from the message, that was close to the target and which would be useful as a means of conveying the message. For example:
Target: Chair. **Possible Acceptable Responses**: Sitting; sofa
Acceptable because sitting and sofa relate to 'chair' and convey useful information related to the item chair.
Target: Table. **Possible Unacceptable Responses**: Wall; toothbrush
Not acceptable because wall and toothbrush do not relate closely to 'table' and do not convey information closely related to the item table.

were awarded two points. This difference in weighting is intended to reflect the level of communicative benefit gained by reaching this response – with an exact match being more beneficial than an acceptable (but less precise) alternative.

7.10.1.4 Inter-rater reliability of scored gesture data

To evaluate the reliability of the scores awarded for the gesture, videos for 22 of 77 gesture assessments (29%) were viewed and scored by second scorer. Selection of these videos was distributed evenly across the participant data. As for the scoring process above, the second scorer was blinded to the design of the project and the time point at which the assessment had been conducted.

A two-way, mixed method intraclass correlation (ICC) was conducted to compare outcomes from the second scorer to those reported by the primary researcher. An adequate degree of reliability was found between the two score sets (Portney & Watkins, 2000). The average measure ICC was .681 with a 95% confidence interval from .636 to .721 ($F(879,879) = 3.14$, $p<.001$).

7.11 Secondary Outcome Measures

7.11.1 Naming Assessment

Again derived from a measure employed by Marshall et al. (2013) to pilot test GeST, this assessment employed the same 40 photograph images of objects as the gesture assessment described above (figure 38).

7.11.1.1 Administration

Within the naming assessment, participants were presented with a photograph of an object and asked to state the name of that object. Responses were transcribed/recorded by the researcher and scored for accuracy.

A maximum score of 40 was available for this assessment.

Assessment was video recorded for subsequent inter-rater reliability measures.

7.11.1.2 Scoring

The primary researcher carried out scoring for this assessment. Those responses that matched the target item exactly or a synonym of the target item were awarded 1 point. Those responses where a scorer or participant had supplied no response or indicated that they did not know the answer, were awarded 0 points.

7.11.1.3 Inter-rater reliability of scored naming data

To evaluate the reliability of the scores awarded for the naming data, videos for 14 of 77 naming assessments (18%) were viewed and scored by a researcher external to the project. These videos had been randomly selected using a computer-based randomisation process. The second scorer was blinded to the design of the project and the time point at which the assessment had been conducted. A two-way, mixed method intraclass correlation (ICC) was conducted to compare outcomes from the second scorer to those reported by the primary researcher. A high degree of reliability was found between the two score sets. The average measure ICC was .907 with a 95% confidence interval from .657 to .972 ($F(13,13) = 27.81$, $p<.001$).

7.11.2 Interactive Gesture Assessment

7.11.2.1 Rationale

An interactive charades assessment was developed for the purposes of this study. It was designed to establish whether gestural communication within the participant's everyday communication environment changes as a result of exposure to GeST+PowerGeST. It aims to capture not only the clarity of participant gesture in isolation, but also the active communicative success of a participant's gestures in their everyday environment. Working on the premise that a family member, friend or carer in the participant's immediate environment is likely to be exposed to the participant's practice and use of gestures, this assessment aims to document changes in the effectiveness of everyday communication as a result of the GeST+PowerGeST intervention. Full details of the development of this assessment can be found in section 6.3.1.

7.11.2.2 Administration

Participants were seated opposite a family member, friend or carer who acted as the *gesture recipient*. In front of them and to their right, was a laptop used to present the assessment stimuli (Figure 40). As is the case for the spoken naming and gesture assessment, tested items in this measure comprise both treated and untreated targets. In order to mimic a more naturalistic stimulus for gesture production than a sole picture confrontation, stimuli consisted of a short video clip of an everyday situation (e.g. a person answering a telephone) – of the format reported in figures 12-13. Video clips were presented on a computer screen using Microsoft PowerPoint presentation software. Immediately following the clip, a still photograph of a relevant object appeared against a white background - of the type shown in Figure 38.

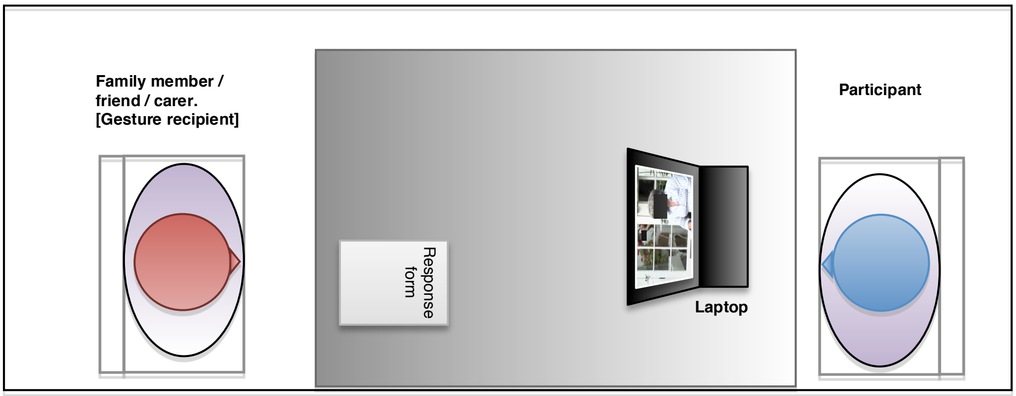


Figure 40. Delivery of IGA

The participant was instructed:

"I'm going to show you a short video. At the end of the video is a picture. Your job is to gesture that picture to X (like a game of charades). X will try to work out who or what it is and write it down."

Where X is the family member, carer or friend's name.

The gesture recipient was then instructed:
"Your job is to work out what the picture is. Write down your answer in the box on your sheet. You can speak to Y or use gesture to check your understanding. When you're ready for the next item, let me know and I'll press the button to move on."
Where Y is the participant's name.

Two practice items were shown with feedback from the assessor as required – prompting the participant to direct their gesture productions towards the gesture recipient and prompting the gesture recipient to check their understanding as necessary.

12 test items were then presented using the above format with no further feedback from the assessor. The computer recorded the length of time between video presentation and progression to the next item – giving a measure of the length of time spent on each item. A maximum score of 24 was available for this assessment.

The interactive charades assessment had four alternate versions: A, B, C and D. Each contained 6 treated and 6 untreated test items. The order of presentation was different for each version – having been separately randomly assigned. Participants undertook alternating test versions across the four time points: i.e. A B C D or B A D C. This alternating presentation aimed to counteract any order presentation or learning effects within participants and across groups.

Each treated item was tested at least once across the four time points. Four items: wife, wine, glasses and tea were tested twice.
Each untreated item was tested at least twice across the four time points. Four items: gloves, rainbow, tap and bed were tested three times.

The reason for this duplication was the availability of video stimuli. A total of 30 video clips were available to use as test stimuli – originating from the scenario vignette videos developed for level 3 of GeST. Of these, 20 were used as treatment items within the therapy interventions. The remaining 10 were not used in intervention and hence represented appropriate controls items.

Whilst an equal number of video stimuli for both treated and untreated conditions would have been desirable, the resources available delimited the choice of untreated items to just 10. To minimise the effects of these repetitions over time, items were distributed across assessments to ensure no one item was repeated within a version. Full item allocation by version is provided in Appendix B.

7.11.2.3 Scoring

In completing the interactive gesture assessment, each participant's communication partner generated a written script of numbered responses for two practice items and 12 test items.

Where multiple responses were written the first response only was scored.

Responses from test items were transcribed into typed text and scored for comparison to a set of synonym items generated using WorldNet 3.1 (Princeton University, 2010). A full list of synonyms provided in appendix H. Unresolved items (i.e. those not matched as correct) were then marked using the scoring criteria previously reported in section 7.10.1.2.6.

7.11.3 Technology Assessment

7.11.3.1 Rationale

The technology use and confidence measure was developed for the purpose of this study. It was designed to assess the impact of Gest + PowerGeST on participants' wider use of technology. Development of the technology use and confidence measure was motivated by reports from participant carers' in the GeST pilot study (Marshall, 2013) where users had demonstrated an increased interest in other, non-computer-based household technologies during their involvement in the intervention protocol. To examine this more methodically, this measure aims to capture the presence of any significant changes in technology use during involvement in the project and allows for an exploration of the effect of computer gesture intervention on broader technology use. In brief, the usage measure comprises a list of 17 items of everyday technology. Participants are asked to state whether they have used the stated item within the preceding month. A confidence scale is additionally provided alongside each examined technology item - allowing the exploration of the effect of computer gesture therapy intervention upon participants' confidence in relation to other items of technology. Full details of the piloting and development of this measure can be found in section 6.3.2.

7.11.3.2 Administration

The technology use and confidence measure comprises a series of 17 items of everyday technology. Participants were shown a picture of each technology and asked to report whether they had used it within the last month (Figure 41) by pointing to the yes or no icon printed beneath the picture.

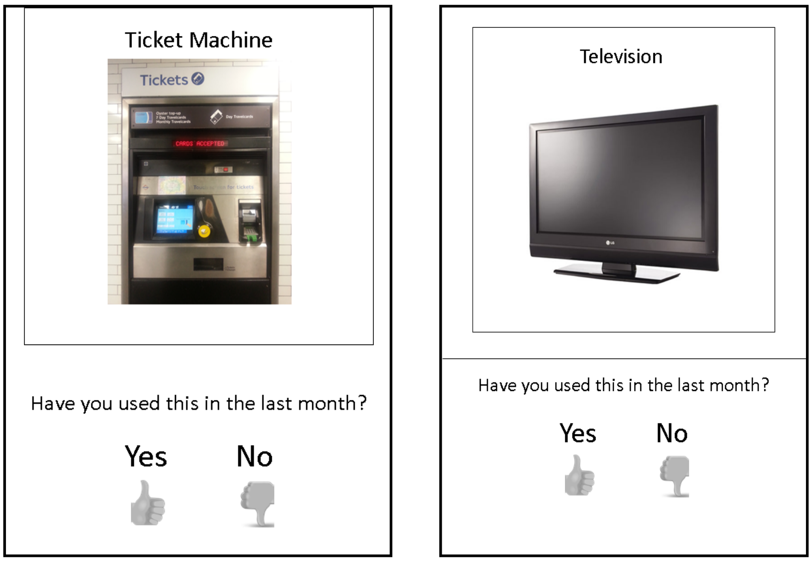


Figure 41. Example 'ticket machine' and 'television' picture stimuli from the technology use and confidence measure

Following each response, participants were then asked to rate how confident they felt in using that piece of technology by pointing to a mark on a visual rating scale adapted, with permission, from an item within the Visual Analogue Scale of Self-Esteem (VASES, Brumfitt and Sheeran, 1999. See Figure 20). Participants were asked to rate confidence regardless of whether they had answered yes or no to having used the item recently.

7.11.3.1 Scoring

At the end of the assessment, participants were awarded a technology use score of up to 17 points – with one point awarded for each item of technology they reported using within the last month.

Participants were additionally awarded a technology confidence score of up to 85 points - with up to five confidence points awarded per item. For each item, a minimum one point was awarded for a report of ++ on the “not confident” end of the scale and the maximum of five points was awarded for a report of ++ on the “confident” end of the scale.

7.12 Additional Outcomes

7.12.1 Data Logging and Usage Records

7.12.1.1 Rationale

Existing research into aphasia intervention (Bhogal, Teasell, & Speechley, 2003) reports evidence of a link between therapy dose and intensity and the scale of therapy gain. A key argument for the use of technology in therapy intervention is that autonomous self-delivered intervention allows individuals to benefit from an increased 'dose' and intensity than might otherwise be available to them through one-to-one therapist support alone.

Administration via computer offers an opportunity to log precise use of intervention even in the absence of a therapist/researcher. Furthermore, it does not ask additional demands of the participant and – if implemented successfully – ensures consistent reporting of use across all participants, avoiding inherent variability introduced through reliance on self-report usage figures. Furthermore, detailed logging can ensure precise capture of fine-grained information such as response times or in the case of GeST – time taken to accurately produce a correctly identified gesture.

7.12.1.2 Data Capture

Each time a participant switched on their computer to practice with GeST, information about a fixed set of criteria was logged by program software. Information was stored as a .csv (comma separated values) file on the host laptop and subsequently exported by the researcher for analysis.

Participants practised a different block of 5 gestures each week, and a separate log file was generated for each block of 5-gesture therapy (creating 4 files in total across the first 4 weeks of therapy with GeST).

Information captured within each log files comprised 7 columns of information:

Level used (e.g. level 1/2/3 or the action application start/quit)	Date	Time	Gesture status (i.e. recognised/ skipped)	Item name	Time taken to recognise Gesture (seconds)	Level 2 score

Data for PowerGeST administration was intended to be logged for analysis in a similar way. The delivery of PowerGeST via PowerPoint however meant that a separate method needed to be employed to capture this data from the PowerPoint program. A PowerPoint Add-in file was developed to capture the following information:

NEW SHOW	01/11/13		
SLIDE NUMBER	START TIME	END TIME	TIME ON SLIDE
1	11:35:01	11:35:04	0 min. 3 sec.
SHOW ENDED		TOTAL	0min. 3 sec.

Although this logging tool appeared effective during pre-intervention testing however, it often failed to capture data from sessions undertaken during the therapy period and therefore did not provide full data regarding the amount of time spent using PowerGeST within autonomous self-directed practice. As an alternative tool to capture levels of use, participants were filmed practising with PowerGeST for the first and last times during their final week of therapy intervention (i.e. the week when they were using PowerGeST). Time taken to complete a practice session from ‘Switch-on’ to application close was calculated by referring to the ‘start’ and ‘finish’ session videos. This will be reported alongside GeST usage figures within the Results section (8.3.2).

7.12.1.3 Analysis

Each participant generated a total of 4 separate GeST data logs – each log documenting the time spent practising and the actions carried out during practice for a different group of 5 of the 20 treated gesture items.

i.e. Log one documented all actions pertaining to Item set A, log two to item set B and so forth (as described in Table 14).

Table 14. Data reported within each GeST log file over four weeks of GeST intervention

Log 1 (week 1)	Log 2 (week 2)	Log 3 (week3)	Log 4 (week 4)
Set A Gestures	**Set B Gestures**	**Set C Gestures**	**Set D Gestures**
Glasses	Hat	Piano	Boy
Tea	Money	Food	Dentist
Telephone	Beer	Stamp	Football
Book	Car	Waiter	Spider
Wife	Dentures	Wine	Walking Stick

Each .csv log file was saved as a Microsoft Excel spreadsheet. Individual sessions within each log file were identified by establishing which rows reported “application start” and “application quit”, and considering all rows between these two indicators to comprise one session.

On occasions where participants did not fully shut down the computer at the end of a practice session, time gaps between actions (i.e. between rows), which were 20 minutes or longer were taken to indicate the end of one session and the beginning of another. To establish the length of a given session, the time stamp from the first action in that session was deducted from the time stamp of the last action of a session to give a session time in minutes and seconds. Information for all participant sessions across 4 time logs each was calculated in this way.

7.13 Changes to outcomes

No major changes to protocol were undertaken during the completion of this study.

7.14 Sample size

As described in 7.3.1, this design describes a feasibility study extending and developing previous pilot research into computer-delivered gesture therapy. It comprises parallel groups, with one group being a treatment group and the second a waitlist control group. The sample size taken was an availability sample generated based upon the time and resources available to the researcher. Intended sample size was initially set at 30 participants; however, due to participant availability, the final sample size achieved was 20 participants.

7.15 Interim analyses and stopping guidelines

No additional interim analyses were undertaken before the completion of the study.

7.16 Randomisation: sequence generation and type

Upon completing the screening process and entering the study, participants were allocated an anonymous project identification number. Participants were allocated to either the immediate or delayed therapy group in one of two ways. For the first group of participants to enter the study – recruited in the South East of England, allocation was carried out as each participant entered the study. A second block of data collection was carried out in the South West of England. For logistical reasons the first five participants recruited here were allocated as a group. In this case, allocation was carried out in blocks of 5 cases at a time. This enabled data collection and therapy delivery to be carried out contemporaneously for this group of participants.

7.17 Randomisation: allocation concealment mechanism

For participants recruited in the South East of England and allocated one at a time to therapy condition, a batch of 10 equally sized paper labels was put into a paper bag. Five labels showed the word "**immediate**" and five showed the word "**delayed**". Following a participant's entry to the study, one label was drawn from the bag and this label was used to allocate the participant to either the **immediate** or the **delayed** therapy group. For participants recruited in the South West of England, block allocation was achieved using the same randomised process with two equally sized labels showing either the word "**immediate**" or "**delayed**".

7.18 Randomisation: implementation

The principal researcher's secondary PhD Supervisor, who was blinded to the participant's identity and communication profile, carried out randomised allocation.

7.19 Blinding

Following randomised allocation to group, both the participant and the principal researcher were aware of the intervention assignment. Scoring for the primary outcome measure (gesture assessment) was completed by judges blinded to the study design, group allocation and time of assessment.

7.20 Similarity of interventions

The delayed therapy group received no intervention between time points 1 and 2. For this reason there is judged to be no similarity between the input received by immediate therapy group and the absence of therapy received by the delayed therapy group between time points one and two.

7.21 Statistical methods

Each of the four main outcome measures was analysed using a similar, two-step process. These are outlined below, beginning with the primary gesture outcome measure and the secondary naming and interactive gesture measures – each was analysed in the same way.

7.22 Primary Outcome Gesture Measure and Secondary Outcome Naming and Interactive Gesture Measures

7.22.1 Analysis 1 – between group analysis

The first analysis compared the results of the immediate and delayed therapy groups between time point 1 and 2. Data were analysed using a mixed model three factor ANOVA with two within-participant factors of time (time point one versus time point two) and item (treated versus untreated items) and a between-participant factor of group (delayed versus immediate).

7.22.2 Analysis 2 – within group analysis

The second analysis examined differences over time for all participants, on treated and untreated items. Data were analysed using a two factor ANOVA with within-participant factors of time (pre-therapy, post-therapy and 5-week maintenance scores) and item (treated versus untreated items).

7.23 Secondary Outcome Technology Use and Confidence Measure

7.23.1 Analysis 1 – between group analysis

The first analysis here again compared the results of the immediate and delayed therapy groups between time point 1 and 2. Data were analysed using a mixed model two factor ANOVA with a within-participant factor of time (time point one versus time point two) and a between-participant factor of group (delayed versus immediate).

7.23.2 Analysis 2 – within group analysis

The second analysis examined differences over time for all participants. Data were analysed using a one factor ANOVA with a within-participant factor of time (pre-therapy, post-therapy and 5-week maintenance scores).

7.24 Additional analyses

Supplementary research questions were examined as follows:
The duration of therapy practice undertaken was examined through descriptive statistical analysis of computer logs, generated automatically

during participant practice. The maintenance of therapy effects over a longer term was addressed through a paired samples t-test comparison of scores before therapy with those achieved 10 weeks after its cessation.

The investigation of potential candidacy measures and the relationship between therapy intensity and size of therapy effect was examined using correlational analyses. Similarly, the investigation of the relationship between gains achieved on one measure in relation to gains achieved on other measures was also examined using a correlational analysis.

Chapter 8. Results

8.1 Introduction

This chapter presents analysis to address the previously identified research questions (section 4.4). Participant demographic data is presented first. This is followed by information regarding the duration and intensity of computer therapy use (research question 3a). Subsequent sections address research questions in order 1a through to 3e. The choice to address 3a earlier in the chapter was made to provide the reader with the relevant background information against which to consider primary and secondary outcomes.

8.2 Participants

8.2.1 CONSORT Flow Diagram

Forty-seven participants were referred to the project. Of these, 22 were allocated to either immediate or delayed therapy condition: 20 took part in the first two repeated measures assessments and 18 in all four repeated measures. See Figure 42 for full details.

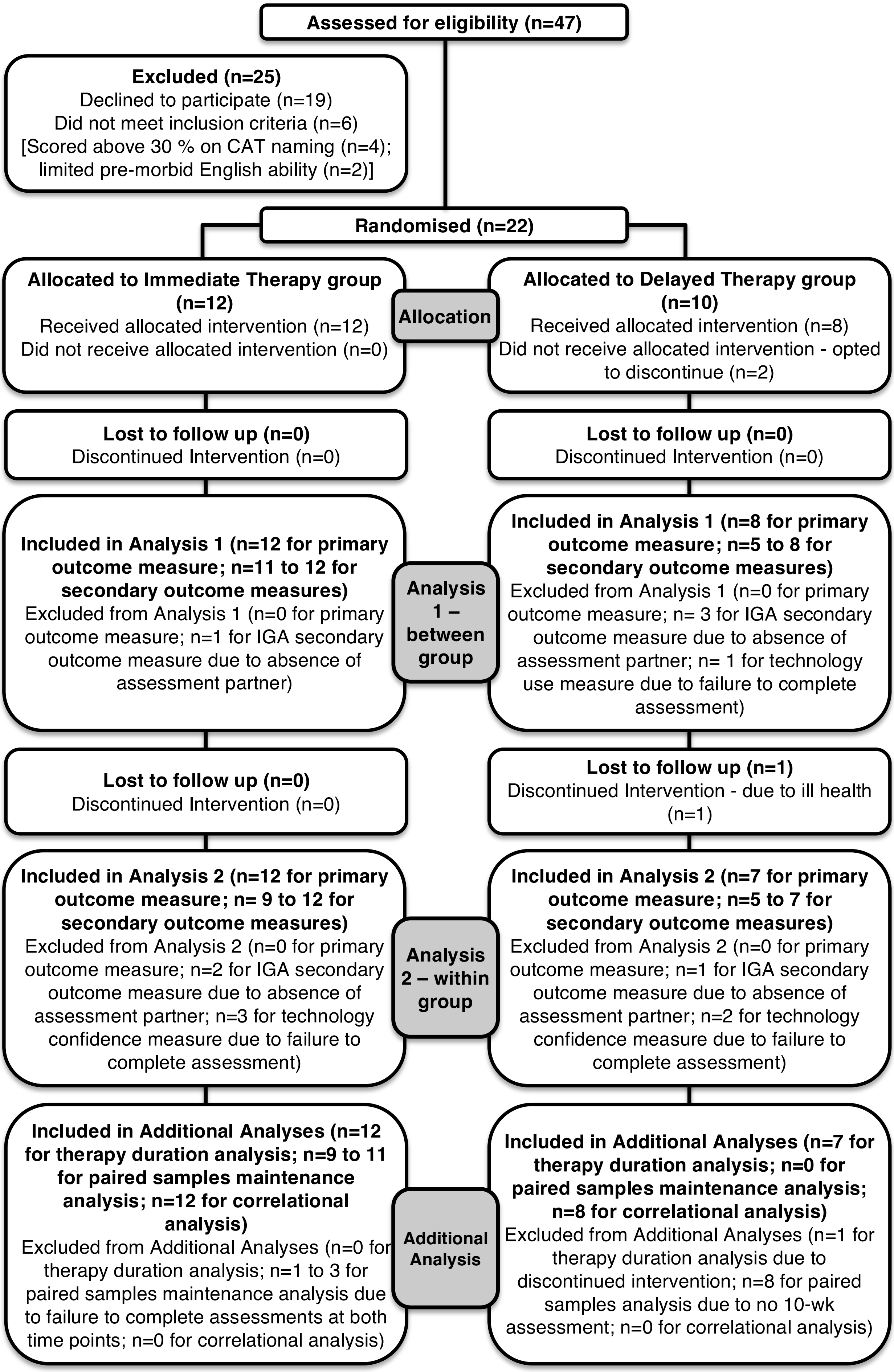


Figure 42. CONSORT flow diagram showing how many participants took part in each stage of the study

8.2.2 Characteristics of Participants

Participant characteristics are reported in Table 15.

8.2.2.1 Age

Participant age range reflects that of the typical stroke population with the majority of participants being aged 65 years or above (current estimates report around 75% of people who suffer a stroke being 65 or above – Royal College of Physicians, 2012). A smaller number of participants (6) were between 18 and 65 years old.

8.2.2.2 Aetiology

Participants all demonstrate severe expressive aphasia following one or more Cerebral Vascular Accidents (CVAs). Due to the fact that many of the participants may be aged 60 years or above, other issues associated with aging – such as the presence of cataracts affecting visual field, or typical age-related hearing loss may also be present – however these were not assessed directly.

8.2.2.3 Handedness

Eighteen of 20 participants demonstrated hemiplegia of the right arm. Ninteen of 20 participants reported a pre-morbid right-hand dominance. Of these, 18 reported a switch to reliance upon their left hand following stroke due to right-sided limb weakness. For these 18 participants then, the hand required for use in gesturing was their non-dominant left hand. Of the two outstanding participants, "R" retained use of her dominant right hand and "I" retained use of his pre-morbidly dominant left-hand.

8.2.2.4 Hearing

No detailed assessment of hearing was carried out for this project. This therapy strategy has a very low dependency on speech perception and therefore auditory discrimination skills are not considered critical to the successful use of the therapy. GeST levels 1 and 2 make use of one spoken instruction, which is consistent and repeated for all items ("Here is the gesture for X. Now it's your turn"). This spoken instruction is supported by visual demonstration of the gesture and a picture version of the target item. GeST level 3 and PowerGeST do make a more active and varied use of spoken language – in the form of brief dialogues, however these are supported by written and picture versions of the target item to be gestured and therefore speech perception alone is not critical to the use of these videos to practice gesture production.[1]

[1] Whilst hearing may not be critical, the ability to see and the ability to

Table 15. Participant details

ID	Gender	Age	Time post-onset (months)	Dominant hand post stroke	Dominant hand prior to stroke	Work History	Computer Use		Mobile Phone Use		Hearing	Vision	Languages Spoken
							Before stroke	Now	Before stroke	Now			
A	M	68	12	L	R	Taxi driver	Never	Never	Almost daily	Never	No reported difficulties	Reports difficulties with peripheral vision	English
B	F	80	25	L	R	Dinner lady	Occasional	Never		Never	No reported difficulties	Cataracts but B reports good view of assessment images	English
C	M	70	22	L	R	Engineer	Once a week	Never	Almost daily	Never	No reported difficulties	No reported difficulties	English; Swahili; Guajarati
D	F	61	21	L	R	Leader of after school club	Once a week	Once a week	Almost daily	Almost daily	No reported difficulties	No reported difficulties	English
E	M	58	30	L	R	Sales director at software company	Almost daily	Once a week	Almost daily	Once a week	No reported difficulties	Corrected with lenses	English
F	M	Not given	Not given	L	R	Data not available	Occasional	Never	About once a week	Never	No reported difficulties	Corrected with lenses	English
G	F	62	32	L	R	Manager within voluntary sector	Almost daily	Once a week	Almost daily	Never	No reported difficulties	No reported difficulties	English
H	M	73	230	L	R	Publisher. Trainee opera singer	Daily word processor use	Never	Never	Never	No reported difficulties	Corrected with lenses	English
I	M	56	120	L	L	TV and film editor	Almost Daily	Almost Daily	Almost Daily	Never	Some reduced hearing reported in R ear. No hearing aid.	Corrected with lenses	English
J	F	81	13	L	R	Secretary, dinner lady.	Almost Daily	Never	Almost Daily	Never	No reported difficulties	Reports difficulties with peripheral vision	English

ID	Gender	Age	Time post-onset (months)	Dominant hand post stroke	Dominant hand prior to stroke	Work History	Computer Use		Mobile Phone Use		Hearing	Vision	Languages Spoken
							Before stroke	Now	Before stroke	Now			
K	M	65	46	L	R	Delivery driver	Never	Almost Daily	Never	Never	No reported difficulties	Corrected with lenses	English
L	M	62	185	L	R	Cargo serviceman for an airline, sales person	Almost Daily	Never	Almost Daily	Never	No reported difficulties	Corrected with lenses	English, Urdu, Punjabi
M	F	46	15	L	R	Housewife	Almost every day – PC	Almost Daily - iPad	Almost Daily - iPhone	Never	No reported difficulties	Some difficulties reported on RH side	English
N	M	87	79	L	R	Machine worker - printing press	Never	Never	Almost Daily	Never	Wears hearing aids in both ears.	Corrected with lenses	English
O	M	66	18	L	R	Train planning/rostering - British Rail	Almost Daily - PC	Never	Once a week	Never	No reported difficulties	Some difficulties reported on RH side	English
P	M	69	37	L	R	Taxi driver / Horticulturalist / Dairy herdsman	Almost Daily - Laptop	Almost Daily - iPad	Almost Daily	Never	Slight loss in one ear - has hearing aid but does not wear it	Corrected with lenses	English
Q	F	71	37	L	R	Nurse, sheltered housing warden.	Once a week - laptop	Almost Daily - iPad	Almost Daily - basic phone	Once a week - basic phone	No reported difficulties	Corrected with lenses	English
R	F	78	55	R	R	Teacher, Brownie Leader, Potter	Almost Daily	Almost Daily	Never	Never	Wears hearing aids in both ears	Corrected with lenses	English, Hebrew, French, Italian
S	M	72	54	L	R	Architect	Almost Daily	Never	Never	Never	No reported difficulties	Corrected with lenses	English
T	M	56	90	L	R	Ran a taxi firm and drove vans	Almost Daily	Almost Daily	Almost Daily	Once a week	Wears hearing aids in both ears	Corrected with lenses	English, Urdu

Participants were pseudo-randomly allocated into either an immediate or delayed therapy group (as described in Methods chapter - sections 6-8). Allocations are reported in Appendix F.

Characteristics of participants in the immediate (n=12) versus delayed (n=8) intervention groups are summarised in Table 16.

Table 16. Group characteristics for Immediate and Delayed therapy groups

	Group Allocation (Immediate or Delayed)	**Participant numbers within each category**		
Gender	Immediate	7 male; 5 female		
	Delayed	6 male; 2 female		
Pre-stroke reported computer use	Immediate	3 never; 4 once a week or less; 5 almost every day		
	Delayed	0 never; 2 once a week or less; 6 almost every day		
	Group Allocation (Immediate or Delayed)	**Mean**	**Std. Deviation**	**Std. Error Mean**
Age (in years)	Immediate	**67.83**	10.18	2.94
	Delayed	**67.00**	10.71	4.05
Total number of months post stroke (at Time point 1)	Immediate	**61.42**	71.18	20.55
	Delayed	**54.86**	38.58	14.58
CLQT Visuospatial Cognitive Domain Score (of 105)	Immediate	**45.67**	24.87	7.18
	Delayed	**48.13**	26.90	9.51
BUPS non-standardised praxis summary (of 42. high score = little or no praxis impairment)	Immediate	**21.25**	9.67	2.79
	Delayed	**18.88**	9.05	3.20
CAT Picture Naming Subtest Raw Score (of 24)	Immediate	**0.42**	1.16	0.34
	Delayed	**0.88**	2.10	0.74
CAT Spoken Word Comprehension Subtest Raw	Immediate	**11.17**	2.72	0.79

	Group Allocation (Immediate or Delayed)	Participant numbers within each category		
Score (of 15)	Delayed	**11.75**	2.82	1.00
CAT Spoken Sentence Comprehension Subtest Raw Score (of 16)	Immediate	**6.17**	2.59	0.75
	Delayed	**6.63**	1.69	0.60
CAT Written Word Comprehension Subtest Raw Score (of 15)	Immediate	**9.08**	3.94	1.14
	Delayed	**8.88**	3.18	1.13
Total Gesture Assessment Score at T1 (of 40)	Immediate (Treated + Untreated Items)	**11.75**	9.54	2.75
	Delayed (Treated + Untreated Items)	**15.25**	10.58	3.74
Total Naming Assessment Score at T1 (of 40)	Immediate (Treated + Untreated Items)	**2.67**	3.94	1.14
	Delayed (Treated + Untreated Items)	**1.75**	1.83	0.65
Total Interactive Gesture Assessment Score at T1 (of 24)	Immediate (Treated + Untreated Items)	**11.50**	6.14	1.77
	Delayed (Treated + Untreated Items)	**7.71**	6.75	2.39
Technology Use Score at T1 (of 17)	Immediate	**4.17**	3.01	0.87
	Delayed	**6.50**	4.38	1.55
Technology Confidence Score T1 (of 85)	Immediate	**36.38**	20.68	5.97
	Delayed	**40.38**	24.57	8.69

Independent samples t-tests reveal no significant differences between immediate and delayed groups in relation to age, time post stroke, Visuospatial score on the CLQT (cognitive assessment), BUPS score (assessment of praxis) or CAT subtests (picture naming, spoken word comprehension, sentence comprehension and written word comprehension).

Furthermore, no significant differences are observed between groups for either the primary or secondary outcome measures. All t-test results were $F \leq 2.99$, $p \geq 0.10$.

Key observations from the above data include the following:
Low naming scores for both the standardised CAT assessment (between 0 and 1 out of 24) and the experimental naming measure (between 1 and 3 out of 40). This contrasts with the relatively superior score of between 12 and 15 out of 40 for the experimental gesture measure. Technology use and confidence scores are also relatively low in comparison to values reported for participants with stroke and no aphasia and those without stroke or aphasia (reported in section 6.3.2.6).

8.3 Research Question 3a

How Much Computer Practice do Participants Undertake and at What Intensity?

Results in this section comprise details of logged use of the GeST therapy technology during the first four weeks of the 5-week therapy period. PowerGeST did not log activity and therefore activity from the fifth week is not reported.

8.3.1 Time Spent Using GeST

Time spent using GeST was computer-logged, automatically, for each participant. Outcomes for each week of practice are reported in Table 17.

GeST log data (Table 17) shows that participants undertook a minimum of 5 hours 20 minutes use across a four-week period (28 days) and a maximum of 26 hours 51 minutes use. On average, participants completed just less than two sessions a day (52/28=1.86 sessions per day) each lasting around 17 minutes.

Table 17. Week-by-week summary of sessions and practice times for all participants' use of GeST

	Week 1		Week 2		Week 3		Week 4		**TOTALS**	
ID	No. of Sessions	Total time spent	No. of Sessions	Total time spent	No. of Sessions	Total time spent	No of Sessions	Total time spent	**No of Sessions**	**Total time spent**
A	35	8:13:58	35	8:30:32	39	6:05:11	23	4:00:58	**132**	**26:50:39**
B	16	2:24:00	8	1:40:56	9	1:50:01	6	1:41:30	**39**	**7:36:27**
C	25	9:12:34	9	2:52:27	8	2:41:11	13	6:51:54	**55**	**21:38:06**
D	7	1:05:35	8	1:13:11	10	2:06:25	9	0:55:13	**34**	**5:20:24**
E	17	8:09:29	5	1:12:06	10	2:23:50	6	1:22:43	**38**	**13:08:08**
F	3	00:21:04	Discontinued therapy due to ill health							
G	10	4:05:32	9	3:13:47	13	3:25:47	10	3:58:55	**42**	**14:44:01**
H	11	3:47:04	12	3:42:30	20	5:26:00	10	3:28:02	**53**	**16:23:36**
I	11	3:27:08	10	2:43:50	11	3:50:11	9	3:57:56	**41**	**13:59:05**
J	15	4:13:46	20	4:27:44	18	1:48:59	9	1:23:52	**62**	**11:54:21**
K	22	5:40:46	20	5:27:30	23	4:14:56	15	5:18:42	**80**	**20:41:54**
L	11	3:55:08	8	2:28:49	13	2:40:01	6	4:51:19	**38**	**13:55:17**
M	10	4:17:36	9	5:13:41	3	2:18:49	10	5:57:43	**32**	**17:47:49**
N	9	11:40:51	3	0:44:41	4	0:18:38	6	1:05:20	**22**	**13:49:30**
O	23	13:18:07	8	3:50:11	7	2:42:35	16	5:31:14	**54**	**25:22:07**
P	54	4:13:54	19	4:48:37	16	4:30:55	14	4:37:40	**103**	**18:11:06**
Q	14	3:09:14	5	2:06:09	5	1:21:50	4	1:00:58	**28**	**7:38:11**
R	14	5:33:57	11	1:49:13	12	1:56:06	6	1:01:33	**43**	**10:20:49**
S	9	1:45:17	13	2:01:38	8	2:22:38	7	1:34:41	**37**	**7:44:14**
T	14	2:48:12	11	3:12:23	13	4:45:02	18	4:11:21	**56**	**14:56:58**
Min	**3**	**00:21:04**	**3**	**00:44:41**	**3**	**00:18:38**	**4**	**00:55:13**	**22**	**5:20:24**
Max	**54**	**13:18:07**	**35**	**08:30:32**	**39**	**06:05:11**	**23**	**06:51:54**	**132**	**26:50:39**
Mean	**16.5**	**05:04:10**	**11.74**	**03:13:41**	**12.74**	**02:59:26**	**10.37**	**03:18:30**	**52.05**	**14:50:40**

A repeated-measures ANOVA with a within-participant factor of time (four levels – weeks 1, 2, 3 and 4) revealed a significant effect of time upon the total time spent practising each week. ($F=5.65$ (3, 54), $p<0.05$). i.e. amount of therapy practice undertaken changed significantly over time. The partial eta-squared value indicates a large effect size ($\eta_p^2 = 0.24$). Bonferroni corrected post hoc tests identify no significant difference between individual time points ($p>0.05$). Despite this, Figure 43 indicates that the change over time is mainly due to a tail off in use after week 1.

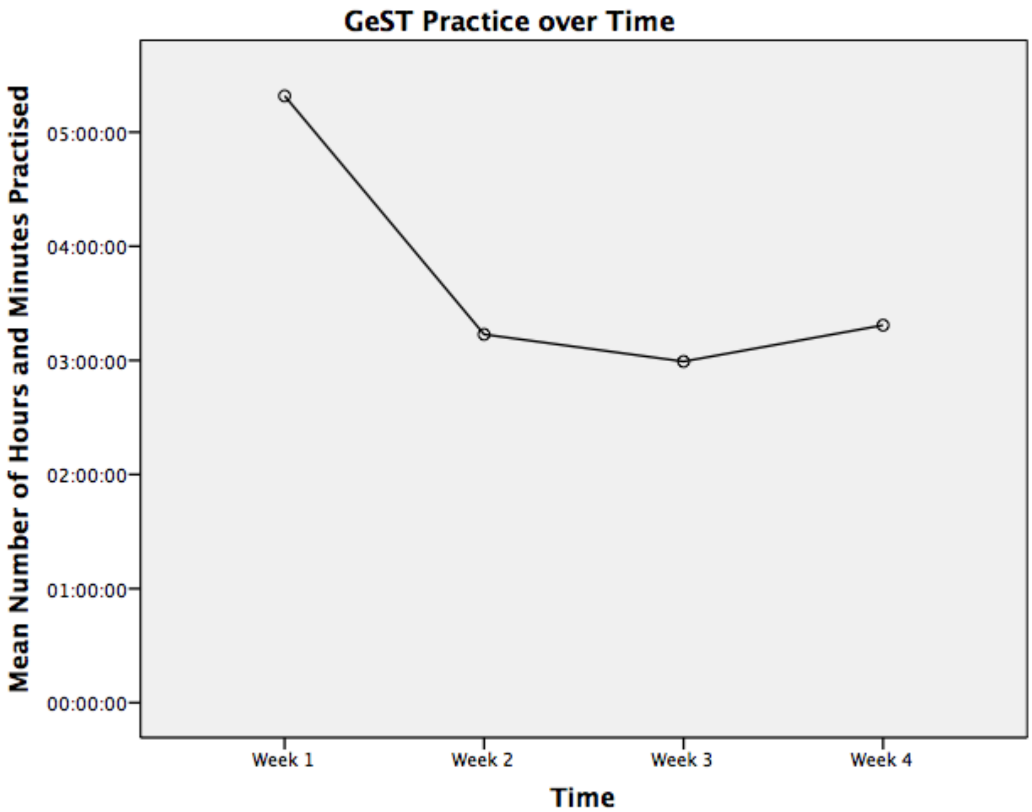


Figure 43. Time spent practising GeST each week

To further examine the use of the GeST system in more depth, log data were sampled at two time points:

1. The first session* from the second day of use.

2. The last session* from the penultimate day of use.

The aim in choosing these two sessions was to assess participants' unaided use of the tool during early and later sessions where the researcher was not present to provide support. At both time points data were analysed to reveal

* Individual sessions were identified as any data listed between an "*application start*" and "*application quit*" record in the data log. (Further details of how this information was summarised are reported in the Methods chapter of this document.)

the session's duration (i.e. from system start up to system shutdown), the number of times a participant navigated to one of the GeST program's three different practice levels (navigation count), the total number of levels accessed (i.e. 1, 2, or 3) the amount of time spent outside of level 1 (i.e. in 2 or 3) and the total number of gestures productions attempted. Average values for these outcomes are reported in Table 18.

Table 18. Mean interaction values for GeST

	First Session from Second Day of Use			Last Session from Penultimate Day of Use		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Session duration	27s	1h 15m 43s	26m 40s	35s	50m 33s	21m 20s
Level navigation count (number of switches to a different level)	0	13	2.47	0	5	1.84
Number of levels accessed (out of 3)	0	3	2.11	1	3	2.26
Time spent outside of level 1	0s	48m 19s	12m 31s	0s	42m 20s	15m 13s
Total number of Gestures attempted	0	83	24.89	1	56	23.58

The session duration data in Table 18 suggest that participants engaged with the technology (practice sessions lasted between 21 and 26 minutes on average). Data regarding the navigation count (average 2 switches between level in a session), the number of levels accessed (2 per session) and the time spent outside of level 1 (between 12 ½ and 15 minutes) indicate a relative mastery of the navigational aspects of the GeST system. Furthermore, the total number of gestures attempted (24-25 gestures per session) indicates a substantial number of gesture interactions being undertaken at both time points. With the exception of a shorter session time and proportionally more time spent outside of level one during the late session when compared to the early session, data in both sessions appear roughly comparable. To investigate this further, paired samples t-tests were conducted to compare the number of navigational switches, the number of gestures attempted and the number of levels accessed in the early versus the late session for each participant. These analyses revealed no significant difference for any of the comparisons, which suggests there was no meaningful change in GeST competency over the four-week period.

A subsequent paired samples t-test was conducted to compare the time taken to complete the GeST exercise in the early session compared to the time taken to complete the GeST exercise during a later session, approximately four weeks later. This revealed no significant difference in session time for first use (*M*= 32:41min, *SD*=23:07min) compared to session time for in final use (*M*= 24:16min, *SD*=14:48min); t (17) =1.59, $p > 0.05$.

8.3.2 Time Spent Using PowerGeST

Unlike GeST, PowerGeST did not automatically log usage data. Data from PowerGeST use were therefore gathered by video recording the participants' first and final use of the PowerGeST system at the start and end of week 5. Video data were then examined to record the session duration (i.e. from start up to shut down) for each participant at each time point. Outcomes are summarised in Table 19 Table 1.

Table 19. Time taken to complete PowerGeST exercise at start & end of week 5

	Minimum	Maximum	Mean	Std. Deviation
Time to complete PowerGeST - First Use (m:s)	06:01	21:48	08:55	03:54
Time to complete PowerGeST - Final Use (m:s)	05:07	16:44	07:29	02:37

A paired samples t-test was conducted to compare the time taken to complete the PowerGeST exercise at first use compared to the time taken to complete the PowerGeST exercise at final use, one week later. There was a significant difference in session time for first use (*M*= 8.9min, *SD*=3.9min) compared to session time for in final use (*M*= 7.4min, *SD*=2.7min); $t(16) = 3.67$, $p < 0.005$. These results indicate that participants became significantly faster (or more competent) at completing the PowerGeST exercise over the course of a week. This change in competency marks a contrast to the lack of change observed in the GeST log data, however it is notable that the comparable session length for the different technologies is strikingly different dropping from an average of 21 minutes and 20 seconds for the penultimate session of GeST compared to 7 minutes and 29 seconds for the final session of PowerGeST.

8.4 Repeated Measures Analyses Methods

Four measures were repeated within the design of this study: gesture (primary outcome measure), naming, interactive gesture and technology use and confidence (all secondary outcome measures). Results from each measure will be analysed separately using the following structure:

- First Analysis: Between groups (immediate therapy versus delayed therapy) and across time points 1 and 2 for all four measures. Between items (treated versus untreated item scores) for gesture, naming and interactive gesture measures. Analysis design and descriptive statistics are reported in section 8.4.1
- Second Analysis: Within group analysis - pooled group analysis across pre-therapy, post-therapy and maintenance time points for all four measures. Between item analysis (treated versus untreated item scores) for gesture, naming and interactive measures. Analysis design and descriptive statistics are reported in section 8.4.2.

8.4.1 First Analysis - Between group and across time analyses for all four measures and between-item analyses for gesture, naming and interactive gesture measures

This analysis contrasted participants from the immediate therapy group (n=11 to 12) with the delayed therapy group (n=5 to 8) across time points 1 and 2 (illustrated below in Figure 44)

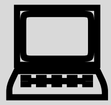
	Time Point 1		←5 weeks→	Time Point 2
Immediate Therapy Group	**Background assessments**: • CAT single word naming • CAT spoken single word and sentence comprehension • CAT written single word comprehension • CLQT non-linguistic executive function • BUPS limb praxis	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence	**Gest + PowerGeST Intervention**  2 x video recordings of PowerGeST being used	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence
Delayed Therapy Group			No input	

Figure 44. Assessment protocol for Analysis 1 of the Immediate and Delayed therapy groups

- comparing individuals who had received the intervention with a control group who had not, as well as items which were treated in the intervention with items which were not. Whilst all 20 participants completed the primary outcome gesture measure and the secondary outcome naming measure, only 16 completed the interactive gesture assessment (due to the absence of a communication partner from one or more testing session). Furthermore, whilst all 20 completed the technology use assessment, only 15 completed the technology confidence measure at both time points (due to a failure to engage with the measurement scale used to report confidence level).

Table 20 summarises assessment outcomes for both groups at time points 1 and 2.

Table 20. Mean (standard deviation) assessment scores at T1 and T2 by group

Assessment		Group	T1 Score (SD)	T-test comparing Immediate vs. Delayed T1	T2 Score (SD)
Gesture Assessment	Treated Items (Max score = 40)	Immediate (n=12)	6.75 (5.86)	t(18)=-0.93, p= 0.36. Not significant	11.33 (6.80)
		Delayed (n=8)	9.38 (6.59)		9.13 (5.46)
	Untreated Items (Max score = 40)	Immediate (n=12)	5.00 (4.07)	t(18)=-0.46, p= 0.65. Not significant	5.75 (4.83)
		Delayed (n=8)	5.88 (4.29)		4.88 (5.41)
Naming Assessment	Treated Items (Max score = 40)	Immediate (n=12)	1.83 (2.79)	t(18)=0.68, p= 0.50. Not significant	2.75 (3.47)
		Delayed (n=8)	1.13 (0.99)		0.63 (0.74)
	Untreated Items (Max score = 40)	Immediate (n=12)	0.83 (1.53)	t(18)=-0.33, p= 0.74. Not significant	0.75 (1.36)
		Delayed (n=8)	0.63 (1.06)		1.00 (1.93)
Interactive Gesture Assessment	Treated Items (Max score = 12)	Immediate (n=11)	6.01 (3.36)	t(17)=0.95, p= 0.36. Not significant	6.27 (3.66)
		Delayed (n=5)	4.60 (2.97)		6.66 (3.13)
	Untreated Items (Max score = 12)	Immediate (n=11)	5.10 (3.08)	t(17)=1.09, p= 0.29. Not significant	4.73 (4.43)
		Delayed (n=5)	4.40 (3.78)		6.20 (3.56)
Technology Use & Confidence	Technology Use (Max score = 17)	Immediate (n=12)	4.17 (3.01)	t(18)=-1.42, p= 0.17. Not significant	4.91 (3.00)
		Delayed (n=8)	6.50 (4.38)		7.63 (4.34)
	Technology Confidence (Max score = 85)	Immediate (n=9)	43.65 (13.03)	t(15)=-0.27, p= 0.79. Not significant	43.77 (11.92)
		Delayed (n=6)	45.64 (17.37)		56.08 (9.94)

Independent samples t-tests revealed no significant differences between groups in relation to T1 scores - suggesting that the groups started at comparable levels of performance for all of the examined outcome measures.

8.4.2 Second Analysis – Within Group Analysis (pooled group across time analyses for all four measures and between-item analyses for gesture, naming and interactive gesture measures)

This analysis collapsed the two participant groups and examined scores for all participants over three time points: pre-therapy, post-therapy and after a 5-week maintenance period, as illustrated in Figure 45 overleaf.

By conflating the scores here, we have a more powerful analysis, now comprising 19 participants (one participant from the delayed group did not continue beyond T2 due to ill health). The within-group factor here is item type, namely: whether items tested were treated or untreated within the intervention.

	Time Point 1	5 weeks	**Time Point 2**	5 weeks	**Time Point 3**	5 weeks	**Time Point 4**
Immediate Therapy Group	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture • Technology Use and Confidence	**Gest + PowerGeST Intervention** 2 x video recordings of PowerGeST being used	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence	No input	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence	No input	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence
Delayed Therapy Group	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence	No input	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence	**Gest + PowerGeST Intervention** 2 x video recordings of PowerGeST being used	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence	No input	**Repeated Measures assessments**: • Gesture • Naming • Interactive Gesture Technology Use and Confidence

Pre-therapy scores

Post-therapy scores

5-week maintenance scores

Figure 45. Conflation of measures for Immediate and Delayed therapy groups

Table 2 summarises assessment outcomes for all participants (n=14 to 19) at pre-therapy, post-therapy maintenance time points. Whilst all 19 remaining participants completed the primary outcome gesture measure and the secondary outcome naming measure, only 16 completed the interactive gesture assessment (due to the absence of a communication partner from one or more testing session). All 19 completed the technology use assessment, however only 14 completed the technology confidence measure at all three time points (due to a failure to engage with the measurement scale used to report confidence level).

Table 21. Mean (standard deviation) assessment scores pre-therapy, post-therapy and at maintenance for all participants

Assessment		Pre – Therapy Score (SD)	Post-Therapy Score (SD)	5 – week Maintenance Score (SD)
Gesture Assessment (n=19)	Treated Items (Max score=40)	7.84 (5.80)	11.32 (6.53)	11.32 (6.96)
	Untreated Items (Max score=40)	5.21 (4.48)	6.00 (5.17)	6.26 (5.56)
Naming Assessment (n=19)	Treated Items (Max score=40)	1.37 (2.31)	2.52 (3.15)	1.58 (2.34)
	Untreated Items (Max score=40)	0.95 (1.68)	0.68 (1.25)	0.68 (1.20)
Interactive Gesture Assessment (n=16)	Treated Items (Max score=12)	6.38 (2.96)	7.31 (3.11)	8.50 (3.18)
	Untreated Items (Max score=12)	5.56 (2.94)	4.75 (3.97)	6.06 (3.94)
Technology Use & Confidence (n=19; n=14)	Technology Use (Max score=17)	5.79 (3.85)	6.37 (3.53)	5.32 (3.80)
	Technology Confidence (Max score=85)	47.38 (13.67)	47.43 (12.97)	46.38 (13.75)

Primary Outcome Measure

8.5 Research Question 1a

Does Practice with GeST + PowerGeST increase the Production of Gestures in Isolation?

*i. **Are gains confined to items trained in therapy or do they generalise to untrained items?***

8.6 Scorers

Data for the primary outcome measure (the gesture in isolation assessment) were scored by judges blinded to the time of the assessment, the identity of the items tested and the experimental design of the study.

All scorers were students of speech and language therapy at City University London. Scoring took place over a series of 7 sessions. 42 student scorers took part in the scoring process. A total of 77 videos were generated for 20 participants (four videos per participant with the exception of participant 6 who discontinued after 2 assessments and participant 14 who discontinued after 3 assessments).

Videos were judged using the procedure outlined in section 7.10.1.2. As described, those answers which were identical or synonymous to the target were awarded two points. Items that elicited no response or "don't know" were awarded zero points. For all other responses, target and response were presented to a further separate scorer – blinded to the time and participant. This scorer judged the acceptability of the response in relation to the target in terms of whether it conveyed: "relevant, useful information from the message, close to the target and useful as a means of conveying the message". Items deemed acceptable against these criteria were awarded one point.

A total of 1683 items were submitted to this additional scoring judgement. From this, a total of 442 items were awarded one point for acceptability. Scores were reallocated to original assessments to establish an overall gesture score at each time point, with subtotals for points awarded to both treated and untreated items.

Outcomes from the repeated measure gesture assessments were subjected to analysis 1 described in Figure 44 and analysis 2 described in Figure 45 above.

8.6.1 Analysis One – Between Group Analysis

8.6.1.1 Between Group Differences on Measures of Gesture at T1 and T2

This analysis looked at within- and between-participant changes over time for 20 people who completed the first two assessments within the experiment. Data were analysed using a mixed-model three factor ANOVA with two within-participant factors of time (time point one versus time point two) and item (treated versus untreated items) and a between-participant factor of group (delayed versus immediate). Descriptive statistics are reported in table 20. For the immediate therapy group, data reveal an average increase of around 5 points (4.58 = 11.33-6.75) for treated items after therapy and of around 1 (0.75 = 5.75-5.00) for the untreated items. Over the same period, the delayed therapy group shows no real change (-0.25 = 9.13-9.38) for treated items and a decrease of 1 point (-1 = 4.88 – 5.88) for the untreated items.

Outcomes from a Shapiro-Wilk test indicate a non-normal distribution of scores for untreated items in the delayed group at time point 1 ($W=0.76$, $p<0.05$). Measures of skewness reveal a positive skew for this data set (skewness z-score for T2 > 1.96), as illustrated in Figure 46.

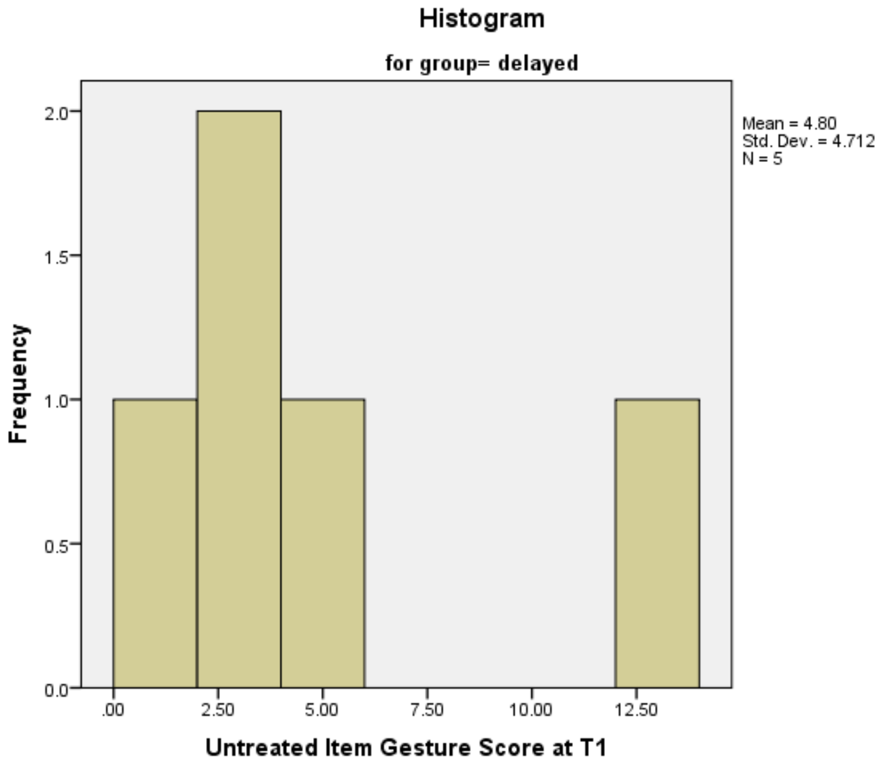


Figure 46. Histogram illustrating the positive skew of untreated item gesture scores for the delayed therapy group at T1

To address the non-normal distribution of the data, a log transformation was applied to all T1 and T2 gesture scores. Due to zero scores within the raw data, a value of 1 was added to each data point prior to transformation. Data were then transformed using a logarithmic transformation of base 10. This transformation method was chosen to compress higher scores and hence adjust for the positive skew of the raw data. As the subsequent values satisfy the assumption of normality, subsequent between-group analysis for gesture data was completed using this transformed data.

The mixed-model ANOVA analysis reveals a main effect of item (F=39.29 (1,18), p<0.05). The partial eta-squared value (η_p^2 = 0.69) indicates a large effect size. There was no effect of time (F=1.97 (1, 18), p>0.05, η_p^2 = 0.10) or group (F=0.06 (1,18), p>0.05, η_p^2 = 0.00). There was a significant interaction between group and time (F=10.88 (1,18), p=<0.005) and between time and item (f=7.77 (1,18), P<0.05). Again, partial eta-squared values (η_p^2 = 0.38; η_p^2 =0.30) indicate a large effect size for both comparisons. However, there was no interaction between group and item (F= 1.81 (1, 18), p>0.05, η_p^2 = 0.09) or between time, group and item (F=0.01 (1, 18), p>0.05).

The interaction between group and time indicates a positive effect of intervention. It shows that the participants who had received therapy between T1 and T2 (the immediate therapy group) improved, whereas those who had not yet received therapy (the delayed therapy group), did not.

The interaction between time and item type indicates a differential effect of time for treated versus untreated items in that treated items improved more than untreated items.

Figure 47 and Figure 48 below show change in performance over time for treated and untreated items respectively.

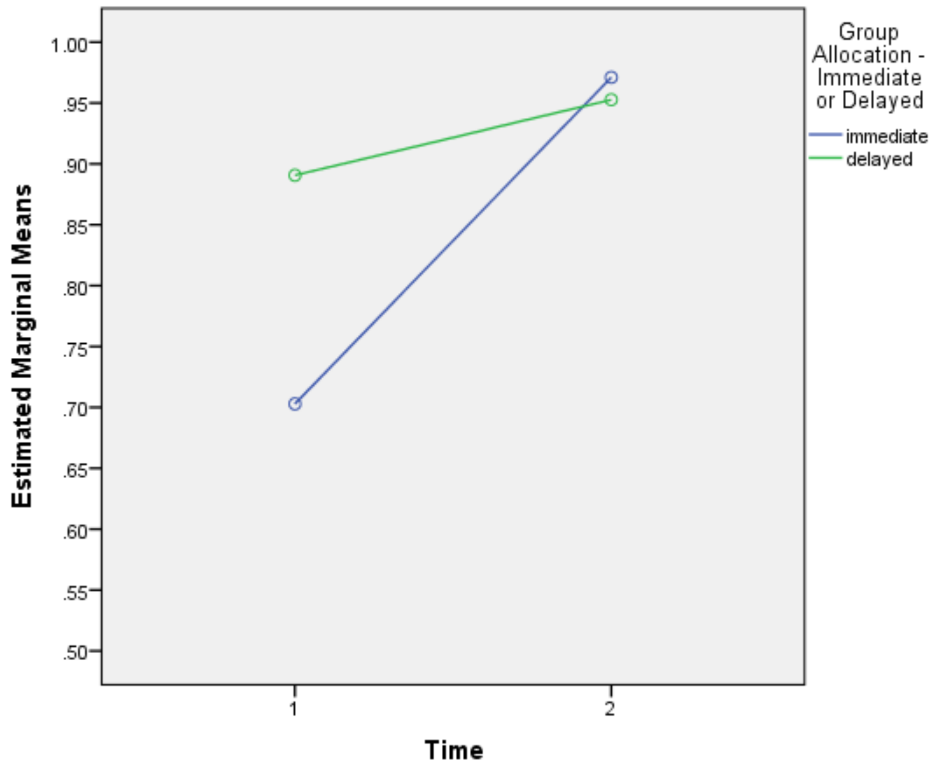


Figure 47. Treated item scores over time and by group

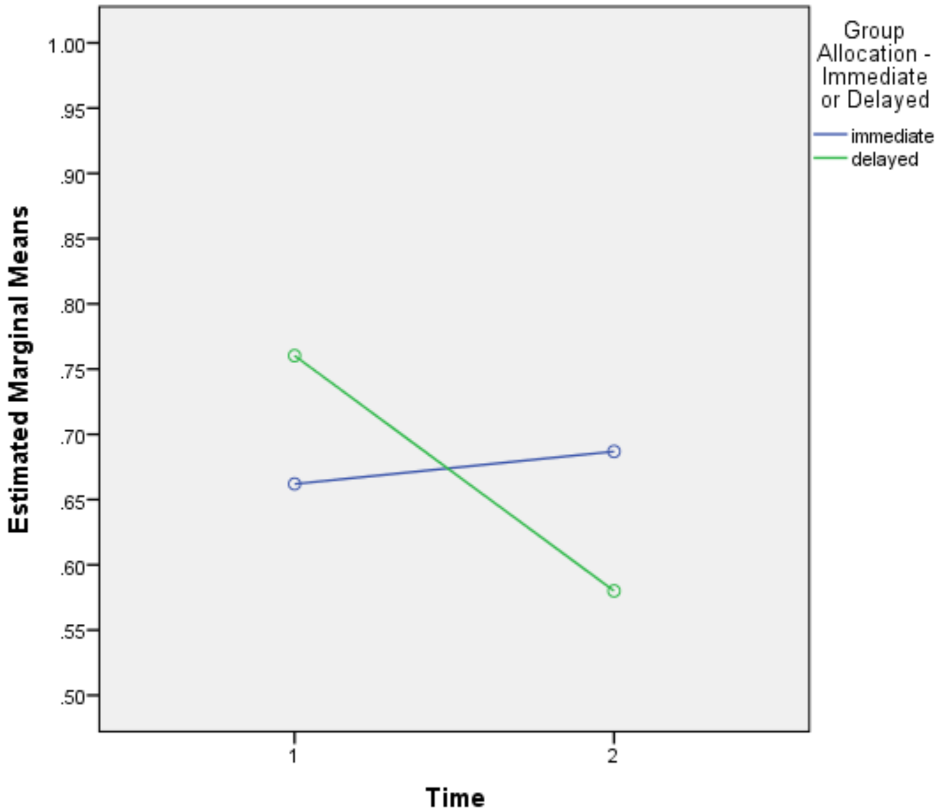


Figure 48. Untreated item scores over time and by group

These graphs illustrate an improvement both treated[2] and untreated scores for the immediate therapy group (blue lines). The delayed therapy group (green lines) demonstrates a slight increase in scores for treated items and a decrease in scores untreated items. The most marked change is the improvement in performance for treated items in the immediate therapy group.

8.6.2 Analysis Two – Within-Group Analysis

8.6.2.1 Gesture score changes over time for conflated participant groups comparing treated and untreated items

As previously described, data from both the immediate and delayed groups were conflated for the within-group analysis. Gesture score data were

[2] It is important to note here that whilst items are referred to as 'treated' for both groups, for the delayed group they might more accurately be described as 'to-be-treated' – acknowledging the fact that the 'treatment' intervention has yet to take place.

analysed using a two-factor ANOVA with the within-participant factors of time (pre-therapy, post-therapy and 5-week maintenance scores) and item (treated versus untreated items). Descriptive statistics are reported in Table 21, above.

Outcomes from Shapiro-Wilk tests of this data indicate a normal distribution of gesture use scores for both treated and untreated items at all three time points. As this therefore satisfies the assumption of normality, original data was used for the subsequent analysis.

The two-factor ANOVA analysis revealed a main effect of time ($F=8.88$ (2, 36), $p<0.005$) and of item ($F=25.02$ (2, 18), $p<0.005$). Partial eta-squared values ($\eta_p^2 = 0.33$; $\eta_p^2 = 0.58$) indicate a large effect size for both factors. There was no interaction between time and item ($F=2.42$ (2, 36), $p>0.05$). The partial eta-squared value ($\eta_p^2 = 0.12$) however, indicates a medium effect size for this interaction.

In combination with the descriptive statistics we can observe that the main effect of time demonstrates a significant improvement in gesture scores and that the main effect of item suggests a significant difference in scores for treated versus untreated items.

Bonferroni post hoc tests indicate a significant difference between pre and post-therapy scores ($p<0.05$) and between pre-therapy and 5-week maintenance scores ($p<0.05$) but not between post-therapy and maintenance scores ($p>0.05$). Thus, the change occurred over the therapy period and was maintained after therapy was withdrawn.

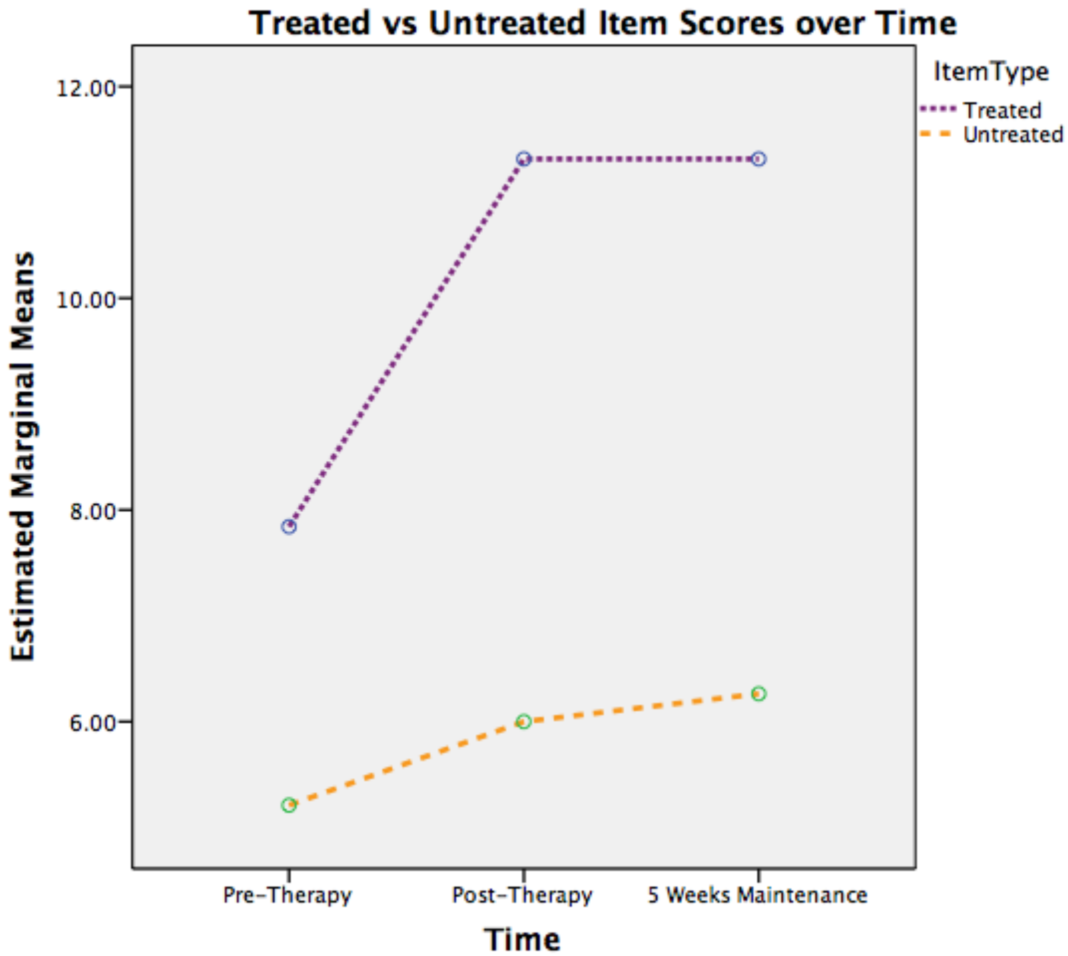


Figure 49. Treated versus untreated gesture scores over time

Figure 49 shows the change in gesture score over time for both treated and untreated items. This graph further illustrates the therapy effect. Despite the non-significant interaction, there was an average 3.5-point improvement in the score for the treated items (3.48 =11.32-7.84), compared to an average 1 point improvement for the untreated items (1.05 = 6.26-5.21). The main effect of item is also evident, since untreated items demonstrate lower scores across all time points.

Individual Change Scores

Participants demonstrated varied outcomes at an individual level. Figure 50 shows change in treated gesture score for each participant – calculated by subtracting each individual's pre-therapy treated item score from their post-therapy treated item score.

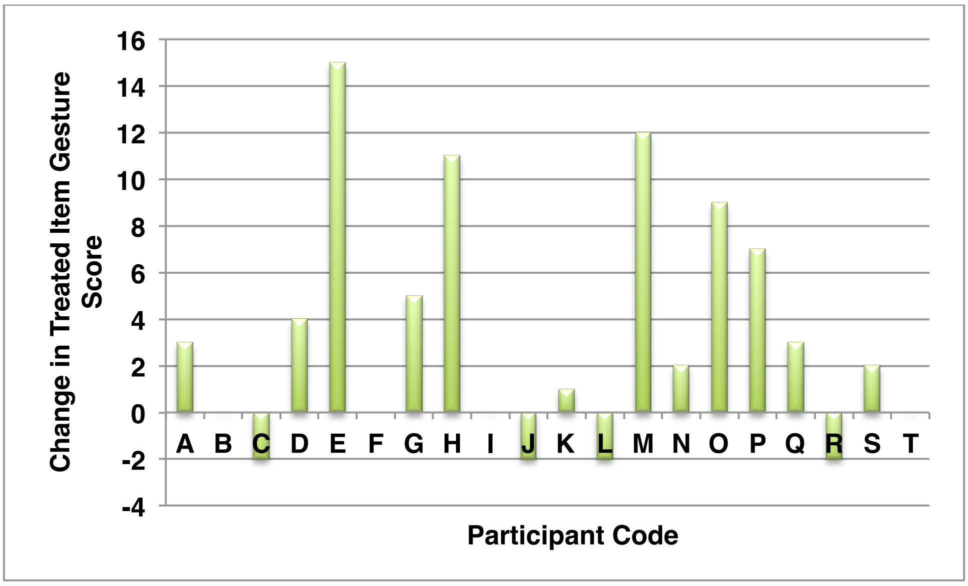


Figure 50. Change in treated item gesture score immediately after therapy for each participant

This indicates that 9 participants of 19 made gains of between 3 and 15 points out of 40 immediately post-therapy. Participants B, H and S demonstrated no change – as evidenced by the lack of green bar linked to these participants' codes along the x-axis.

8.6.1 Item-by-item analysis of gesture production for treated items

Figure 51 presents gesture production scores for individual gesture items at time point 1 and post-intervention (i.e. at T2 for the "immediate" therapy group and T3 for the "delayed" therapy group). The first 20 items on the x-axis form the treated gesture group, the second 20 form the untreated group. Both item sets indicate similar distribution patterns at T1, although performance for treated items appears slightly higher than for untreated items. (See Appendix A for details of how treated versus untreated items were selected).

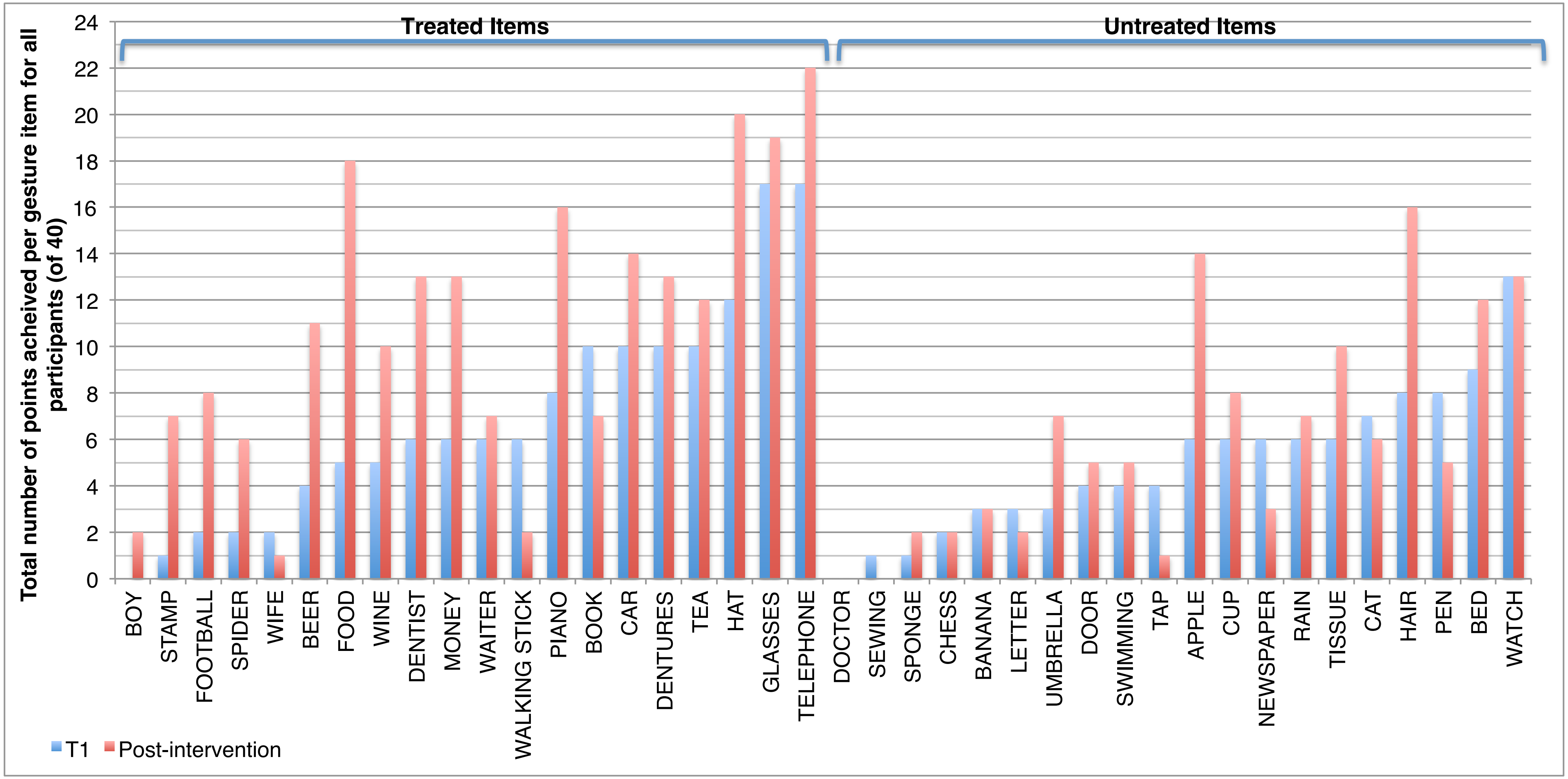


Figure 51. Number of points achieved per gesture item at T1 and post-intervention – treated and untreated items

The red, post-intervention score bars indicate a general increase in gesture performance following therapy, with more consistent improvement observed across the treated item set. Seventeen of 20 treated items demonstrate improvements post-intervention (items which did not improve were: "wife", "walking stick" and "book"). Nine of 20 items from the untreated group demonstrate improvement post-intervention. Considering just the treated group now, ten items made gains of five points or more post-intervention. These were: "stamp", "football", "beer", "food", "wine", "dentist", "money", "piano", "hat" and "telephone".

Secondary Outcome Measures

8.7 Research Question 2a

Does practice with GeST + PowerGeST improve spoken naming?

i. ***Are gains confined to items trained in therapy or do they generalise to untreated items?***

Participants completed the naming assessment at four time points. At each time point, they were awarded a score of one for each item correctly named during the assessment. There were 40 items in the assessment: twenty that were treated using GeST + PowerGeST and 20 that were untreated. Sections 8.7.1 and 8.7.2 present outcomes from the group analysis of participant data from these assessments. An additional, supplemental analysis of individual performance is presented later, in section 8.12.1.

8.7.1 Analysis One – Between Group Analysis

8.7.1.1 Between Group Differences on Measures of Naming at T1 and T2

Naming score data were analysed using a mixed-model three-factor ANOVA with two within-participant factors of time (time point one versus time point two) and item (treated versus untreated items) and a between-participant factor of group (delayed versus immediate). Naming scores for treated and untreated items for both the immediate (n=12) and delayed (n=8) therapy groups are summarised in Table 20. For the immediate therapy group, these reveal an average increase of around 1 point (0.92= 2.75-1.83) for treated items after therapy and a decrease of 0.5 (0.5 = 0.63-1.13) for the untreated items. Over the same period, the delayed therapy group shows no real change (-0.08 = 0.75-0.83) for treated items and a decrease of around 0.5 points (0.37 = 1.00 – 0.63) for the untreated items.

Although the groups begin at slightly different starting points, t-tests reveal that this difference is not significant.

Outcomes from a Shapiro-Wilk test indicate a non-normal distribution of naming scores for treated and untreated items in the immediate group at time point 1 (W=0.68, p<0.05; W=0.68, p<0.05). Measures of skewness reveal a positive skew (skewness z-scores for T1 > 1.96). Shapiro-Wilk test results also reveal a non-normal distribution of naming scores for untreated items in the immediate group at time point 2 (W=0.66, p<0.05). Again a positive skew is observed (skewness z-scores for T1 > 1.96).

To address the non-normal distribution of the data, a log transformation was applied to all T1 and T2 naming scores. Due to zero scores within the raw data, a value of 1 was added to each data point prior to transformation. Data were then transformed using a logarithmic transformation of base 10. As for the gesture data, this transformation method was chosen to compress higher scores and hence adjust for the positive skew of the raw data. Between-group analysis for naming data was then completed using this transformed data.

The mixed-model ANOVA analysis demonstrates no main effect of time (F=0.26 (1,18), p>0.05, $\eta_p^2 = 0.01$) item (F=5.23 (1,18), p>0.05, $\eta_p^2 = 0.23$), or group (F=0.22 (1,18), p>0.05, $\eta_p^2 = 0.01$). There are no interactions between group and time (F=1.89 (1,18) p>0.05, $\eta_p^2 = 0.10$), group and item

(F=1.10 (1,18) p>0.05, $\eta_p^2 = 0.06$) or time and item (F=0.09 (1,18), p>0.05, $\eta_p^2 = 0.01$). Further, there is no interaction between group, item and time (F=2.82, (1,18) p>0.05, $\eta_p^2 = 0.14$). This indicates no significant change in naming scores either by group or by item with respect to their performance over time.

8.7.2 Analysis Two – Within-Group Analysis

8.7.2.1 Naming Score Changes Over Time for Conflated Participant Groups and Change over Time for Treated and Untreated items

As previously described, data from both the immediate and delayed groups were conflated for the within-group analysis. This analysis of naming scores comprised a two-factor ANOVA with within-participant factors of time (three levels - pre-therapy, immediately post-therapy and 5-week maintenance) and item (two levels - treated items and untreated items). Descriptive statistics are reported in Table 21, above.

Outcomes from a Shapiro-Wilk test indicate a non-normal distribution of treated item naming scores at the pre-therapy time point (W=0.73, p<0.05). The measure of skewness reveals a positive skew (skewness z-scores for T1 > 1.96), illustrated in Figure 562

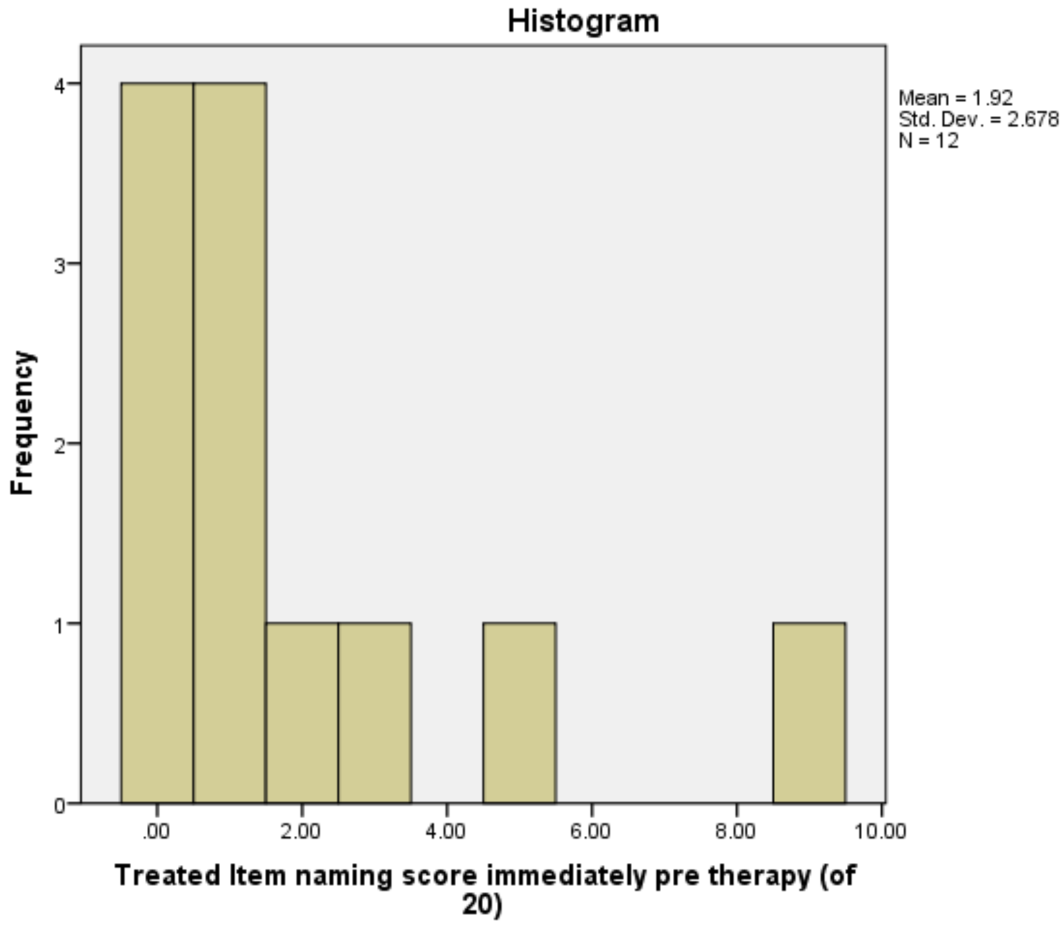


Figure 52. Histogram illustrating the positive skew of treated items naming scores at the pre-therapy time point

To address the non-normal distribution of the data, a log transformation was applied to all pre-therapy, post-therapy and maintenance naming scores. Due to zero scores within the raw data, a value of 1 was added to each data point prior to transformation. Data were then transformed using a logarithmic transformation of base 10. As for the previous data, this transformation method was chosen to compress higher scores and hence adjust for the positive skew of the raw data. Within-group analysis for naming data was then completed using this transformed data.

The two-factor ANOVA analysis reveals a main effect of item (F=7.07 (1, 18), $p<0.05$, $\eta_p^2 = 0.28$), but no effect of time (F=3.06 (2, 36), $p>0.05$, $\eta_p^2 = 0.15$). Analysis also reveals an interaction between time and item (F=3.63 (2, 36),

$p<0.05$, $\eta_p^2=0.17$). Partial eta-squared values indicate a large effect size for both significant findings.

Figure 53 illustrates the difference in performance over time for the treated and untreated items.

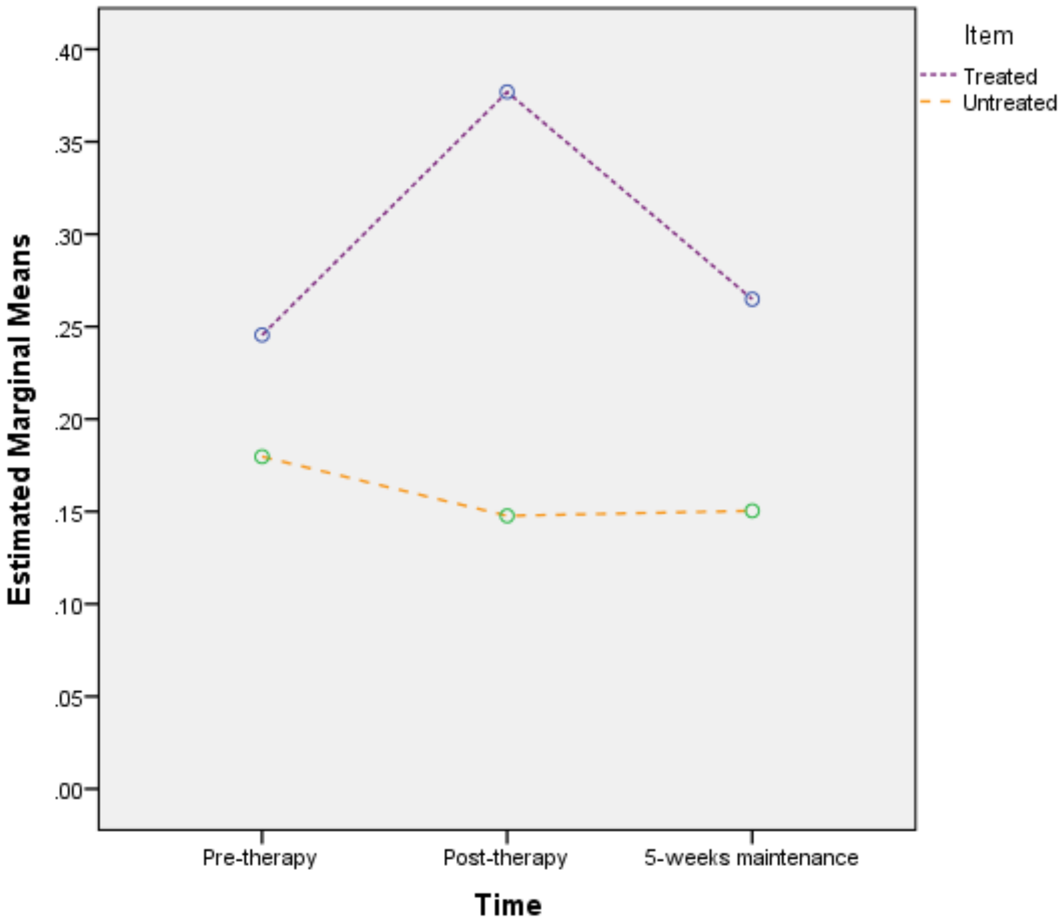


Figure 53. Mean naming scores pre-therapy, post-therapy and at maintenance for treated and untreated items (n=19)

Bonferroni post hoc tests for pooled treated and untreated items indicated no significant difference between pre and post-therapy scores ($p>0.05$), pre-therapy and 5-week maintenance scores ($p>0.05$) or post-therapy and maintenance scores ($p>0.05$). This is due to no main effect of time.

8.8 Research Question 2b

Can people with severe aphasia use learned gestures within interactive communication?

This question was addressed using the Interactive Gesture Assessment (IGA). Participants completed the IGA at four time points. At each time point, participants were awarded a score out of 24 points. Two points were awarded for each item correctly identified by the participant's communication partner, and one point was awarded for each response deemed to be related to that target by a second marker (see section 7.11.2.3). Test stimuli for each IGA consisted of six treated and six untreated items, contributing to a maximum possible score of 12 for both treated and untreated items.

8.8.1 Participant numbers

To complete the IGA, it was necessary for both the participant and their key communication partner to be present for assessment. Unfortunately, not all communication partners were able to attend all assessment sessions and for this reason data for some participants is incomplete. Analyses were carried out for all full data sets allowing analysis of outcomes for 16 participants only. This reduces the power of the analysis and may affect the measurable observation of any effects.

8.8.2 Analysis One – Between Group Analysis

8.8.2.1 Between Group Differences on measures of Interactive Gesture at T1 and T2

This analysis looked at change in IGA scores between Time 1 and Time 2 and between participants in the immediate and delayed therapy groups. Data for 11 of 12 participants from the immediate group and 5 of 8 participants from the delayed group were included in the analysis[3].

Outcomes from Shapiro-Wilk tests of this data indicate a normal distribution of treated and untreated IGA scores for both groups at time points 1 and 2. As this satisfies the assumption of normality, original data was used for the subsequent analysis.

Scores out of 24 were contrasted using a three-factor mixed ANOVA with the within-participant factors of time (time point one versus time point two) and item (treated versus untreated items) and a between-participant factor of group (delayed versus immediate).

Descriptive statistics (Table 20) reveal a mean score of 11.18 out of 24 at T1, reducing slightly to 11.00 at T2 for the immediate therapy group. The delayed therapy group show a mean score of 8.00 out of 24 at T1, rising to 12.8 at T2. Although the groups show different scores at T1, t-tests reveal that the difference between groups is not significant.

The three-factor mixed ANOVA analysis demonstrates no main effect of time ($F=1.53$ (1,14), $p>0.05$, $\eta_p^2 = 0.10$), item ($F=2.84$ (1,14), $p>0.05$, $\eta_p^2 = 0.17$), or group ($F=0.03$ (1,14), $p>0.05$, $\eta_p^2 = 0.00$). Partial eta-squared values however indicate a medium effect size for time, a large effect size for item and no measurable effect size for group. There are no interactions between group and time ($F=1.85$ (1,14) $p>0.05$), group and item ($F=1.09$ (1,18) $p>0.05$) or time and item ($F=0.19$ (1,14), $p>0.05$), however partial eta squared values ($\eta_p^2 = 0.12$; $\eta_p^2 = 0.07$; $\eta_p^2 = 0.01$) indicate a medium, large

[3] **As previously noted, not all participants completed this assessment at each time point – due to the absence of their communication partner from one or more of the assessment sessions**

and small effect size for these comparisons respectively. Further, there is no interaction between group, item and time (F=0.04, (1,14) p>0.05) and partial eta squared (η_p^2 = 0.00) indicates no measurable effect size.

ANOVA analysis revealed no evidence of therapy gain for the immediate therapy group. Descriptive data suggest a positive trend over time for the delayed group (who did not receive GeST + PowerGeST therapy between T1 and T2). This trend, however, is non-significant.

8.8.3 Analysis Two – Within Group Analysis

8.8.3.1 Interactive Gesture Score Changes Over Time for Conflated Participant Groups and Change over Time for Treated and Untreated items

As for the previously reported assessments, data from both the immediate and delayed groups were conflated for the within group analysis. This analysis of IGA scores comprised a two factor ANOVA with within-participant factors of time (three levels - pre-therapy, immediately post-therapy and 5-week maintenance) and item (two levels - treated items and untreated items). Descriptive statistics are reported in Table 21, above.

Outcomes from Shapiro-Wilk tests of this data indicate a normal distribution of treated and untreated IGA scores for all three time points. Therefore, original data was used for the subsequent analysis.

Outcomes reveal a main effect of time (F=4.31 (2, 30), p<0.05, η_p^2 = 0.22) and of item (F=90.09 (1, 15), p<0.01, η_p^2 = 0.57). Partial eta-squared values indicate a large effect size for both time and item. There was no interaction between time and item (F=2.58 (2, 30), p>0.05) however a large effect size is indicated (η_p^2 = 0.22).

Figure 54 illustrates the difference in performance over time for the treated and untreated items.

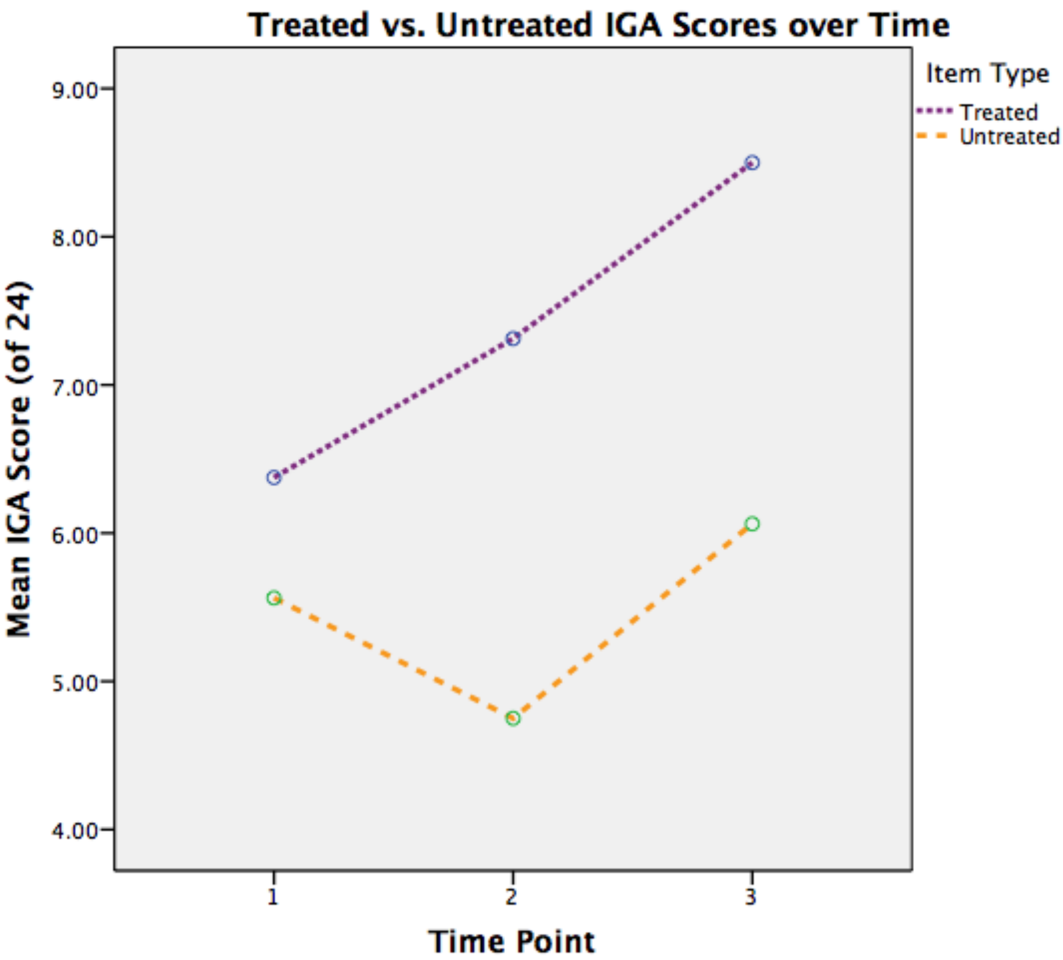


Figure 54. Mean IGA scores pre-therapy, post-therapy and at maintenance for treated and untreated items (n=16)

This graph illustrates the significant change in performance over time and the significant difference between performances for treated vs. untreated items. However, no interaction between time and item group can be observed. Participants' ability to convey gesture in an interactive situation was significantly improved at 5-weeks maintenance. Both treated and untreated items began at a comparable point (with participants scoring around 5 out of 24). Items in the treated category increased to around 7 after intervention, where items in the untreated category decreased to around 5. After a further

5-week period, scores for treated items increased to 8.5 and untreated items to 6.

Bonferroni post hoc tests indicate a significant difference between pre-therapy and 5-week maintenance scores ($p<0.05$). Thus, the change occurred over the 10-week period including therapy and no therapy. As was the case for the naming scores previously reported, a significant difference between item type is indicated ($p>0.005$) for the IGA outcomes with treated items performing consistently better than untreated items.

8.8.4 Summary – interactive gesture assessment items

For the initial analysis presented here, 16 participants were separated into two groups – those who had received therapy between time points 1 and 2 (n=11) and those who had not (n=5). Neither group demonstrated a significant change in overall IGA scores between these two time points. Specifically, we see no effect of GeST + PowerGeST gesture therapy intervention upon performance within this comparison. Expanding the analysis to compare participants against themselves over time (N = 16), we see an average improvement of approximately three points for items which have been treated between pre-therapy and maintenance measures – although the interaction between item and time was not significant. We see virtually no significant change in items whose gestures have not been treated.

8.9 Research Question 2c

Does access to GeST + PowerGeST affect participants' use of technology and confidence in its use?

Results in this section comprise the outcomes from the technology questionnaire. Pre-intervention levels of technology use are reported first, followed by between- and within-group comparisons.

8.9.1 Levels of technology use reported at T1

Figure 55 reports the levels of technology use reported by participants at time point 1. This demonstrates relatively low levels of technology use across the group for most items tested – with the exception of television and the television remote control. This contrasts with the level of technology use reported by individuals without aphasia in Figure 21.

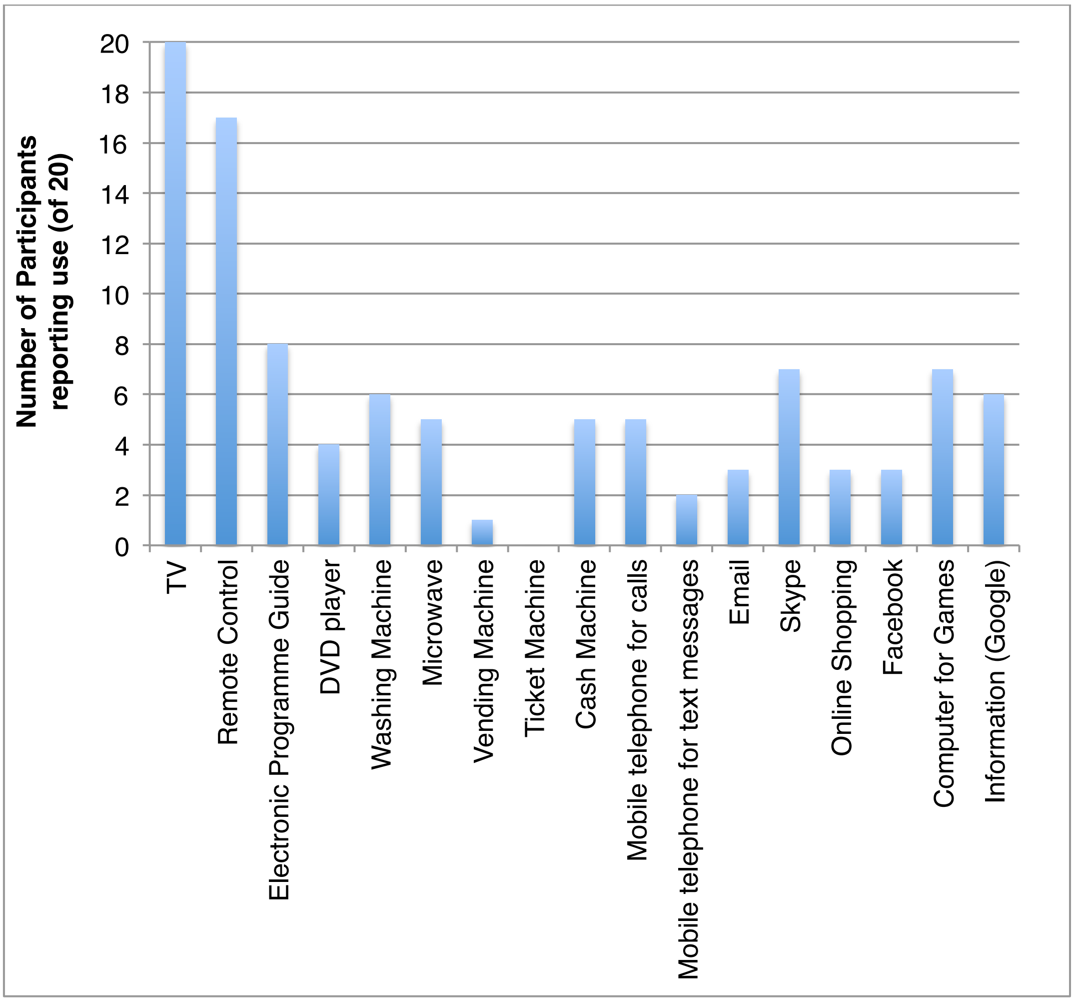


Figure 55. Number of participants reporting use of individual technology items at T1

8.9.2 Analysis One – Between Group Analysis

8.9.2.1 Between Group Differences on Measures of Technology Use and Confidence at T1 and T2

Technology use and confidence scores for both the immediate (n^{use}=12; $n^{confidence}$=12) and delayed (n^{use}=8; $n^{confidence}$=7) therapy groups are summarised in Table 20. Fewer participants completed the confidence score measure than the use score measure. [Those who did not complete the confidence score failed to do so because they demonstrated a difficulty in comprehending the confidence scale used in the confidence score and were therefore unable to engage with this particular measure.] For the immediate therapy group, use scores reveal an average increase of around 0.5 points (0.74 = 4.91 - 4.17) after therapy. Over the same period, use scores for the delayed therapy group shows an increase of around 1 point (1.13 = 7.63 – 6.50). Confidence scores for the immediate group decrease by around 3.5 points (-3.55 = 32.83 – 36.38) and for the delayed group increase by around 7.75 points (7.69 = 48.07 – 40.38).

Technology use and confidence score data were analysed using a two-factor mixed-methods ANOVA with a within-participant factor of time (time point one versus time point two) and a between-participant factor of group (delayed versus immediate).

Outcomes from a Shapiro-Wilk test indicate a non-normal distribution of technology use scores in the immediate group at time point 1 ($W=0.63$, $p<0.05$). Measures of skewness reveal a positive skew (skewness z-scores for T1 > 1.96).

To address the non-normal distribution of the data, a log transformation was applied to all T1 and T2 technology scores. No zero scores were present within the raw data, therefore it was not necessary to add a value of 1 to each data point prior to transformation. Data were transformed using a logarithmic transformation of base 10. As for the naming and gesture data, this transformation method was chosen to compress higher scores and hence adjust for the positive skew of the raw data. As the data now conformed to the assumption of normality, between-group analysis for technology use was subsequently completed using this transformed data.

The mixed-methods ANOVA analysis for technology use scores demonstrates no effect of time (F=3.16 (1,18), p>0.05, η_p^2 = 0.15), or group (F=0.94 (1,18), p>0.05, η_p^2 = 0.05). Furthermore, there are no interactions between group and time (F=0.06 (1,18) p>0.05, η_p^2 = 0.00).

Outcomes from a Shapiro-Wilk test indicate a normal distribution of technology confidence scores for both groups and at both time points. As this satisfies the assumption of normality, original data were used for the technology confidence analysis.

ANOVA analysis for technology confidence scores demonstrates no effect of time (F=0.29 (1,17), p>0.05), or group (F=0.95 (1,17), p>0.05). Partial eta-squared values indicate a small effect size for time (η_p^2 = 0.02) and group (η_p^2 = 0.05). There are no interactions between group and time (F= 2.07 (1,17) p>0.05) although a medium effect size is observed (η_p^2 = 0.11).

8.9.3 Analysis Two – Within Group Analysis

8.9.3.1 Technology Use and Confidence Score Changes Over Time for Conflated Participant Groups

As for the previously reported assessments, data from both the immediate and delayed groups were conflated for the within group analysis. This analysis of technology use and confidence scores comprised a one-factor ANOVA with a within-participant factor of time (three levels - pre-therapy, immediately post-therapy and 5-week maintenance). Descriptive statistics are reported in Table 21, above.

Outcomes from a Shapiro-Wilk test indicate a normal distribution of technology use and confidence scores at all three time points. As this satisfies the assumption of normality, original data were therefore used for the subsequent technology use and confidence analyses.

The one-factor ANOVA analysis for technology use reveals no effect of time ($F=2.24$ (2, 36), $p>0.05$). The partial eta-squared value indicates a medium effect size ($\eta_p^2 = 0.11$). Figure 56 illustrates the difference in performance over time.

ANOVA analysis for technology confidence reveals no effect of time ($F=0.14$ (2, 26), $p>0.05$). The partial eta-squared value indicates a medium effect size ($\eta_p^2 = 0.11$). Figure 57 illustrates the difference in performance over time.

These findings suggest that a competency in two purpose-built technologies does not increase use of, or confidence with other, everyday technologies.

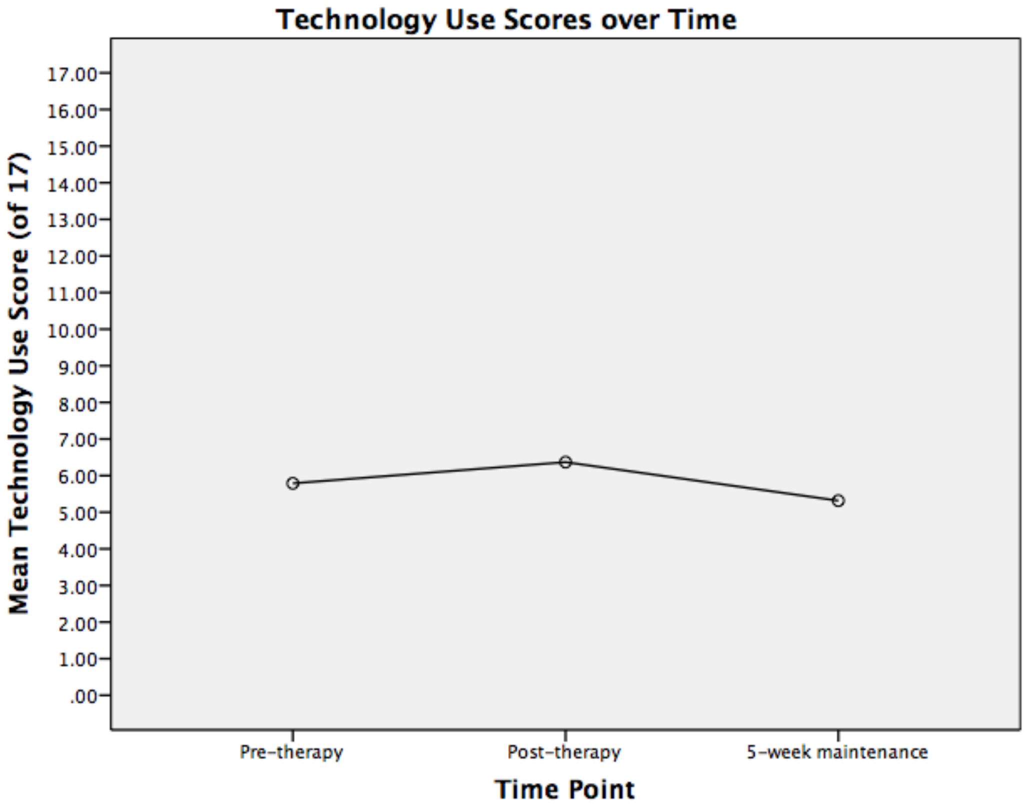


Figure 56. Mean technology use scores pre-therapy, post-therapy and at maintenance for treated and untreated items (n=19)

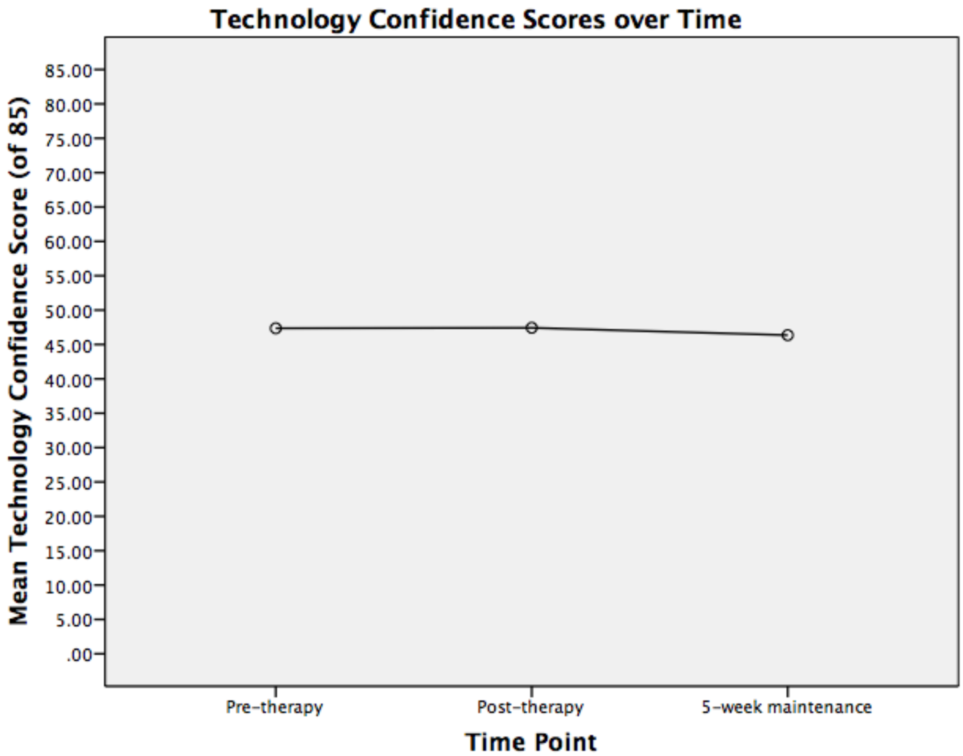


Figure 57. Mean technology confidence scores pre-therapy, post-therapy and at maintenance for treated and untreated items (n=14)

8.9.4 Item-by-item analysis of technology use

Figure 58 presents levels of technology use reported by participants at time point 1 alongside those reported by participants post-intervention (i.e. at T2 for the "immediate" therapy group and T3 for the "delayed" therapy group). Data from one participant (previously reported in figure 59) was excluded from this graph due to missing post-intervention data. Outcomes indicate a slight trend towards increased reports of use for some items. Increased use was reported by three or more people for the following four items: "electronic programme guide", "microwave", "DVD player" and "mobile telephone for call use". Other items either did not vary at all or varied only by an increase or decrease of one or two points following intervention. High levels of use were seen before and after therapy for the items "television" and "remote control" (16-19 participants out of 19 reporting use). The item "ticket machine" received no reported use either before or after intervention.

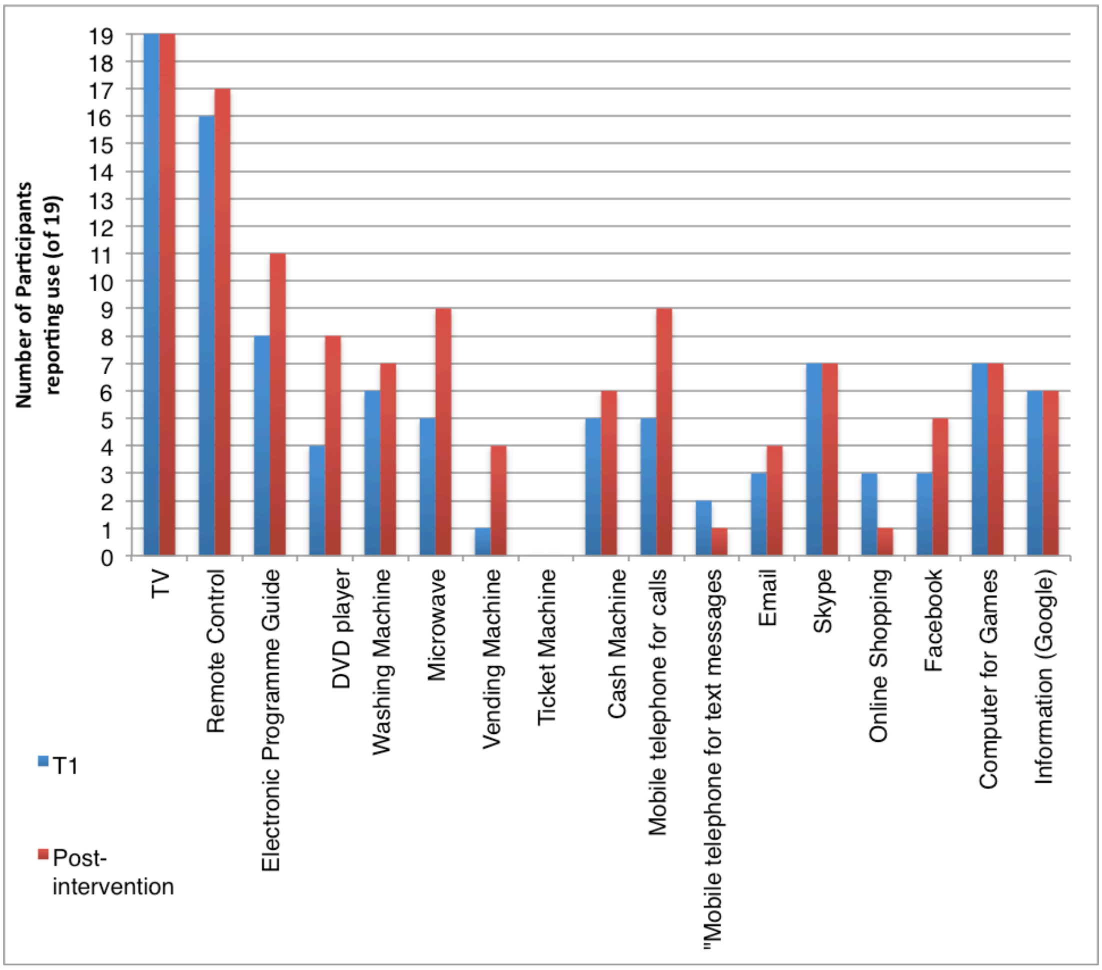


Figure 58. Number of participants reporting use of individual technology items at T1 and then post-intervention (n=19)

Additional Analyses

8.10 Research Question 3b

Are therapeutic gains maintained after therapy has been withdrawn for 10 weeks?

All participants in this study undertook a follow-up assessment 5 weeks after therapy cessation. Those in the immediate therapy group additionally undertook a further follow up assessment 10 weeks after therapy cessation. This additional measure allows us to examine the longer-term effect of GeST + PowerGeST for the participants in the immediate therapy group.

Table 22 presents mean assessment scores for this group at T1 (pre-therapy) and T4 (10 weeks post-therapy-cessation).

Table 22. Mean assessment scores for immediate therapy group at time points 1 & 4

Assessment		T1 Score (SD)	T4 Score (SD)
Gesture Assessment	Treated Items (Max score = 40; n=11)	7.36 (5.73)	9.82 (7.39)
	Untreated Items (Max score = 40; n=11)	5.45 (3.93)	5.91 (4.66)
Naming Assessment	Treated Items (Max score = 40; n=11)	2.00 (2.86)	2.36 (3.26)
	Untreated Items (Max score = 40; n=11)	0.91 (1.58)	1.36 (1.86)
Interactive Gesture Assessment	Treated Items (Max score = 12; n=11)	6.45 (3.08)	8.82 (3.12)
	Untreated Items (Max score = 12; n=11)	5.73 (3.10)	5.91 (3.73)
Technology Use & Confidence	Technology Use (Max score = 17; n=11)	4.00 (3.01)	3.90 (2.59)
	Technology Confidence (Max score = 85; n=9)	43.06 (13.68)	41.39 (13.30)

It should be noted here that T1 figures reported in Table 22 differ from those in Table 20. Pairwise comparisons have only been calculated for participants who completed *both* T1 and T4 assessments. Due to the reduced number of participants who completed the final assessment measure T4, the T1 values reflect an n-value of between nine and 11. These figures were compared using paired samples t-tests.

Paired samples t-tests revealed no significant differences between scores for any assessment at T1 versus scores at T4 ($p > 0.05$) with the exception of the treated items within the interactive gesture assessment (t(10)=-2.47, p < 0.05). This indicates no significant change in performance for the primary measure of gesture or the secondary naming or technology assessments between pre-intervention and 10 weeks post-intervention for the immediate therapy group. A significant improvement is observed for treated items in the interactive gesture assessment across this time period.

8.11 Research Questions 3c and 3d

Can we identify clinically relevant prognostic indicators for those who might be good candidates for GeST + PowerGeST therapy? Specifically in relation to executive function, praxis and language.

Is there a relationship between therapy intensity and the size of therapy effect?

Research questions 3c and 3d were addressed by conducting correlational analyses. These aimed to establish whether gains achieved for the primary and/or secondary outcome measures during the intervention period were related to the candidacy measures conducted with participants, demographic factors or time spent using GeST. Gain scores for treated items were calculated by subtracting post-therapy treated item scores from pre-therapy treated item scores for the naming, gesture and IGA assessments. Gain scores for technology use and confidence were calculated by subtracting total pre-therapy use and confidence scores from post-therapy use and confidence scores for each participant. These gain scores were then examined for correlation in relation to one another and also to background measures. Results of these correlations reported below in Table 23.

Pearson's bivariate two-tailed correlation reveals no significant relationship between change scores for primary or secondary measure assessments and any of the stated candidacy measures or measures of age, time post-stroke or time spent practising with GeST. The data from these measures cannot, therefore, be used to predict which individuals were most and least likely to benefit from GeST + PowerGeST intervention.

The following background measures were however found to correlate significantly with one another:

A. CLQT Visuospatial Cognitive Domain Score.

WITH

- CAT spoken word comprehension score
- CAT spoken sentence comprehension score
- CAT written word comprehension score
- BUPS praxis score
- Gesture Comprehension score

B. CAT Spoken word comprehension score

WITH

- CAT spoken sentence comprehension score
- CAT written word comprehension score
- BUPS praxis score
- Gesture comprehension score

C. CAT spoken sentence comprehension score

WITH

- Gesture comprehension score

D. BUPS Praxis Score

WITH

- CAT written word comprehension score
- Gesture comprehension score

Significance p-values for all correlational analyses are reported, in **bold**, in Table 23.

Table 23. Pearson correlations with significance for primary and secondary outcome change scores, demographic information and candidacy measures. ** p < 0.01, * p < 0.05 (2-tailed)

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Age	.														
2. Time post-stroke in months	-0.01	.													
3. Time spent practising with GeST (hrs:min)	-0.18	-0.01	.												
4. CLQT visuospatial cognitive domain score	-0.28	-0.26	-0.30	.											
5. CAT Naming score	0.10	-0.12	0.06	0.22	.										
6. CAT Spoken word comprehension score	-0.40	-0.08	-0.12	**.59****	0.34	.									
7. CAT Spoken sentence comprehension score	-0.13	-0.25	0.21	**.54***	0.28	**.52***	.								
8. CAT written word comprehension score	-0.03	-0.12	-0.12	**.68****	0.26	**.68****	0.44	.							
9. BUPS Praxis Score	-0.39	0.13	-0.25	**.77****	0.02	**.64****	0.35	**.72****	.						
10. Gesture comprehension score	-0.34	-0.13	0.06	**0.61****	0.22	**0.64****	**0.67****	**0.59****	**0.61****	.					
11. Treated gesture change Score	-0.38	-0.03	0.24	0.15	0.06	0.42	0.25	0.22	0.12	0.08	.				
12. Treated Naming change Score	-0.23	0.20	0.17	0.21	0.12	0.14	0.15	0.02	0.21	0.30	0.11	.			
13. Treated IGA change Score	-0.31	0.00	-0.12	0.26	-0.02	0.19	0.26	0.33	0.25	0.21	0.01	0.23	.		
14. Technology Use change score	0.16	0.13	-0.22	-0.08	0.13	0.11	0.28	0.01	0.08	0.16	0.16	-0.10	0.18	.	
15. Technology Conf. change score	0.07	-0.22	0.07	0.33	0.06	-0.09	-0.16	0.48	0.49	0.29	-0.14	0.08	0.24	0.03	.
Mean	67.53	59.00	0.60	46.65	0.60	11.40	6.35	9.00	20.30	8.35	3.47	1.16	0.76	0.58	-4.00
Standard Deviation	10.09	60.03	0.22	25.02	1.57	2.70	2.23	3.57	9.26	1.79	5.14	2.34	2.70	1.84	18.22
N	19	19	19	20	20	20	20	20	20	20	19	19	17	19	15

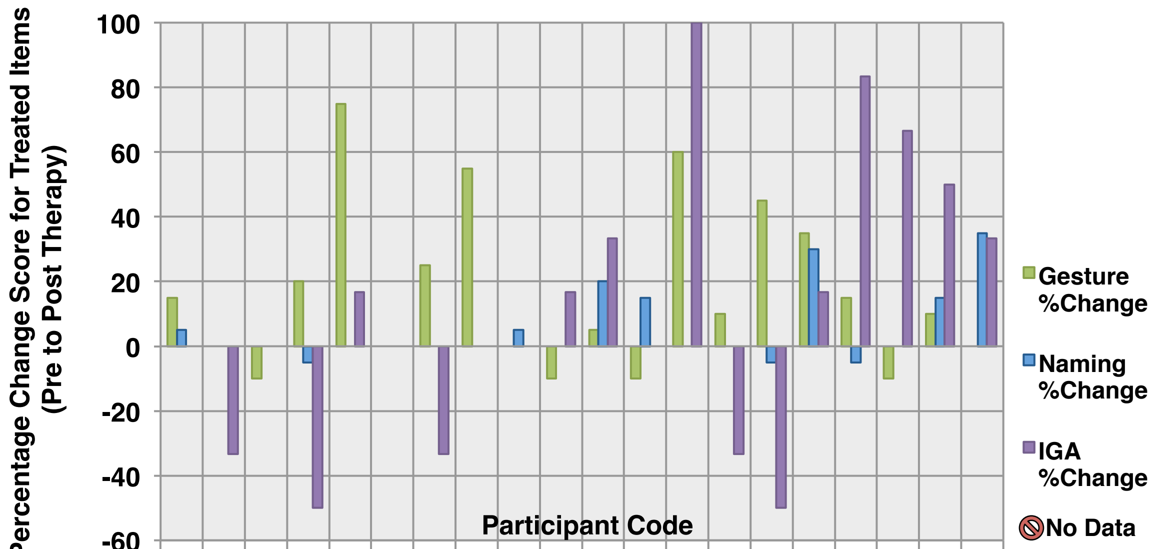
8.12 Research Question 3e

Is there a relationship between gains on assessment of gesture in isolation and gains on assessment of other skills? Specifically, spoken naming and gesture in interactive communication.

Figure 59 shows the percentage change score for treated items within the primary and secondary outcome measures previously reported.

Percentage scores for each assessment were calculated by dividing the change score observed for each participant by the total possible change score for that assessment (20 points for Gesture and Naming and 12 points for IGA) and then multiplying the result by 100.

Bars below the axis show negative changes in performance and bars rising above the axis indicate a positive change in performance. The absence of a bar indicates no data or no change.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
Gesture%Change	15	0	-10	20	75	🚫	25	55	0	-10	5	-10	60	10	45	35	15	-10	10	0
Naming%Change	5	0	0	-5	0	🚫	0	0	5	0	20	15	0	0	-5	30	-5	0	15	35
IGA%Change	0	-33	0	-50	17	🚫	-33	0	🚫	17	33	🚫	100	-33	-50	17	83	67	50	33

Figure 59. Percentage change in treated item score pre to post therapy for all participants

This plot indicates the variability in performance across measures demonstrated by individual participants.

Overall, the data suggest no clear relationship between improvements made in one assessment when compared to improvements (or deterioration) made in another. Similarly, those who make greater gains in their naming demonstrate no consistent association with benefits to gesture performance. Further analyses confirm this, revealing no significant correlations between percentage change scores for primary and secondary outcome measures (Table 24).

Table 24. Pearson correlations for percentage change scores in primary and secondary outcome measures

Variables	15	16	17	18	19
15. Percentage Change in Treated Gesture Score	.				
16. Percentage Change in Treated Naming Score	-0.20	.			
17. Percentage Change in Treated IGA Score	0.01	0.23	.		
18. Percentage Change in Technology Use Score	0.18	-0.10	0.18	.	
19. Percentage Change in Technology Conf Score	-0.33	0.08	0.24	0.03	.
Mean	17.37	5.79	12.75	3.41	-4.71
Standard Deviation	25.68	11.70	45.08	10.80	21.43
N	19.00	19.00	17.00	19.00	15.00

8.12.1 Exploration of factors associated with individual participant profiles - high and low responders

The preceding analyses have enabled us to look at the group as a whole in order to explore research questions 3c and 3e. If we consider the variability outcomes for the group however (as illustrated in figure 63), we can see that there are a number of participants whose treated gesture scores improved by 50% or more in contrast to others whose score decreased slightly (by a margin of 2 points). To further explore the issues of candidacy and the relationship between gains in measures of gesture in isolation, we can examine cross assessment performance for groups of individuals from either end of the improvement scale. Participants E, H and M each improved by a margin of 10 or more points. We can refer to this group as the “high responders”. Participants C, J, L and R, in contrast, each showed a reduction in gesture score of 2 points following treatment. We can refer to this group as the “negative responders.” Table 25 shows individual scores on background measures and other outcomes for the identified participants. Column 11 demonstrates the percentage change in treated gesture scores when comparing scores from immediately before and immediately after therapy. For participants in the immediate therapy condition this reflects T2 minus T1 values. For participants in the delayed therapy condition, this reflects T3 minus T1 values.

Items where figures deviate from the group’s mean score by one or more standard deviation, are shaded in grey and suffixed with “+” or “-“ symbols to illustrate whether the given figure is above or below the group mean. Columns 1-10, 11, 12 and 13 reflect column headings used in table 23. Additional columns 10a, 10b and 10c have been added to further explore the presence of any factors which might relate to either ‘high’ or ‘negative’ response patterns. Aside from the classification criteria (treated gesture change score – column 11), in general, no clear blocks of shading can be observed and hence no indications of a strong and consistent deviation from the mean for either group on any of the reported measures. Indeed, figures appear consistently varied across both groups with little evidence to indicate that any of the factors explored provide a strong predictive value for identifying candidates who are likely respond most or least successfully to the examined intervention. There are two potential exceptions to this lack of pattern. Column 1 shows a trend towards slightly higher than average ages for the non-responders when compare to the high responders. Column 10a shows participant gesture production scores at the outset of treatment being skewed both positively and negatively away from the mean.

Table 25. Individual assessment scores and usage figures for “high” responders and negative responders to treated gestures

	ID	1. Age (yrs)	2. Time post-onset (months)	3. Time spent practising with GeST (hrs:min)	4. CLQT visuospatial cognitive domain score	5. CAT Naming score	6. CAT Spoken word comprehension score	7. CAT Spoken sentence comprehension score	8. CAT written word comprehension score	9. BUPS Praxis Score	10. Gesture comprehension score	10a. T1 treated gesture scores (out of 40)	10b. T1 technology use score (out of 17)	10c. Estimated Time spent in gesture levels 2&3 (hrs:min)	11. Treated gesture percentage change Score	12. Treated naming percentage change Score	13. Treated IGA percentage change Score
	Group MEAN (SD)	**67.83 (10.08)**	**59 (60.03)**	**14:50:40 (5:17:23)**	**46.65 (25.02)**	**0.60 (1.57)**	**11.40 (2.70)**	**6.35 (2.23)**	**9.00 (3.57)**	**20.30 (9.26)**	**8.35 (1.79)**	**7.80 (6.14)**	**5.10 (3.70)**	**07:55:25 (5:12:52)**	**17.37 (25.68)**	**5.79 (11.70)**	**12.75 (45.08)**
High Responders	**E**	58	30	13:08:08	45	0	13	8	6	13	8	7	7	05:52:36	**75++**	0	16.67
	H	73	**230++**	16:23:36	34	0	12	**4-**	11	24	**6-**	2	2	10:07:02	**55+**	0	0
	M	**46 --**	15	17:47:49	**61-**	0	**12--**	6	11	29	8	7	3	**14:42:21+**	**60+**	0	**100+**
Negative Responders	**C**	70	22	**21:38:06+**	13	0	6	5	**4-**	**1--**	**6-**	**0-**	**11+**	**14:13:46+**	**-10-**	0	0
	J	**81+**	13	11:54:21	66	0	13	7	12	25	8	**18+**	**1-**	05:33:20	**-10-**	0	16.67
	L	62	**185++**	13:55:17	25	0	10	5	**5-**	23	9	**14+**	4	**02:11:59-**	**-10-**	15	-
	R	**78+**	55	10:20:49	71	0	10	8	12	27	9	**18+**	**10+**	06:28:15	**-10-**	0	**66.67+**

Key to suffixes for shaded cells

\+ = 1 standard deviation above mean score

++ = 2 standard deviations above mean score

\- = 1 standard deviation below mean score

-- = 2 standard deviations below mean score

--- = 3 standard deviations below mean score

8.12.2 Exploration of Individual Change Scores for Naming

Despite limited naming effects observed at a whole group level, participants involved in this study demonstrated varied outcomes at an individual level. This is illustrated in Figure 63, with five participants (K, L, P, S & T) having made gains of between three and seven items out of 20 (15-35%) immediately post-therapy. As identified in section 2.7, some evidence of gestures being used to cue speech has been found (see Rose, Raymer, Lanyon, & Attard, 2013, for a review) – though typically in therapies which explicitly train gesture as a method of promoting spoken language. Observations made during data collection - that a number of participants undertook spontaneous and unprompted spoken repetition during computer gesture practice – motivated the following post-hoc analysis in relation to the present study of compensatory gesture therapy.

During practice, a number of participants verbally repeated spoken gesture labels, whereas others did not. Measurement of the level of repetition observed was captured from videos taken during the final week of PowerGeST therapy, where each individual was filmed undertaking an unsupported therapy session with PowerGeST both at the outset and the end of their one week's practice. Within the sessions, participants practised gestures for all 20 treated items in sequence. Some participants additionally spoke the name of the item aloud whilst practising the gesture, whilst others practised the gestures in silence. The number of words repeated during each participant's final PowerGeST session was counted to give each participant a score out of 20 for treated word repetition. There is a significant positive correlation between an individual's change score on the naming assessment and their repetition score out of 20 ($r = 0.71$, $n = 19$, $p < 0.05$), as illustrated in Figure 60.

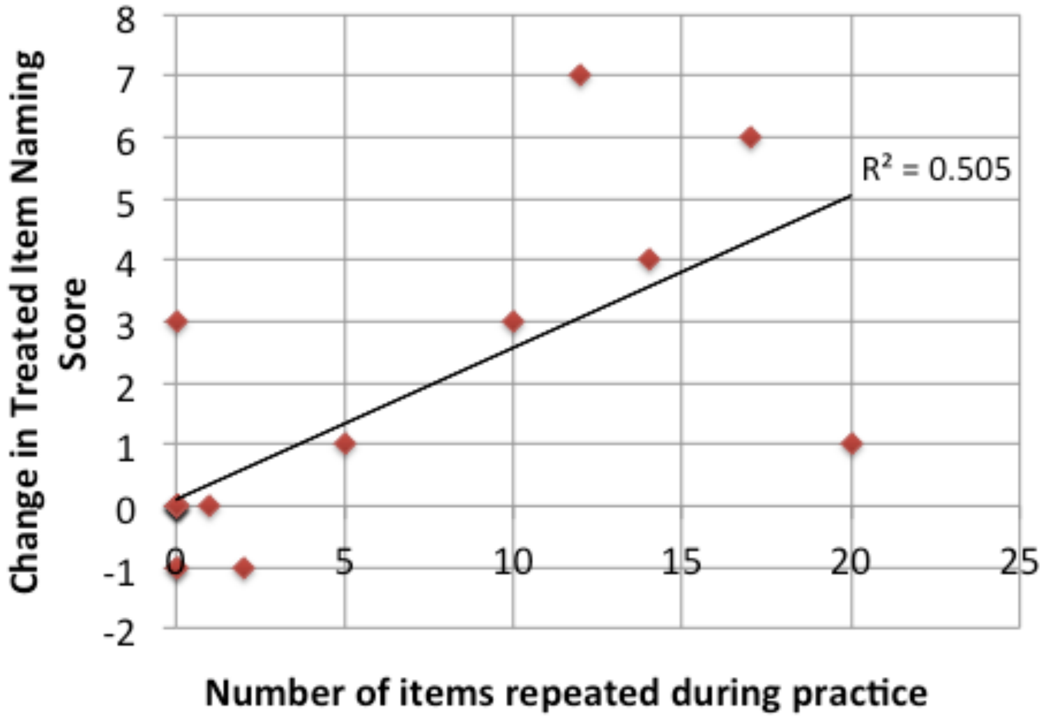


Figure 60. Scatter plot showing the link between change in treated item naming score and the number of items spontaneously repeated during the final, filmed PowerGeST session

8.12.3 Can people with severe aphasia acquire a vocabulary of spoken words using GeST and PowerGeST?

Outcomes from naming analyses (8.7.1.1, 8.7.2.1, 8.12.1) provide no evidence to support a link between the examined gesture therapy and benefits for spoken naming in individuals with severe aphasia. Correlations however, indicate a link between the repetition of items and temporary improvements in naming observed. These improvements are limited only to items that have been treated and do not show any generalisation to untreated items. Furthermore, gains observed during the period immediately after therapy, were not maintained during follow up assessment, a further five weeks later.

8.13 Psychometric Properties of Novel Assessments

In addition to the main research questions of this thesis, data from the current study also provides further opportunity for examination of the psychometric properties of the two novel assessments developed for the purposes of this research – namely the IGA and the Technology Use and Confidence Measure. The presence of repeated measures in the absence of intervention between T1 and T2 for the delayed group, provides an opportunity to examine the test-retest reliability for both measures across this group of eight participants.

8.13.1 Test- retest reliability of the IGA and the Technology Use and Confidence Measure

Due to the study design, test / retest sessions were separated by a period of five weeks. A two-way, mixed method intraclass correlation (ICC) was conducted to compare outcomes between test totals for both measures on both occasions.

8.13.1.1 IGA

A good degree of reliability (>0.7) was found between the T1 and T2 score sets for the total IGA score. The average measure ICC was .716 with a 95% confidence interval from -.497 to .968 ($F(4,4) = 4.44$, $p<.01$). This analysis was limited by the lack of data for three of the eight participants in the group.

8.13.1.2 Technology Use and Confidence Measure

An excellent degree of reliability (>0.9) was found between the T1 and T2 score sets for the technology use score. The average measure ICC was .948 with a 95% confidence interval from .719 to .990 ($F(7,7) = 24.47$, $p<.001$). Technology confidence scores, in contrast, demonstrate very poor reliability (<0.7), with an average measure ICC of .561 with a 95% confidence interval from -1.04 to .93 ($F(5,5) = 2.47$, $p<.17$). This indicates a very poor reliability for the confidence measure of this test over time for those participants included in the analysis. This analysis is limited by the lack of data for two of the eight participants in the group.

8.14 Summary of Results

Results were presented for both primary and secondary measures and additional analyses. Therapy usage information was presented first, followed by an analysis of repeated measure assessments. Analysis of repeated measures focussed firstly on between group differences over time and then on between item differences over time.

Research Question 1a

Does practice with GeST + PowerGeST improve gesture production in isolation?

Analysis revealed an effect of GeST + PowerGeST therapy on gesture. The between group analysis comparing the immediate therapy group with the delayed therapy group showed a critical interaction between time and group, indicating an effect of treatment versus no treatment. When results across 19 participants were collapsed there was further evidence that gesturing of treated items improved after therapy. These gains occurred during the therapy period and were maintained over time.

i. ***Are gains confined to items trained in therapy or do they generalise to untrained items?***

Gains were confined to trained (treated) items only.

Research Question 2a

Does practice with GeST + PowerGeST improve naming?

i. ***Are gains confined to items trained in therapy or do they generalise to untreated items?***

The effects of GeST + PowerGeST therapy on naming were marginal. The between group analysis comparing the immediate therapy group with the delayed therapy group showed no effect of treatment for treated or untreated items. When results across 19 participants were collapsed there was some evidence that naming of treated items improved after therapy. However, the gain was numerically very small and was not maintained. A further analysis revealed a positive correlation between repetition of treatment targets during GeST practice and individual naming gains.

Research Question 2b

Can people with severe aphasia utilise learned gestures within interactive communication?

In contrast to the analysis of the primary gesture assessment, outcomes from these analyses indicate no effect of GeST + PowerGeST therapy upon interactive gesture communication. Analysis 2 revealed a modest but significant improvement in performance of interactive gesture score (1.3 points) over time. This effect was specific to treated items, however, it did not clearly flank the therapy period.

Research Question 2c

Does access to GeST + PowerGeST affect participants' use of technology and confidence in its use?

Neither between-group or within-group analyses provided any evidence of an effect of GeST + PowerGeST therapy upon either technology use or technology confidence. Baseline data indicated comparatively low levels of technology use at the outset of the study (Figure 60) when contrasted against people without aphasia (Figure 21).

Research Question 3a

How much computer practice do participants undertake and at what intensity?

Descriptive data revealed evidence of consistent use of GeST's therapy software throughout the five-week therapy protocol. ANOVA analysis revealed reduced overall levels of use over time for GeST. T-test comparisons of individual session times revealed a significant reduction in session duration for early versus late session use of PowerGeST but not GeST. These findings indicate a reduced number of instances of use over time and reduced session duration over time for PowerGeST but not GeST.

Research Question 3b

Are therapeutic gains maintained in the longer term?

No evidence of change was observed between performance at ten weeks post-therapy and that at pre-therapy for any measure except the interactive gesture assessment. This suggests that gains made within the previous time period were not maintained at 2 ½ months post-therapy for those 12 participants in the immediate therapy group who completed the repeated measures assessments at 10 weeks post intervention. It provides tentative evidence of a belated therapy effect upon interactive gesture.

Research Questions 3c and 3d

Can we identify clinically relevant prognostic indicators for those who might be good candidates for GeST + PowerGeST therapy? Specifically in relation to executive function, praxis and language.

Is and is there a relationship between therapy intensity and the size of the therapy effect?

Response to therapy proved variable, however, analyses provided no evidence of correlation between the examined background measures and changes in either the primary or secondary outcome measures. This indicates no predictive power for the examined factors in identifying participants who will respond most effectively to GeST + PowerGeST therapy.

Research Question 3e

Is there a relationship between gains on one assessment and gains on another?

Data presented here suggest no association between gains made on one assessment in relation to gains made on others. No evidence was found to link improvements in measures of gesture in isolation with improvements in naming or gesture in interactive communication.

Chapter 9. Discussion

9.1 Introduction

This thesis sheds light on the question of whether people with severe aphasia can use a therapy technology, and, if they can, whether they benefit from its use. The nature of these benefits are elucidated and factors relating to gains are explored. The study described aimed to address each of the research questions identified in section 4.4. The following chapter first examines study outcomes for each research question in relation to the previously established evidence for that specific topic. Next, the strengths and limitations of the methods and findings are discussed. Recommendations for future research are then suggested and the implications for clinical intervention are examined. The chapter closes with conclusions from the study and the thesis overall.

9.2 Discussion of Results

9.2.1 Research Question 1a.

Does practice with GeST + PowerGeST improve gesture production in isolation?

Yes. Practice with GeST + PowerGeST does improve performance on a primary outcome measure of gesture ability. Between group analysis (8.6.1) confirmed that this effect occurs for those participants who undertook the therapy intervention between time points one and two and not for those who did not undertake the therapy intervention during this period. Gains were observed for trained items only.

Within-group analysis (8.6.2) revealed that practice with GeST + PowerGeST created a significant improvement in scores across the whole group following therapy and that therapeutic effects were maintained after a five-week period without further therapy. By subtracting pre-therapy totals from maintenance assessment totals we see that participants gained an average of 4.5 points on scores of gesture over time. This equates to a gain of two full gestures or an improvement in the quality of four gestures (precisely correct gestures being awarded two points in scoring and semantically related but not precisely correct gestures being awarded one point). Item-by-item analysis of gesture scores revealed that ten items made gains of five or more points post-intervention ("stamp", "football", "beer", "food", "wine", "dentist", "money", "piano", "hat" and "telephone"). These items could be considered as relatively responsive to intervention when contrasted to the three items whose performance decreased post-intervention ("wife", "walking stick" and "book").

i. Are gains confined to items trained in therapy or do they generalise to untrained items?

Turning to the secondary question regarding the generalisation of gains, descriptive statistics suggested that the gains were almost entirely confined to treated gestures. However, the between-group mixed ANOVA failed to find a three way interaction, between time, group and item. Such an interaction would confirm an item specific effect of therapy. Similarly, no interaction was found between time and item on the within-group analysis. The absence of such interactions may reflect differences in stimuli or could be attributed to the relatively low powered design and the variability of data. An item-by-item analysis of gestures also revealed a generally more consistent pattern of gains for the treated items than for untreated items – in further support of the findings of a therapy effect. Both treated and untreated gesture item sets indicated similar distribution patterns at T1, supporting the notion that the sets were *roughly* comparable at outset - although performance for treated items appeared to be slightly higher than for untreated items. (See Appendix A for details of how treated versus untreated items were selected). This slight difference could be seen to militate against the potential for subsequent comparative gains in this group as the treated items therefore had less room for improvement. The finding of gains in spite of this difference therefore, suggests the presence of a robust and meaningful effect of therapy for the treated items.

Overall, these results suggest that GeST + PowerGeST can have a modest but significant effect on robust external measures of gesture. As acknowledged by Howard, Best, & Nickels in 2015, modest therapy effects can still be seen as significant where it has the potential to make a real difference to a person's life. Participants with severe aphasia in this study scored, on average, one out of 24 on a standardised assessment of naming (CAT naming subtest - Swinburn, Porter, & Howard, 2004) indicating very limited means of verbal expression. A modest gestural gain in this context may represent an important increase in an individual's communicative repertoire. Furthermore, the validity of this effect is underscored by the fact the gain was established through a blinded scoring process where markers were unaware of the time of assessment or the design of the study. Nor had they had any prior interaction with the individual whose gestural abilities they were assessing.

Comparison to existing literature

Gains of two to four communicative items here can be seen to represent a meaningful improvement for a group who have demonstrated limited levels of success in other computer therapy research (Palmer et al., 2012) and who have been omitted from other studies all together (Bartlett, Fink, Schwartz, & Linebarger, 2007; Linebarger, Schwartz, & Kohn, 2001) due to a perceived lack of benefit of the specific therapy tool being tested for those unable to produce single words or short phrases.

Looking at outcomes in relation to other off-line gesture treatments, the outcomes here mirror the successes reported by several authors in the systematic review of gestures conducted by Rose, Raymer, Lanyon, & Attard in 2013 (Code & Gaunt, 1986; Coelho, 1991; Ferguson, Evans, & Raymer, 2012; Hoodin & Thompson, 1983; Marshall et al., 2012; Raymer S. et al., 2011; A. Raymer et al., 2006; A. M. Raymer et al., 2007; Rodriguez, Raymer, & Gonzalez Rothi, 2006), demonstrating improvements for trained gestures. Turning to the outcomes observed for gestures in other computer therapies targeting gesture skills in aphasia, we find just one example for comparison – that of Marshall et al. (2013). This study reported significant gains for items treated by therapist in combination with computer practice with GeST. The present study replicates this finding, demonstrating evidence of a therapy effect for those items trained by computer with some therapist input but no evidence of generalisation to treated items.

In the context described above, the present study provides valuable evidence to support the effects of gesture therapy for training gesture in aphasia. The specific methods used here also indicate the feasibility of implementing such practice using a computer delivered therapy.

9.2.2 Research Question 2a

Does practice with GeST + PowerGeST improve naming?

i. ***Are gains confined to items trained in therapy or do they generalise to untrained items?***

No. Practice with GeST + PowerGeST does not improve performance on the secondary outcome measure of naming ability for either trained or untrained items. The between group analysis (8.7.1) found no evidence of a naming effect for those participants who undertook the therapy intervention between time points one and two compared to those who did not undertake the therapy intervention during this period. Within-group analysis (8.7.2) found no significant effect of GeST + PowerGeST on naming scores across the whole group following therapy or at follow-up, five weeks later.

The main purpose of gesture therapy in this intervention was to support practice of gestures in adults with severely limited expressive output due to aphasia. The rationale for such intervention might be that participants have proven resistant to spontaneous recovery of spoken output and are seeking an alternative means of supporting their communication. As described in section 7.8.4 participants were advised to focus on the production of gesture during practice with no explicit instruction to speak aloud the associated word. There is however, some exposure to the auditory and written form of the word during practice with GeST and PowerGeST. Participants do hear the item name repeatedly over time during practice. With this in mind, whilst no effects were observed for the group overall, additional analyses reported in section 8.12.1 revealed a correlation between change in treated item naming scores and the number of items participants' spontaneously repeated during the final filmed PowerGeST session. This association indicates that those participants who made spontaneous use of verbal repetition during gesture practice demonstrated a trend towards improved naming performance in the short term. There may be a number of explanations for this. One explanation could be that repetition of words during gesture practice actively contributes to an improvement in naming. This could be explored in future research by actively encouraging all participants to engage in naming repetition during gesture practice. Alternatively, as repetition skills are known to vary in aphasia (Baldo, Katseff, & Dronkers, 2012) there is a possibility that those participants who are *more* able to repeat are *more* able to make gains on naming. It could be that those participants in this study who *were* more able to repeat actively adopted this as a strategy. This suggests a possible candidacy consideration for future potential users – i.e. by taking a measure of verbal repetition at the outset, we might predict short term naming gains. We are unable to explore this further within the present data however, as no measure of repetition was undertaken at time point 1.

Comparison to existing literature

Looking at outcomes in relation to other off-line gesture treatments, the overall outcomes here do not replicate the consistent naming successes reported for trained items by several authors in Rose et al.'s (2013) systematic review of gesture therapy (Attard, Rose, & Lanyon, 2013; Boo & Rose, 2011; Crosson et al., 2007; Daumuller & Goldenberg, 2010; Hoodin & Thompson, 1983; Pashek, 1997; A. M. Raymer et al., 2007; Richards, Singletary, Gonzalez-Rothi, Koehler, & Crosson, 2002; Rose M., Douglas, & Matyas, 2002). It should be noted however, that – in contrast to the present study, whose primary focus was gesture intervention - many of the above studies focussed first and foremost on the intervention of spoken outcomes, with gesture typically being used as an additional cue on top of spoken cues such as word repetition or phonological prompting. Marshall et al. (2012) shed further light on this by actively training different sets of words within a naming-only intervention and a gesture-only intervention. Authors found no effect of cross modality generalisation from items trained in the naming-only condition to those trained in the gesture-only condition or vice versa. This is consistent with the present study where gesture practice did not appear to prompt naming gains except for in participants who included an additional verbal repetition component of their own accord.

Turning to the outcomes observed for naming in other computer therapies targeting gesture skills in aphasia, we find just one example for comparison – that of Marshall et al. (2013). This study reported no gains for naming items which received gesture therapy delivered by a therapist in combination with computer practice with GeST. The present study replicates this finding, demonstrating no evidence of a naming therapy effect for those items trained by computer across the group analysis. As observed previously however, evidence of naming gains is observed for some individuals although these gains are not maintained over time.

The findings for naming in this study do not support the use of the described protocol as a direct approach for improving naming skills for individuals with severe aphasia. Future investigations may reveal a benefit for users identified with strengths in repetition, potentially further enhanced by active encouragement to incorporate verbal repetition in practice. This increased verbal component may induce a greater impact on spoken naming gains although without further investigation, it is unclear which factors will contribute to benefits for spoken naming or the persistence of any gains observed.

9.2.3 Research Question 2b

Can people with severe aphasia use learned gestures within interactive communication?

The present study provides mixed evidence to support the notion that practice with GeST + PowerGeST improves performance on a secondary outcome measure of interactive gesture ability. Between group analysis (8.8.2) found no evidence of an improvement in performance on a measure of interactive gesture communication for participants who undertook the therapy intervention between time points one and two compared to those who did not undertake the therapy intervention during this period. Whilst we have previously observed evidence of a therapy effect in measures of single items in isolation (9.2.1), a between group analysis does not see those gains translate into a change in the interactive gesture assessment (IGA) measure. Further insight into the therapy's effects for interactive communication is provided through within-group analysis. Within-group analysis (8.8.3) found a significant improvement for interactive gesture over time across the whole group, although this did not take effect immediately after therapy but instead between pre-therapy and at a maintenance period five weeks after its cessation. This analysis also found that performance for items that were treated was significantly different to performance for those that had not been treated - although the lack of interaction between time and item provides no evidence of a therapy-specific effect. One possible explanation for this lack of interaction could be that the analysis was underpowered and hence failed to pick up the presence of a small but existent effect.

One explanation for the consistent improvement of treated and untreated items across time can be attributed to a flaw in the measure itself. The improvement may illustrate a learning effect within the IGA and not that of a therapeutic gain. Looking to the gradient of change over time for treated items in figure 54 we observe that the rate of this increase appears consistent both in the presence of therapy (pre-therapy versus post-therapy outcomes) and in the absence of therapy post-therapy vs. five-week maintenance outcomes). This consistent trend suggests that the improvement may not be accounted for by the presence of the therapy intervention but instead perhaps by a learning effect of the test process itself. A separate trend we observe is for the consistent advantage for treated items above untreated items at all assessment points – which does not change over time. This suggests that the treated items may simply have been easier than the untreated items. The IGA is a novel measure that has not been fully trialed (for example to examine its internal consistency and test-re-test reliabilty). We must therefore interpret its findings with caution. Further validation of this measure is necessary in order to establish its test retest reliability and to better understand the validity of its outcomes. (Additional discussion of the strengths and limitations of this assessment are presented in section 9.3.3.)

Comparison to existing literature

Putting this limited evidence into context, we can see that it tentatively contrasts with findings from Caute et al., 2013 who demonstrated significant improvement message delivery in their interactive message task – to accompany significant improvements in gesture production in isolation. Why might the gesture training in this instance have effected a measurable change in interactive message delivery? Caute et al.'s study involved individual face-to-face gesture training with a therapist and additional communication strategy training of the communication partner. This contrasts with the computer delivery of gesture of GeST + PowerGeST in the present study. Although a limited amount of face-to-face therapy was provided within the examined protocol (10-15 minutes each week in addition to regular computer practice), this did not entail any training of the communication partner.

An alternative explanation for this lack of translation from gesture production in isolation to interactive gesture success could be that the findings mirror the outcomes commonly reported amongst other aphasia therapies such as naming, where – with the occasional exception (Best et al., 2013) - single word training gains frequently fail to translate into conversational or sentence level gains (Meier, Johnson, Villard, & Kiran, 2017; Nickels, 2002; Woolf et al., 2016).

Whilst results here are limited by the infancy of the measure employed, they do at least begin to contribute to the evidence base regarding the transfer effects of gesture therapy – suggesting that gains in gesture in isolation may not automatically generalise into interactive communication following training with computer practice. As is apparent from the lack of other research available against which to contextualise findings, any outcome regarding this topic is an important contribution to our understanding of the issue.

The IGA was included in this study due to the comparative paucity of other assessments exploring the interactive use of gesture. Its tentative findings may give us an indication that observed treatment effects for gesture are limited to their production in isolation. Alternatively, the measure may not give us any accurate indication of the interactive use of gesture. Whatever the case, findings indicate a need for further research in this area in order to identify both how to effectively capture change in this skill and also to establish which 'active' ingredients in therapy give rise to such a change.

9.2.4 Research Question 2c

Does access to GeST + PowerGeST affect participants' use of technology and confidence in its use?

Not for the group reported here. Practice with GeST + PowerGeST did not promote an observable improved performance on the secondary outcome measure of technology use and confidence. The between group analysis (8.9.2) found no evidence of a change in technology use or confidence for those participants who undertook the therapy intervention between time points one and two compared to those who did not undertake the therapy intervention during this period. Furthermore, within-group analysis (8.9.3) found no significant effect of GeST + PowerGeST on technology use scores or confidence ratings across the whole group following therapy. Item-by-item analysis suggested a small trend towards improved levels of use for the four items "electronic programme guide", "microwave", "DVD player" and "mobile telephone for call use" following intervention. These gains however, did not translate into significant group effects when considered alongside the remaining, comparatively stable, items.

These analyses suggest no meaningful wider impact of GeST + PowerGeST upon levels of technology use or confidence ratings for items of everyday technology. Outcomes are perhaps unsurprising given that there was no targeted training of more general technology access skills within the intervention protocol.

Comparison to existing literature

In relation to previous evidence, the above finding does not support anecdotal reports from participants of Marshall et al.'s 2013 study that the use of GeST led to increased levels of technology engagement for participants with severe aphasia (with the exception of perhaps one or two items). It could be that the increase in technology use and confidence reported by carers within Marshall et al.'s study reflected a shift in their *perceptions* of the technical abilities of participants with aphasia and not an objective change of use as experienced by the participants themselves. Initial measures of technology use (8.9.1) reported low levels of engagement prior to therapy for all participants. This contrasts with the subsequent level of use demonstrated when practising with GeST + PowerGeST (reported in section 8.3). It is encouraging, therefore, that a group experiencing previous technological exclusion can nevertheless practise successfully with a computer tool. The presence of training and support for the use of GeST and PowerGeST may account for its comparative success when contrasted to other technologies, as well as the careful design decisions taken to achieve its accessibility for people with severe aphasia.

Evidence from the current study indicates that, given training and support, purpose-built technology such as GeST and PowerGeST can be accessed successfully by those with severe aphasia, even when access to more mainstream technologies is limited. Findings indicate that, for the small group examined here, the successful use of such a technology however did not significantly impact upon participants' abilities or confidence levels with other devices.

9.2.5 Research Question 3a.

How much computer practice do participants undertake and at what intensity?

Participants utilised two separate therapy tools in sequence within the intervention examined in this thesis. They first completed four weeks of practice with GeST, followed by one week of practice with PowerGeST. Although all were advised to use the computer for around one hour a day (section 7.8.4), participants practised with varying levels of intensity and duration. Using evidence from computer logs, we can see that, on average, participants practised with GeST for around 15 hours across four weeks, with practice sessions being carried out just less than twice a day (an average of 13 sessions per week) for an average of around 17 minutes. We have less data regarding the duration and intensity of PowerGeST use. We cannot be certain of the total time spent practising nor the typical numbers of sessions undertaken per week, however, those data that are available (reported in section 8.3.2) suggest that sessions here lasted between seven and a half and nine minutes on average.

Given the discussions around the level of therapist resource available (Code & Petheram, 2011) versus level of practice required to effect change (Bhogal, Teasell, & Speechley, 2003), we can explore the relative therapist input for participants involved in this study. As described in section 7.8.2, during the four-week GeST intervention, participants within the present study were given 10-15 minutes of offline, one-to-one therapist time each week to familiarise themselves with the gestures that would be practised using GeST computer therapy. This represents around one hour offline therapist input in total. They were additionally given around 15 minutes of assisted computer use per week. This represents around one hour of therapist-supported computer practice input in total. The logged times include the one hour of assisted computer use but do not include the one hour of offline therapy. Therefore, if we add one hour (of offline therapy) to the total time spent using GeST (around 15 hours), this means that two hours of therapist investment resulted in an average of 16 hours of gesture practice per participant over four weeks.

Comparison to existing literature

The 16-hour duration of therapy achieved within this study falls within the observed range of gesture therapy duration reported by Rose et al. in their 2013 review (6.5 to 32 hours), sitting comfortably above the mean value of 11.2 hours. It substantially exceeds the average therapy intensity observed by Rose et al. (2013) achieving an average of 13 sessions per week in contrast to two or three. This indicates that sessions in the present study were carried out 'little (17 minutes) and often (13 times per week)' in contrast to a smaller number of longer-lasting sessions per week.

We can compare these figures further to the duration and intensity of therapy reported in other computer therapies targeting spoken or gesture skills in aphasia by looking to the findings of the systematic literature review conducted within chapter three of this thesis. Here, the 16-hour duration of therapy achieved within this study again just falls within the observed range (14 – 100 hours) of practice as logged by computers, this time sitting well below the mean value of 40 hours. Duration of practice in many studies took place for longer than the five weeks provided with GeST + PowerGeST. For example,four studies (Fridriksson et al., 2009; Linebarger et al., 2001; Marshall et al., 2013; Pedersen, Vinter, & Olsen, 2001) took place over a longer therapy period (ranging from six to 28 weeks).

Taken together, the findings suggest that the present study demonstrated a relatively more intense gesture therapy than is commonly seen in other off-line, gesture therapy research but with a duration at the lower end of that typically observed within other computer therapy research. The presence of significant gesture gains within this context is a positive outcome indicating that comparatively low levels of computer practice delivered at a higher intensity than is typical for offline therapy can offer significant, if modest, improvements in gesture performance for adults with severe aphasia.

9.2.6 Research Question 3b

Are therapeutic gains maintained after therapy has been withdrawn for 10 weeks?

Treated gesture gains achieved within therapy are not maintained beyond five weeks. All participants in this study undertook follow-up assessment five weeks after the cessation of intervention. The within group analyses indicated the maintenance of gains in gesture in isolation at this point and the possible additional achievement of IGA gains. No gains or maintenance of gains were observed for naming or technology measures at this five-week follow-up.

Participants in the immediate group additionally went on to undertake an additional follow-up assessment at ten weeks post intervention. The outcomes from t-tests comparisons of the gesture in isolation assessment suggest that although there was a slight increase in the observed scores for treated gesture items between the pre-therapy assessment and 10-week follow-up (7.36 out of 40 rising to 9.82 out of 40), this difference was not significant – indicating no persistence of the improvements in isolated gesture production. In contrast, the same comparison for scores on the IGA at pre-therapy time point (T1) versus 10-week follow-up (T4) indicates a significant improvement in performance (6.45 out of 12 rising to 8.82 out of 12). Analysis for remaining data indicated no significant change between these two time points for any other assessment. The reduction from significant improvement for isolated gestures at five weeks post-therapy to non-significant improvement at 10-weeks post-therapy is disappointing as it hints that those therapeutic gains achieved in the short term may not persist indefinitely following cessation of therapy input. The rise in IGA score however, is interesting as it may indicate some delayed level of transfer from gestures in isolation into interactive gesture use for those participants in the immediate therapy group. An alternative proposal – as explored in section 9.2.3 is that this comparative gain represents a learning effect for IGA itself – in place of a delayed reaction treatment effect.

Outcomes will now be examined in relation to background literature.

Comparison to existing literature

The finding of a maintenance of gesture gains at five weeks marries with the majority of evidence collated in the systematic literature review of computer therapies, where eight of the eleven studies that examined maintenance effects reported a significant improvement at maintenance assessment when compared to pre-therapy assessment, an average of 7.3 weeks following therapy cessation. Two of the remaining 11 studies reported maintenance effects for some but not all participants, and one (Palmer et al., 2012) reported no evidence of maintenance. Critically, the follow up measure in this final study was taken eight months after therapy cessation – by far the longest period between the end of therapy and maintenance assessment of any of the studies examined. This echoes the finding established in the present study which also found initial evidence to support maintenance of therapeutic (gesture in isolation) gains in the short term but for these to reduce to non-significant levels in the long term. It does not match the tentative outcomes observed for the IGA assessment which indicated evidence of a growing gain in interactive gesture performance over time, however – as previously observed in 9.2.3 – findings from this measure may represent a flawed measure rather than a definite therapy effect.

A separate point to note in the comparison of the present study against existing literature is that the focus of therapy for the majority of other computer interventions was towards spoken production. Marshall et al. (2013) – the pre-cursor to the present study is the only other research that examines computer gesture therapy. Authors here also found that gains were maintained at six weeks post therapy. This aligns with the findings observed in the present study.

Taken together, we see that outcomes here largely mirror those established in previous studies of computer intervention for aphasia. It is evident that the outcomes generated within this thesis, support the findings for maintenance over a period of five weeks, as we observe in other related computer therapy literature. The persistence of their effects beyond this timeframe however, appears limited and in tandem with the findings from Palmer et al, indicates a need for further evidence to examine the long-term maintenance of therapy goals achieved using existing computer intervention protocols. If gains are found to tail off over time, there could be a case for re-introducing computer practice for a periodic 'top up' of learnt skills. The benefit of the computer delivery system in this instance is that individuals wishing to 'top up' their skills should not necessarily require high levels of additional input from a speech and language therapist.

9.2.7 Research Question 3c

Can we identify clinically relevant prognostic indicators for those who might be good candidates for GeST + PowerGeST therapy? Specifically in relation to executive function, praxis and language.

No. Correlational analysis of background measures of cognition, level of praxis and level of aphasia impairment revealed no association with observed gains in performance on gesture, naming or technology measures (section 8.11). We do, however, observe strong internal correlation between individual candidacy measures of cognition, language and praxis. Indeed, the homogeneity observed for these measures may serve to explain the lack of identification of prognostic indicators established to predict future candidacy for GeST + PowerGeST intervention.

Comparison to existing literature

We can look to specific previous studies to examine individual factors of cognition, limb praxis, and language as predictors for success.

Cognition

Beginning with cognition, evidence from Nicholas, Sinotte, & Helm-Estabrooks (2011) indicated that performance on the visuo-spatial assessments within the Cognitive Linguistic Quick Test (CLQT, Helm-Estabrooks, 2001) were positively linked with levels of success in using an augmentative communication device (C-Speak Aphasia) to communicate a message following six months use of the system. The same cognitive assessment tool was employed for the current study, and although a range of performance on this measure was observed, no association between this assessment and the scale of gains achieved for primary or secondary measures was found.

Praxis

Looking to praxis, results from the present study repeat the finding by Marshall et al. (2012) and Caute et al. (2013) in that there was no link between performance on a measure of limb praxis and the size of gains reported following gesture intervention. All studies, including the current one, identified the presence of limb apraxia in many of the participants, yet reported overall gains as a result of the therapy. It seems, therefore, that the presence of limb apraxia does not prohibit improvement on the training of gesture.

In contrast to this outcome, evidence from Hogrefe, Ziegler, Weidinger, & Goldenberg (2012), who conducted a descriptive examination of gesture performance in adults with aphasia (in the absence of any intervention) established that praxis *could* be found to be predictive of gestural comprehensibility (i.e. how easily a participant's gestures could be understood by an external observer). However, their study design did not address the question of therapy gain and so does not directly contrast with the outcome observed here.

Language

Within the previously mentioned Hogrefe et al., study, performance on standard measures of aphasia (The Aachen Aphasia Test, Huber, Poeck, & Willmes, 1984) was *not* found to be predictive of gestural comprehensibility. The present study equally found no evidence of a link between gesture outcomes (or naming outcomes) and subtests taken from a standard measure of aphasia (The Comprehensive Aphasia Test, Swinburn et al., 2004). This should be interpreted with the caveat however, that correlations in Hogrefe et al.'s study used different assessments and examined the static skill of gesture at one point in time and not the ability to develop this gesture through targeted intervention.

The present study has not found any evidence to link the scale of therapeutic outcomes to measures of cognition, praxis or language. This suggests we are unlikely to be able to use the measures of these skills employed here to predict candidacy for GeST + PowerGeST gesture therapy. From a wider perspective, it may mean that these skills are not indicative of an individual participant's ability to benefit from such therapy. An alternative explanation may be that the measures used here have not effectively captured the aspects of cognition, praxis and/or language which can serve to predict such gains. Finally, it may be that the relatively low number of participants included here give rise to a somewhat underpowered analysis and that an increased sample size would reveal more subtle associations which have not been identified here.

9.2.8 Research Question 3d

Is there a relationship between therapy intensity and the size of therapy effect?

Not within the present design. It should be observed however, that this factor was not explored systematically within the current study (i.e. by contrasting two or more identified treatment intensities). Instead, the natural variation amongst duration of self-administered therapy across a

fixed period of four weeks use of GeST, was examined in relation to the scale of any therapy gains observed for each participant.
Correlational analysis of these factors revealed no association between time spent practising and the scale of gains.

Comparison to existing literature

This outcome provides interesting evidence in relation to findings from Bhogal et al., 2003, whose literature review established a positive association between the number of hours practised per week and the presence of a positive therapy effect (described in 2.5.2). Bhogal et al. (2003) found that studies with comparatively greater treatment intensity were linked to significant gains in language outcomes, specifically those with an average of 8.8 hours per week persisting for around 11 weeks. Those studies with lower therapy intensity (average two hours per week for 23 weeks) demonstrated no significant gains. Returning to the present study, due to the lack of logging data for the fifth, PowerGeST week of intervention, we are only able to examine therapy intensity for the initial four weeks when participants used GeST. Here we see an average of 16 hours of intervention – equating to around four hours per week. We can see that this is somewhat closer to the therapy intensity reported for the non-significant group in Bhogal's review. However, outcomes reported for the present study do show significant improvement and – interestingly, also demonstrate a high intensity if we consider not the number of hours but the number of sessions. Participants here undertook an average of 52 sessions across 28 days – that's around 13 sessions per week. For the studies reported by Bhogal et al, the two hours of therapy comprised two hour-long sessions per week. The present model represents nearly two, 17-minute sessions per day every day of the week. Studies reported by Bhogal et al. describe face-to-face therapy. This is typically delivered in sessions lasting around one hour at a time. The present study utilises computer-delivered intervention methods which enable participants to practice independently for shorter, more frequent periods. The current study therefore, appears to support the case for high intensity intervention, with intensity dictated not by the total number of hours per se but rather the total number of independent sessions undertaken.

9.2.9 Research Question 3e

Is there a relationship between gains on assessment of gesture in isolation and gains on assessment of other skills? Specifically, spoken naming and gesture in interactive communication.

No apparent relationships are evident in the presented data. Correlational analyses comparing gesture, naming and interactive gesture outcomes reveal no link between the scale of a gain in one assessment compared to the scale of a gain in another. For instance, treated gesture gains do not predict gains for naming or interactive gesture performance. This is in interesting outcome, suggesting that the stimulation of one skill does not automatically benefit growth in another. In relation to the lack of association between naming and gesture outcomes, it is worth observing that the learning of naming in this situation represents the re-establishment of a skill which participants can reasonably be expected to have had prior to the onset of their aphasia. The learning of pantomime gesture as a key means of expression however, is likely to be a de novo activity – requiring them to newly establish a skill they have not previously used in this capacity. Outcomes from individual participants suggest there might be a link between a comparatively high or low level of gesture performance prior to therapy and a lack of gain on gesture assessment subsequent to intervention – with those four participants who demonstrated a slight reduction in gesture scores after intervention having shown a gesture score that was either well above average or well below average at T1 assessment (section 8.12.1). There is also a slight trend towards those in the non-responder group being older than those in the high responder group. It should be noted however, that these trends are observed only for a very small group. Future studies may wish to explore factors of age and gesture ability at outset with larger numbers of participants in order to identify whether they present a possible prognostic indicator.

Comparison to existing literature

If we look to Rose et al.'s 2013 review of gesture therapies as well as the systematic review of computer therapies presented in chapter 3, we see that this independence of skill development is a common trend. With the exception of Caute et al. (2013) we see little evidence elsewhere of gains in one task transferring to another. Furthermore, the transfer observed in that particular study could well be attributed to the additional training of the participant's communication partner. This finding indicates the importance of identifying the key skill to be targeted in an intervention and ensuring that the individual with aphasia is supported to work towards gains in that skill without necessarily anticipating gains in an associated but non-targeted aspect of communication. For example, the therapy within this project appears to demonstrate a positive improvement for gestural abilities in isolation. There is no clear evidence that the approach confers improvement in other spheres such as naming or interactive gesture. Future interventions for severe aphasia may benefit from a series of targeted activities working in a focused way on independent skills in turn.

9.3 Strengths and limitations of the present study

9.3.1 Specific strengths of this research

The present study addresses an under researched population, and provides evidence into the efficacy of gesture. It provides evidence around technology use in speech and language therapy and evidence into the effect of gesture training on naming. The design of the study, whilst not a full, clinical, randomised control trial does utilise a number of key facets to eliminate bias – for example the blinding of scoring for the primary measure and the quasi-randomisation of participants into immediate and delayed therapy group.

9.3.2 Study design

The validity of the current study is greatly enhanced by the use of blinded scorers in the analysis of the primary gesture outcome measure. This separates it from a number of the other therapy studies reviewed in chapter three (systematic literature review of computer-based intervention). Like many other therapy studies however, it was not double-blinded. The participants knew when they were and were not receiving an active treatment, and the treating therapist was aware of group assignment. As observed by Rose et al. (2013) however, this type of blinding is very difficult to achieve in trials comparing clinical behavioral interventions as opposed to drug interventions and so it is difficult to propose an alternative study design that might offer the ability to achieve this.

Some outcomes in the present study may have benefitted from having a higher number of participants and hence a more highly powered analysis:

- In spite of a lack of differences between the immediate and delayed therapy groups on measures of demographic data and candidacy it should be acknowledged that the power to detect differences between groups was low. Larger participant numbers would enable more confident assertion of the similarity between the immediate and delayed therapy groups and hence increased validity of the findings from between group analyses.
- The format of the IGA led to some data points being lost (as further elucidated in 9.3.4). An increased sample size would overcome this lack of reliability in communication partner attendance.

9.3.3 Gesture and naming outcome measures

Gesture is difficult to assess. Unlike spoken names or formalised signs there are no right and wrong responses, rather success depends on whether or not the meaning of the gesture can be interpreted. The outcome measure adopted to examine gesture production in isolation for this study enabled blinded scoring of this difficult-to-capture skill. The study adapted gesture assessment methods previously used within a study conducted by Marshall et al. (2013). It thus benefitted from a stimuli set and administration process previously established as robust, practical and acceptable. Similarly, the related naming measure enabled the use of comparable stimuli, thus facilitating capture of production abilities in both the gesture and spoken naming modalities.

Turning to the reliability of these measures, whilst excellent inter rater reliability was observed for the secondary outcome naming measure (as reported in section 7.11.1.3; ICC = .681), a slightly lower (but still adequate) level of agreement was observed for the primary gesture outcome measure (section 7.10.1.4; ICC = .907). Given the number of scorers involved in the scoring of the latter measure (42) in comparison to the two who undertook the naming scoring, this outcome is perhaps unsurprising. Indeed, having taken this factor into account, it is encouraging to see an adequate level of agreement. Moreover, it could be argued that a lower level of reliability would militate against the discovery of a therapy effect. However, the significant and systematic response to intervention that was observed for the current study, suggests the presence of a robust finding in spite of the challenges of meaningfully capturing and measuring the skill of communicative gesture production.

9.3.4 Novel assessments developed for this project

Two novel assessment tools were developed for the purposes of this study: the interactive gesture assessment and the technology use and confidence measure. The merits and limitations of both will be examined next.

The Interactive Gesture Assessment (IGA)

The IGA is a novel assessment developed to capture the effects of GeST + PowerGeST upon a participant's ability to successfully convey a message to a regular communication partner. Other limited examples of interactive gesture assessment do exist (Caute et al., 2013; Purdy, Duffy, & Coelho, 1994). However, these have not been standardised and pertain directly to items trained in the interventions described within those studies. Due to the lack of existing tools already available to assess communicative gesture, the IGA was created solely for the purposes of this study. It was piloted briefly as described in section (as described in section 6.3.1) to assess feasibility and found to be practical for the desired purpose and acceptable to a participant with aphasia and his regular communication partner. Examination of its psychometric properties found evidence of face and content validity and test-retest reliability. It should be acknowledged however that the other measures of reliability (such as internal consistency) and validity (such as construct validity and sensitivity to change) were not assessed. This constitutes a shortcoming of the measure, indicating that, whilst outcomes may be indicative of a valid finding, we cannot be entirely assured of the replicability of results or the validity with which they have captured the intended construct – interactive gesture production. It should be noted however, that this same problem exists for other related measures used in alternative research (Caute et al., 2013; Purdy et al., 1994).

A further limiting factor for our interpretation of outcomes from IGA is the number of participants who were able to complete the assessment at all key time points. To complete the IGA, it was necessary for both the participant and their key communication partner to be present for assessment. Unfortunately, not all communication partners were able to attend all assessment sessions and for this reason data sets for some participants are incomplete. Analyses could only be carried out for full data sets – limiting examination of finding to data from 16 participants only. This reduces the power of the analysis and may have further affected the measurable observation of any effects.

It is also important to consider the nature of the task that was used within the IGA. Unlike the primary assessment of gesture in isolation, the IGA comprised an interactive communication task and, as such, participants received real-world assistance from a regular communication partner. Due to this task format, outcomes inevitably reflected a measure of skill for both parties –i.e. the participant AND their communication partner. For this reason, any documented changes may arise from the communication partner as well as, or in place of, the participant. No additional training was provided to the carers - in comparison to the one other study that documents a transfer to interactive communication (Caute et al., 2013)

Overall, the IGA was imperfect. However, interactive gesture is a challenging skill to assess and the IGA, as it stands here, represents a novel, early stage assessment tool that begins to address this. It can be seen as equivalent in validity to previously developed tools investigating the domain of interactive gesture. Within the present study, the IGA specifically aimed to measure whether therapeutic gains in gesture achieved through computer practice can be transferred into real-world interactive communication with a participant's regular communication partner. Further development of this tool will help to establish whether the lack of carry over observed in the present study can be attributed to the absence of an effect or whether it is due to an insufficiently sensitive measure. Future development of this assessment would benefit from a focussed examination of as-yet untested psychometric properties such as internal consistency, construct validity and sensitivity to change.

The Technology Use and Confidence Measure

The technology use and confidence measure is a novel assessment developed to capture the effects of GeST + PowerGeST upon a participant's reported use and confidence with a number of everyday items of technology. Again, due to the lack of existing tools to address this question with adults with severe aphasia, it was created solely for the purposes of this study. (Other tools such as the one developed by Finch & Hill, 2014 were not felt to be appropriate due to their heavy reliance on written information.) The measure demonstrated good test-re-test reliability within the typical population (6.2.3.5.2) and was found to be practical for the desired purpose and acceptable to a participant with aphasia and his regular communication partner. However, subsequent use within the current project revealed that whilst the technology use component was accessible to all participants, some found the associated confidence scale difficult to engage with (only 14 of 19 participants who completed the research protocol completed the confidence scale measure). This not only means that less data was available to examine the effect of GeST + PowerGeST on the construct of confidence in more general technology use, but also that the measure employed may not effectively capture this concept for participants with severe aphasia. Examination of its psychometric properties found evidence of face and content validity and, for an age matched population without aphasia - discriminative validity and test-retest reliability. Additional analyses identified excellent test-retest reliability of technology use scores for participants with severe aphasia but poor test-retest reliability for technology confidence scores. It should be acknowledged however that the other psychometric properties (such as construct validity and sensitivity to change) were not assessed. This constitutes a shortcoming of the measure.

Returning to consider the technology use component of the measure, other factors outside of aphasia may further account for participant's low levels of technology use – for example mobility issues associated with stroke (such as wheelchair use) may prevent access to some of the examined items (e.g. vending machines) and free public access schemes for older adults may exclude the need for use of the item of ticket machine. These factors may indicate a requirement for further refinement of items used within the measure.

Overall, this measure represents a novel, early stage assessment tool that offers us an ability to capture self-reported levels of everyday technology use within the population of adults with severe aphasia. It enables participants to provide their own responses overcoming the reliance upon carer report, which is an alternative sometimes adopted in technology studies to capture information regarding participants with aphasia (as reported in Galliers et al., 2012). The technology confidence component of this measure however, proved less acceptable for this participant group. Future refinement could be made to replace confidence with a more tangible and scalable construct such as amount of use. Furthermore, future development of this assessment would benefit from a focussed examination of as-yet untested psychometric properties such as internal consistency, construct validity and sensitivity to change.

Within the present study, the technology use and confidence measure aimed to capture the effects of GeST + PowerGeST on individuals' use and confidence in the use of everyday technology. The measure was validated with other adults without aphasia and with and without stroke and demonstrates good reliability for these groups. However, further development of this tool will help to establish whether the lack of change observed in the present study can be attributed to the absence of an effect or whether it is due to problems with the measure.

9.3.5 Novel therapy adjunct developed for this project

To extend the practice opportunities available to participants in the present study above those who had taken part in previous pilot research for GeST (Marshall et al., 2013), a novel therapy adjunct – PowerGeST – was created.

PowerGeST

PowerGeST is a standalone tool comprising a software file that presents participants with a sequence of familiar video clips of gesture to repeat. Though it utilises video stimuli from one strand of GeST's therapy exercises, it is different from GeST software in that it does not make use of gesture recognition technology and operates using a flat navigation structure (sections 5.3.4 and 5.4.3). Evidence from video footage of participants practising with PowerGeST indicates that all participants were able to operate the tool autonomously from start to finish using the laptop and external keyboard which had been familiarised during the previous four-week's use of GeST. The utilisation of PowerPoint as a mode of delivery mirrors successful examples of computer practice materials demonstrated in previous naming therapy research (Choe, Azuma, Mathy, Liss, & Edgar, 2007; Choe, Azuma, & Mathy, 2010; Choe & Stanton, 2011; Routhier, Bier, & Macoir, 2016), with all studies demonstrating positive therapy outcomes for participants using such tools.

9.4 Recommendations for future research

Findings discussed above indicate a number of directions for future investigation. Key topics are explored in more detail below.

9.4.1 Interactive gesture assessment

The field of gesture assessment in aphasia would benefit from further development of an assessment tool which can effectively measure gesture in interactive communication. Refinement of the IGA would include assessment of the tool's validity and establishment of norms. Such development would enable us to better examine whether gains achieved for gestures in isolation translate to wider benefits within interactive communication.

9.4.2 Naming

Some limited naming effects were observed within the present study for participants who spontaneously repeated spoken words for the gestures they were practising. Future study iterations using the same therapy protocol might explore the effect of actively encouraging all participants to repeat verbal labels during the practice of gesture items. There is a risk that this might serve to inhibit progress in gesture or naming production however by focussing attention on the known challenges of word-finding for some participants whose verbal skills prohibit them from effective repetition. Within the existing study it is difficult to know whether individual strengths in the ability to repeat words play a part in the ability to make naming gains or instead whether the mere presence of verbal repetition might improve naming for all participants. Future investigations where naming is encouraged might also include a background assessment of individual's naming abilities at the outset of therapy in order to tease these issues apart further.

9.4.3 Identifying the active components of therapy

Using the data from the current study, we are unable to ascertain whether practice with PowerGeST offers a differential effect to that achieved by using GeST alone. Moreover, we are currently unable to establish which components of the explored protocol are those which give rise to therapeutic gains – for example, does the presence of gesture recognition in GeST make a meaningful contribution to its effect? Do any specific 'gaming' aspects of the therapy (such as the point-scoring or virtual world aspects of level 2 of GeST) offer an advantage over alternative types of practice? Some insight into use may be gathered through further analysis of existing log data reporting time spent within individual levels during GeST practice. Beyond this, an experimental comparison of the gains achieved using different software components would shed further light on the unique contribution of specific software designs within computer therapy for aphasia.

9.4.4 Maintenance of effects

A persistence of gesture therapy effects over five weeks is encouraging. However, analysis of outcomes at 10-weeks post therapy suggest that such gains may not be preserved in the absence of further input. As observed in the review of background literature, there is currently very little known about the endurance of computer therapy gains in the longer term. Future research to investigate this in more depth and with a larger number or participants would be a helpful means of informing clinical practice. Should it prove correct that therapy gains taper off after an extended period without intervention, future examination of how much additional practice is required to maintain gains over time could prove invaluable to our understanding of computer therapy service-delivery.

9.5 Clinical implications of findings

9.5.1 Use of technology in therapy

Evidence provided here indicates that severe aphasia limits an individual's access to much, everyday technology. In spite of this, the present study demonstrates a strong case in support of the autonomous use of specialised technology by adults with severe aphasia. The adherence to intervention is particularly meaningful in light of the findings that this group accesses comparatively little technology overall. The development of a technology that adults with severe aphasia can access autonomously represents important progress as a means of providing access to the many aspects of daily living that this group might not otherwise have access to. However, the finding established herein, that this success does not impact upon participants' more general technology use - suggests that many existing technologies presently remain inaccessible to adults with severe aphasia.

9.5.2 Use of technology by adults with severe aphasia

Future work will be required to further develop our understanding of the specific barriers posed by mainstream technologies and the steps that can be taken to overcome them – potentially some of those adopted to allow competency in the purpose-built technologies described within this thesis. The ultimate aim for such research will be to open up access to mainstream technologies so that individuals with severe aphasia may use them with equal or greater success than has presently been achieved for GeST and PowerGeST. The outcomes from this study imply that barriers to technology use may be overcome through the application of considered principles to address language limitations to technology access and through appropriate training and support.

9.5.3 Use of GeST and PowerGeST in wider practice

As outlined in section 9.4 future investigations could help us to better understand the 'active ingredients' of the two therapy tools investigated within the present study. In the absence of this however, we do have evidence of a positive therapy effect upon gesture for the examined protocol. Given the relative lack of alternative computer intervention available there is a case for sharing the established technology within the wider world of clinical practice. Constraints exist regarding the current prototype nature of the tools' software and these would need to be addressed before the tools could be distributed more widely. Additionally, the ability to customise target gesture items for individual users would be a desirable development in order to ensure that the modest gains which can be achieved are of maximum benefit to the individuals with aphasia who have worked to achieve them. Nonetheless, an addition to the toolkit of computer therapy intervention which can benefit the comparatively under-resourced population of people with severe aphasia is a crucial target to which this research should aspire.

9.6 Conclusion

This thesis examined the effects of computer-delivered gesture therapy for adults with severe aphasia. Outcomes indicated that, in spite of comparatively limited access to other forms of technology, participants here were able to effectively and autonomously utilise a purpose-built computer gesture therapy technology. Moreover, they demonstrated a significant gain in their production of treated gesture items. Gains were not observed for items which were untreated or for spoken items, however, there is some indication that naming for specific individuals was improved following self-instigated verbal repetition during gesture practice. Additionally, although gesture gains were maintained in the short-term, there is less evidence of their persistence beyond a period of five weeks after therapy. No gains were seen in the measure of interactive gesture although the validity of this measure has not been fully established and so it is unclear whether the lack of gains observed represent a lack of therapy effect or simply an unreliable measurement tool. Finally, the study established no indicators of prognostic benefit amongst measures of background language, cognition or praxis. Future research will help to establish which specific therapy components may be of benefit as well as details regarding how to preserve therapy effects in the long term. In the meantime, evidence presented within this thesis provides an encouraging case for the use of well-designed computer therapy by a typically under-researched population – those with severe aphasia.

References

Ali, M., Lyden, P., & Brady, M. (2015). Aphasia and dysarthria in acute stroke: Recovery and functional outcome. *International Journal of Stroke, 10*(3), 400-406.

Attard, M. C., Rose, M., & Lanyon, L. (2013). The comparative effects of multi-modality aphasia therapy and constraint-induced aphasia therapy-plus for severe chronic broca's aphasia: An in-depth pilot study. *Aphasiology, 27*(1), 80 111. doi:10.1080/02687038.2012.725242

Baldo, J. V., Katseff, S., & Dronkers, N. F. (2012). Brain regions underlying repetition and auditory-verbal short-term memory deficits in aphasia: Evidence from voxel-based lesion symptom mapping. *Aphasiology, 26*(3-4), 338-354.

Bartlett, M. R., Fink, R. B., Schwartz, M. F., & Linebarger, M. (2007). Informativeness ratings of messages created on an AAC processing prosthesis. *Aphasiology, 21*(5), 475-498.

Bek, J., Blades, M., Siegal, M., & Varley, R. (2010). Language and spatial reorientation: Evidence from severe aphasia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 646.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke; a Journal of Cerebral Circulation, 34*(4), 987-993. doi:10.1161/01.STR.0000062343.64383.D0

Bilda, K. (2011). Video-based conversational script training for aphasia: A therapy study. *Aphasiology, 25*(2), 191-201.

Blomert, L., Kean, M., Koster, C., & Schokker, J. (1994). Amsterdam—Nijmegen everyday language test: Construction, reliability and validity. *Aphasiology, 8*(4), 381-407.

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

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie.

Brumfitt, S., & Sheeran, P. (1999a). *VASES: Visual analogue self-esteem scale, user's manual* Winslow.

Brumfitt, S., & Sheeran, P. (1999b). The development and validation of the visual analogue Self-Esteem scale (VASES) 1. *British Journal of Clinical Psychology, 38*(4), 387-400.

Caute, A., Pring, T., Cocks, N., Cruice, M., Best, W., & Marshall, J. (2013). Enhancing communication through gesture and naming therapy. *Journal of Speech, Language, and Hearing Research, 56*(1), 337-351.

Caute, A., Pring, T., Cocks, N., Cruice, M., Best, W., & Marshall, J. (2012). doi:10.1044/1092-4388(2012/11-0232)

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.

Cherney, L. R., Kaye, R. C., & van Vuuren, S. (2014). Acquisition and maintenance of scripts in aphasia: A comparison of two cuing conditions. *American Journal of Speech-Language Pathology, 23*(2), S343-S360.

Choe, Y., Azuma, T., & Mathy, P. (2010). The use of computers and augmentative and alternative communication devices in independent naming practice: Three single-case studies. *Journal of Medical Speech-Language Pathology, 18*(2), 12-27.

Choe, Y., Azuma, T., Mathy, P., Liss, J. M., & Edgar, J. (2007). The effect of home computer practice on naming in individuals with nonfluent aphasia and verbal apraxia. *Journal of Medical Speech-Language Pathology, 15*(4), 407-422.

Choe, Y., & Stanton, K. (2011). The effect of visual cues provided by computerised aphasia treatment. *Aphasiology, 25*(9), 983-997.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern assessment of verbs and sentences (NAVS). *Aphasiology, 26*(10), 1250-1277.

Code, C., & Gaunt, C. (1986). Treating severe speech and limb apraxia in a case of aphasia. *International Journal of Language & Communication Disorders, 21*(1), 11 20. doi:10.3109/13682828609018540

Code, C., & Petheram, B. (2011). Delivering for aphasia. *International Journal of Speech Language Pathology, 13*(1), 3-10. doi:10.3109/17549507.2010.520090

Coelho, C. (1991). Manual sign acquisition and use in two aphasic subjects. *Clinical Aphasiology, 19*, 209-218.

Crosson, B., Fabrizio, K., Singletary, F., Cato, A., Wierenga, C., Parkinson, B., & Rothi, L. J. (2007). Treatment of naming in nonfluent aphasia through manipulation of intention and attention: A phase 1 comparison of two novel treatments. *Journal of International Neuropsychological Society, 13*, 582.

Cubelli, R., Marchetti, C., Boscolo, G., & Della Sala, S. (2000). Cognition in action: Testing a model of limb apraxia. *Brain and Cognition, 44*(2), 144-165.

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (CREATE). *Psychology and Aging, 21*(2), 333.

Daumuller, M., & Goldenberg, G. (2010). Therapy to improve gestural expression in aphasia: A controlled clinical trial. *Clinical Rehabilitation, 24*(1), 55-65. doi:10.1177/0269215509343327

De Luca, R., Calabrò, R. S., Reitano, S., Digangi, G., Bertè, F., Sergi, G., & Bramanti, P. (2014). Should individuals with chronic aphasia be treated with dedicated PC-based training? considerations about a case study. *NeuroRehabilitation, 35*(4), 711-717 7p. doi:10.3233/NRE-141175

De Ruiter, J. (2000). The production of gesture and speech. *Language and Gesture,*

Doesborgh, S., van de Sandt ‐ Koenderman, M., Dippel, D., van Harskamp, F., Koudstaal, P., & Visch ‐ Brink, E. (2004). Cues on request: The efficacy of multicue, a computer program for wordfinding therapy. *Aphasiology, 18*(3), 213-222.

Donkervoort, M., Dekker, J., van den Ende, E., Stehmann-Saris, J. C., & Deelman, B. G. (2000). Prevalence of apraxia among patients with a first left hemisphere stroke in rehabilitation centres and nursing homes. *Clinical Rehabilitation, 14*(2), 130-136.

Doyle, P. J., McNeil, M. R., Mikolic, J. M., Prieto, L., Hula, W. D., Lustig, A. P., . . . Elman, R. J. (2004). The burden of stroke scale (BOSS) provides valid and reliable score estimates of functioning and well-being in stroke survivors with and without communication disorders. *Journal of Clinical Epidemiology, 57*(10), 997-1007.

Druks, J. (2000). Object and action naming battery.

Enderby, P., Pickstone, C., John, A., Fryer, K., Cantrell, A., & Papaioannou, D. (2009). Resource manual for commissioning and planning service for SLCN. *London: Royal College of Speech and Language Therapists,*

Enderby, P., & Petheram, B. (2002). Has aphasia therapy been swallowed up? *Clinical Rehabilitation, 16*(6), 604-608.

Ferguson, N., Evans, K., & Raymer, A. (2012). A comparison of intention and pantomime gesture treatment for noun retrieval in people with aphasia. *American Journal of Speech-Language Pathology, 21*, S126.

Finch, E., & Hill, A. J. (2014). Computer use by people with aphasia: A survey investigation. *Brain Impairment, 15*(02), 107-119.

Fink, R., Brecher, A., Montgomery, M., & Schwartz, M. F. (2001). Moss talk words [computer software manual]. *Philadelphia: Albert Einstein Healthcare Network,*

Fink, R., Brecher, A., Schwartz, M. F., & Robey, R. R. (2002). A computer-implemented protocol for treatment of naming disorders: Evaluation of clinician-guided and partially self-guided instruction. *Aphasiology, 16*(10-11), 1061-1086.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975a). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975b). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Fridriksson, J., Baker, J. M., Whiteside, J., Eoute, D.,Jr, Moser, D., Vesselinov, R., & Rorden, C. (2009). Treating visual speech perception to improve speech production in nonfluent aphasia. *Stroke; a Journal of Cerebral Circulation, 40*(3), 853-858. doi:10.1161/STROKEAHA.108.532499

Galliers, J., Wilson, S., Muscroft, S., Marshall, J., Roper, A., Cocks, N., & Pring, T. (2011). Accessibility of 3D game environments for people with aphasia: An exploratory study. Paper presented at the *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility,* 139-146.

Galliers, J., Wilson, S., Roper, A., Cocks, N., Marshall, J., Muscroft, S., & Pring, T. (2012). Words are not enough: Empowering people with aphasia in the design process. Paper presented at the *Proceedings of the 12th Participatory Design Conference: Research Papers-Volume 1,* 51-60.

Garrett, Z., & Thomas, J. (2006). Systematic reviews and their application to research in speech and language therapy: A response to TR pring's ‘Ask a silly question: Two decades of troublesome trials’(2004). *International Journal of Language & Communication Disorders, 41*(1), 95-105.

Gingrich, L., Hurwitz, R., Lee, J., Carpenter, J., & Cherney, L. (2013). Quantifying naming and oral reading performance in aphasia: The NORLA-6 scale. Paper presented at the *Annual Convention of the American Speech-Language-Hearing Association, Chicago, IL,*

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail making test: Normative values from 287 normal adult controls. *The Italian Journal of Neurological Sciences, 17*(4), 305-309.

Goldenberg, G., Hartmann, K., & Schlott, I. (2003). Defective pantomime of object use in left brain damage: Apraxia or asymbolia? *Neuropsychologia, 41*(12), 1565-1573.

Goldenberg, G., Hermsdorfer, J., Glindemann, R., Rorden, C., & Karnath, H. O. (2007). Pantomime of tool use depends on integrity of left inferior frontal cortex. *Cerebral Cortex (New York, N.Y.: 1991), 17*(12), 2769-2776. doi:bhm004

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders* Lippincott Williams & Wilkins.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*(4), 404.

Grant, D. A., & Berg, E. A. (1993). *Wisconsin card sorting test (WCST).* PAR.
Hammelrath. (2001). *DVL-38: Test de dénomination des verbes lexicaux*

Hedlund, J. L., & Vieweg, B. W. (1979). The hamilton rating scale for depression: A comprehensive review. *Journal of Operational Psychiatry, 10*(2), 149-165.

Helm-Estabrooks, N. (1992). *ADP: Aphasia diagnostic profiles* Pro Ed.

Helm-Estabrooks, N. (2001). *Cognitive linguistic quick test: Examiner's manual* Psychological Corporation.

Helm-Estabrooks, N. (2002). Cognition and aphasia: A discussion and a study.

Hogrefe, K., Ziegler, W., Weidinger, N., & Goldenberg, G. (2012). Non-verbal communication in severe aphasia: Influence of aphasia, apraxia, or semantic processing? *Cortex, 48*(8), 952-962.

Hoodin, R., & Thompson, C. (1983). Facilitation of verbal labeling in adult aphasia by gestural, verbal or verbal plus gestural training. *Clinical aphasiology* (pp. 62)

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology, 29*(5), 526-562. doi:10.1080/02687038.2014.985884

Huber, W., Poeck, K., & Willmes, K. (1984). The aachen aphasia test. *Advances in Neurology, 42*, 291-303.

Jones, C. W., Handler, L., Crowell, K. E., Keil, L. G., Weaver, M. A., & Platts-Mills, T. F. (2013). Non-publication of large randomized clinical trials: Cross sectional analysis. *BMJ (Clinical Research Ed.), 347*, f6104. doi:10.1136/bmj.f6104

Kaplan, E., Goodglass, H., Weintraub, S., Segal, O., & van Loon-Vervoorn, A. (2001). *Boston naming test* Pro-ed.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). The boston naming test, 2nd. *Philadelphia: Lea & Febiger,*

Kelly, H., Brady, M., & Enderby, P. (2010). Speech and language therapy for aphasia following stroke. *Cochrane Systematic Review,*

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication, 25*, 207-227.

Kendon, A. (2000). Language and gesture: Unity or duality. *Language and Gesture,* (2)

Kertesz, A. (1982). *Western aphasia battery test manual* Psychological Corp.

Kurland, J., Wilkins, A. R., & Stokes, P. (2014). iPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatment. *Seminars in Speech and Language, 35*(1), 51-63. doi:10.1055/s-0033-1362991

Kurland, J., Wilkins, A. R., & Stokes, P. (2014) iPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatement. *Seminars in Speech and Language,* 35(1), 51-63. Doi:10.1055/s-0033-1362991

Laganaro, M., Di Pietro, M., & Schnider, A. (2003). Computerised treatment of anomia in chronic and acute aphasia: An exploratory study. *Aphasiology, 17*(8), 709-721.

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

Laine, C., Horton, R., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Godlee, F., . . . Marusic, A. (2007). Clinical trial registration—looking back and moving ahead. *New England Journal of Medicine, 356*(26), 2734-2736.

Lane Fox, M. (2015). **Digital recommendations for the NHS**. Retrieved from https://doteveryone.org.uk/2015/12/08/martha-lane-foxs-digital-recommendations-for-the-nhs/

Layfield, C. A., Ballard, K. J., & Robin, D. A. (2013). Evaluating group therapy for aphasia: what is the evidence? EBP Briefs, 7(5), 1–17. Bloomington, MN: Pearson.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine, 151*(4), W-65-W-94.

Linebarger, M., Schwartz, M., & Kohn, S. (2001). Computer-based training of language production: An exploratory study. *Neuropsychological Rehabilitation, 11*(1), 57-96.

Linebarger, M. C., Schwartz, M. F., Romania, J. R., Kohn, S. E., & Stephens, D. L. (2000). Grammatical encoding in aphasia: Evidence from a "Processing prosthesis". *Brain and Language, 75*(3), 416-427. doi:10.1006/brln.2000.2378

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy, 83*(8), 713-721.

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.

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., & Woll, B. (2004). Aphasia in a user of british sign language: Dissociation between sign and gesture. *Cognitive Neuropsychology, 21*(5), 537-554.

Marshall, J., Best, W., Cocks, N., Cruice, M., Pring, T., Bulcock, G., . . . Matthews, N. (2012). Gesture and naming therapy for people with severe aphasia: A group study. *Journal of Speech, Language, and Hearing Research, 55*(3), 726-738.

Marshall, J., Roper, A., Galliers, J., Wilson, S., Cocks, N., Muscroft, S., & Pring, T. (2013). Computer delivery of gesture therapy for people with severe aphasia. *Aphasiology, 27*(9), 1128-1146.

Marshall, J. (2006). The roles of gesture in aphasia therapy. *Int J Speech Lang Pathol, 8*(2), 110-114. doi:10.1080/14417040600672384

Mayo Foundation for Medical Education and Research. (2015). Aphasia: Definition. Retrieved from http://www.mayoclinic.org/diseases-conditions/aphasia/basics/definition/con-20027061

McCall, D., Virata, T., Linebarger, M. C., & Berndt, R. S. (2009). Integrating technology and targeted treatment to improve narrative production in aphasia: A case study. *Aphasiology, 23*(4), 438-461.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* University of Chicago press.

McNeill, D. (2006). Gesture and communication. *Brown (Ed.),* , 58-66.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P., . . . Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*(8), e1-e37.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269.

Mortley, J., Wade, J., Davies, A., & Enderby, P. (2003). An investigation into the feasibility of remotely monitored computer therapy for people with aphasia. *Advances in Speech Language Pathology, 5*(1), 27-36.

Mortley, J., Wade, J., Enderby, P., & Hughes, A. (2004). Effectiveness of computerised rehabilitation for long-term aphasia: A case series study. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners, 54*(508), 856-857.

Meier, E. L., Johnson, J. P., Villard, S., & Kiran, S. (2017). Does naming therapy make ordering in a restaurant easier? dynamics of co-occurring change in cognitive-linguistic and functional communication skills in aphasia. *American Journal of Speech-Language Pathology,* , 1-15.

Nicholas, M., Sinotte, M. P., & Helm-Estabrooks, N. (2011). C-speak aphasia alternative communication program for people with severe aphasia: Importance of executive functioning and semantic knowledge. *Neuropsychological Rehabilitation, 21*(3), 322-366.

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

Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences, 8*(6), 537-548.

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., . . . Hughes, H. (2012). Computer therapy compared with usual care for people with long-standing aphasia poststroke: A pilot randomized controlled trial. *Stroke; a Journal of Cerebral Circulation, 43*(7), 1904-1911. doi:10.1161/STROKEAHA.112.650671

Pashek, G. (1998). Gestural facilitation of noun and verb retrieval in aphasia: A case study. *Brain and Language, 65*, 177.

Pashek, G. V. (1997). A case study of gesturally cued naming in aphasia: Dominant versus nondominant hand training. *Journal of Communication Disorders, 30*(5), 349 366. doi:10.1016/S0021-9924(96)00079-2

Pedersen, P. M., Vinter, K., & Olsen, T. S. (2001). Improvement of oral naming by unsupervised computerised rehabilitation. *Aphasiology, 15*(2), 151-169.

Petheram, B. (2004). Editorial computers and aphasia: A means of delivery and a delivery of means. *Aphasiology, 18*(3), 187-191.

Porch, B. (1981). *Porch index of communicative ability (PICA).* Palo Alto: Consulting Psychologists Press.

Porteus, S. D. (1959). The maze test and clinical psychology.
Purdy, M., & van Dyke, J. (2011). Multimodal communication training in aphasia: A pilot study. *Journal of Medical Speech-Language Pathology, 19*, 45.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* Prentice Hall Upper Saddle River, NJ.

Purdy, M., Duffy, R. J., & Coelho, C. A. (1994). An investigation of the communicative use of trained symbols following multimodality training. *Clinical Aphasiology, 22*, 345-356.

Purdy, M. (2002). Executive function ability in persons with aphasia. *Aphasiology, 16*(4-6), 549-557.

Purdy, M., & Koch, A. (2006). Prediction of strategy usage by adults with aphasia. *Aphasiology, 20*(02-04), 337-348.

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

Raven, J. (1977). Court jH, raven J. *Manual for Raven's Progressive Matrices and Vocabulary Scales,*

Raymer S., McHose, B., Smith, K., Iman, L., Ambrose, A., & Casselton, C. (2011). Contrasting effects of errorless naming treatment and gestural facilitation for word retrieval in aphasia. *Neuropsychological Rehabilitation, 22*, 235.

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

Raymer, A. M., & Thompson, C. K. (1991). Effects of verbal plus gestural treatment in a patient with aphasia and severe apraxia of speech. In T. E. Prescott (Ed.), *Clinical aphasiology* (pp. 285)

Raymer, A. M., Kohen, F., Blonder, L. X., Douglas, E., Sembrat, J. L., & Rothi, L. J. G. (2007). Effects of gesture and semantic-phonologic treatments for noun retrieval in aphasia. *Brain and Language, 103*(1-2), 219 220. doi:10.1016/j.bandl.2007.07.014

Raymer, A. M., Rowland, L., Haley, M., & Crosson, B. (2002). Nonsymbolic movement training to improve sentence generation in transcortical motor aphasia: A case study. *Aphasiology, 16*(4-6), 493 506. doi:10.1080/02687030244000239

Rehabilitation Institute of Chicago,. (2007). AphasiaScripts [computer software].

Richards, K., Singletary, F., Gonzalez-Rothi, L. J., Koehler, S., & Crosson, B. (2002). Activation of intentional mechanisms through utilization of nonsymbolic movements in aphasia rehabilitation. *Journal of Rehabilitation Research and Development, 39*, 445.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The philadelphia naming test: Scoring and rationale. *Clinical Aphasiology, 24*, 121-134.

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

Rodriguez, A. D., Raymer, A. M., & Gonzalez Rothi, L. J. (2006). Effects of gesture+verbal and semantic-phonologic treatments for verb retrieval in aphasia. *Aphasiology, 20*(2-4), 286 297. doi:10.1080/02687030500474898

Rose M., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology, 16*, 1001.

Rose, M., & Douglas, J. (2008). Treating a semantic word production deficit in aphasia with verbal and gesture methods. *Aphasiology, 22*(1), 1.

Rose, M. L., Raymer, A. M., Lanyon, L. E., & Attard, M. C. (2013a). A systematic review of gesture treatments for post-stroke aphasia. *Aphasiology, 27*(9), 1090-1127.

Rose, M. (2006). The utility of arm and hand gestures in the treatment of aphasia. *International Journal of Speech Language Pathology, 8*(2), 92-109. doi:10.1080/14417040600657948

Rose, M. L., Attard, M. C., Mok, Z., Lanyon, L. E., & Foster, A. M. (2013). Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study. *Aphasiology, 27*(8), 938 971. doi:10.1080/02687038.2013.810329

Rose, M. L., Raymer, A. M., Lanyon, L. E., & Attard, M. C. (2013b). A systematic review of gesture treatments for post-stroke aphasia. *Aphasiology, 27*(9), 1090 1127. doi:10.1080/02687038.2013.805726

Rose, M., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology, 22*(7-8), 691 706. doi:10.1080/02687030701800800

Routhier, S., Bier, N., & Macoir, J. (2016). Smart tablet for smart self-administered treatment of verb anomia: Two single-case studies in aphasia. *Aphasiology, 30*(2-3), 269-289.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440-479.

Schwartz, M. F. (2013). Theoretical analysis of word production deficits in adult aphasia. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences, 369*(1634), 20120390. doi:10.1098/rstb.2012.0390

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences, 298*(1089), 199-209.

Sica, C., Magni, C., Ghisi, M., Altoè, G., Sighinolfi, C., Chiri, L. R., & Franceschini, S. (2008). Uno strumento aggiornato per la misura degli stili di coping: Il coping orientation to the problems experienced-nuova versione italiana (COPE-NVI). *Psicoterapia Cognitiva E Comportamentale, 14*(1), 27-53.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174.

Spinnler, H., & Tognoni, G. (1987). Italian group on the neuropsychological study of ageing: Italian standardization and classification of neuropsychological tests. *Ital J Neurol Sci, 6*(suppl 8), 1-120.

Streiner, D. L., & Norman, G. R. (2008). Health measurement scales: A practical guide to their development and use. *Health measurement scales: A practical guide to their development and use* () Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643.

Swinburn, K., Porter, G., & Howard, D. (2004). *CAT: Comprehensive aphasia test* Psychology Press.

Tate, R., Mcdonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the single-case experimental design (SCED) scale. *Neuropsychological Rehabilitation, 18*(4), 385-401.

The Stroke Association. (2016). *State of the nation: Stroke statistics.* (). https://www.stroke.org.uk/: The Stroke Assocation.

Thompson, C. K., Choy, J. J., Holland, A., & Cole, R. (2010). Sentactics®: Computer-automated treatment of underlying forms. *Aphasiology, 24*(10), 1242-1266.

U.S. Army. (1944). Army individual test battery, 1944: Manual of directions and scoring.

van de Sandt-Koenderman, W. (2011). Aphasia rehabilitation and the role of computer technology: Can we keep up with modern times?

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*(1), 1.

van de Sandt-Koenderman, M. (2004). High‐tech AAC and aphasia: Widening horizons? *Aphasiology, 18*(3), 245-263.

Varley, R. (2011). Rethinking aphasia therapy: A neuroscience perspective. *International Journal of Speech-Language Pathology, 13*(1), 11-20.

Varley, R. A., Klessinger, N. J., Romanowski, C. A., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences of the United States of America, 102*(9), 3519-3524. doi:0407470102

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.

Welsh, M. C., Pennington, B. F., Ozonoff, S., Rouse, B., & McCabe, E. R. (1990). Neuropsychology of early‐treated phenylketonuria: Specific executive function deficits. *Child Development, 61*(6), 1697-1713.

Wenke, R., Lawrie, M., Hobson, T., Comben, W., Romano, M., Ward, E., & Cardell, E. (2014). Feasibility and cost analysis of implementing high intensity aphasia clinics within a sub-acute setting. *International Journal of Speech-Language Pathology, 16*(3), 250-259.

Wernicke, C. (1874). *Der aphasische symptomencomplex: Eine psychologische studie auf anatomischer basis* Cohn.

Wieczorek, R., Huber, W., & Darkow, R. (2011). Tense/aspect category in fluent and nonfluent german aphasia: An experimental training programme for verb production. *Aphasiology, 25*(8), 851-871.

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.

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

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

World Health Organization. (2001). *International classification of functioning, disability and health: ICF.* World Health Organization.

Youmans, G., Holland, A., Muñoz, M., & Bourgeois, M. (2005). Script training and automaticity in two individuals with aphasia. *Aphasiology, 19*(3-5), 435-450.

Zheng, C., Lynch, L., & Taylor, N. (2016). Effect of computer therapy in aphasia: A systematic review. *Aphasiology, 30*(2-3), 211-244.

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language, 39*(1), 14-32.

Acronyms and Abbreviations

Acronym/Abbreviation	Meaning
AAC	Augmentative or Alternative Communication
AAT	Aachen Aphasia Test (Huber, Poeck, & Willmes, 1984)
ABA	Apraxia Battery for Adults (Dabul, 2000)
ADP	Aphasia Diagnostic Profiles (Helm-Estabrooks, 1992)
AM	Attentive Matrices (Spinnler & Tognoni, 1987)
ANELT-A	Amsterdam-Nijmegen Everyday Language Test (Blomert, Kean, Koster, & Schokker, 1994)
ANOVA	Analysis of Variance
AO	Audio Only
AoS	Apraxia of Speech
AV	Audio Visual
BDAE	Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2001)
BNT	Boston Naming Test (Kaplan et al., 2001)
BUPS	Birmingham University Praxis Scale (as cited in Bickerton et al, 2012)
CAT	Comprehensive Aphasia Test (Swinburn et al., 2004)
CG	Clinician guided
CIAT	Constraint Induced Aphasia Therapy
CINAHL	Cumulative Index to Nursing and Allied Health Literature database
CLQT	Cognitive Linguistic Quick Test
CONSORT	Consolidated Standards of Reporting Trials (Schulz, Altman, & Moher, 2010)

Acronym/Abbreviation	Meaning
COPE	Coping Orientation to Problems Experienced – New Italian Version. (Sica et al., 2008)
CVA	Cerebrovascular Accident
DST	Digit Span Test (Orsini et al., 1987)
DVL-38	Test de dénomination des verbes lexicaux. (Hammelrath, 2001)
EBSCO	Elton B. Stephens Co – refers to EBSCO Information Services
EBSCOHOST	EBSCO Information Services database
EF	Executive Function
GeST	Computer Gesture Therapy Tool investigated within this thesis
GReAT	Gesture Recognition in Aphasia Therapy project
HCID	Human Computer Interaction Design
HD	High Definition
HRSD	Hamilton Rating Scale for Depression (Hedlund & Vieweg, 1979)
ICC	Intraclass correlation
ID	Identity Code
IGA	Interactive Gesture Assessment
ILAT	Intensive Language Action Treatment
M	Mean
MEDLINE	Medical Literature Analysis and Retrieval System Online
MMSE	Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975)
NORLA-6	Naming and Oral Reading for Language in Aphasia 6-Point Scale (Gingrich, Hurwitz, Lee,

Acronym/Abbreviation	Meaning
	Carpenter, & Cherney, 2013
NR	Not Reported
NVNT	Noun/Verb naming test (Zingeser & Berndt, 1990)
OS	Operating System
PC	Personal Computer
PEDro-P	Physiotherapy Evidence Database - Rating Scale for Randomised and Non-Randomised Controlled Trials
PICA	Porch Index of Communicative Ability (Porch, 1981)
PICOS	Participants, Interventions, Comparisons, Outcomes, Study Design
Power AFA	Italian “AFASIA” software program developed by Powerwolf Software Solutions (reported in De luca et al 2014)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSG	Partially self-guided
pts	Participants
QPA	Quantitative Production Analysis (Berndt, & Schwartz, 1989)
RCT	Randomised Control Trial
REACT-2	Online speech and language therapy tool providing exercises for aphasia
RH	Right Hand
RML	Reversal Motor Learning, (Giovagnoli et al., 1996)
SCED	Single Case Experimental Design
SD	Standard Deviation

Acronym/Abbreviation	Meaning
SLP	Speech and Language Pathology/Pathologist
SLT	Speech and Language Therapy/Therapist
SPTA	Speech Pathology Therapy Assistant
TBI	Traumatic Brain injury
tDCS	Transcranial Direct Current Stimulation
TMS	Transcranial Magnetic Stimulation
TMT	Trial Making Test A & B (Giovagnoli et al., 1996)
TR-PR	Treated and Practised
TR-UNPR	Treated and Unpractised
TV	Television
UK	United Kingdom
UNTP-UNPR	Untreated an Unpractised
UNTR-PR	Untreated and Practised
URL	Uniform Resource Locator (web page address)
VASES	Visual Analogue Scale of Self-Esteem (Brumfitt & Sheeran, 1999a)
WAB	Western Aphasia Battery (Kertesz, 1982)
WAB AQ	Western Aphasia Battery Aphasia Quotient
WNL	Within Normal Limits

Appendices

Appendix A - Gesture Item Selection

Story #	Random # 1 - 100
4	20
3	26
5	30
1	44
2	90
6	100

Lowest 4 numbers are included in research study, highest 2 are excluded.

Chosen stories are therefore 1, 3, 4, 5
Untreated items are matched by the baseline recognition score from pilot study (max score 9)

Therefore Items as follows

TREATED

Category	Target	Total number correctly identified at T1 of GeST pilot (of 9)
T1	GLASSES	5
T1	TEA	4
T1	TELEPHONE	4
T3	HAT	3
T4	PIANO	3
T1	BOOK	2
T3	MONEY	1
T4	FOOD	1
T4	STAMP	1
T1	WIFE	0
T3	BEER	0
T3	CAR	0
T3	DENTURES	0
T4	WAITER	0
T4	WINE	0
T5	BOY	0
T5	DENTIST	0
T5	FOOTBALL	0
T5	SPIDER	0
T5	WALKING STICK	0

UNTREATED

Category	Target	Total number correctly identified at T1 of GeST pilot (of 9)
UT	HAIR	5
F3	WATCH	4
F1	SWIMMING	4
T6	UMBRELLA	3
UT	DOOR	2
UT	RAIN	2
UT	NEWSPAPER	1
F3	APPLE	1
F1	BANANA	1
T6	TAP	0
T6	BED	0
F3	CAT	0
F2	PEN	0
F2	TISSUE	0
UT	CUP	0
F2	CHESS	0
UT	DOCTOR	0
F1	SPONGE	0
T2	SEWING	0
UT	LETTER	0

Appendix B - Interactive Gesture Assessment Item Choice

Order	Set A	Set B	Set C	Set D
P1	**Hat**	**Food**	**umbrella**	**Iron**
P2	**Book**	**Bed**	**Beer**	**Piano**
1	Iron	**Tap**	book	Scissors
2	Piano	**wife**	spider	**Rainbow**
3	Beer	Iron	Dentures	Money
4	Umbrella	**WINE**	Sewing	**TEA**
5	Dentist	Camera	waiter	**Glasses**
6	**WIFE**	Stamp	Scissors	Sewing
7	Camera	umbrella	**Rainbow**	**Gloves**
8	Food	**Gloves**	**tea**	Boy
9	**Bed**	walking stick	**TAP**	**BED**
10	**Wine**	car	**GLASSES**	Remote Control
11	**Tap**	**RAINBOW**	Remote Control	Hat
12	**GLOVES**	telephone	Bed	Football

Treated items, marked in black (Items in bold occur across two sets)
Set A = Food, Dentist, Beer, Piano, Wine & WIFE
Set B = Stamp, telephone, walking stick, wife, car & WINE
Set C = Dentures, waiter, book, tea, spider & GLASSES
Set D = Football, Hat, Money, Glasses, Boy & TEA

Untreated items, marked in red (Items in bold occur across three sets)
Set A = Iron, Camera, Tap, Umbrella, Bed & GLOVES
Set B = Gloves, Tap, Camera, Iron, umbrella & RAINBOW
Set C = Sewing, Remote Control, Bed, Scissors, Rainbow & TAP
Set D = Gloves, Rainbow, Scissors, Remote Control, Sewing & BED

Appendix C - List of Items Included in the Technology Use and Confidence Assessment

Item Number	Item Name
1	Television
2	Remote Control
3	Electronic Programme Guide
4	DVD player
5	Washing Machine
6	Microwave
7	Vending Machine
8	Ticket Machine
9	Cash Machine
10	Mobile telephone for calls
11	Mobile telephone for text messages
12	Email
13	Skype
14	Online Shopping
15	Facebook
16	Computer for Games
17	Internet for information (Google)

Appendix D – details of participants without aphasia who undertook the Technology Use and Confidence measure (section 6.3.2.6)

Participant Number	Group	Gender	Age	Ethnicity	Occupation	Number of years in education	Highest qualification
1	Healthy	Male	65+	Black British	Skilled trade	16	Aleve/trade/vocational
2	Healthy	Male	55-64	Black British	Skilled trade	12	Alevel/trade/vocational
3	Healthy	Female	55-64	Black British	Professionals/highly skilled/managerial	16	Alevel/trade/vocational
4	Healthy	Female	55-64	White British	Professionals/highly skilled/managerial	17.5	Degree level & above
5	Healthy	Female	55-64	Black British	Professionals/highly skilled/managerial	20	Degree level & above
6	Healthy	Male	55-64	Black British	Skilled trade	.	GCSE/equivalent
7	Healthy	Female	55-64	Black British	Professionals/highly skilled/managerial	24	Degree level & above
8	Healthy	Female	55-64	Black British	Skilled trade	15	Alevel/trade/vocational
9	Healthy	Female	55-64	Black British	Professionals/highly skilled/managerial	14	Alevel/trade/vocational
10	Healthy	Female	65+	Black British	Professionals/highly skilled/managerial	16	Alevel/trade/vocational
11	Healthy	Female	55-64	White British	Professionals/highly skilled/managerial	18	Degree level & above
12	Healthy	Male	55-64	White British	Professionals/highly skilled/managerial	18	Degree level & above
13	Healthy	Male	65+	White British	Professionals/highly skilled/managerial	11	Alevel/trade/vocational

Participant Number	Group	Gender	Age	Ethnicity	Occupation	Number of years in education	Highest qualification
14	Healthy	Male	55-64	White British	Skilled trade	14	Alevel/trade/vocational
15	Healthy	Male	55-64	White British	Professionals/highly skilled/managerial	12	Alevel/trade/vocational
16	Healthy	Male	55-64	White British	Professionals/highly skilled/managerial	16	Degree level & above
17	Healthy	Female	55-64	White British	Professionals/highly skilled/managerial	18	Degree level & above
18	Healthy	Female	65+	White British	Professionals/highly skilled/managerial	13	Alevel/trade/vocational
19	Healthy	Male	55-64	White British	Professionals/highly skilled/managerial	17	Alevel/trade/vocational
20	Healthy	Male	65+	White British	Professionals/highly skilled/managerial	11	GCSE/equivalent
21	Healthy	Male	65+	White British	Professionals/highly skilled/managerial	10	Alevel/trade/vocational
22	Healthy	Female	65+	White British	Skilled trade	14	Degree level & above
23	Healthy	Female	65+	White British	Skilled trade	11	Alevel/trade/vocational
24	Stroke no aphasia	Male	65-74	White British	Skilled trade	11	Alevel/trade/vocational
25	Stroke no aphasia	Male	75-84	White British	Skilled trade	16	Alevel/trade/vocational
26	Stroke no aphasia	Male	65-74	Other	Professionals/highly skilled/managerial	11	Alevel/trade/vocational
27	Stroke no aphasia	Female	75-84	White British	Skilled trade	12	Alevel/trade/vocational
28	Stroke no aphasia	Female	75-84	White British	Skilled trade	9	Alevel/trade/vocational

Participant Number	Group	Gender	Age	Ethnicity	Occupation	Number of years in education	Highest qualification
29	Stroke no aphasia	Female	65-74	White British	Professionals/highly skilled/managerial	13	Alevel/trade/vocational
30	Stroke no aphasia	Male	85 plus	White British	Professionals/highly skilled/managerial	22	Degree level & above
31	Stroke no aphasia	Female	85 plus	White British	Skilled trade	9	Alevel/trade/vocational
32	Stroke no aphasia	Male	55-64	White British	Skilled trade	24	Degree level & above
33	Stroke no aphasia	Female	65-74	White British	Skilled trade	11	GCSE/equivalent
34	Stroke no aphasia	Female	55-64	Other	Professionals/highly skilled/managerial	15	Degree level & above
35	Stroke no aphasia	Female	55-64	White British	Professionals/highly skilled/managerial	18	Degree level & above
36	Stroke no aphasia	Female	55-64	Black British	Skilled trade	10	GCSE/equivalent
37	Stroke no aphasia	Male	65-74	Black British	Professionals/highly skilled/managerial	14	GCSE/equivalent
38	Stroke no aphasia	Male	65-74	White British	Skilled trade	10	Alevel/trade/vocational
39	Stroke no aphasia	Male	55-64	White British	Skilled trade	11	GCSE/equivalent
40	Stroke no aphasia	Male	75-84	White British	Skilled trade	11	GCSE/equivalent
41	Stroke no aphasia	Male	65-74	White British	Skilled trade	11	Alevel/trade/vocational

Appendix E - Letter of Ethical Approval

CITY UNIVERSITY LONDON

School of Health Sciences

Research Office
Northampton Square
London EC1V 0HB

Tel: +44 (0) 20 7040 5704

www.city.ac.uk

Ref: PhD/12-13/10

22 January 2013

Dear Abi / Jane / Stephanie

Re: Computer Delivery of Gesture Therapy for People with Aphasia

Thank you for forwarding amendments and clarifications regarding your project. These have now been reviewed **and approved** by the Chair of the School Research Ethics Committee.

Please find attached, details of the full indemnity cover for your study.

Under the School Research Governance guidelines you are requested to contact myself once the project has been completed, and may be asked to complete a brief progress report six months after registering the project with the School.

If you have any queries please do not hesitate to contact me as below.

Yours sincerely

[redacted]

Alison Welton
Research Governance Officer

[redacted]

CITY UNIVERSITY LONDON

School of Health Sciences

Please note that City University has extensive insurance cover in place for the academic year 2012/2013, relevant details of which currently are:

1. Employers Liability
This is cover for legal liability to employees for death, injury or disease arising out of the business of the University. The limit of indemnity is £50,000,000 for any one claim.

2. Public and Products Liability
This is cover for legal liability to third parties for accidental loss of or damage to property or for death, injury, illness or disease arising out of our business and including liability arising from goods sold or supplied. The limit of indemnity is £50,000,000 for any one claim.

3. Professional Indemnity
This is cover for legal liability to third parties for breach of professional duty due to negligent act, error or omission in the course of our business. The limit of indemnity is £25,000,000 for any one claim.

Clinical trials cover is included within the above insurances in place.

David

David Evans
[redacted]

T: [redacted]
F: [redacted]
E: [redacted]

Appendix F

Case History and Demographic Information

PARTICIPANT : ___________ Date of Interview: ___________

MALE / FEMALE DATE OF BIRTH:

AGE:

WHEN DID YOU HAVE YOUR STROKE?

………………………………………………………………………………

DO YOU HAVE A RECORD OF YOUR BRAIN SCAN THAT YOU ARE HAPPY TO SHARE?

CAN YOU USE BOTH HANDS? YES NO

WHICH HAND IS BEST? LEFT RIGHT

DO YOU HAVE PROBLEMS WITH:

MOVING?………………………………………………………………………

SEEING?………………………………………………………………………

HOLDING THINGS?………………………………………………………………………

HEARING?………………………………………………………………………

OTHER?………………………………………………………………………

Do you live alone?: Yes No

If yes, who is your most regular contact?

How often do they visit?

DO YOU USE A COMPUTER: (B = Before; N = Now)

NEVER	ONCE IN A WHILE	ABOUT ONCE A WEEK	ALMOST EVERY DAY

DO YOU USE A MOBILE PHONE: (B = Before; N = Now)

NEVER	ONCE IN A WHILE	ABOUT ONCE A WEEK	ALMOST EVERY DAY

Languages spoken:

Handedness:
To write with (before stroke):
To play tennis with (before stroke):

Work history:

Appendix G – Participant allocation

Participant group allocation, therapy order allocation, gesture and naming assessment allocation and IGA order allocation.

Participant Code	Group Allocation	Therapy block practice order				Order of Gesture and Naming assessments	Order of Interactive Gesture assessments
		Week 1	Week 2	Week 3	Week 4		
A	Immediate	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
B	Immediate	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
C	Delayed	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	BADC
D	Immediate	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	ABAB	BADC
E	Delayed	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	BABA	ABCD
F	Delayed	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
G	Immediate	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	ABAB	BADC
H	Immediate	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	ABAB	BADC
I	Delayed	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
J	Delayed	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	BABA	ABCD
K	Immediate	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	BADC
L	Immediate	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	BADC
M	Immediate	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
N	Immediate	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	ABCD
O	Immediate	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	BADC
P	Immediate	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	BADC
Q	Immediate	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
Did not continue to study	Delayed	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	ABCD
R	Delayed	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	BADC
Did not continue to study	Delayed	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	BADC
S	Delayed	Block 1 (Story 1)	Block 2 (Story 3)	Block 3 (Story 4)	Block 4 (Story 5)	ABAB	ABCD
T	Delayed	Block 3 (Story 4)	Block 4 (Story 5)	Block 1 (Story 1)	Block 2 (Story 3)	BABA	ABCD

Appendix H – Acceptable synonyms for gesture scoring

Synonym items generated using WorldNet 3.1 (Princeton University, 2010)

Item	Synonyms
Apple	apple, orchard apple tree, Malus pumila
Banana	banana
Bed	bed
Beer	beer
Book	book, volume
Boy	male child, boy, son
Camera	camera, photographic camera
Car	car, auto, automobile, motorcar, machine
Cat	cat, true cat
Chess	chess, chess game
Cup	cup
Dentist	dental practitioner, dentist, tooth doctor
Dentures	denture, dental plate, plate
Doctor	doctor, doc, physician, MD, Dr., medico
Door	door
Food	nutrient, food, solid food
Football	football, football game
Glasses	specs, spectacles, eyeglasses, glasses
Gloves	glove
Hair	hair
Hat	hat, chapeau, lid
Iron	iron, smoothing iron
Letter	letter, missive
Money	money
Newspaper	newspaper, paper
Pen	pen
Piano	piano, pianoforte, forte-piano
Rain	rain, rainfall
Rainbow	rainbow
Remote Control	remote, remote control
Scissors	scissors, pair of scissors
Sewing	sewing, stitching, stitchery
Spider	spider
Sponge	sponge
Stamp	stamp, postage, postage stamp
Swimming	swimming, swim
Tap	water faucet, water tap, tap, hydrant
Tea	tea
Telephone	telephone, telephone set, phone

Tissue	tissue, tissue paper
Umbrella	umbrella
Waiter	waiter, server
Walking Stick	walking stick
Watch	watch, ticker
Wife	wife, married woman
Wine	wine, vino