iReadMore: optimisation and personalisation of a therapy app for acquired reading impairments

Thomas Richard Daniel Langford

2025

UCL Queen Square Institute of Neurology
University College London

This Thesis is submitted for the degree of Doctor of Philosophy

Declaration of Authorship

I, Thomas Langford, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

First and foremost, I would like to thank Professor Alex Leff for granting me the opportunity to pursue this PhD. Your guidance, encouragement, and dedication have been invaluable, profoundly shaping both my research and my development as a researcher.

I am also deeply thankful to my Second Supervisor, Professor Jenny Crinion, for her sage advice, expert insights and reassurance throughout this endeavour.

I am grateful to everyone in the Neurotherapeutics Lab for the friendships, memories and making this experience that much richer. Special thanks to my PhD partner in crime, Dr Emily Upton.

I would also like to express my appreciation for my family and Hugo, whose unwavering encouragement and support I am immensely grateful for.

My sincere thanks go to the Medical Research Council for funding my place in the UCL-Birkbeck Doctoral Training Programme.

Finally, and perhaps most importantly, I am deeply grateful to every participant in my research and all users of the iReadMore app. Without your insights and contributions, none of this would have been possible.

Abstract

This thesis explores the development, evaluation and optimisation of a digital reading therapy for individuals with an acquired reading impairment. The app aims to improve single-word reading speed and accuracy. This research builds on prior research into the clinical efficacy of the iReadMore therapy mechanism, previously delivered via a laptop-based prototype.

In Chapter 1 of the thesis, I present a qualitative study of the development of the iReadMore App, addressing the need for improved accessibility and user engagement to enable independent, home-based therapy. A co-design process involving individuals with alexia, with or without aphasia, informed the creation of the app's release version. Design recommendations for digital therapies for persons with alexia were derived from the co-design process and analysed using a framework analysis.

In Chapter 2, I report preliminary findings from the ongoing clinical effectiveness trial investigating the therapeutic effects on reading accuracy and reaction time in real-world app users. This chapter highlights early indications of therapy effects. Additionally, I explore the unique challenges associated with conducting research in real-world contexts.

In Chapter 3, I present two studies investigating the potential of machine learning to predict treatment outcomes for users of the iReadMore app and another digital therapy for improving speech comprehension, the Listen-In app. In these studies, I focused on training models using data that can feasibly be collected within an app, the studies

provide insights into the viability of using routine, easily collected in-therapy data to support treatment outcome prediction.

In the general discussion, I connect the qualitative findings, preliminary trial results, and treatment prediction insights, contextualising the findings, and exploring potential directions for future research.

Impact Statement

Approximately, a third of all stroke survivors experience aphasia, a language disorder that can lead to impairments across all language modalities and frequently persists into the chronic phase of recovery (Feigin, Norrving and Mensah, 2017). With the added context of an estimated 1.2 million stroke survivors currently living in the UK (Stroke Association, 2018), and a projected 27% increase by 2047 (Wafa *et al.*, 2020), the demand for effective and accessible aphasia therapies is expected to rise significantly.

Evidence has demonstrated that treatments for aphasia can effectively improve language functioning even many years post-stroke, but meaningful recovery often requires substantial therapy doses, typically between 20–100 hours (Bhogal, Teasell and Speechley, 2003; Brady *et al.*, 2022). Despite this, limited rehabilitation resources within the National Health Service (NHS) mean that individuals with aphasia typically receive just 12 hours of therapy on average (Clarke *et al.*, 2018; Palmer, Witts and Chater, 2018). The Development of effective digital therapies offers a promising solution, enabling scalable, high-dose therapy delivery to meet the growing needs of this population.

The first chapter of my thesis explores the co-design process of developing iReadMore, a digital reading therapy designed for individuals with acquired reading impairments. The goal was to improve accessibility and therapy engagement, enabling at-home, independent use to achieve the high therapy doses necessary for significant treatment effects. An iterative co-design process was conducted, and a framework analysis was employed to identify key design recommendations. These recommendations were published to guide similar co-design efforts and provide a detailed methodology of work

(Langford *et al.*, 2022). The outcomes of this process, combined with funding achieved from Research England, facilitated the release of the iReadMore app.

Chapter 2 presents preliminary findings from the iReadMore Rollout Trial, which suggest early indications of a treatment effect for improving reading accuracy in trained words. However, further data collection is required to validate these findings. This chapter also offers valuable insights into the differences between how individuals with alexia or aphasia engage with self-led digital therapies in real-world contexts compared to controlled research settings, highlighting practical challenges and opportunities for improvement in real-world implementation.

Finally, Chapter 3 investigates the development of treatment prediction models for digital therapies, proposing the use of in-app data, particularly data collected during therapy sessions, as an alternative to traditional reliance on clinical assessments. The findings suggest that this in-therapy data could serve as a viable alternative or complement to clinical assessments, offering a means of generating predictions for digital therapy users when clinical assessments are not feasible. These results open new pathways for future research in aphasia rehabilitation and digital therapy development.

Statement of Publications

The data presented in Results Chapter 1 of this thesis have been published in JMIR Neurotechnology (Langford *et al.*, 2022). A report on the output of this work, the iReadMore app, has been published in ASSETS '21: Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Langford, Leff and Romano, 2021).

Design Innovation for Engaging and Accessible Digital Aphasia Therapies: Framework Analysis of the iReadMore App Co-Design Process. 2022. Langford T, Fleming T, Upton E, Doogan C, Leff A, Romano D, JMIR Neurotechnology.

iReadMore: A Reading Therapy App Co-Designed by People with Aphasia and Alexia. 2021. Langford T, Leff A, Romano D, Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility.

The preliminary data presented in Results Chapter 2 will be incorporated into future work evaluating the iReadMore therapy app.

The data on the Listen-In app presented in Results Chapter 3 is being prepared for submission.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AA	Attentional Alexia
ACT	Auditory Comprehension Test
AOS	Apraxia of Speech
ARCS	Attentive Reading and Constrained Summarisation
BPVS	British Picture Vocabulary Scale
CA	Central Alexia
CART	Copy and Recall Treatment
CAT	Comprehensive Aphasia Test
CI	Confidence Interval
cSART	Children's Sustained Attention to Response Task
DA	Deep Alexia
DRC	Dual Route Cascaded
ENVASA	Environmental Sound Discrimination Test
HA	Hemianopic Alexia
ICAP	Intensive Comprehensive Aphasia Program
LOOCV	Leave-One-Out Cross-Validation
lvPPA	Logopenic Variant Primary Progressive Aphasia
ML	Machine Learning
MMOR	Modified Multiple Oral Rereading
MRI	Magnetic Resonance Imaging
NA	Neglect Alexia
nfvPPA	Nonfluent Variant Primary Progressive Aphasia
NHS	National Health Service

ORLA	Oral Reading for Language in Aphasia
PA	Pure Alexia
PCA	Posterior Cerebral Artery
PDT	Phoneme Discrimination Test
PLORAS	Predicting Language Outcome and Recovery After Stroke
PPA	Primary Progressive Aphasia
PROMs	Patient-Reported Outcome Measures
RCT	Randomised Controlled Trial
REML	Restricted Maximum Likelihood Estimation
RFR	Random Forest Regression
RHH	Right Homonymous Hemianopia
RMSE	Root Mean Square Error
RT	Reaction Time
SA	Surface Alexia
SD	Standard Deviation
SDT	Self-Determination Theory
SLT	Speech and Language Therapy
svPPA	Semantic Variant Primary Progressive Aphasia
SVR	Support Vector Regression
VBM	Voxel-Based Morphometry
VFT	Visual Fields Test
WAB-R AQ	Western Aphasia Battery-Revised Aphasia Quotient
WRT	Word Reading Test

1 Introduction

1.1 Background

There are currently 1.2 million stroke survivors in the UK (Stroke Association, 2018). This number is expected to rise to well over 2 million by 2035. This increase can be attributed to stroke survival rates having doubled in the last decade, in part due to improvements in acute stroke medicine and greater public awareness of early stroke symptoms. Over the same period, prevalence of strokes in adults aged 35-54 has increased by 20% (Seminog *et al.*, 2019). These factors will result in greater clinical demand for post-stroke rehabilitation services.

Alexia is an acquired reading impairment that often arises due to a stroke or other brain injury affecting specific brain regions. Individuals with alexia have impaired reading abilities compared to before the injury. Alexia can occur with or without aphasia, depending on whether other aspects of language are affected. Aphasia is a common neuropsychological impairment caused by a stroke affecting a third of all stroke survivors (Berthier, 2005). It is a generalised language disorder that can affect all areas of expressive (e.g. speaking or writing) and receptive (e.g. listening or reading) communication. Aphasic symptoms are highly variable and dependent on stroke lesion size and location. In a study of the negative impact on quality of life of 60 highly-prevalent diagnoses, aphasia had the largest negative impact on quality of life, larger than a diagnosis of cancer or Alzheimer's disease (Lam and Wodchis, 2010). The negative implications for stroke survivors' quality of life arise from how aphasia disrupts

activities of daily living and life participation. Even in milder forms, it can be detrimental to quality of life (Hilari, 2011; Northcott and Hilari, 2011).

Bhogal *et al.* (2003) found that on average 100 hours of post-stroke aphasia therapy targeting one language domain (speaking, listening, reading or writing) is required to achieve a clinically meaningful improvement. Currently, NHS patients receive around 12 hours of therapy over 6-8 sessions (Clarke *et al.*, 2018; Palmer, Witts and Chater, 2018). However, it was also found that actual therapy time could be as low as 20 minutes per session (Clarke *et al.*, 2018). Insufficient therapy doses and outcomes are contributing to the statistic that 45% of stroke survivors surveyed by the Stroke Association reported that they felt 'abandoned after leaving hospital' (Stroke Association, 2016).

Although, some patients do receive higher doses, it is unlikely that stroke survivors outside of a controlled research setting would receive the dose suggested by Bhogal *et al.* However, recent research studies show the majority of stroke survivors (including those who had their stroke many years ago) would benefit from receiving high-dose and/or intensive therapies (Cherney, 2012; Dignam *et al.*, 2015; Doogan *et al.*, 2018). Consequently, new therapies that can be delivered in large doses and frequently are required to meet the current unmet clinical need. One solution is to develop digital therapies that can be used independently with unlimited access.

Digital therapies provide users with the flexibility to engage in self-paced therapy, offering unlimited access and enabling higher intensity practice. This accessibility can help achieve the substantial doses of practice often required for measurable improvements. However, digital therapies frequently have a narrow scope, confined by

the specific training content and therapeutic mechanisms they employ. For instance, many are rooted in mass-practice principles, which focus on repetition within a single language domain. While this approach can enhance trained items within that domain, there will be little to no improvement observed in untrained areas (Howard *et al.*, 2006). As such, the potential for digital therapies to provide broader, more comprehensive benefits remains an area of active development and research.

iReadMore is a digital therapy for alexia that was previously found to significantly improve reading speed and accuracy for individuals with central alexia (Woodhead *et al.*, 2018) and reading speed for individuals with pure alexia (Woodhead *et al.*, 2013). The digital therapy provided mass training exposure with significant therapeutic effects demonstrated in these trials. However, it was clear that the therapy needed to be developed further for real world use by individuals with alexia and potentially additional comorbidities, independently at home, while maintaining engagement and independent use.

Reading is an area that has been less focused on in terms of aphasia rehabilitation, and the development of the iReadMore app aligns well with my research interests in digital therapies, co-design, and machine learning. This PhD project seeks to bridge the gap between evidence-based therapies and real-world application, focusing on how digital interventions can be personalised and applied outside of clinical trials.

In this thesis, I aim to investigate the following research questions:

1 How can co-design and gamification principles be used to develop a reading rehabilitation app that ensures engagement, accessibility, and supports independent use by real-world users?
2 What are the preliminary treatment effects of the iReadMore therapy app for real-world users in a digital rollout trial?
3 Can treatment outcome prediction algorithms be trained using solely data collected through the iReadMore app?

By completing the co-design process and releasing the iReadMore app publicly will provide an important step in translating years of research, conducted prior to my joining the project into practice, fulfilling the ambition of providing an evidence-based therapy for individuals with alexia to use independently and without restriction. In addition, it is my hope that through a framework analysis of the co-design process, a contribution to our understanding of how to empower individuals with alexia to engage better with digital therapies and other software might be identified. Finally, the app's release will also enable the data collection for research questions 2 and 3, subject to timing and data availability.

The online rollout trial will generate valuable data on the therapy's effectiveness for real-world users, as well as detailed insights into the realistic usage of a digital therapy outside of the confinements of a typical research study. This data will be analysed to explore optimal therapy utilisation and strategies for maintaining engagement.

The final research chapter will explore the feasibility of developing treatment outcome prediction algorithms based solely on data that can realistically be collected through a digital application. This may offer a potential alternative or complement to existing practices that rely on brain scan data and behavioural tests conducted in clinical or research settings; an approach previously explored for the iReadMore therapy (Aguilar *et al.*, 2018).

In this introduction, I will explore the literature on reading therapies for alexia, with a particular focus on digital interventions, and the factors that influence treatment outcomes, ranging from therapy designs to user considerations. I will then outline the iReadMore therapy mechanism and review previous research as the foundation for my research. As a starting point, I will first define typical reading processes through cognitive models and examining the neurological basis for reading to provide a context for the therapy.

Given the diverse methodologies and subject areas covered in each study, dedicated subject-specific introductions are also included at the start of each methods chapter to provide appropriate context.

1.2 Typical Reading

1.2.1 Cognitive Models of Typical Reading

In the last half century, numerous models of word identification in reading have been proposed, three of which have reached considerable prominence. They are; the Interactive Activation Model (McClelland and Rumelhart, 1981), the Dual Route Cascaded (DRC) Model (Coltheart *et al.*, 1993, 2001; Coltheart and Rastle, 1994) and the Triangle Model (Seidenberg and McClelland, 1989; Plaut *et al.*, 1996; Harm and Seidenberg, 1999, 2004). All of these models are representation of reading single words aloud.

The Interactive Activation Model (Figure 1a) assumes that information processing during reading consists of distinct levels of processing beginning with the activation of nodes relating to sub-letter visual features (lines, dots, curves), these activate letter level nodes that correspond to those features and inhibits others that do not. This is turn activates word level nodes and lead to word identification (Figure 1b). Several letters are identified in parallel with inhibitory feedback from higher order levels (identification of whole words) able to inhibit nodes that do not fit the predictions. This model represents the visual processing of words effectively. However, it oversimplifies the cognitive processes involved in reading by excluding the influences of semantic and phonological processes.

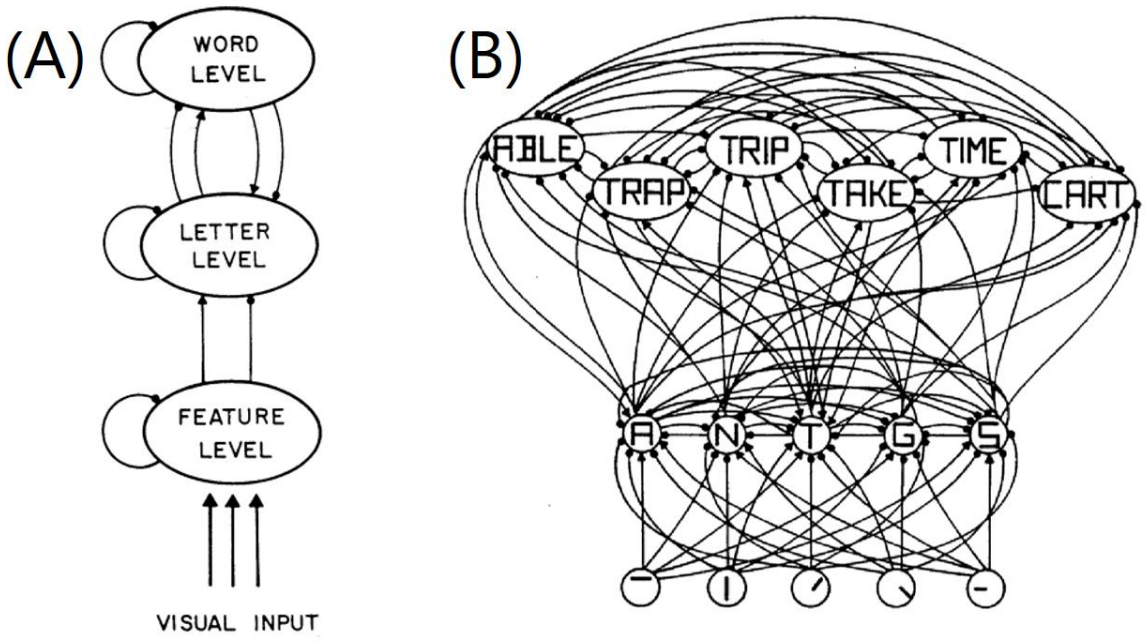

Figure 1 - Interactive Activation Model of Reading. a) Arrows represent excitatory connections and dot-ended connections represent inhibitory connections in the model. b) Example of node interconnections in the Interactive Activation Model from feature to word level (adapted from McClelland & Rumelhart, 1981).

The Dual Route Cascaded model (Figure 2) has two routes, a direct, unidirectional grapheme-phoneme route (for familiar, regular words) and an indirect, bidirectional route (containing excitatory and inhibitory connections for irregular, low-frequency, and exception words) that uses knowledge of pronunciation of letter sequences and semantics to inform word identification.

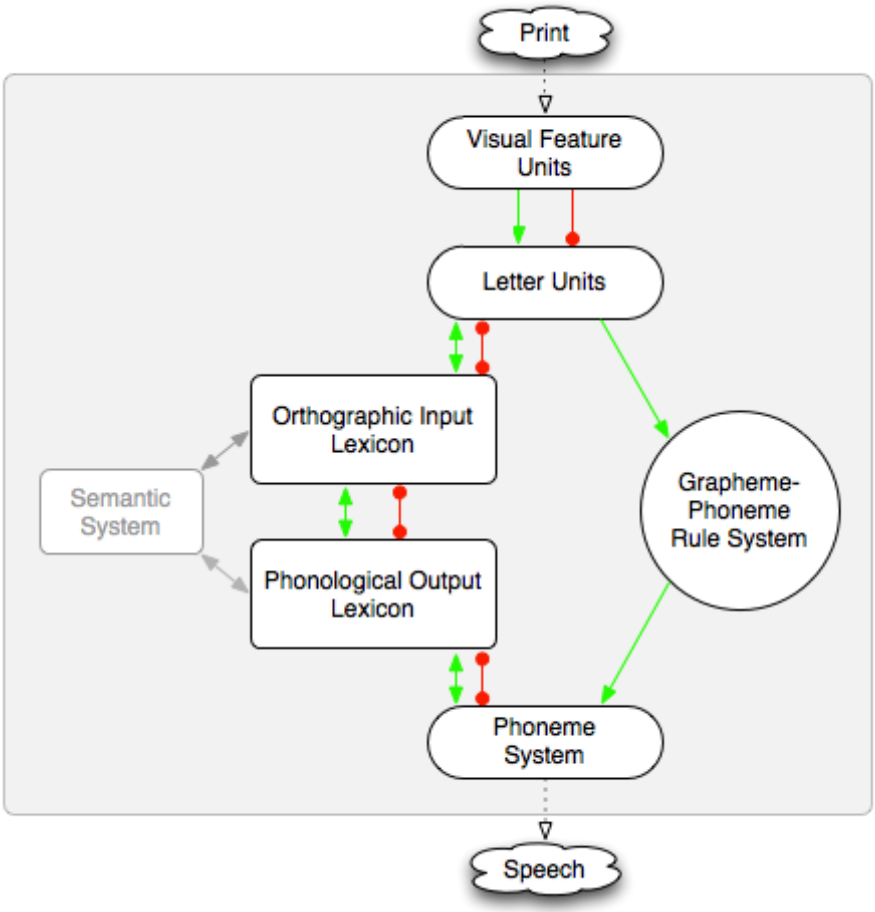

Figure 2 - Dual Route Cascade Model of Reading. Green arrows indicate excitatory connections, red dot-ended lines indicate inhibitory connections (obtained from http://www.cogsci.mq.edu.au/~ssaunder/DRC/).

The Triangle Model (Figure 3) is a connectionist model of word reading, comprising of three interconnected domains; orthographic (O), phonological (P) and semantic (S) representations. Word identification can be achieved through direct O-P or indirect O-S-P routes. Both routes are activated during word reading and the feedback/feedforward is strengthened through experience, where one route will become dominant and lead to quicker word identification. Both routes can read all types of words, but connectivity will be differently weighted for different types of words (and different individual reading

styles). The Triangle Model uses the assumption that frequent or regular words are identified more accurately and quicker as the direct connections should be stronger.

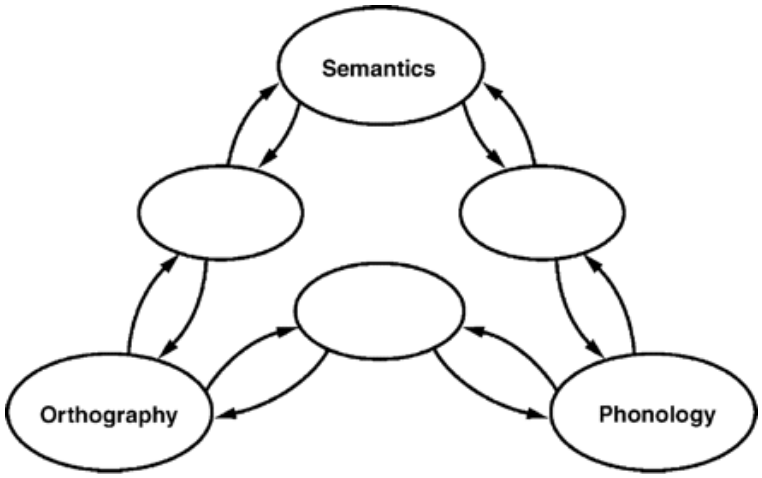

Figure 3 - The Triangle Model of Reading. The empty circles represent hidden units where weightings between the connections occur (obtained from Seidenberg, 2005).

1.2.2 Neurological Underpinnings of Typical Reading

Various neuroimaging studies of able readers have revealed that reading activates a predominantly left-lateralised network of occipitotemporal, temporal, and inferior frontal areas (Price, 2012; Carreiras *et al.*, 2014; Perrone-Bertolotti *et al.*, 2017; Zhou and Shu, 2017). Reading is thought to be achieved via two complementary pathways, the dorsal and ventral streams (Jobard, Crivello and Tzourio-Mazoyer, 2003; Cohen *et al.*, 2008; Taylor, Rastle and Davis, 2013). The dorsal reading stream supports phonological decoding for grapheme-to-phoneme reading (Hickok and Poeppel, 2007; Saur *et al.*, 2008; Taylor, Rastle and Davis, 2014), while the ventral reading stream comprises regions involved in visual word recognition (Binder *et al.*, 2006; Glezer, Jiang

and Riesenhuber, 2009; Ben-Shachar *et al.*, 2011) and access to the meaning of a word (Binney *et al.*, 2010; Visser *et al.*, 2012; Carlson *et al.*, 2014). When sensory input arrives at the occipital cortex via thalamic relay, simple concrete feature analyses are undertaken (e.g. unimodal) (Mumford, 1992). Information is then passed on to the higher-order ventral occipitotemporal cortex and inferior frontal gyrus where more abstract (e.g. multimodal) features analyses take place. These processes occur in parallel and higher order analyses are feedback down the hierarchy to lower order regions resulting in suppression of word predictions that are incongruent with the feedback, ultimately leading to word identification (Friston, 2005).

Cognitive models of reading have provided an understanding of the labour involved in word identification, but they do not contribute to our understanding of where these processes are taking place anatomically. Two prominent models of reading have emerged from neuroimaging literature; the Local Combination Detector model (Dehaene *et al.*, 2005) and the Interactive Account model (Price and Devlin, 2011). The Local Combination Detector model of visual word recognition suggests that as neurons are tuned to progressively larger fragments of a word as their location moves anteriorly, word reading is achieved primarily through feedforward processing along the visual ventral stream. This implies a bottom-up approach reliant on visual (orthographic) processing and not the semantic or phonological forms of a word. Subsequent studies have suggested that feedback from the frontal cortex plays a role in facilitating visual processing in word reading (Cornelissen *et al.*, 2009; Wheat *et al.*, 2010). Building on this concept, the Interactive Account model suggests that efficient word recognition relies on sensory input being fed forward (bottom-up) and predictions learnt by

experience being fed backwards (top-down) within this network. This model therefore accounts for how orthographic, semantic and phonological representations interact within the reading network. If considered in the context of the triangle model, phonological and semantic representations may form the higher-level components predicting orthographic and visual representation (Kerry *et al.*, 2019).

1.3 Acquired Impaired Reading

1.3.1 Post-Stroke Alexia

Alexia is a generic term to describe an acquired reading impairment. Alexia causes slow, effortful or inaccurate reading in people who were previously able readers. Stroke is the most common cause, other causes include brain injuries and neurodegenerative disorders. The presentation of symptoms is highly variable based on the location and severity of the injury and neurological or behavioural individual differences. Even in milder forms, alexia can be a debilitating disorder that restricts life participation, social relationships, employment and living independently (Hilari, 2011; Northcott and Hilari, 2011; Woodhead *et al.*, 2013).

There are a number of subtypes of alexia which fit into two subdivisions; peripheral and central alexia. Peripheral alexia result from disruptions in visual word processing in the occipital cortex prior to written information reaching language comprehension areas (Leff *et al.*, 2006; Starrfelt *et al.*, 2013), typically caused by a stroke affecting the

posterior cerebral artery in the left-hemisphere. Peripheral alexia has 4 subtypes; pure, neglect, attentional and hemianopic alexia. These are described in Table 1.

By contrast, central alexia results from disruptions in the language and semantic pathways, affecting how words are understood and interpreted after they have been visually recognised. Central alexia is defined as a reading impairment in the context of a general language disorder (known as aphasia). In central alexia, damage to the middle cerebral artery region of the temporal lobe disrupts word reading by impairing grapheme-phoneme conversion or word comprehension. Central alexia affects two-thirds of individuals with post-stroke aphasia (Leff and Starrfelt, 2014). Central alexia has 3 subtypes; surface, phonological and deep alexia, described in Table 1 on the next page. In addition, 'mixed-type' central alexia is described below for presentations commonly observed that do not fit distinct, classical subtypes.

Table 1- Peripheral and Central Alexia Subtypes (data adapted from Riley, Brookshire and Kendall, 2017).

Alexia category	Alexia Subtype	Definition	Point of breakdown (DRC model)	Point of breakdown (Triangle Model)
Peripheral Alexia	Pure	Impaired single word reading with word length effect, preserved single letter reading.	Visual feature and letter analyses	N/A
	Neglect	Neglect of visual field on the left or right, impairs reading of the initial or last letters of a word.	Letter analysis	N/A
	Attentional	Poor visual attention, Difficulty reading due to visual crowding of letters and words.	Letter analysis	N/A
	Hemianopic	Partial loss of vision on one side of the visual field. Leads to slowed reading and omissions.	Visuo-motor control of reading eye movements	N/A
Central Alexia	Phonological	Difficulty reading unfamiliar words or nonwords.	Non-lexical route (Grapheme-Phoneme Rule System)	O-S pathway
	Surface	Difficulty with irregular words.	Lexical route (orthographic input lexicon and phonological output lexicon)	P-O pathway
	Deep	Makes semantic errors and difficulty with function words.	Non-lexical and lexical routes	S-P and O-P pathways

1.3.1.1 Peripheral Alexia

Pure Alexia

Pure Alexia (PA), also known as alexia without agraphia, is a rare and severe form of reading impairment that typically arises from damage to the left ventral occipitotemporal cortex, often following an occlusion of the posterior cerebral artery. This impairment was first documented by Dejerine, who described a case of Monsieur C, a patient unable to read following a stroke (Dejerine, 1892). PA specifically impairs the ability to recognise whole words as distinct visual units during reading, despite the preservation of other cognitive functions such as spelling and writing. Patients with PA commonly exhibit a right homonymous hemianopia (RHH) in addition to significant difficulties with single-word reading, which cannot be solely explained by the visual field defect. Reading speed is notably slow on word-reading tasks, with some patients requiring up to ten seconds to read a six-letter word (Cohen *et al.*, 2004). A hallmark feature of PA is the 'word length effect' (WLE), where reading time increases with the length of the word, leading to a disproportionately greater difficulty with longer words. Although patients with PA often retain accurate spelling and writing abilities, indicating intact word knowledge, they struggle to access this knowledge from the visual modality. Letter reading is typically preserved, allowing patients to adopt a compensatory "reverse-spelling" strategy, where they read each letter of a word sequentially and combine them to form the complete word. This strategy, while effective, is effortful and slow, further contributing to the overall impairment in reading fluency.

A severe variant of pure alexia, in which even the ability to read single letters is impaired, is known as global alexia. This is the most profound form of peripheral alexia

and typically results from extensive damage affecting both the left occipital cortex and the interhemispheric white matter tracts. Dejerine's original patient, Monsieur C, was diagnosed with alexia without agraphia (or pure alexia) but also exhibited an inability to recognise or read individual letters, which aligns with the modern understanding of global alexia.

Neglect Alexia

Neglect Alexia (NA) is a reading disorder typically resulting from damage to the parietal or occipitotemporal regions of the brain. It is characterised by an inability to read words accurately when they are flanked by distracting stimuli, particularly when these stimuli interfere with the individual's visual attention. This condition is associated with a spatial neglect, where the individual fails to attend to one side of the visual field, often the left hemisphere in right-handed individuals. As a result, patients may omit or misread critical visual information, such as the initial or final letters of a word, which can disrupt word recognition and comprehension. In more severe cases, this attentional bias can lead to skipping entire words, which can contribute to the misinterpretation of the text's meaning.

Attentional Alexia

Attentional Alexia (AA) similarly results from parietal damage in regions responsible for visual attention. It is characterised by an inability to direct attention to individual letters of words within a visual field, causing interference and crowding among letters and words that often lead to omissions or other errors.

Hemianopic Alexia

Hemianopic alexia (HA) is a reading disorder typically resulting from damage to the primary visual pathways in the left occipital cortex, leading to a right homonymous hemianopia, where the individual experiences visual loss in the right visual field. HA is the most common form of peripheral alexia. People with this condition often exhibit reduced reading speed, especially when the impairment involves the right parafoveal or foveal regions, which are crucial for efficient word recognition and fixation. The primary difficulty in HA arises from the inability to make accurate saccades—rapid eye movements—along lines of text, which leads to issues such as omissions of words, challenges in maintaining the left-to-right reading saccades, and an increased sense of reading fatigue. Despite these challenges, individuals with HA are able to access both lexical and non-lexical routes for word recognition, and testing often finds preserved letter and word reading abilities. However, the ability to read text fluently is compromised by the impaired reading speed. This is largely due to the difficulty in planning and executing saccades across the text, as upcoming words fall within the blind hemifield which disrupts smooth reading progression (Leff *et al.*, 2001).

1.3.1.2 Central Alexia

Phonological Alexia

Phonological Alexia is marked by a deficit in reading unfamiliar or non-words, while the reading of actual, known words remains relatively intact. Phonological alexia is associated with damage to the left inferior frontal gyrus, insula and Rolandic operculum. Patients with this subtype struggle with the translation of print to sound as utilisation of the phonological representations or the direct orthographic-to-phonological pathway is impaired. Errors often include the lexicalisation of nonwords, where a nonword is read as a real word, such as reading "soof" to "soot".

Surface Alexia

Surface Alexia (SA) is characterised by the inability to read irregularly spelled words (such as "yacht") correctly, but can read regular or phonetically-spelled words with ease. Irregular spellings tend to lead to regularisation errors. For example, the word "pint" might be read as if it rhymed with "mint". This subtype often co-occurs with semantic dementia, where there is a gradual deterioration of the anterior temporal poles, lead to a progressive breakdown in the ability to access semantic information, further impairing word recognition and comprehension. Patients with surface alexia rely heavily on grapheme-to-phoneme conversion rules and have difficulty recognising whole words. The impairment is typically linked to damage in brain areas responsible for "whole-word" reading route, such as the left posterior middle and inferior temporal gyrus.

Deep Alexia

Deep Alexia (DA) is characterised by multiple impairments in word reading due to extensive damage in the left hemisphere language network. It can be considered as a severe form of phonological alexia which compromises experiencing semantic errors in addition to phonological errors. People with DA make errors where the target word is replaced with a semantically related word, such as “cat” and “dog”, or “yes” and “no”. They also exhibit visual and phonological errors and struggle with reading function words. People with DA may also have difficulty with abstract and function words. The Triangle Model of Reading suggests that DA results from a severe deficit in phonological representations affecting both the orthographic-to-phonological and orthographic-to-semantic pathways.

Mixed Alexia

Not all cases of central alexia (CA) fit neatly into the traditional categories of surface, phonological, and deep alexia, and the suitability of the models that explain the impairments associated with these subtypes has been the subject of ongoing debate (Woollams *et al.*, 2007). More recently, the term *mixed central alexia* has been used to describe cases where patients exhibit symptoms that span multiple subtypes and cannot be easily classified into any one category. A study of 64 chronic CA cases found that 78% of the patients presented with a mixed pattern of impairments, further challenging the utility of discrete subtypes (Leff and Starrfelt, 2014). In response to this complexity, a continuum-of-subtypes theory has been proposed, which suggests that

alexia symptoms exist along a spectrum rather than in isolated categories (Crisp and Lambon Ralph, 2006), visualised in Figure 4. This framework underscores the need for more flexible, individualised therapeutic approaches to address the varied and sometimes undefined nature of reading impairments. By utilising data from real world users of the iReadMore therapy, a therapy designed to be potentially useful in all CA subtypes, this study aims to further refine our understanding of central alexia subtyping and its relevance to clinical applications. Additionally, the continuum-based theory will form the foundation for a novel alexia subtyping test aimed at assessing reading impairments within the phonological–semantic space.

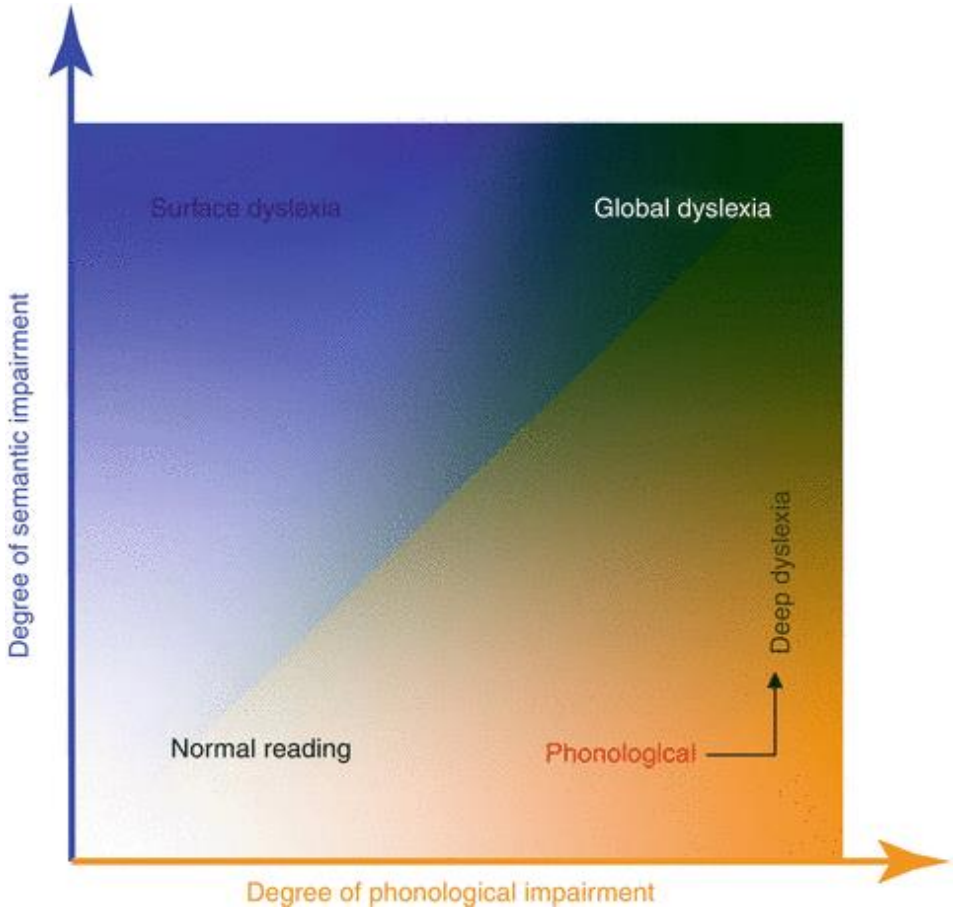

Figure 4 - The positioning of central alexia within a phonological–semantic space (adapted from Crisp and Lambon-Ralph, 2006).

1.3.2 Alexia in Primary Progressive Aphasia

Symptoms of aphasia also occur in neurodegenerative disorders. Most notably in a group of dementias known as primary progressive aphasias (PPAs). This a term for dementias distinguished by a progressive decline in language comprehension and production initially experienced in the absence of other neurological impairments (Mesulam, 1987). As the disorder progresses, other cognitive, behavioural or motor impairments may also begin to develop. Cases of PPA are rare with an estimated prevalence ranging from 3-11 cases per 100,000 (Coyle-Gilchrist *et al.*, 2016; Magnin *et al.*, 2016). Presentations of symptoms vary and there are presently three recognised subtypes of PPA (Gorno-Tempini *et al.*, 2011); a semantic variant (svPPA), a nonfluent variant (nfvPPA) and a logopenic variant (lvPPA).

svPPA involves the loss of the ability to interpret the semantic content of language and tends to be accompanied by severe anomia (Marshall *et al.*, 2018). People with nfvPPA tend to present a more diverse range of symptoms. However, agrammatism and apraxia of speech (AOS) are common and lead to speech production impairments including telegraphic speech (lacking function words) and slurring respectively. The logopenic variant, lvPPA is characterised clinically by word finding difficulties, phonological errors and impaired sentence repetition and comprehension (Beber *et al.*, 2014; Rohrer and Warren, 2016). Single word processing also declines with disease progression (Rohrer *et al.*, 2013).

People with nfvPPA tend to present a more diverse range of symptoms at initial clinical assessments when compared with the other PPA subtypes. However, agrammatism and

apraxia of speech (AOS) are common and lead to speech production impairments including telegraphic speech (lacking function words) and slurring respectively. Cases of pure agrammatism or AOS are also common (Rohrer and Warren, 2016). Speech comprehension is also affected, with particular difficulties arising from the processing of syntactically complex sentences; which is linked to agrammatism (Peelle *et al.*, 2008). In nfvPPA, atrophy occurs asymmetrically in the left inferior frontal lobe (particularly around Broca's area), insular and anterior-superior temporal gyrus atrophy (Brambati *et al.*, 2009; Wilson *et al.*, 2010; Marshall *et al.*, 2018). Many people with nfvPPA will go on to develop Parkinsonism and some develop motor neurone disease.

The logopenic variant, lvPPA is characterised clinically by word finding difficulties, occasional phonological errors and impaired repetition and comprehension of sentences (Beber *et al.*, 2014; Rohrer and Warren, 2016). Single word processing remains intact initially but declines with disease progression (Rohrer *et al.*, 2013). Initially, people with lvPPA may experience 'tip of the tongue' phenomenon; in which they strongly feel they know the word they want to produce but are unable to recall it. This interrupts speech production, but verbal output tends to be free from agrammatism or AOS demarcating this variant from nfvPPA. Many other cognitive domains can also be affected including mood regulation, working and episodic memory, praxis and spatial awareness (Butts *et al.*, 2015; Piguet *et al.*, 2015). Initially degeneration in lvPPA is mainly limited to the left posterior temporo-parietal area, in particular the posterior-superior temporal and inferior parietal cortices (Gorno-Tempini *et al.*, 2004; Marshall *et al.*, 2018). However, as the disease progresses the variability of brain areas affected increases and thus, symptom patterns become more

heterogeneous. lvPPA has an association with the pathology of Alzheimer's disease (AD) and most cases (but not all) can be considered an atypical variant of AD.

The biological validity of the clinical subtypes has gained support from recent MRI and genetics studies supporting the concept that these subtypes have distinct pathologies (Marshall *et al.*, 2018). However, as with classification of post-stroke alexia, people diagnosed with PPA in practice rarely fall neatly into one subtype. The result is some cases being termed as 'atypical' (or mixed) PPA, which accounts for between 17% (Harris *et al.*, 2013) and 41% (Sajjadi *et al.*, 2012). Atypical cases may present with a single impairment such as pure alexia.

1.3.3 Neurological Underpinnings of Alexia

Cognitive models of typical reading have been primarily developed from insights gained from reading impairments, particularly alexia, which occur when the reading system fails. However, these models often neglect how the brain compensates for damage and adapts in response to impairment. Following a stroke, neuroplasticity enables the potential for recovery, but this process often involves the reorganisation of reading functions, rerouting them through alternative neural pathways.

From a neurological perspective, there are three primary hypotheses regarding language reorganisation after a stroke that impairs reading abilities. The first hypothesis suggests that the brain compensates by activating the right hemisphere homologues of the damaged left hemisphere regions, which are typically responsible for reading. This shift represents a transition from a predominantly left-lateralised reading network to a more

bilateralised network. The second hypothesis posits that perilesional regions adjacent to the damage within the left hemisphere may become increasingly involved in supporting reading functions. Lastly, a combination of both mechanisms may be employed, with the brain engaging both right hemisphere regions and perilesional areas to recover reading abilities.

The extent of recovery in reading abilities is influenced by several factors. Crucially, the location and severity of brain damage play a significant role in the likelihood of regaining functional reading. Recovery is more probable when the damage is limited and when key areas, such as the Visual Word Form Area (VWFA) or the arcuate fasciculus remain intact. In cases where these regions are not completely destroyed, individuals are more likely to recover their pre-stroke reading abilities.

Age is another significant factor in the recovery process. Younger patients tend to exhibit better functional recovery of reading abilities, owing to greater brain plasticity and more robust compensatory mechanisms. The role of education level in recovery, however, remains ambiguous. Some studies suggest that higher education levels may contribute to improved recovery, whilst other research questions the strength of this correlation. The precise influence of education on reading rehabilitation continues to be a topic of debate in stroke rehabilitation literature.

Understanding the mechanisms of brain reorganisation and the factors influencing recovery is essential for refining rehabilitation strategies for individuals with alexia. These insights highlight the brain's remarkable capacity for adaptation and compensation, offering hope for more effective interventions in the restoration of

reading abilities following neurological damage. Reading therapy studies have observed changes in language lateralisation, with improvements in reading abilities linked to a shift from disorganised, bilateral reading networks to more structured, left-lateralised networks (Saur *et al.*, 2006, 2008; Kurland *et al.*, 2008; Kerry *et al.*, 2019).

1.4 Reading Therapies

1.4.1 Rehabilitation Approaches and Current Practice

Presently, there is no agreed, best-practice treatment for central (Leff and Starrfelt, 2014) or pure alexia (Starrfelt *et al.*, 2013). Only relatively recently has the clinical effectiveness of Speech and Language Therapy (SLT) for treating aphasia been supported by small, but significant effect sizes in Cochrane reviews (Brady *et al.*, 2012, 2016). Whether treatment was delivered one-to-one, in a group or via a computer did not influence the effectiveness. However, therapies with a higher dose or intensity had improved outcomes. Although, the higher intensity studies also had higher dropout rates.

There are 5 broad types of reading therapies; strategy-based, cognitive, oral reading, impairment-based and compensatory strategies. These are outlined below, digital therapies are detailed thereafter.

1.4.1.1 Strategy-based therapies

For single word reading, strategies include letter-by-letter reading and tactile-kinaesthetic reading (letter tracing), these are for people with pure alexia (Coltheart, 1998). When applied, these strategies cause slow and effortful reading with only small gains in comprehension and therefore, should only be used for people with poor letter recognition. However, tactile-kinaesthetic reading has been demonstrated to generalise to improved reading speed and accuracy without letter tracing in one study (Lott *et al.*, 2010).

Other strategies can be used for text comprehension including Attentive Reading and Constrained Summarisation (ARCS; Rogalski and Edmonds, 2008) and Proposition Identification and Constrained Summarisation (PICS). In a systematic review of reading comprehension therapies, 4 out of 6 patients in 4 strategy-based case studies/series had significant improvements (Purdy *et al.*, 2019).

1.4.1.2 Cognitive Treatments

Non-linguistic cognitive impairments, such as impaired attention or working memory, may also contribute to reduced reading ability. Some treatments focus on these cognitive impairments to indirectly improve reading in people with cognitive impairments and mild aphasia/alexia. Notable examples include Sequenced Exercises for Working Memory (Mayer and Murray, 2002) and Attention Process Training-2. With appropriate patient selection, they can be effective in improving reading comprehension (Lee *et al.*, 2018; Purdy *et al.*, 2019).

1.4.1.3 Oral Reading Therapies

Two oral reading therapies have evidence bases, Oral Reading for Language in Aphasia (ORLA) and Modified Multiple Oral Rereading (MMOR). ORLA uses repetitive, shared reading aloud with a clinician of increasing difficulties to improve reading comprehension. MMOR is similar to ORLA, but focused on independent reading by the patient with feedback from the clinician (Mayer and Murray, 2002; Kim and Russo, 2010). Thus far, ORLA and MMOR have shown non-significant improvements in reading comprehension (Purdy *et al.*, 2019).

1.4.1.4 Compensatory Aids

Compensatory interventions aim to improve reading only when the intervention is in use. In this sense, they are aids and not therapies. An example of an aid being a pair of glasses; they improve reading ability only when they are worn. Compensatory aids for alexia include text-to-speech software, contextual picture support or using an e-reader to support reading, such as by enlarging the font or reducing length of text on display. A study found e-readers can significantly improve technology-assisted reading comprehension (Caute *et al.*, 2018). Compensatory strategies can be effective with quick results, but limited in real-world applications.

1.4.1.5 Impairment-based therapies

Impairment-based therapies aim to improve reading speed and accuracy by directly targeting the impaired processes. As a result, therapy progress can be slow and require considerable effort from the patient. However, long-term outcomes for these therapies can be significant (Woodhead *et al.*, 2018). Copy and Recall Treatment (CART) is a therapy for writing found to improve reading additionally (Beeson, 1999). In particular, combinations of oral reading and CART have found positive results in 3 participants (Orjada and Beeson, 2005; Beeson *et al.*, 2010).

iReadMore is also an impairment-based therapy and detailed in section 1.4.3.

1.4.2 Review of Digital Therapies and Evidence Base

1.4.2.1 Stroke

Presently, numerous digital therapies are available. Therapies that train reading and have been evaluated in at least one peer-reviewed publication are detailed below. Cost to the user for these programmes range from £20 per month to a one-off payment of £275 not including hardware costs.

4-in-1 Language Therapy is an app for impairment-based training in 4 domains; naming, auditory comprehension, reading and writing. Only one study on this therapy has been published (Stark and Warburton, 2018). This study found CAT expressive scores improved significantly in a group of 10 participants with expressive aphasia and intact

comprehension. There is no evidence supporting improvements in reading comprehension.

Constant Therapy is a cognitive therapy for people with aphasia and/or cognitive impairments. It has tasks targeting various language and cognitive impairments. In a non-randomised, non-blinded comparative study of Constant Therapy and weekly SLT sessions versus SLT sessions alone (control group), participants in the intervention group significantly improved on the WAB-R AQ (a composite measure of language ability) compared to the control (Des Roches *et al.*, 2015). This study has a number of limitations including no follow-up, inclusion of participants without aphasia and significant differences between study groups. There is no evidence on the efficacy of Constant Therapy for improving reading comprehension specifically.

A computerised version of ORLA has been developed for self-led home use. A study of 25 stroke survivors with chronic, non-fluent aphasia compared low dose, computerised ORLA with SLT-led ORLA (Cherney, 2010). They found equivalent small, but significant gains on the WAB-R AQ for both interventions compared to no therapy. Neither version of ORLA improved WAB reading scores.

React-2 is an online resource of SLT exercises including reading comprehension. One case study on an individual with alexia found within-task improvements for paragraph reading and qualitatively-reported reading improvements (Palmer, 2015).

StepByStep is a therapy for training expressive, rather than receptive, language. However, a case series reported that one participant 'reported benefit to reading skills, evidence that could not be corroborated' (Mortley *et al.*, 2004).

To summarise, there is minimal evidence to support the clinical efficacy of currently available, digital reading therapies. In general, there is evidence to support the implementation of digital therapies and a number of associated benefits. Digital interventions can substantially increase therapeutic dose (Katz, 2010), feelings of self-worth and independence (Petheram, 2004). However, further research is needed to support the efficacy of these technologies.

1.4.2.2 Primary Progressive Aphasia

Interruptions to reading abilities occur in all variants of PPA. People with svPPA tend to have difficulty with irregular words, which reflects surface alexia (Woollams *et al.*, 2007; Wilson *et al.*, 2009). People with nfvPPA or lvPPA, more often have difficulties with reading non-words or unfamiliar words implying phonological alexia (Brambati *et al.*, 2009; Matías-Guiu *et al.*, 2017). Across the three variants of PPA, it has been demonstrated that single word reading speed and accuracy of people are significantly impaired compared to healthy controls or people with Alzheimer's disease (Brambati *et al.*, 2009; Matías-Guiu *et al.*, 2017).

There is a paucity of evidence supporting treatments for reading impairments in PPA. There is currently no evidence of effectiveness for any curative or pharmacological interventions for people with PPA (Volkmer *et al.*, 2019). SLT therapies used clinically with people with PPA tend to have been developed for post-stroke reading impairments. The efficacy of translating therapies from post-stroke aphasia to PPA has also not been

well studied. No studies on self-managed digital therapies or digital reading therapies were found.

The lack of evidence on reading interventions reflects the wider picture of the current clinical situation for people with PPA. A recent, survey-based study illuminated the current lack of therapeutic interventions available to people with PPA in the UK (Volkmer *et al.*, 2018). 90.5% of respondents reported that there was no SLT care pathway for people with PPA within their Trust. Only (11.4%) reported that reading and/or writing therapies were 'often' or 'always' available.

1.4.2.3 Factors influencing therapy success

The success of reading therapies for individuals living with alexia can be influenced by several key factors that interact to shape the outcomes of treatment. These factors include therapy dose, intensity, timing, motivation and engagement, and personalisation of the therapeutic mechanism. Each plays a significant role in determining the effectiveness of rehabilitation and the degree of recovery a patient can achieve.

Dose

The dose of therapy refers to the overall amount of an intervention that a patient receives. Research has shown that a higher therapy dose generally leads to better outcomes in cognitive and linguistic rehabilitation (Brady *et al.*, 2016). Bhogal et al

found that high dose aphasia therapies (defined as a dose of around 100 hours) lead to significant improvements. However, the RELEASE Consortium found that 20-50 hours led to optimal improvements from SLT (Brady *et al.*, 2022). Both meta-analyses found that the optimal dose of therapy was higher than what people with aphasia typically receive in the UK from NHS care, where they receive approximately 12 hours of SLT-led therapy (Clarke *et al.*, 2018; Palmer, Witts and Chater, 2018). An estimated 68% of people with chronic aphasia will have alexia (Brookshire *et al.*, 2014). However, reading therapy is not always a part of their treatment protocol, and data on the proportion of therapy time dedicated to reading interventions, particularly within the NHS, remains limited.

Intensity

Therapy intensity considers the dose received within a specific time period, such as hours per week. A network meta-analysis looking for optimal therapy intensity found 2 hours per week led to optimal therapeutic effects that did not increase further with increased intensity of practice (Brady *et al.*, 2022). When considering therapy intensity, it is important to achieve a realistic balance, as excessive therapy without adequate rest may lead to fatigue or burnout, which could lead to therapy disengagement or demotivation. Higher intensity studies have also been found to have higher dropout rates (Brady *et al.*, 2016).

Frequency

Consistent practice is recommended for optimal outcomes from reading therapies, a higher frequency of sessions with sufficient exposure to reading tasks increases the brain's opportunity for neuroplasticity, allowing for more extensive reorganisation of neural pathways involved in reading. Therefore, optimal dosing should ensure consistent and frequent exposure to therapeutic exercises, while considering individual tolerance. Practicing on 5 days per week has been found to lead to optimal therapeutic outcomes, with no additional gain from increasing the weekly frequency (Brady *et al.*, 2022).

Challenge

The difficulty of the therapy, or challenge, often measured by the difficulty level of tasks and the effort required from the therapy user, is another critical factor in alexia rehabilitation. Higher-challenge interventions, such as those that push cognitive limits or challenge reading processes in progressively complex ways, can facilitate neuroplasticity and improve functional outcomes. Intensity must be calibrated to each individual's ability to engage and process information without overwhelming them and offering the appropriate level of therapy (such as letter-based versus text-based). Too little difficulty may not provide enough challenge to stimulate recovery, while too high a level may lead to frustration or demotivation. The challenge of the therapy should be tailored to the patient's current capabilities, gradually increasing as they show

improvement. Depending on the type of alexia, different therapy content may be needed, such as focussing on abstract, irregular or function words.

Timing

The timing of therapy plays a crucial role in rehabilitation success, particularly in the context of stroke recovery. Neuroplasticity is most pronounced in the weeks and months following a stroke, making early and timely intervention crucial for optimal recovery outcomes if feasible in this timeframe. However, significant improvements can also be achieved in the chronic phase of recovery. Studies have shown that years after a stroke, targeted interventions can continue to promote functional recovery (Fridriksson and Hillis, 2021).

Motivation and Engagement

Motivation and engagement are essential to the success of alexia therapy, as they directly affect the patient's willingness to participate and persist with the treatment. Patients who are more motivated are likely to put forth greater effort, attend more sessions, and maintain a positive attitude toward rehabilitation (Watila and Balarabe, 2015; Harrison, Palmer and Cooper, 2020). Engaging therapeutic activities that are meaningful and relevant to the individual’s everyday life can help sustain motivation. Additionally, setting small, achievable goals can enhance a patient’s sense of progress and accomplishment, further boosting motivation. Incorporating patient preferences,

such as tasks related to their personal interests or daily activities, can also increase engagement and improve therapy outcomes.

Personalisation

Personalisation of the therapeutic approach is crucial in addressing the unique needs and challenges of each individual with alexia. Factors such as the type of alexia (e.g. surface or phonological alexia), the extent and location of brain damage, the individual's cognitive strengths and weaknesses, and their learning style should all inform the design of the therapy. Co-morbidities may also need to be considered and the therapy choice adapted to find therapies that are accessible and appropriate. A personalised treatment plan and therapy contents allows for focus on the specific areas that need improvement, whether it's letter or word recognition, reading comprehension or speed, or a focus on particular areas of language, such as irregular, abstract or function words. Personalised interventions are more likely to be effective because they target weaknesses and accommodate therapy users' pace of learning. Additionally, ongoing assessment and adjustment of the therapy plan are important to ensure that the approach remains relevant and effective as recovery progresses.

1.4.3 iReadMore Therapy App

1.4.3.1 Therapy Design

iReadMore is a self-led therapy app for improving single word reading accuracy and reaction time (RT) in individuals with acquired reading impairments. iReadMore has been demonstrated to improve reading speed and accuracy in pure and central alexias (Woodhead *et al.*, 2013, 2018). iReadMore therapy is not expected to improve reading for the remaining alexia subtypes.

The therapy contains two phases, Exposure and Challenge. In the Exposure phase, ten words are displayed, read aloud and visualised by an image indicating the semantic representation (Figure 5). These pairings are always congruent.

Figure 5 - Therapy Exposure Phase (initial iReadMore prototype before co-design enhancements).

In the Challenge phase, a word from the Exposure phase is read aloud and either the same or a different word is displayed on the screen simultaneously. The user must decide whether they read and heard the same word (congruent trial) or two different

words (incongruent trial) and respond by pressing one of two buttons accordingly (Figure 6). If the words are different, the written word will be a distractor word that is phonologically similar, such as ‘cat’ and ‘car’. Immediate feedback is provided to the user on whether the trial was answered correctly or incorrectly to facilitate error-reducing learning. Each Challenge phase lasts 30 trials. The therapy has an adaptive algorithm to adjust the difficulty of the training to suit individual user’s reading abilities. Difficulty parameters include adjusting exposure time, word complexity and similarity between the written distractor words and spoken words.

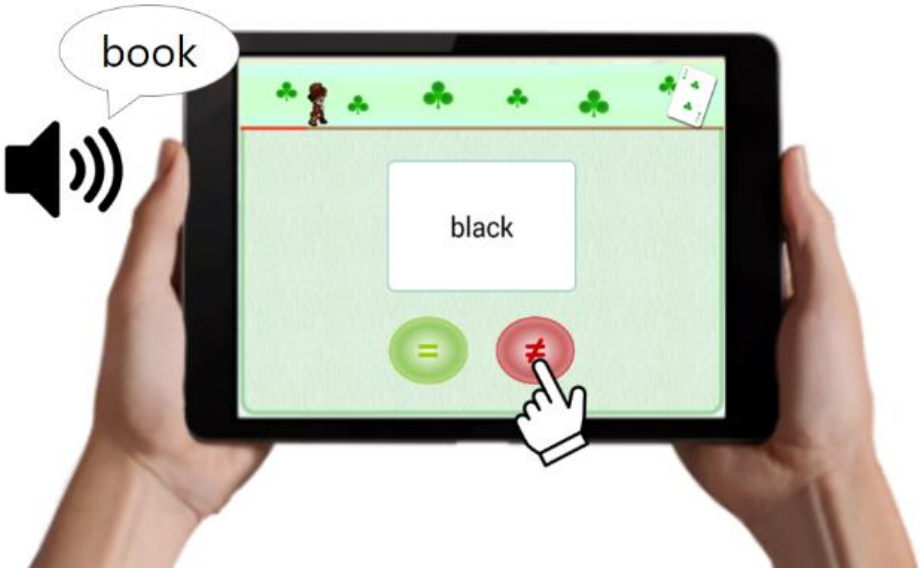

Figure 6 - Therapy Challenge Phase (initial iReadMore prototype before co-design enhancements).

iReadMore includes 590 of the most frequently written words from the SUBTLEX lexical database (Brysbaert and New, 2009). High frequency words were chosen to maximise the relevance (ecological utility) of the training to everyday life. The therapy mechanism relies on mass practice of cross-modal, paired associate learning (Holcomb and Anderson, 1993) based on the Triangle Model approach to reading. Untrained words will

not be affected, and thus word reading improvements are not expected to generalise beyond trained items.

The therapy utilises gamification and is designed to be accessible in order to drive motivation and engagement for the user group. iReadMore is downloadable via the Google Play store on Android tablet devices. iReadMore is a CE-marked class 1 medical device that was developed by the Neurotherapeutics Group, University College London.

1.4.3.2 Clinical Research

iReadMore was first evaluated in a study of 9 participants with chronic pure alexia (Woodhead *et al.*, 2013). Participants were asked to use iReadMore independently for at least 20 minutes a day over 6 weeks. Immediately following the therapy, reading speeds for trained words had improved significantly. Reading accuracy did not improve; but this was expected as individuals with pure alexia tend to display slow, but accurate reading abilities. The average reading speed increased by 149ms, a training effect size of 11.5%. Importantly for people living with pure alexia, the largest improvements in reading speed were seen for longer words, implying that the treatment was reducing the word length effect. However, improvements were specific to reading single words and did not generalise to letter or text reading. Improvements in reading speed were also not maintained at 2-4 week follow up. This may be due to the low dose of therapy in the study (10-14 hours).

Following on from this study, an randomised controlled trial (RCT) was conducted to evaluate the clinical efficacy of iReadMore with 21 people with chronic central alexia

(Woodhead *et al.*, 2018). This trial had a crossover design with participants receiving therapy over two 4-week blocks with separate word lists being trained. On average, participants used iReadMore for 70 hours (range: 60-78 hours). At therapy cessation, an average improvement in reading accuracy of 9% for trained words was demonstrated (Cohen's *d* = 1.38 – large effect size). Reading speed also improved by an average of 100ms for trained words (*d* = 0.98 – large effect size). As expected, improvements did not generalise to untrained words.

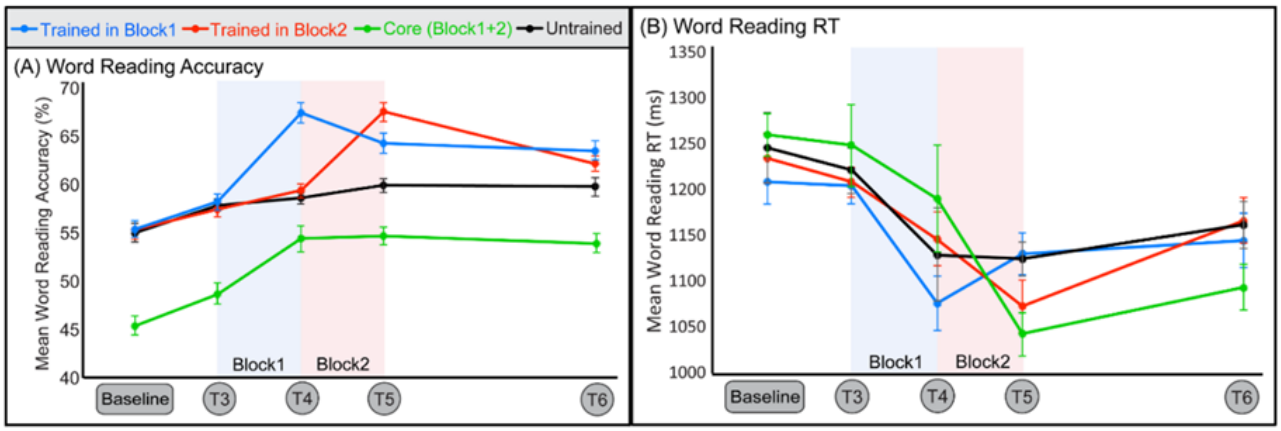

Figure 7 – iReadMore Therapy effects on: a) word reading accuracy and b) reaction time. There are four word lists: Trained in Block 1 (blue), Block 2 (red), Untrained (black) and Core words (green). (from Woodhead et al., 2018)

At 3-month follow-up (T6 in Figure 7), significant improvements in reading accuracy were maintained, although some gains were slightly diminished. Improvements in RT were not significant at follow-up, but remained below baseline. Considering core words only (a list of highly common function words such as 'and', 'the' and 'it'), the significant gains in both accuracy and RT were maintained at follow-up. Core words were trained in Blocks 1 and 2 suggesting a larger dose or longer therapy duration may be required to maintain gains in reading speed. The core words tended to be abstract and have low imageability making them harder to learn. However, in post hoc paired t-tests, it was noted that word imageability did not influence the efficacy of the therapy.

2 Overview of Thesis

The three results chapters of this thesis explore distinct but related aspects of the iReadMore app and its application in aphasia therapy. The first chapter focuses on the co-designing the iReadMore app with persons with alexia or aphasia, aimed at enhancing user engagement and accessibility. This process included a series of 5 co-design focus groups with 25 people with aphasia and carers/family members, followed by remote beta testing sessions with an additional 25 participants, conducted during the COVID-19 pandemic. The findings were evaluated using a qualitative framework analysis approach to investigate design recommendations for digital alexia therapies.

The second chapter explores preliminary findings from a rollout trial conducted following the release of the iReadMore app. Real-world therapy users completed reading tests every 5 hours of therapy. This preliminary study evaluated data from 14 participants who completed more than 20 hours of therapy between 1st March 2021 and 1st April 2023. A mixed linear regression model was employed to analyse the outcomes, accounting for the small sample size and timepoint attrition beyond the 20-hour mark. The analysis focused on changes in reading accuracy and reaction time.

The final results chapter investigates the potential for treatment outcome prediction models to be trained using therapy data. This study, conducted in two parts, draws on data from the iReadMore RCT and a more extensive dataset of therapy data from the RCT of Listen-In, a digital therapy targeting speech comprehension. The intention is that once sufficient data is collected in the rollout trial, that this work could be developed further using real-world data.

3 Main Aims and Research Questions

This thesis includes work conducted in co-designing the iReadMore app and preparing it for public release as well as the preliminary analyses of the data collected from real-world users post-release. The results in this thesis are presented in three chapters. The aims and hypotheses are outlined below.

3.1 Research Chapter 1 – Development of the iReadMore App

3.1.1 Aim

To develop a reading rehabilitation application utilising co-design and principles of gamification to ensure engagement and accessibility of the therapy app and support independent use by real-world users.

3.1.2 Rationale

The iReadMore therapy mechanism was developed to improve single-word reading accuracy and reading speed for individuals with chronic central or pure alexia. In initial studies, a laptop was used to deliver the therapy within research settings, with participants conducting therapy under the observation of researchers. Anecdotally, it was noted that participants encountered difficulties using the technology and maintaining motivation throughout the study duration. This highlighted the need for further development to enhance the accessibility and engagement of the therapy,

enabling individuals with aphasia and alexia to feasibly engage with the iReadMore therapy at home and without the support of a therapist.

To address these challenges, a co-design approach was adopted to ensure that the prototypes developed would be suitable and user-friendly for the end users. Additionally, gamification, an established strategy in digital applications and therapies, was incorporated to enhance the appeal of the therapy and improve motivation. By integrating game-like elements, the design aimed to foster consistent use and sustained attention, key factors associated with improved therapeutic outcomes.

The development of a usable and engaging app was a central component of my thesis, as it would provide the data collection for the analysis presented in Chapter 2.

3.1.3 Objectives

1 Carry out a series of co-design focus groups with individuals with alexia to develop and refine iReadMore prototypes.
2 Utilise the qualitative data from the co-design sessions to inform a framework analysis exploring themes associated with usability and engagement for digital alexia therapies.
3 Implement design recommendations into the iReadMore app release version.

3.2 Research Chapter 2 – Preliminary Findings of the iReadMore Rollout Trial

3.2.1 Aim

Investigate preliminary treatment effects of the iReadMore therapy app for real world users in a digital rollout trial.

3.2.2 Rationale

This study outlines the ongoing roll-out trial of the therapy's clinical effectiveness with real-world app users and include analyses of the data available up to April 2023. Insights into how the digital therapy is used in everyday settings, beyond the controlled environment of a typical research study will also be explored. This data will be analysed to explore potential contributions to dialogues on optimal therapy utilisation and the impacts of strategies implemented for sustaining user engagement.

3.2.3 Objectives

1 Explore group level, within-patient treatment effects in terms of reading speed and accuracy for therapy users.

2 Explore therapy utilisation in terms of dose, frequency and intensity patterns.

3 Explore quantitative PROMs and qualitative data using appropriate methodology depending on data quality and quantity.

3.3 Research Chapter 3 – Prediction of Therapy Outcomes Using Therapy Data

3.3.1 Aim

To explore approaches to therapy outcome prediction utilising only data that could be collected through the iReadMore app.

3.3.2 Rationale

Treatment outcome prediction algorithms have been explored for aphasia treatments. Typically, these have relied on researcher-collected training data, including brain scans and behavioural testing. In this chapter, I aim to investigate whether therapy data collected through a digital platform, specifically the iReadMore app, can serve as a viable alternative for predicting treatment outcomes. Digital therapy platforms offer the advantage of collecting data continuously and seamlessly, which presents a unique opportunity to monitor patient progress in real-time without the need for external tests or measurements.

As a starting point, data from the iReadMore clinical trial was used to test this approach. However, due to limited data availability, the approach was replicated using data from the randomised trial of the Listen-In speech comprehension therapy app, which provided a more extensive dataset, including detailed in-app therapy challenge data. By expanding the scope to a more comprehensive dataset, this study aims to determine whether similar treatment outcome predictions can be made using data

collected directly from digital therapy apps. If sufficient data is gathered from the real-world users of the iReadMore app upon its public release, the methodology will be replicated to further explore the feasibility of creating an accurate and scalable treatment prediction algorithm.

3.3.3 Objectives

1 Explore the potential for using therapy data from the iReadMore clinical trial to predict treatment outcomes.

2 Use data from the Listen-In speech comprehension app clinical trial to further refine and test therapy data-based treatment outcome prediction.

3 If sufficient real-world data is collected from the iReadMore app following its public release, investigate the creation of a treatment prediction algorithm based on this real-world user-generated dataset.

4 Methods

4.1 Study 1: Development of iReadMore

4.1.1 Background on Co-design Approach and Theoretical Underpinning

4.1.1.1 Co-design

Therapies that utilise mass practise, by nature, involve highly repetitive tasks that can lead to frustration and boredom (Kurland, Wilkins and Stokes, 2014). Therefore, it can be difficult for patients to sustain motivation and engagement over many weeks or months. For a self-managed therapy, such as iReadMore, sustaining user motivation will be vital. Further, for a digital therapy, there are a number of barriers to accessing the therapy for people with aphasia relating to their communication impairment, co-morbidities and level of experience with digital technologies (Menger, Morris and Salis, 2016; Munsell *et al.*, 2020). One approach that can be utilised to improve the acceptability and accessibility of a therapy is to design it with the target user demographic; known as co-design. Co-design has been applied to a number of digital post-stroke aphasia therapies (Marshall *et al.*, 2013, 2016; Wilson *et al.*, 2015; Messamer, Ramsberger and Atkins, 2016). EVA Park is an example of a co-design SLT-led therapy for people with stroke-related speech impairments delivered in an online virtual environment. It was found that users responded positively to the novelty of the codesigned therapy, evaluated both in terms of a 0% therapy dropout rate (Marshall *et al.*, 2016) and high acceptability deduced from qualitative interviews (Amaya *et al.*, 2018). A number of techniques can enable people with aphasia to participate fully in co-design research by supporting total communication. Techniques can include

drawing, writing, gesturing, visual aids, emotion scales and inviting carers and partners to attend (Neate *et al.*, 2019). It can also be beneficial to know the communication profiles of participants ahead of time in order to effectively support their communication (Wilson and Kim, 2019). Video recording the sessions is also recommended to pick up on nonverbal communication.

4.1.1.2 Alpha and Beta Testing

In the development of digital therapies, testing phases such as alpha and beta testing play a critical role in ensuring the functionality, usability, and effectiveness of the application (Sarzynski *et al.*, 2017). These testing stages are particularly significant for co-designed applications intended for individuals with alexia or aphasia, where the target user group may face specific accessibility and engagement challenges (Roper *et al.*, 2018).

Alpha testing is the initial evaluation phase carried out within a controlled environment. This phase is typically conducted by the development team, such as software developers and researchers familiar with the intended use and goals of the therapy. The focus during alpha testing is to identify and rectify technical issues, validate core functionalities, and assess basic usability. For a therapy like the iReadMore app, alpha testing involves confirming that the app's gamified elements, interface design, and therapeutic mechanisms operate as intended. Crucially, this phase ensures that the application aligns with the abilities of its intended users, accounting for challenges faced by individuals with language impairments and comorbidities, and to ensure the

prototypes align with design contributions from participants who are involved in the co-design activities.

Beta testing, by contrast, extends the testing process to a broader, real-world user base, allowing the application to be assessed in diverse and naturalistic settings. This phase is essential for understanding how users with alexia or aphasia interact with the therapy in their daily lives and naturalistic settings, such as in their own homes and using their own devices. Beta testing generates valuable insights into user perspectives and unforeseen challenges that arise in practical usage. For example, beta testing of iReadMore might explore patterns in therapy adherence, user feedback on accessibility features, and identify unclear or unused features.

Both alpha and beta testing are integral to the iterative development process, ensuring that the final therapy is robust, user-centred, and capable of meeting the needs of its intended audience. These stages also provide data that inform broader research questions related to therapy engagement and real-world effectiveness, forming a foundation for the subsequent evaluation and refinement of the digital therapy.

4.1.1.3 Gamification

Gamification is an overarching term used to denote applying a diverse array of game design elements in non-game tasks to increase motivation and engagement. Motivation is a key contributing factor in the success of stroke rehabilitation . Increased levels of motivation can improve therapeutic outcomes for people with aphasia (Hallams and Baker, 2009; Cahana-Amitay *et al.*, 2011). A number of studies with positive clinical

findings for gamified aphasia therapies (Katz and Wertz, 1997; Des Roches *et al.*, 2015; Marshall *et al.*, 2016; Conroy *et al.*, 2018). Conroy et al. (2018) reported anecdotally that users found their gamified therapy 'especially engaging and motivating' and the authors believed gamification contributed to the significant clinical gains by engaging users' executive and attentional functions, in addition to the speech production system, resulting in improved learning and retention. More generally, a number of studies have found commonly-applied game design elements do not tend to appeal to older populations, and can be regarded as either valueless or pressurising (Gerling, Schild and Masuch, 2010; Altmeyer, Lessel and Krüger, 2018). However, the same game design elements will have different effects in different applications. Therefore, it is recommended to conduct context-specific research on gamification (Deterding, 2013). Despite the positive clinical findings mentioned previously, there is a lack of studies on the views of people with aphasia on gamification.

4.1.1.4 Self-Determination Theory (SDT)

SDT proposes a framework for the study of motivation reliant on the fulfilment of three innate psychological needs; autonomy (feeling free to choose), competence (feeling capable of doing something), and relatedness (feeling connected and supported by others) (Deci and Ryan, 1985; Ryan and Deci, 2000). In terms of aphasia therapy, patients are more likely to engage with a therapy when they feel they have a choice in doing so (autonomy), they feel the therapy will help them achieve their therapy goals (competence) and they feel supported in this activity by family members, support groups and/or clinicians (relatedness) (Coppens and Patterson, 2018). However,

considering people with aphasia have a communication impairment it can be difficult to establish a sense of autonomy, competence and relatedness.

SDT proposes that motivation towards a particular behaviour can be explained along a continuum of three types of motivation that regulate behaviour in order to fulfil the innate psychological needs, these are; intrinsic motivation, extrinsic motivation and amotivation (Deci and Ryan, 1985; Ryan and Deci, 2000). Multiple motivational drivers can coexist and influence behaviour. Intrinsic motivations are self-derived, extrinsic motivations are derived from external sources and amotivation is a lack of motivation towards an action. In the context of rehabilitation, an intrinsic motivation would be undertaking a therapy with the intention of making a valuable therapeutic gain, an extrinsic motivation would be undertaking a therapy to seek external reward (e.g. praise from a family member or clinician) or to avoid punishment (e.g. being discharged from a therapy programme) and amotivation would be feeling incapable of undertaking a therapy or viewing the therapy as unbeneficial (Vansteenkiste, Niemiec and Soenens, 2010). This theory is often applied to gamification research and has significant parallels with theories of motivation proposed in aphasia rehabilitation literature such as the person-centred, life participation (Chapey *et al.*, 2012) and social approaches (Simmons-Mackie, 2012) to aphasia intervention.

4.1.2 Ethics

Ethical approval for this study was granted by the UCL Research Ethics Committee (Project ID: 15423/001). All participants provided written informed consent prior to sessions commencing.

4.1.3 Study Design

Five in-person co-design sessions were held between June 2019 and January 2020 at the Institute of Cognitive Neuroscience, University College London in an accessible location. The sessions were facilitated by a multidisciplinary team of speech and language therapists (VF, EU), a clinical psychologist (CD) and a medical design engineer (TL). All facilitators had completed professional training in qualitative health research at University College London and/or had prior experience in facilitating focus groups with people with aphasia. An app developer also observed the sessions. Sessions were limited to 4-6 participants to allow for group discussions without restricting each participants' time to contribute. The number of sessions conducted was based on the iterative framework analysis process which was conducted after each session to reflect on whether subsequent sessions would be beneficial to further investigate the areas of interest. Sessions lasted between one and two hours including breaks and time for refreshments; further details are provided in Box 1.

Group discussions were held in a communal meeting room. When participants were testing the app prototypes, they could decide to do this in the meeting room using headphones or in a private side room which provided less distractions. Semi-structured questions were used to guide the discussions and were provided to all facilitators prior to the session. A framework analysis was conducted after each session to reflect on the discussions and develop the session guide and materials for the next session.

Session Structure

The content of the sessions varied, but all contained the following core structure:

1. **Welcome and introductions** (5-10 minutes) – participants were welcomed and introduced to one another. Facilitators introduced themselves and basic participation tips for the sessions were provided.
2. **iReadMore instructions** (5 minutes) – instructions for using the therapy were delivered by a member of the research team using a presentation and live demonstration, followed by answering questions from the group. In later sessions, this was replaced by an instruction video co-designed by participants which was tested for inclusion in the app.
3. **Independent use of the app** (10-15 minutes) – following this, the latest prototype version of iReadMore therapy was tested on an Android tablet device, followed by an open discussion of first impressions of the therapy.
4. **Group discussion/ideation** (20-40 minutes) – afterwards, discussions would lead into a problem and idea generation session using a pre-planned, semi-structured session guide.
5. **Refreshments and open discussion** (20-40 minutes) – Finally, participants were offered refreshments and were able to talk freely. This gave participants the time to make any further points they would like and ask further questions in a less structured manner.

Box 1 - Co-design focus group session structure

4.1.3 Procedure & Co-Design

After participants were welcomed and provided informed consent, the aims of the co-design process were presented along with participation tips for the group discussion. Following this, participants tested the latest app prototypes by independently using the therapy with provided instructions. Facilitators would observe one or two participants' interactions with the app. Facilitators assisted participants if required and made notes on any difficulties they were encountering.

Discussions began by asking participants about their experience of testing the therapy prototype. This would then lead into a semi-guided discussion based on pre-selected topics targeting key aspects of the therapy design, settings, functionality, interface, accessibility issues and motivational/gamification concepts. Issues or difficulties raised during the interaction with the app acted as starting points for the co-design process and participants then collaborated with each other and the facilitators to generate potential design solutions to address these issues. Where participants had a difference of opinion on the value of a design concept, an effort was made by the facilitators to see whether it could be refined in a way that led to a consensus. In addition, the mechanism of action of the therapy was not altered in the co-design process as this was previously demonstrated to be clinically efficacious. If a co-design concept could potentially preclude therapy effectiveness or participation for other users (e.g.; for those with visual or hearing impairments), it was highlighted and withdrawn from the process. The participants' co-designed ideas were then developed further in collaboration with the research team and app developer using mock-ups and prototyping software, and taken to the following co-design session for the next group to try out.

In order to facilitate total communication and analysis of nonverbal output, the sessions were video recorded by two video cameras and a variety of resources were available to participants including paper, pens, visual analog mood scales and printed visuals of the app. Questions to participants were also presented with visual aids to support comprehension. All notes and drawings made in the sessions were scanned and used alongside the video recordings and transcripts in the data analysis. To support the inclusion of participants with moderate to severe communication impairments, participants could bring a partner/carer or be paired with a speech and language therapist to help facilitate participation. After the session, participants were contacted via phone or email to enquire if they had any further comments they wished to contribute.

Following the completion of the co-design group sessions, one-to-one beta testing sessions were held to further refine the outcome of the co-design process and prepare the app for public release. This phase was conducted remotely due to the coronavirus pandemic. A further twenty-five participants were recruited through our mailing list and social media for the remote testing phase. Participants were either provided with a tablet containing the iReadMore app or they downloaded iReadMore onto their personal device using the TestFlight App on iOS. Participants in this phase tested the app for a period ranging from 5 to 14 weeks and provided feedback on subsequent versions at monthly catch-ups and in-between when issues arose.

4.1.5 Data Collection and Analysis

Video recordings, notes and drawings from participants and facilitators were analysed using framework analysis; which utilises a process of iterative refinement of themes in a data-driven approach(Richie and Spencer, 1994). Transcripts were developed from the session videos for annotation purposes. Both the videos and transcripts were analysed to ensure non-verbal data (such as gestures and expressions) were not lost in the transcription process. Framework analysis was selected for its suitability in analysing qualitative data at a group-level in research that has a specific goal-based intention, such as co-design. There are five interconnected stages to framework analysis and these were conducted in this study as described in Box 2. The analysis was conducted in Microsoft Excel by two researchers (TL, VF). Where disagreements occurred over codes, the two researchers would discuss their conflicting interpretations and aim to reach a consensus, potentially generating new codes as a result. Data saturation was discussed by the two researchers coding the data who jointly decided when saturation had been achieved based on no further themes and codes being generated after the focus groups.

Framework Analysis

1. **Familiarisation** – the data was studied in order to gain an insight into key concepts and recurrent themes. After each session, new data was analysed. This allowed for initial codes and themes to be generated. After all sessions were complete, the dataset was analysed again in full.

2. **Identifying a Thematic Framework** – emerging themes and subthemes were established and developed through discussions between the researchers. Data summaries were produced to represent the data in a succinct format.

3. **Indexing** – the generated codes and themes were applied to the data summaries. Although not part of the framework analysis, related quotes were also identified and sorted.

4. **Charting** – data summaries were reorganised under the generated themes in the framework and rewritten in a more abstract manner to reflect the theme.

5. **Mapping and Interpretation** – after charting, theme summaries were generated to represent the findings at a high-level in the context of the research question. Descriptions and interpretations of the themes are presented below. Explanations and insights into the themes are considered in the Discussion section.

Box 2 - Framework analysis methodology

4.1.6 Participants

Participants were recruited using stratified purposive sampling with convenience sampling through the research group and institutional mailing lists and other individuals known to participants in this study. Participants included people with chronic alexia and their partners/carers. We aimed to get a diverse group of participants by stratifying for age, gender, experience with digital devices, and commonly co-occurring stroke morbidities; such as physical, visual, auditory and cognitive impairments.

Table 2 reports the participant demographics. Twenty-five participants took part in 1 of 5 co-design sessions (4-6 participants per group). Participants varied in age from 29 to 78 (M = 57, STD = 12.0) and 52% were female. Nineteen had central alexia (alexia and aphasia), three had pure alexia and hemianopia, and three were partners or carers to someone with acquired alexia. Prior experience with technology varied, 10 participants had gained substantial experience using one of our digital therapies in a previous clinical trial, while another six participants had never previously owned a smartphone or tablet.

Table 2 – Demographics of 25 participants in the co-design.

Participants (females)	25 (13)
Mean age (range)	57 (29 - 78)
Diagnosis: - Central Alexia (Alexia and aphasia) - Pure Alexia - No Alexia (Partner/Carer)	 19 3 3
Prior technology experience: - Has a smartphone or tablet - Has never owned a smartphone or tablet - Previous participant in digital therapy app research	 19 6 10

4.2 Study 2: Preliminary Findings of the iReadMore Rollout Trial

4.2.1 Ethics

Ethical approval for the iReadMore Roll-Out Study was obtained UCL Research Ethics Committee (ID: 7609/001). The trial protocol was pre-registered on www.clinicaltrials.gov (NCT04849091). The participant information sheet was provided in written and auditory forms. All participants gave informed written consent in accordance with the Declaration of Helsinki.

4.2.2 Study Design

An online roll-out trial is being conducted to evaluate the clinical effectiveness of iReadMore in real-world users for improving single-word reading accuracy and speed. This work was delayed by the COVID-19 pandemic and will continue after the completion of this PhD. Preliminary findings are presented here in this thesis.

Participants self-enrolled in to the study via the app, participation was not mandatory and users could utilise the app fully without enrolment. The self-registration and testing processes were co-designed and beta-tested to enable independent, remote testing. A within-participant evaluation of the co-designed iReadMore therapy app was conducted. The primary outcomes are reading accuracy and reaction time for matched trained and untrained words. Interval testing was conducted after the completion of every 5 hours of therapy. The quantitative secondary outcomes are self-reported measures of impairment and activities of daily living.

Qualitative feedback was also analysed using a thematic analysis to investigate participants' perspectives on their own reading abilities, experiences of using the app and any other general comments they wanted to add. Qualitative feedback is provided by typing which may limit participation from participants with aphasia or alexia, however developing an audio recording feature or alternative functionality was not feasible within the timeframe and funding for this project due to the technical complexity.

Only participants who conducted more than 20 hours of therapy were included in the quantitative data analysis. Users who have reported a childhood reading disorder (e.g. dyslexia), are in the acute phase (first 6 months after diagnosis) or are under 18 were excluded from the analyses. Further analyses were conducted on the influence of therapy frequency and intensity on the therapeutic outcomes.

4.2.3 Sample Size

The sample size calculation aimed to estimate the number of real-world users required to complete a comparable therapy dose to the iReadMore trial (35 hours) to demonstrate effectiveness. Calculations were based on a significance level of 5% (alpha) and a power of 90% (beta), using data from previous studies with iReadMore prototypes in stroke patients with chronic pure alexia (Woodhead *et al.*, 2013) and chronic aphasia (Woodhead *et al.*, 2018).

For participants with pure alexia, improvements in word reading speed of 149.0ms (SD = 214.5, Cohen's d = 1.38) were observed in the previous study. For those with chronic

aphasia, improvements included an 8.7% increase in word reading accuracy (SD = 6.3) and a 100ms reduction in reading time (SD = 97.8, Cohen's d = 0.98).

Using an online sample size calculator (obtained from https://www.dssresearch.com/resources/calculators/sample-size-calculator-average/), a required sample size of 18 was determined for word reading reaction time (test value = 149ms, SD = 214.5). For word reading accuracy, the required sample size was 204 (test value = 8.7%, SD = 6.3).

4.2.4 Participants

Participants are entirely self-selecting. Real-world users of the iReadMore app could opt-in to allow their in-app data to be collected pseudonymously for this study. The app is downloadable onto Apple and Android phones and tablet via the Apple App Store and Google Play Store with a one-week free trial. Continued use costs £5 per month. Previous participants in any study from the UCL Neurotherapeutics Group including participants of the National Hospital for Neurology and Neurosurgery Intensive Comprehensive Aphasia Programme are able to use the app freely for a period of 3 months. In order to use the app, participants require access to a tablet or phone device and an internet connection. Members of our mailing list and other groups' mailing lists were notified about the app and study as well as through social media. Further recruitment activities included online and in-print advertisements and presenting at stroke support groups and patient and public awareness events. The app was featured in articles in UK national newspapers as well as specialist subject-specific publications.

4.2.5 Inclusion and Exclusion Criteria

The inclusion criteria in this study are:

- Participants must be over 18 years old
- Diagnosed with stroke, dementia, PPA, brain injury or brain tumour
- Diagnosed or self-diagnosed with an acquired impairment of reading ability (alexia)
- Willing and able to give informed consent for participation in the study

The exclusion criteria are:

- History of developmental reading or speech and language disability (such as dyslexia)
- Participants who are in the acute phase of their stroke (<6 months after diagnosis)

4.2.6 Intervention: iReadMore Word Reading Therapy

iReadMore is a self-led therapy for improving single word reading accuracy and reaction time in individuals with acquired reading impairments. It is intended to be used at home independently by persons with aphasia or alexia. iReadMore has been demonstrated to improve reading speed and accuracy in pure and central alexias (Woodhead *et al.*, 2013, 2018). Thirty minutes of daily practice is recommended to users; this is based on the findings of the iReadMore RCT, however, the optimal therapy doses are not currently known and will be explored in the present work. iReadMore is available on Android and

Apple mobile phone and tablet devices. It is publicly available to download on Apple App and Google Play Stores. iReadMore is registered as a CE-marked Class I medical device.

The therapy contains two phases, Exposure and Challenge. Instructions for both phases are provided before the first therapy session and are accessible at any point via the Help section of the app and on YouTube.

4.2.6.1 Therapy Part 1: Exposure Phase

In the Exposure phase, ten words are successively displayed, read aloud and visualised by an image providing a semantic representation (Figure 5). These pairings are always congruent. The user must click on a card to reveal the word to ensure active participation and attention to the task. Previously, the words appeared without prompting making the participant's involvement passive. The exposure phase primes participants to the correct pairings of written and spoken words that will be tested in the subsequent Challenge Phase. This is hypothesised to reinforce the correct associations between the orthographic, phonological and semantic representations of words, resulting in improved word reading abilities.

In both phases of the therapy, after words are displayed, a visual backwards mask appears to suppress afterimages of the word remaining in the participants' visual field. This allows for the precise control of the exposure time that the user has to read the word.

4.2.6.2 Therapy Part 2: Challenge Phase

In the Challenge Phase, the 10 words from the Exposure Phase are tested. Simultaneously, a written word and a spoken word are presented and unlike the Exposure Phase, these trials are not always congruent (Figure 6). Instead, the word written may be a distractor word that is phonologically similar to the spoken word; an example being [throw] and /through/. The user must decide whether the words they read and heard are the same (a congruent trial) or different (an incongruent trial) by clicking one of two buttons. Once the user has responded, audio and visual feedback are provided to inform the user of correct and incorrect responses. This feedback is provided to support error-reducing learning.

When a user correctly responds to a challenge item they gain a point, an additional point is granted for fast responses and for five correct responses consecutively. Points are needed to successfully complete a Challenge Phase; the number of points required depends on the user's therapy level. Challenge Phases end if the therapy points are achieved or if 30 trials have been completed without the score being obtained and this is considered an unsuccessful Challenge Phase.

There are two algorithms determining the difficulty of the therapy content, an item-specific difficulty algorithm and a non-item-specific difficulty algorithm.

The item-specific difficulty algorithm determines how difficult the distractor words are for a given target word. This is determined by user performance on that specific target word in the previous Challenge Phases. In this way, words that the user is getting consistently correct progress to harder difficulty levels faster than a word that the same

user is performing less well on. There are three levels of difficulty (easy, medium and hard). Each target word has multiple distractor words (between 4 and 9) with varying degrees of similarity to the target word. More orthographically similar words and words where the initial and final phonemes of the distractor match the target word (such as ‘than’ and ‘thin’) are deemed higher difficulty. If a user performs consistently well on hard distractors, the word will be removed from the therapy (which we call ‘word mastery’). If a user is performing badly at a medium or hard difficulty distractor, the level can drop down and will remain in the therapy word list for longer. If the word seen and heard are both the target word (congruent trial), correct responses do not affect the difficulty algorithm or word mastery, but incorrect responses can decrease the difficult level. The minimum possible number of exposures of a word before word mastery is achieved is 4 times in the therapy (not including congruent trials).

A separate, non-item-specific difficulty algorithm increases and decreases based on Challenge Phase performance to determine how many points are needed to complete a Challenge Phase, how long the user has to read a word (exposure duration), and what is defined as a fast or slow response. The number of points required to complete a level starts at 20 and goes up to a maximum of 59. If a user is performing badly, they will go down a level in order to maintain that the therapy is challenging without becoming so difficult that it is frustrating and risks them disengaging from the therapy. The exposure time and response time decrease with increased therapy levels to promote faster reading, ranging from 2000ms to 100ms for exposure time and 4000ms to 2000ms for the threshold for a fast response.

4.2.6.3 Therapeutic Mechanisms

The therapeutic mechanisms of the intervention are twofold. Firstly, through repeated exposure to congruent pairings of written and spoken words, and pictures (representing the semantic meaning of a word) through paired-associate (or Hebbian) learning (Holcomb and Anderson, 1993) related to the Triangle Model approach to reading (Plaut *et al.*, 1996). Secondly, feedback on correct and incorrect trials in the Challenge Phase reinforces learning of correct responses in line with reinforcement learning and error-reducing learning (Middleton and Schwartz, 2012).

The therapy also utilises a mass practice of an impairment-based therapy approach with unlimited access to enable high therapy doses. Gamification and co-design are employed to ensure accessibility of the therapy and drive motivation and engagement for the self-led user group; an approach rooted in the self-determination theory.

4.2.6.4 Training and Testing Stimuli

iReadMore includes 590 of the most frequently written words (SUBTLEXWF > 50) from the SUBTLEX lexical database (Brysbaert and New, 2009). High frequency words were chosen to maximise the ecological utility (or relevance) of the training to everyday language use. This was chosen as word reading improvements from the therapy are not expected to generalise beyond trained items. All words are between 3 and 6 letters to ensure they can be read in one fixation. All word classes (nouns, adjectives, verbs, function words etc.) are represented in the word list, including words that have low imageability or concreteness. Words with hyphenation or punctuation have been

excluded and only one morphological variant of the same word (such as only including ‘run’ and not ‘ran’ and ‘running’ additionally).

From the full word list, three matched lists of 180 words were generated, these are referred to as Lists A, B and C. For each word on list A, there is a corresponding word on lists B and C matched for letter and syllable length as well as imageability and frequency. In the iReadMore RCT, all 590 words were tested at two baseline timepoints and a customised set of matched words from A, B and C were generated based on their individual profile. However, this was not deemed feasible in the release version, therefore it was decided that words would be pre-matched across the lists and users would begin training on all List A words in a randomised order, so each user starts with a different word list. List B words are added into the therapy when word mastery is achieved for a List A word, this would be the corresponding matched word (e.g. item B01 always replaces item A01). Therefore, the lists remained matched for psycholinguistic variables. An additional list of 50 ‘Core’ words includes the highest frequency function words are included as a separate list which remains in the therapy throughout due to high ecological utility. These are not tested in the Word Reading Test due to low imageability and abstractness (e.g. ‘The’), which was deemed inappropriate for inclusion in a word-to-picture matching task.

All words have multiple images relating to the semantic meaning of the word that are used in the exposure phase of the therapy and separate images that are used for the test (expect for core words, which are not tested). Different images are used to prevent the occurrence of users learning to recognise the correct images and not the correct lexical item.

4.2.7 App Registration

When the user downloads and opens the app for the first time, they must complete a comprehensive battery of language and cognitive assessments, and complete demographic data prior to initiating the therapy. This process takes around an hour to complete and breaks can be taken at any point in-between tests on the hold screens, with reminders to consider taking a break and returning to the testing when the user feels they are ready. The registration process was extensively co-designed and beta-tested prior to release as this was an area of particular challenge in making the process intuitive for the user and in order to provide reliable data collection. Only one baseline time-point was used as it was deemed that real-world app users would not tolerate having to complete multiple, spaced baseline assessments, when their intentions of downloading the app would be to using the app for therapeutic purposes and not research participation.

The tests completed at baseline are described in detail in the following sections. The data collected at baseline is used in Chapter 4 as well as the current chapter of my thesis.

4.2.8 Baseline Behavioural Assessment

4.2.8.1 Demographics

The following demographic data is collected through the in-app registration process:

- Age (year and month only)

- Gender (categorical)
- Cause of reading impairment (categorical); users with PPA are further asked to provide their PPA subtype (if known)
- The date that their reading impairment began or was diagnosed (year and month only)
- Presence of childhood reading impairment (yes or no)

4.2.8.2 Word Reading Test (WRT)

The WRT is used to assess reading of written single words at each time point. The WRT implemented in the app is based on a testing protocol devised by Dr Zoe Woodhead for the iReadMore RCT (Woodhead *et al.*, 2018). This outcome measure has been adapted for self-completed, digital delivery by people with aphasia.

The WRT is a word-to-picture matching task where a word is displayed and users are instructed to click on the screen as soon as they have finished reading in order to collect reaction time data. After the screen is clicked, a backwards mask is displayed to control for the duration of stimulus exposure to the written word. Following this, four pictures are displayed, one corresponding to the target word that was just presented and three unrelated distractor images. Distractors for a target word are pre-set and not randomised to ensure that they are not phonological or semantic distractors. The location of the target words and distractor on the screen is randomised. The participant taps the screen to select an image and is not timed for this task. Participants are given up to four seconds to read the word before the trial moves on to the picture matching

task. This time limit was implemented to prevent participant fatigue and to limit the amount of time required to complete the registration phase. Accuracy for the trial is determined by correct or incorrect picture matching. The order of multiple choice responses on the screen is randomised.

The baseline test contains 132 trials. 100 trials are used to establish word reading ability as described in this thesis section. An additional 32 words are used in a WRT subtest to investigate alexia subtypes in a real-world population; this is detailed further in Section 4.2.8.3 below.

At baseline, 50 pairs of matched words (100 trials in total) are selected from List A (that will be trained in the therapy) and List C (words that will not appear in the therapy); such that word A01 is matched to word C01. The word list for the test is selected randomly for each user from the full word lists. The word pairs are matched for psycholinguistic variables. The original test at baseline included all 590 words in the study. This was abridged to 100 words due to the substantial length of time that is required to complete the testing which was concluded would not be tolerated by participants who are downloading an app to conduct therapy independently at home outside of a controlled, research environment.

Picture matching precludes the inclusion of low imageability words (such as function words), therefore, 50 function words (referred to as Core Words) were removed from the testing, but remain in the therapy. In the iReadMore RCT, reading accuracy and speed improvements for core words were in-line with the other general lists of trained words (Word Lists A and B). In addition, there was no evidence for the influence of word

imageability or regularity on the therapy effects, suggests that the therapy can be effective for all word types (Woodhead *et al.*, 2018).

To ensure that the participant has a good understanding of the task prior to the test commencing, instructions are delivered in written and audio format, followed by a reactive practice test of up to 10 trials or until 4 trials have been answered correctly without prompts in a row. The practice test incorporates errorless learning as users must click on the correct image to progress and provides feedback after each trial to ensure understanding of the task.

4.2.8.3 WRT subtest - Alexia Subtyping Test

The additional 32 trials in the WRT were designed to investigate alexia subtyping and are only completed at baseline. These trials are randomly distributed through the baseline reading test. Unlike the 100 trials of Lists A and B items, two of the three distractors in these trials are semantic and phonological distractors, with the third being unrelated. The 32 target words (List D) and 96 distractor words are distinct and do not appear in the other WRT word lists or the therapy content. The target words were selected for their high error-production rates in individuals with alexia, excluding low-imageability words for compatibility with the picture-matching task.

List D was validated using healthy age-matched controls via Gorilla Experiment Builder and Amazon Mechanical Turk. This ensured the clarity of images, appropriate matching to corresponding words, and the absence of outliers in error rates or reaction times for healthy, control individuals. Validation began with a practice round where participants

had to correctly respond to at least 6 out of 8 trials which included only unrelated distractors before proceeding. Unsuccessful participants repeated the practice session after reviewing instructions.

The experiment was completed three times with revised word and image lists until the error rate and reaction times were satisfactory. In each experiment, twenty control participants completed the 32 trials, with randomisation in terms of both the task order and the order of response options on the screen. In between experiments the results were reviewed and the list was amended according to reduce error rates and reaction times. Tasks generating multiple errors or average reaction times exceeding 2000ms were revised, either in terms of the images used or the distractor words being amended. Qualitative feedback provided by participants at the end of each experiment further informed these refinements. Two qualitative responses were collected, one participant noted their slow internet connection may have been delaying their responses times and another was aware of a mis-selection that they made. These examples were therefore excluded from the analysis.

Errors were categorised as semantic, phonological, or unrelated. Across all iterations, semantic errors were the more common, but were deemed acceptably uncommon. The overall error rate decreased from 3.8% to 2.0% over the three experiments, with no repeated incorrect responses for the same tasks across participants in the final iteration. Reaction times remained stable across tasks and the repeated experiments with no outliers collected aside from one individual reporting slow loading times.

4.2.8.4 Children's Sustained Attention to Response Task (cSART)

The cSART is a non-verbal version of the Sustained Attention to Response Task (Manly *et al.*, 2000) that is more suitable for people with alexia or aphasia. It is a domain general test of sustained attention was used to assess each participant's ability to concentrate. It was also used as a non-language control measure at interval testing time points. It was not expected to change with therapy usage, and was therefore used as a control measure to investigate specificity of treatment effects and to explore for test-retest effects from users repeating the same tests multiple times.

This Go/No-Go task contained pictures displaying one of two different people (see Figure 8), one of which was displayed in each trial. Participants were instructed to tap anywhere on the screen whenever the Go trial image was displayed and to withhold tapping for the alternative No-Go trial image. 200 Go trials and 25 No-Go trials were presented over approximately 4 minutes in a pseudorandomised order. As No-Go trials were infrequent and unpredictable, to be accurate, the patient was required to maintain attentive across the duration of the test in order to successfully inhibit a response on the No-Go trials.

Due to the complexity of delivering this task remotely, two practice rounds were implemented in the registration process. The first had 10 Go trials, and the second had 8 Go trials and 2 No-Go trials. A correct response rates of 60% was needed to progress to the test. Otherwise, the practice test and instructions were repeated.

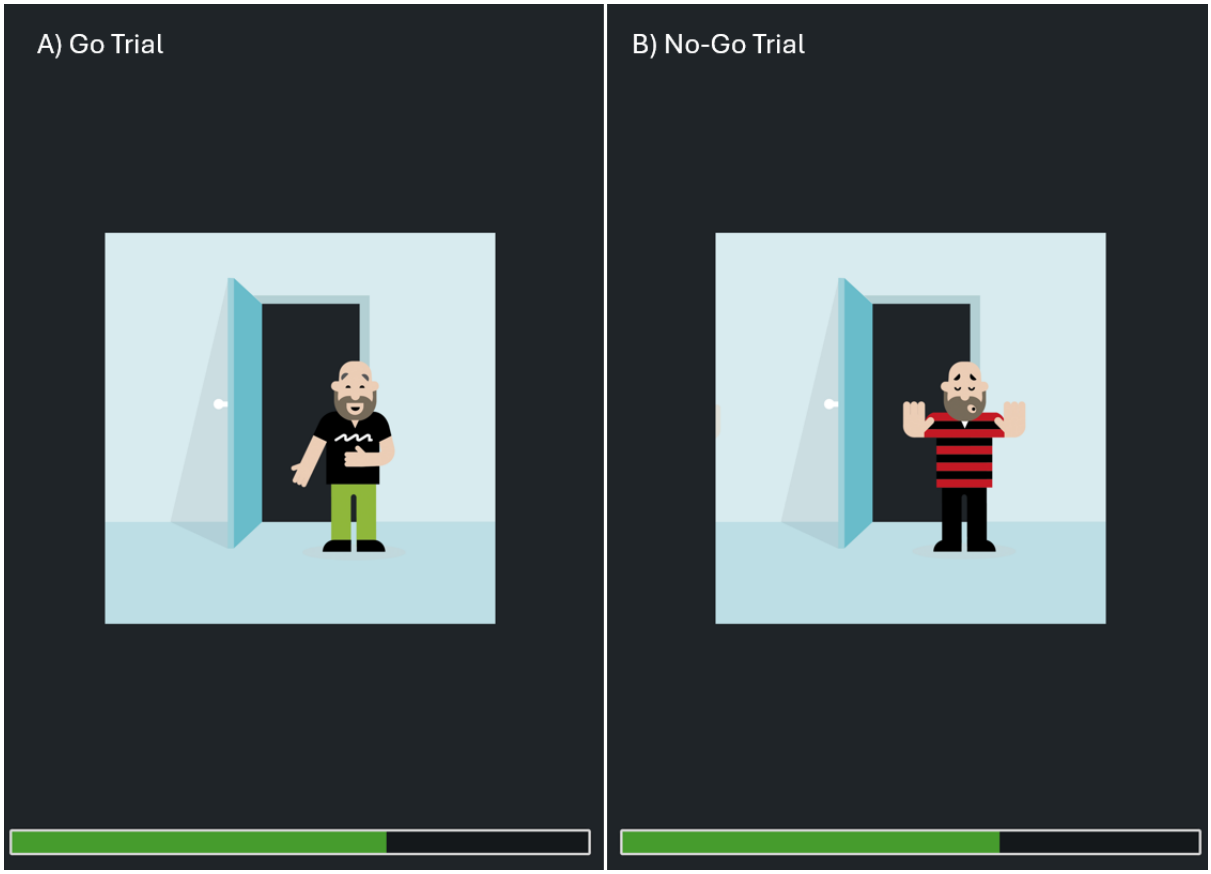

Figure 8 - cSART Go and No-Go trials

4.2.8.5 Spelling Test

The presence of a writing disorder can distinguish pure and central alexia (Leff and Starrfelt, 2014). To test participant's spelling ability, the stimuli and trials from the writing subsection of the CAT (Swinburn, Howard and Porter, 2004) have been adapted for use within the app. This test is scored according to the instructions in the CAT.

4.2.8.6 Written Picture Description

This connected writing test assesses the users' appropriate use of grammar and syntax as well as individual word spelling. Participants are presented with a line drawing of a detailed scene with several elements and asked to provide a written description of it. The stimuli and scoring system is taken from the CAT (Swinburn, Howard and Porter, 2004).

4.2.8.7 Visual Fields Test

Pure alexia is often accompanied by a loss of vision to the right side of space (known as right homonymous hemianopia) which impairs reading ability. This phenomenon is not associated with central alexia. Therefore, Visual Fields Test (VFT) test can be useful for differentiating between pure and central alexia types. The VFT is a validated, digital visual field test has been developed for participants with hemianopia that does not need to be completed unaided by a clinician (Koiava *et al.*, 2012).

Participants are asked to hold the tablet or phone at approximately arm's length in a comfortable position, either being held or placed on a table. Participants are instructed to focus on a red cross in the centre of the screen at all times during stimulus presentation (see Figure 9). A pattern of 1-4 dots will appear momentarily at predetermined locations on the screen during stimulus presentation. Participants are presented with 21 stimuli. Each is followed by 4 images, one of which will reflect the stimuli presented. The other three are close distractor variations based on the stimuli.

Users are then instructed to pick from the 4 options and select the option that matches the pattern that they saw. Participants can view the dot pattern again once if needed.

The red cross is outlined in yellow to improve visibility. The background is a light grey and the dots are presented in darker greys; the contrast increases as dots appear further from the cross in the centre of the screen according to a colour matrix.

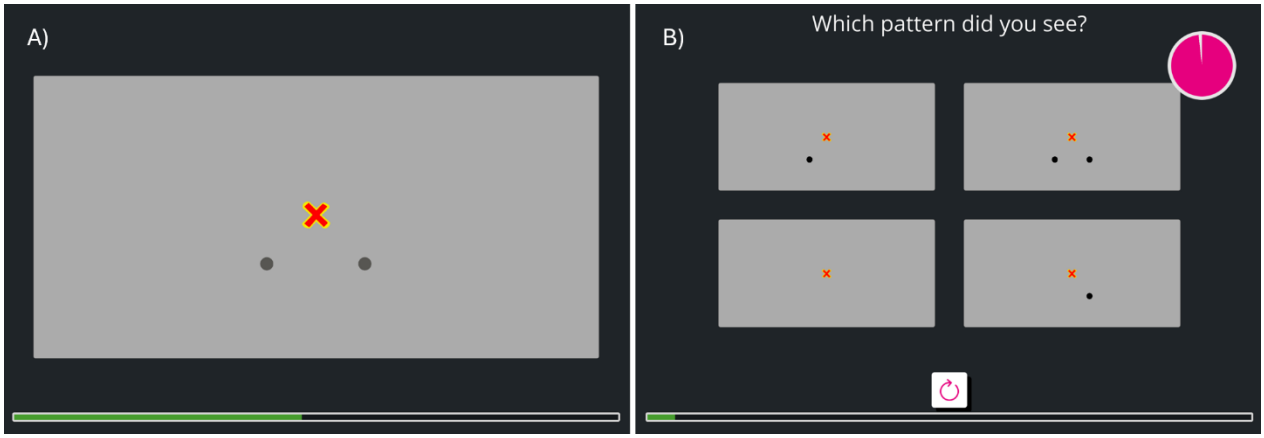

Figure 9 - Visual Fields Test. A) Presentation of the stimuli. B) Users are asked to select the option that matched the to the stimuli.

4.2.8.8 Patient-reported Outcome Measures (PROMs)

In order to assess the users' perceptions on their own reading abilities and performing selected activities of daily living, participants are asked to rate how easy they find:

- Reading words
- Reading sentences
- Reading text
- Writing
- Speaking

- Using your phone or tablet
- Remembering where you left something (such as your glasses)
- Moving around at home
- Travelling
- Shopping

Responses are reported using a 7-point Likert scale from ‘Easy’ to ‘Hard’ with visual support (Figure 10). A second question asks users to select which items they have been able to read recently from a list of functionally-relevant written item categories including:

- Packaging or instructions
- Street signs
- Letters, postcards, or emails
- Newspaper or magazine articles
- Books

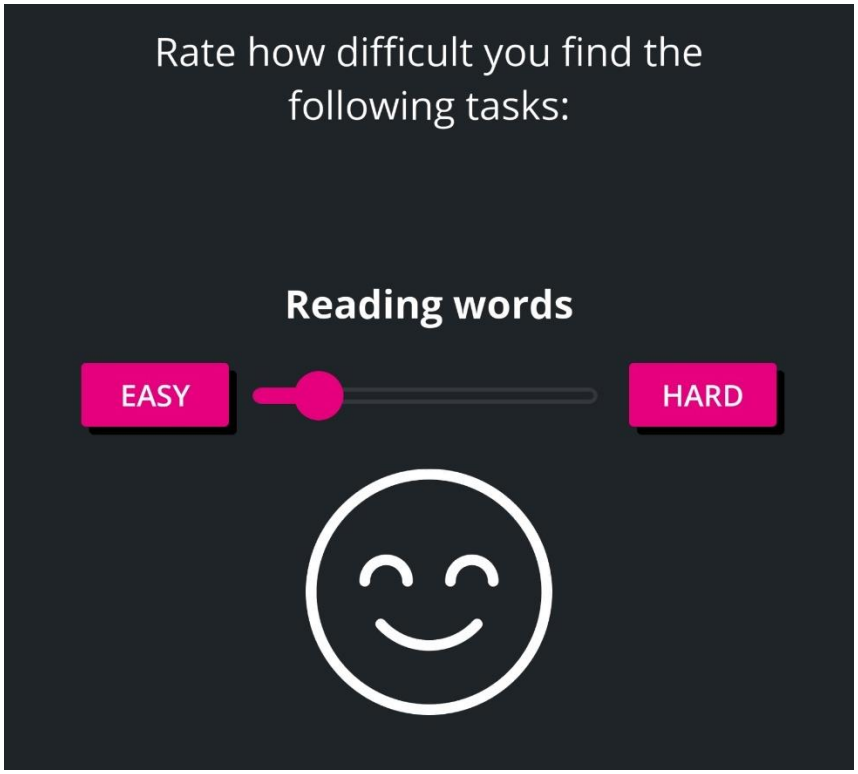

Figure 10 - PROMs Likert scale

4.2.9 Interval assessments

4.2.9.1 Word Reading Test

The WRT (as described previously) is the primary outcome measure for the study. The outcomes from the test are change in word reading accuracy and reaction time comparing trained and untrained items. Reaction times are calculated using correct trials only, and excluding trials where the RT was more than 2 standard deviations from the subject's mean. After every 5 hours of therapy are completed, the WRT is repeated. The interval tests contain 50 trained words from Lists A and B matched with 50 untrained words from List C. Words on Lists A and B are ranked to ensure the most trained words from the last 5 hours of therapy are used in the test. In the rare

circumstance that less than 50 words are trained in the therapy, the test is shortened to ensure only trained words are tested.

4.2.9.2 cSART

The cSART is delivered at every interval time point as a control measure to the treatment effect as this outcome is not expected to improve. The outcome measures were the number of false negative (when the user withholds on a Go trial) and false positive responses (when the user presses on a No-Go trial), and the mean RTs on correct Go trials only.

4.2.9.3 Quantitative PROMs

The Quantitative PROMs are completed after every 5 hours of therapy The Quantitative PROMs are analysed within-participant for change over the therapy time-points as a secondary outcome measure using mixed linear regression.

4.2.9.4 Qualitative PROMs

The qualitative PROMs are completed every 10 hours to provide further detail, however, these are optional in case the participant is unable to answer. The qualitative outcomes are analysed using a qualitative thematic analysis to provide further insights into user

experiences of the app and perceptions of the therapy and potential translation to participants' functional reading abilities.

Three questions are asked in these sections with participants responding using free-text boxes:

- Has your reading ability changed since starting iReadMore therapy?
- What are your thoughts on the therapy?
- Any other comments?

4.2.9.5 Mixed Linear Regression Models

Mixed Linear Regression was employed to analyse the data obtained from the ongoing rollout trial and evaluate the significance and therapy effects of the iReadMore digital reading therapy intervention in individuals with acquired reading impairments. The primary outcomes assessed are changes in reading accuracy and reaction time on the Word Reading Test across multiple timepoints. Two groups of items, trained and untrained, were evaluated in a within-participant analysis. Trained items consist of words directly practiced within the iReadMore therapy, while untrained items are matched words that do not appear in the therapy.

Mixed Linear Regression Modelling was applied because of its suitability to aspects of the dataset, including a small sample size, influence of random effects, and repeated measurements in a within-participant design. Mixed Linear Regression effectively accounts for repeated measures, capturing the inherent dependencies in longitudinal data. Additionally, it accommodates unbalanced datasets, allowing the inclusion of

participants with missing data at specific timepoints as seen in self-led digital therapy usage.

Analyses were conducted in Python, using Restricted Maximum Likelihood Estimation (REML) to estimate parameters. REML was chosen over Maximum Likelihood because of the ability to provide unbiased estimates of variance, particularly relevant to smaller sample sizes or models with random effects, such as real-world datasets.

4.3 Study 3: Prediction of Therapy Outcomes Using Therapy Data

This chapter presents two studies on therapy outcome prediction using data collected from the RCTs of iReadMore and Listen-In (a digital therapy for spoken word comprehension therapy). The intention of both studies was to explore the feasibility and suitability of using only variables collected in the apps currently or which could be collected remotely (i.e. excluding variables requiring MRI scanning or in-person behavioural testing). The first study uses data from the iReadMore trial and contains limited therapy data. The second study is an expansion of the first using data from the Listen-In trial containing substantially more data on participant's therapy progress.

4.3.1 Background on Treatment Outcome Prediction

4.3.1.1 Introduction

The prediction of language outcomes for people with post-stroke aphasia has explored using clinical data, such as neuroimaging scans and behavioural assessments (Lambon Ralph *et al.*, 2010; Plowman, Hentz and Ellis, 2012; Wang *et al.*, 2013; Seghier *et al.*, 2016; Aguilar *et al.*, 2018; Hope, Leff and Price, 2018; Nouwens *et al.*, 2018; Kristinsson *et al.*, 2021; Billot *et al.*, 2022). A predictive algorithm of this type would have great clinical utility in supporting decision-making on treatment options for aphasic patients with greater certainty. Aguilar et al. (2018) investigated prediction of therapy outcomes for users of iReadMore. Predictions of therapy responses for out-of-sample individuals were generated from a combination of demographic, behavioural and lesion location data. The predictions were significantly correlated with the actual therapy responses

(Figure 11). However, the model was overfitting, likely due to a small sample size of 23 participants. The authors reported that they were unable to constrain multivariable models or effectively evaluate the predictive power of the model due to the sample size. This study was only intended as a preliminary investigation, however it was the first study on prediction of therapy outcomes for a reading rehabilitation therapy.

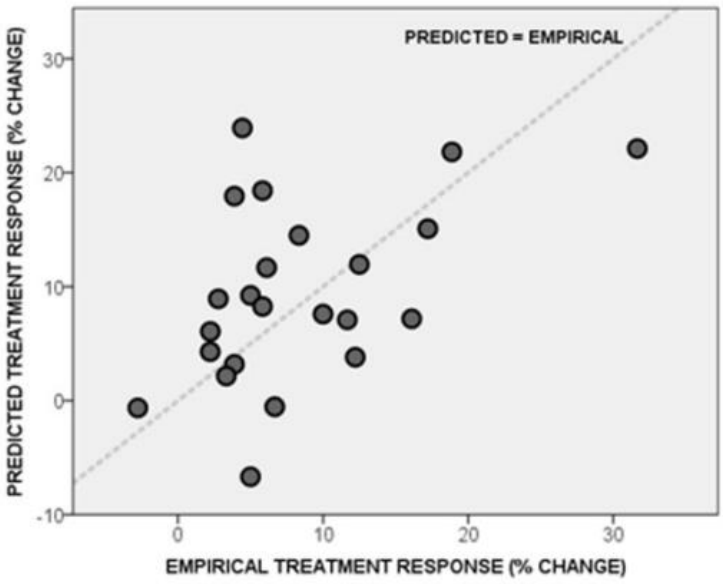

Figure 11 - Predicted treatment responses versus the observed treatment responses to therapy. A perfect prediction would along the diagonal dotted line (y=x). (Obtained from Aguilar et al., 2018)

4.3.1.2 Machine Learning

Machine learning (ML), a branch of artificial intelligence, develops algorithms that learn patterns and make decisions from data without explicit programming. Unlike traditional rule-based systems, ML adapts dynamically, offering transformative potential in fields like healthcare by supporting personalised treatment planning, enhancing diagnostic accuracy, and optimising clinical outcomes.

ML encompasses several core paradigms, each suited to specific data types and problem-solving approaches. Supervised learning, widely used in tasks such as disease diagnosis and treatment outcome prediction, involves training algorithms on labelled data with known input-output pairs. In contrast, unsupervised learning analyses unlabelled data to uncover hidden patterns, such as clustering patients with similar profiles or detecting anomalies. Reinforcement learning focuses on optimising decision-making through interaction with an environment and feedback in the form of rewards or penalties. Hybrid approaches, like semi-supervised and self-supervised learning, combine labelled and unlabelled data to improve efficiency, especially when annotated datasets are limited. Deep learning, inspired by neural networks, excels at capturing hierarchical data representations and has been transformative in areas like image recognition, natural language processing, and genomics.

The success of ML algorithms hinges on several considerations. Data quality and representativeness are critical, as biases or inaccuracies can lead to flawed predictions. Interpretability and ethical implications are particularly significant in healthcare, where decisions impact patient outcomes. The "black box" nature of many algorithms complicates accountability, while privacy and data security risks can arise from handling sensitive information. Overfitting, which limits generalisability, and the growing computational demands of ML models also present challenges, including environmental sustainability concerns. Addressing these issues requires robust testing, ethical oversight, and rigorous regulatory frameworks to ensure responsible deployment.

4.3.1.3 Treatment Outcome Prediction

In the context of treatment outcome prediction, machine learning holds significant promise. By analysing large datasets comprising patient demographics, clinical history, genetic profiles, and treatment records, ML models can identify predictive features associated with therapeutic outcomes. This can enable the development of personalised treatment strategies, optimising interventions, minimising adverse effects and supporting shared decision-making. As such, this chapter explores the application of machine learning to predict treatment outcomes, delving into its methodologies, challenges, and potential to transform patient care.

In treatment outcome prediction, various machine learning models can be employed, each with its strengths and limitations. These models include linear models, decision trees, random forest regression (RFR), elastic net, lasso, and support vector regression (SVR), all of which are commonly used for regression and classification tasks in healthcare applications.

4.3.1.4 ML Regression Models

Linear

Linear regression is one of the simplest machine learning models, assuming a linear relationship between input features and the target variable (Barbur, Montgomery and Peck, 1994). It is widely used for its interpretability and efficiency, especially when the relationship between predictors and outcomes is approximately linear. However, it may

not perform well in complex datasets where relationships between variables are non-linear.

Lasso

Lasso (or Least Absolute Shrinkage and Selection Operator) is a regression technique that applies L1 regularisation to promote sparsity in the model, effectively performing feature selection by shrinking the coefficients of less important features to zero (Tishbirani, 1996). Lasso is beneficial when there is a need to identify and retain only the most relevant features, improving model interpretability and reducing the risk of overfitting.

Elastic Net

Elastic Net is a regularised regression model that combines both ridge regression (L2 regularisation) and lasso (L1 regularisation) (Zou and Hastie, 2005). It is particularly useful when there are many correlated features in the dataset. Elastic Net helps prevent overfitting by imposing a penalty on the size of the coefficients, ensuring a balance between model complexity and predictive power.

Decision Trees

Decision trees are non-linear models that split the data based on feature values to make predictions (Breiman *et al.*, 2017). These trees create a series of decisions,

leading to a prediction at each leaf node. While decision trees can capture complex relationships in data, they are prone to overfitting, especially when deep trees are created without sufficient pruning.

Random Forest Regression (RFR)

An ensemble method based on decision trees, RFR builds multiple trees and combines their predictions to reduce overfitting and improve accuracy (Breiman, 2001). RFR is robust to noise and captures complex, non-linear relationships, making it a powerful choice in many real-world problems. However, it can become computationally expensive with large datasets.

Support Vector Regression (SVR)

SVR is a type of regression model based on support vector machines (Drucker *et al.*, 1997). It is effective in capturing complex, non-linear relationships by using kernel functions to map input features into higher-dimensional spaces. SVR is particularly useful when the data has outliers or the relationship between predictors and outcomes is non-linear, but it requires careful tuning of parameters such as the kernel and regularisation parameters.

Each of these models offers unique advantages depending on the nature of the data and the specific goals of treatment outcome prediction. While linear models are simple and interpretable, non-linear models like decision trees and random forests can handle more complex relationships. Elastic Net and Lasso provide regularisation to prevent

overfitting, while SVR is powerful in cases where non-linear relationships exist in the data.

4.3.1.5 Model Evaluation

The performance of predictive models is typically assessed using several quantitative metrics. Root Mean Square Error (RMSE) is a commonly used measure of the average magnitude of prediction errors, reflecting how closely the model's predictions align with observed outcomes. RMSE values are often accompanied by their standard deviation (SD), which provides insight into the variability of prediction errors across different samples. A lower RMSE and narrower SD suggest that the model's predictions are both more accurate and consistent.

The Coefficient of Determination (R^2) is another key metric, representing the proportion of variance in the observed data that the model explains. Higher R^2 values, closer to 1, indicate better model fit and stronger predictive capability. In addition, 95% confidence intervals (CIs) are used to quantify the precision of model performance estimates. Narrower CIs reflect higher confidence in the reliability and stability of the model's predictions.

By examining these metrics together, researchers and clinicians can gain an understanding of a model's accuracy, variability and reliability. This multifaceted evaluation approach ensures that models are not only statistically robust but also clinically meaningful and applicable in real-world healthcare settings.

While metrics like these are essential for evaluating predictive models, they have limitations. RMSE does not differentiate between systematic bias and random error, potentially masking meaningful patterns in the data. R^2 can be misleading when applied to non-linear models or datasets with outliers, as it may overestimate model performance. Additionally, confidence intervals rely on assumptions about data distribution and can become less reliable in small or highly variable datasets. These limitations highlight the need to interpret performance metrics within the context of the specific modelling approach and clinical application.

4.3.1.6 Prediction of Aphasia Treatment Outcomes

Aphasia treatments exhibit significant variability in how different individuals respond, leading to significant differences in outcomes. Some therapy users may show substantial improvement, known as high responders, while others may experience only marginal gains or no improvement at all, referred to as low responders. The underlying mechanisms driving this variability are complex and require further investigation (Kristinsson *et al.*, 2021). Factors such as the integrity of preserved brain regions, age and educational background have been identified as potential predictors of treatment outcomes, albeit to varying extents. This heterogeneity in treatment outcomes affects the overall effectiveness of aphasia rehabilitation and the ability to identify how potential users will respond could enhance the efficacy of treatment protocols and patient experiences. Additionally, accurate treatment outcome prediction could facilitate evidence-based clinical decision-making (Sutton *et al.*, 2020) and help set realistic patient expectations (Kristinsson *et al.*, 2023).

Several studies have explored treatment response prediction in aphasia rehabilitation (Lambon Ralph *et al.*, 2010; Plowman, Hentz and Ellis, 2012; Wang *et al.*, 2013; Seghier *et al.*, 2016; Aguilar *et al.*, 2018; Hope, Leff and Price, 2018; Nouwens *et al.*, 2018; Kristinsson *et al.*, 2021; Billot *et al.*, 2022). These studies typically use data from pre-therapy behavioural assessments and neuroimaging, including factors like initial aphasia severity, structural integrity, lesion size, demographics, and performance on pre-therapy language-based and cognitive tasks. However, research has also highlighted the potential need for additional data sources to achieve accurate and reliable predictions suitable for clinical use (Price, Seghier and Leff, 2010; Harvey, 2015).

The development of digital therapies offers new avenues for data collection at scale, providing unique opportunities to improve predictive modelling for aphasia treatment outcomes. Digital therapies, such as app-based interventions, can capture a wealth of in-therapy data, including user engagement metrics, response accuracy, progression through therapy levels, and session timing. These data sources, when combined with pre-therapy assessments, offer the potential to create richer, multimodal datasets for analysis. Such an approach could address the limitations of relying solely on static, pre-therapy variables, enabling more dynamic and individualised prediction models.

By leveraging these novel data sources, this research seeks to advance our understanding of the factors influencing therapy responsiveness and improve the overall effectiveness of aphasia rehabilitation.

In this chapter, I explore the potential of predicting treatment outcomes only using variables which can be collected via digital therapy apps (for exampling, excluding brain scan-derived variables), to investigate the feasibility of implementing a prediction algorithm into the app.

4.3.2 iReadMore

4.3.2.1 Study Design

The application of training a prediction algorithm using in-therapy data is a novel approach that has not been previously explored in aphasia therapy. This study utilises data from the iReadMore trial (Woodhead et al., 2018) to replicate the findings of Aguilar et al. (2018). However, this replication was constrained to variables that could feasibly be collected via the iReadMore app, thereby excluding neuroimaging data. Additionally, this study incorporates a broader set of in-therapy variables to enhance the predictive model.

4.3.2.2 Ethics

All participants gave written informed consent. The ethics approval for the iReadMore RCT (NCT02062619) was obtained from the London Queen Square Research Ethics Committee. Approval was granted for anonymised data from the trial to be utilised further in future research studies.

4.3.3.3 Participants

The data analysed in this study was obtained from participants who completed the Phase II RCT of iReadMore (Woodhead et al., 2018). The study included 21 participants (13 female) with chronic aphasia, including alexia. A summary of baseline statistics is presented in Table 3.

Table 3 - Baseline summary of iReadMore trial participants.

	Mean	**SD**
Age (years)	53	11
Time since Stroke (months)	59	39
CAT Naming (%)	61	28
WRT Word Reading (%)	57	30
Lesion Volume (cm^3)	163	99

4.3.2.4 Measure of Treatment Effect

Two measures of treatment effect were used in this study. These were reading accuracy and reading speed obtained from the WRT as delivered in the iReadMore randomised controlled trial. This test consists of Lists A, B and C from the WRT, but not List D which was developed for the rollout trial. In the RCT, the WRT test, which has now been incorporated into the app, was delivered on a laptop in a research setting.

4.3.2.5 Data Completeness

Across 28 variables and 588 data entries, there were 32 missing data entries (see Table 4). The most affected were variables from the Comprehensive Aphasia Test (CAT), with Picture Description and CAT Average Score missing 23.8% of entries. The overall missing data across all variables was 5.5%.

Table 4 - Data completeness for iReadMore randomised trial data.

Category	Variable	Missing Data Entries	Data Completeness (%)
Demographics	Sex	0	100
	Type of stroke	0	100
	Time post stroke (months)	0	100
	Age	0	100
	Calculated age of stroke	0	100
	Lesion volume	0	100
Word Reading Test	Accuracy	0	100
	Reaction Time Average	0	100
	Reaction Time SD	0	100
Comprehensive Aphasia Test (CAT)	Word Repetition	4	81.0
	Naming	4	81.0
	Picture Description	5	76.2
	Word Reading	4	81.0
	Average Score	5	76.2
Written Semantic Matching (Pyramids & Palm Trees)	Accuracy	1	95.2
	Reaction Time Average	1	95.2
	Reaction Time SD	1	95.2
Neale Analysis of Reading Ability	Accuracy	1	95
	Comprehension	1	95
	Words per minute	1	95
children's Sustained Attention to Response Task (cSART)	Hits	0	100
	Omission Errors	0	100
	Go Trial RT	0	100
	False Positives	0	100
	Correct Rejection	0	100
Sentence Reading	Accuracy	2	90.5
	Words per minute	2	90.5
iReadMore Initial Word Duration	Initial Word Duration	0	100
	Total	32	94.5

4.3.2.6 Data Imputation

Various imputation methods were tested to address missing data, including replacing missing values with the lowest value, mean, or median values. Removing participants with missing data was deemed unsuitable as it would lead to a loss of valuable

information in a limited dataset. Consequently, imputation techniques were applied to preserve as much of the data as possible.

Initially, missing values were imputed with the median of the respective variable. However, exploratory analysis revealed that missing data primarily resulted from participants' inability to complete tests, rather than from random factors such as missed appointments. Therefore, imputing with the median could lead to biased results and adding lower scores may provide a more realistic method of imputation. Consequently, several imputation methods were tested to improve the quality of the dataset. A first approach was to replace missing values with zeros, but this led to larger error scores and SDs, as this method assigned zero to missing reaction times or SDs—values that were implausibly perfect.

A more nuanced approach was then trialled as detailed in Table 5.

Table 5 - Data imputation strategy.

Variable	Imputation
WRT Accuracy (%), Words per minute, CAT and Neale test scores	Missing values replaced by zeros
Standard deviation variables	Medians
Reaction time variables	Highest recorded value

This strategy resulted in higher error scores and SDs compared to using the median values, suggesting that the most appropriate imputation method for this dataset is to use the median value for all missing data points. This approach balances data preservation with minimising potential bias in the dataset (Little and Rubin, 2014).

4.3.2.7 Model development and Evaluation

Data preparation, model training, hyperparameter optimisation and metric scoring were performed using Python and Scikit-Learn (Pedregosa *et al.*, 2011). The data was partitioned into three different sets: training, validation and test sets. Model development was conducted using the training and validation sets which generate the models and optimisation of hyperparameters respectively. The test set was used to evaluate the final optimised performance of the models. Multiple models were generated using linear regression, logistic regression, Random Forest Regression (RFR) and Support Vector machine Regression (SVR). For RFR and SVR models, a grid search optimised hyperparameters to generate a model of best fit. Finally, the test set was used to evaluate the performance of the models with their respective optimally-performing hyper-parameters. In order to generate more conclusive evidence of model performance for out-of-sample data, leave one-out cross-validation (LOOCV) was used. LOOCV creates a test set of one data point, while the remaining data (N-1) is used for training. This process is repeated N times (or folds), therefore a test set is created for each data point in the data set. The average error rate is then calculated to evaluate the model.

4.3.3 Listen-In

4.3.3.1 Introduction to Listen-In Therapy

Listen-In is a digital speech comprehension therapy for aphasia. It focuses on improving auditory comprehension and speech processing skills for single words and short phrases, supporting users to regain the ability to understand spoken language and

communicate more effectively. The drivers for developing the Listen-In app are aligned with those for developing the iReadMore app; to provide unrestricted, independent therapy for people with aphasia in an accessible and gamified manner. Listen-In therapy involves massed practice of spoken word-to-picture matching tasks with target words delivered as single words and in phrases or whole sentences, and incorporating both phonological and semantic foils (see Figure 12). Users receive real-time feedback and encouragement based on their performance. Approximately 870 trained words appearing in over 3000 unique challenges. The app uses an adaptive training algorithm to individualise the therapy content and difficulty to suit users' impairment profile and severity.

Figure 12 - Listen-In app therapy challenge.

The efficacy of Listen-In was tested through a cross-over randomised controlled trial involving 35 participants with chronic aphasia. The trial compared 12 weeks of Listen-In therapy (averaging 85 hours of therapy) with 12 weeks of standard care. The randomised controlled trial (RCT) of the therapy found large and significant improvements in speech comprehension when compared to standard of care for trained words (11%, Cohen's d = 1.12) (Fleming *et al.*, 2021). Therapy gains were maintained at 12 and 24 week follow-ups. These gains were item-specific, suggesting that the therapy facilitated the strengthening of particular neural networks associated with the trained words. Furthermore, baseline structural integrity, analysed using voxel based morphometry (VBM), was found to contribute to treatment outcomes. Greater volumes of white matter in distributed regions of the right hemisphere were predictive of better response to the therapy. This work found evidence for therapy-driven structural neuroplasticity. Longitudinal VBM identified improvements in speech comprehension associated with tissue changes in the bilateral temporal lobes, particularly in regions associated with speech processing. This demonstrates that targeted, digital therapies can induce measurable changes in brain structures in addition to behaviour in chronic phases of aphasia. By providing an evidence-based, self-administered therapy option, Listen-In addressed a critical need for accessible and high-dose speech comprehension treatment. It empowers patients to take control of their own rehabilitation journey and provide a feasible means for therapy delivery without overburdening limited healthcare resources.

Listen-In differs from iReadMore in permitting the user the ability to customise the therapy content based on their preferences. Users are prompted to select modules of

words that are subject-related, such as food, adapting the therapy material to the user's interests. Listen-In also incorporates different aspects of gamification to iReadMore such as mini-game breaks in between therapy sessions where users collect puzzle pieces as an additional non-therapy aspect with collectables aimed at driving long-term therapy engagement.

4.3.3.2 Study Design

Building on the approach used in the iReadMore Therapy Prediction Study that I previously conducted, the Listen-In dataset offers a richer source of therapy usage data. This dataset enables a more in-depth investigation into the use of in-therapy data for predicting therapy outcomes. Furthermore, the data structure more closely aligns with what is available in the rollout versions of the iReadMore and Listen-In therapy apps, enhancing the applicability of the findings to real-world implementation.

4.3.3.3 Ethics

Ethical approval was obtained from the National Research Ethics Service, Hampstead Committee (15/LO/0569), and the trial protocol was pre-registered on ClinicalTrials.gov (NCT02540889). All participants provided written informed consent before study participation in accordance with the Declaration of Helsinki. Approval was granted for anonymised data from the trial to be utilised further in future research studies.

4.3.3.4 Participants

Data analysed in this study was obtained from participants who completed a Phase II randomised controlled trial (RCT) of Listen-In, a spoken word comprehension digital therapy (Fleming *et al.*, 2021). Of the 35 participants who completed the trial, 32 had the data required for participation in the present study; 2 were excluded due to missing therapy data caused by hardware malfunctions and 1 due to a corrupted therapy data file. All participants had a left-hemisphere stroke affecting the left perisylvian middle cerebral artery territory that resulted in a diagnosis of aphasia. In Table 6, data are presented for the 32 individuals (8 female) with post-stroke chronic aphasia participating in this study with a mean age of 59 years (SD = 12 years) and a mean time since stroke of 74 months (SD = 61.4 months).

The participants were recruited from the Predicting Language Outcome and Recovery After Stroke (PLORAS) database (Seghier *et al.*, 2016), an outpatient aphasia clinic at the National Hospital for Neurology and Neurosurgery, University College London Hospitals, and previous UCL Neurotherapeutics Group study participants.

The inclusion criteria were: (i) more than six months post-stroke; (ii) English as a dominant language; (iii) scores below 26/30 for comprehension of Spoken Words and 28/32 for Spoken Sentences on the Comprehensive Aphasia Test (CAT) (Swinburn, Howard and Porter, 2004). The exclusion criteria were: (i) a premorbid significant neurological (e.g., degenerative brain disease) or psychiatric (e.g., major depression) disorder (self-report at screen); (ii) unable to give informed consent.

Table 6 - Baseline summary of Listen-In trial participants.

	Mean	SD
Age (years)	59.4	12.4
Time since Stroke (months)	73.9	61.4
CAT Spoken Words (%)	59.2	21.0
CAT Spoken Sentences (%)	30.5	13.2
Lesion Volume (cm^3)	209.1	61.0

4.3.3.5 Measure of Treatment Effect

The measure of the treatment effect is percentage change on the Auditory Comprehension Test (ACT). This is a bespoke outcome measure that was developed for testing the corpus of lexical items that are trained in the Listen-In therapy. The ACT measures comprehension of trained and untrained spoken word/phrase/ sentence-to-picture matching items with one target picture, and five foils (phonological, semantic, and unrelated). There are 110 trained and 110 untrained psycho-linguistically matched items. Treatment effect is deduced using pre-therapy to post-therapy scores for trained items only. Only trained items are used for the treatment effects outcome as the therapy effect generalise to untrained words as is expected for impairment-based aphasia therapies. The full treatment and testing protocol have previously detailed in full in *Fleming et al.* (2021) and the associated supplementary information.

4.3.3.6 Baseline Assessment

Prior to treatment, participants provided demographic information and completed an extensive battery of 10 language and cognition-based assessments, yielding a total of 19 pre-treatment behavioural and demographic variables (Table 7).

Table 7 - Pre-treatment behavioural and demographic assessment battery.

Assessment	Variables
Demographics	Sex Age Handedness Time since stroke Type of stroke Use of a hearing aid
Audiometry	Hearing level Free field Audiometry (dB)
Auditory Comprehension Test (ACT)	ACT (all word) (%) ACT (trained words) (%)
Comprehensive Aphasia Test (CAT) (Swinburn, Howard and Porter, 2004)	Naming (%) Repetition (%) Spoken Words, % Spoken Sentences, %
British Picture Vocabulary Scale (BPVS) (Dunn *et al.*, 1997)	BPVS (%)
Cattell Culture Fair Intelligence Test (Cattell and Cattell, 1960)	Cattell Culture Fair intelligence (%)
Environmental sound discrimination test (ENVASA) (Adapted from Leech *et al.*, 2009)	ENVASA (%)
Phoneme Discrimination Test (PDT) (Adapted from Robson *et al.*, 2012)	PDT (%)
Semantic Association Test (SAT) (Visch-Brink, E. G., Stronks, D., & Denes, 2005)	SAT (%)

4.3.3.7 Therapy Progression

Listen-In therapy is broken down into therapy blocks (equating to each level in the therapy app), each consisting of 15 therapy challenges. At the end of each therapy

block, the therapy data is uploaded to the server. For the purposes of this study, the initial therapy blocks were segregated to investigate the impact of adding additional therapy data on the predictive performance of the models. Please see Table 8 for a summary of the therapy blocks.

Table 8 - Therapy progress by blocks completed.

Blocks Completed	**Challenges Completed**	**Average days to complete (IQR)**	**Estimated Therapy Hours**
10	150	1 (1-2)	1
50	750	5 (2-6)	5.5
100	1500	9 (5-14)	10
500	7500	29 (20-36)	31.5
Total: 1865	**27975**	**79 (77-83)**	**85**

Participants completed therapy doses ranging from 14 to 158 hours and 7605 to 68991 therapy challenges, with averages of 66.5 hours and 27082 therapy challenges.

Table 9 summarised participant's therapy performance. The majority of study participants had no missing therapy data (n=28), one participant had 24% missing challenge data for the therapy at 1500 therapy blocks, however at 500 and 1000 blocks, this therapy user had only 10% missing data. It is not known why this user had such high rates of missing data in the therapy. The remainder had less than 1% missing challenge data (n=3).

Table 9 - Therapy progress summary.

	Mean	SD
Therapy Dose (hours)	66.5	28.3
Challenge Accuracy (%)	86.0	7.0
Challenges Completed	27082	11912
Total number of errors	3191	1936
Mean Challenge Difficulty	1.11	0.13
Missing Therapy Data (%)	0.85	4.31

4.3.3.8 Therapy Performance

In addition to the demographic and behavioural variables, 27 variables were derived from 3 overarching features of the therapy data; dose, challenge outcomes and lexical information (Table 10). Due to the adaptive nature of the therapy, participants progress through the therapy at different paces and with more challenges focusing on lexical items that they are underperforming on. Therefore, it is hypothesised that this variability in therapy performance between participants may have predictive value. Six training datasets were developed from the In-therapy data based on 6 time point cut-offs (all starting at baseline). The six datasets included a no-therapy condition, as well as datasets containing 10, 50, 100, 250, 500, and 1000 therapy blocks. This was to investigate how minimising the quantities of in-therapy data available to generate predictions influenced the predictive value. In practice, this means users will be required to complete a certain number of blocks in order to generate a prediction, therefore it is advantageous to minimise the amount of therapy time required to produce a prediction while still maintaining an acceptable predictive value.

Table 10 - Variables derived from in-therapy performance data.

Therapy features	Variables derived from therapy features
Dose	Therapy Challenges Therapy Blocks Therapy Duration (Days)
Challenges	**Performance:** Therapy Accuracy Mean (%) Therapy Accuracy SD Therapy Reaction Time Mean (seconds) Therapy Reaction Time SD **Difficulty:** Mean challenge difficulty (%) Easy Challenges (%) Medium Challenges (%) Hard Challenges (%) **Repeats:** Challenge Stimuli Repeats (Mean) Challenge Stimuli Repeats (SD) **Errors:** Total Number of Errors Relative Number of Errors (%) Phonological Errors (%) Semantic Errors (%) Unrelated Errors (%) Incorrectly Recorded Errors (%)
Target Lexical items	Adjectives (%) Nouns (%) Prepositions (%) Pronouns (%) Tense (%) Verbs (%) Target Word Frequency Target Word Concreteness

4.3.3.9 Data Completeness

The Listen-In randomised trial dataset demonstrated exceptionally high data completeness across 19 baseline variables (see Table 11) and 27 in-therapy variables (see Table 12). Among all variables, only one, therapy challenges, contained missing data, attributed to technical issues such as app crashes, lost internet connection or incomplete therapy blocks. This variable achieved a data completeness rate of 99.15%, while all other variables had no missing data, ensuring a robust dataset for analysis and minimising the need for imputation or data reconstruction.

Table 11 - Data completeness for demographics and baseline in the Listen-In randomised trial.

Assessment	Variables derived from assessment	Missing Data Entries	Data Completeness (%)
Demographics	Sex Age Handedness Time since stroke Type of stroke Use of a hearing aid	0 0 0 0 0 0	100 100 100 100 100 100
Audiometry	Hearing level Free field Audiometry (dB)	0 0	100 100
Auditory Comprehension Test (ACT)	ACT (all word) (%) ACT (trained words) (%)	0 0	100 100
Comprehensive Aphasia Test (CAT) (Swinburn, Howard and Porter, 2004)	Naming (%) Repetition (%) Spoken Words (%) Spoken Sentences (%)	0 0 0 0	100 100 100 100
British Picture Vocabulary Scale (BPVS) (Dunn *et al.*, 1997)	BPVS (%)	0	100
Cattell Culture Fair Intelligence Test (Cattell and Cattell, 1960)	Cattell Culture Fair intelligence (%)	0	100
Environmental sound discrimination test (ENVASA) (Adapted from Leech *et al.*, 2009)	ENVASA (%)	0	100
Phoneme Discrimination Test (PDT) (Adapted from Robson *et al.*, 2012)	PDT (%)	0	100
Semantic Association Test (SAT) (Visch-Brink, E. G., Stronks, D., & Denes, 2005)	SAT (%)	0	100

Table 12 - Data completeness for therapy data in Listen-In randomised trial.

Therapy features	Variables derived from therapy features	Missing Data Entries	Data Completeness (%)
Dose	Therapy Challenges	4120	99.15
	Therapy Blocks	0	100
	Therapy Duration (Days)	0	100
Challenges	**Performance:**		
	Therapy Accuracy Mean (%)	0	100
	Therapy Accuracy SD	0	100
	Therapy Reaction Time Mean (seconds)	0	100
	Therapy Reaction Time SD	0	100
	Difficulty:		
	Mean challenge difficulty (%)	0	100
	Easy Challenges (%)	0	100
	Medium Challenges (%)	0	100
	Hard Challenges (%)	0	100
	Repeats:		
	Challenge Stimuli Repeats (Mean)	0	100
	Challenge Stimuli Repeats (SD)	0	100
	Errors:		
	Total Number of Errors	0	100
	Relative Number of Errors (%)	0	100
	Phonological Errors (%)	0	100
	Semantic Errors (%)	0	100
	Unrelated Errors (%)	0	100
	Incorrectly Recorded Errors (%)	0	100
Target Lexical items	Adjectives (%)	0	100
	Nouns (%)	0	100
	Prepositions (%)	0	100
	Pronouns (%)	0	100
	Tense (%)	0	100
	Verbs (%)	0	100
	Target Word Frequency	0	100
	Target Word Concreteness	0	100

4.3.3.10 Data Imputation

Missing data entries were imputed using median values to reduce their impact on the predictive algorithm. This approach was selected to minimise potential biases while maintaining the overall distribution of the dataset, ensuring consistency and reliability in the model's performance, as based on the findings of the iReadMore data imputation exercise in section 4.3.2.6.

4.3.3.11 Model Development and Evaluation

Multiple prediction models were trained to predict percentage change in the ACT spoken word comprehension outcome measure for trained words from baseline to therapy completion. The 5 models developed were; Linear Regression, Lasso Regression, Elastic Net Regression, Support Vector Regression (SVR) and Random Forest Regression (RFR). These models were selected due to their suitability to smaller, noisy dataset and computational efficiency. The former three models are linear regression models and the latter two are non-linear. The Scikit-Learn library within Python programming language was utilised for the development and evaluation of the machine learning models (Pedregosa *et al.*, 2011). Missing values were handled using multiple imputation by chained equations with Light Gradient Boosting Machine Random Forests(Ke *et al.*, 2017). Models were trained using 46 pre-therapy features (variables) and 29 in-therapy variables (covering aspects of therapy dose, task parameters and participant responses; see Table 9 and 10.

Data from the trial was segregated into 6 datasets containing variable amounts of initial therapy training data. Models were trained using 90% of the participants' data (n=28) in

a test-train split. Model validation was conducted using LOOCV to maximise the data available for training. Prediction training performance was evaluated using error minimisation metrics (RMSE, RMSE SD, R^2), with RMSE being the primary metric, RMSE SD was used to control for model biases and the coefficient of determination (R^2) was used to assess the model's ability to explain variance in the dependent variable, with a target range of 0.6–0.8 considered indicative of a well-performing model (Iorga *et al.*, 2021). However, exceeding this threshold was not necessarily indicative of superior model performance in this context.

Due to the large number of features and the limited sample size, dimensionality reduction and feature selection techniques were applied. For this, Linear Regression, Lasso Regression and Elastic Net Regression were tuned manually with feature selection derived from the data exploration. SVR and RF models were tuned and trained using hyperparameter optimisation. Additionally, in an effort to address a dataset imbalanced in terms therapy outcome (only 4 out of 35 participants did not improve or had worse speech comprehension post-therapy), under-sampling and upregulation were trialled to balance the dataset relative to the minority subgroup of those who did not improve.

The final model evaluation was performed using the cross-validation function in Python SciKitLearn and evaluated in terms of RMSE, R^2 and F1 score.

5 Results

5.1 Study 1: Development of the iReadMore Therapy App

5.1.1 Overview

In order to co-design the iReadMore app, I collaborated with individuals with aphasia and software developers in an iterative process of design and development. This involved conducting alpha testing, focus groups, and remote beta testing to gather user feedback and design insights. Using qualitative research methods, I analysed user perspectives and identified key themes in their design recommendations. These insights shaped the prototypes, and subsequently, the final release version of the iReadMore app. This data also informed the development of a comprehensive framework of design principles for digital reading therapies tailored to the needs of individuals with alexia.

A series of five co-design focus groups were conducted with 22 individuals with alexia and aphasia and 3 carers/family members. The framework analysis on the focus group was conducted to develop themes from the qualitative data. The analysis generated seven distinct themes of key considerations for the design of a digital intervention for aphasia rehabilitation. The themes generated were; Agency, Intuitive Design, Motivation, Personal Trajectory, Recognisable and Relatable Content, Social and Sharing, and Widening Participation. Figure 13 displays a thematic map of the themes and key subthemes. Design recommendations were generated iteratively throughout

the focus groups and framework analysis process, these were incorporated into successive prototypes and the final version of the application.

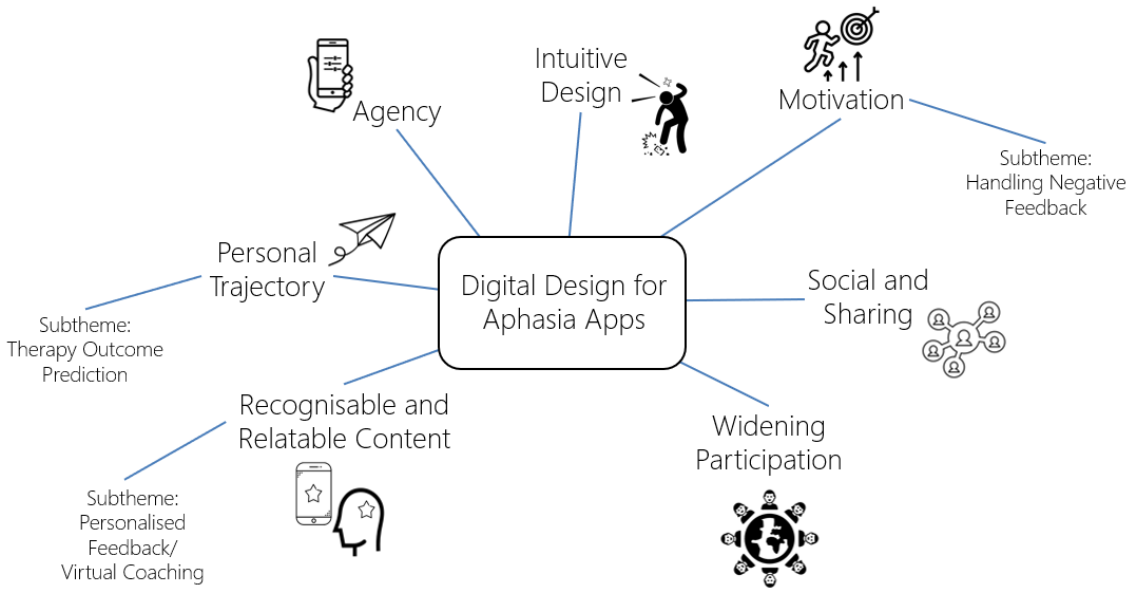

Figure 13 - Thematic map displaying the themes and subthemes generated in the framework analysis of the iReadMore co-design process.

5.1.2 Results

5.1.2.1 Theme 1 – Agency

A prominent theme generated from the co-design process was to utilise functionalities that promote establishing a stronger sense of agency for therapy users. Many participants mentioned the lack of control they have felt in other aspects of their life as a result of their communication impairment and emphasised that restoring feelings of agency, even in small ways, would be of significant value to the users. This theme highlights the idea that the benefits of therapy can potentially extend beyond its primary purpose, offering broader rehabilitative impacts that enhance not only reading abilities but also overall quality of life.

> [on self-managed therapy] *"I think iReadMore is good because it gives X something for himself, something he can complete and be in control of, and I think that gives a big boost to his confidence"* – Partner of stroke survivor with aphasia, female, 70

In practical terms, ways to increase agency that were suggested included giving users more control over therapy parameters and settings. Participants were interested in the workings of the therapy progression algorithm and suggested an additional mechanism that allows users to adjust the therapy difficulty themselves would be valued so they can progress more easily to a difficulty level that suited them. Participants also

preferred to decide their own therapy duration each day rather than have sessions of a fixed length.

Further, it was raised that making the therapy easy to use without assistance would be empowering. The ability to engage with the therapy independently, without the assistance of caregivers or professionals, was frequently cited as a critical factor. This autonomy not only promotes agency but also reduces the burden on family members or caretakers, making the therapy more sustainable for long-term use. While participants wanted control, they also acknowledged the need for optional guidance or support when required. For example, users suggested the inclusion of accessible tutorials, help sections, and optional coaching features that could assist without undermining their autonomy.

Notifications and pop-up reminders were viewed as superfluous and an annoyance as the user should know when to use the therapy and that doing a therapy is a significant activity in their daily lives, motivated intrinsically by a desire to improve on their reading impairments. In specific circumstances, infrequent reminders would be more tolerable as long as they were providing useful information.

5.1.2.2 Theme 2 – Intuitive Design

Simplicity of the app design and ease of use were important considerations to participants. This theme focuses particularly on the uptake of a new digital technology and the need for intuitive design to reduce access barriers and user disengagement. Regardless of whether participants were experienced technology users or not, there

was a unanimous preference for an easy to pick up application. Participants reported that difficulty in starting with a new therapy can lead to feelings of frustration and helplessness. For a therapy app that is intended to be use independently, and particularly for users who may have previously relied on assistance for other technologies or therapies, ensuring confidence in the user is a prerequisite to therapy engagement.

In terms of iReadMore, the initial lack of clarity around where to tap on the screen during the Exposure Phase of the therapy led some participants to doubt their ability to use the therapy unassisted while others felt frustrated. Highlighting the importance of reducing cognitive load for users with alexia or aphasia by streamlining design elements, such as consistent button placement and clear navigation paths, to reduce barriers to the uptake of a new digital technology. To resolve this, it was decided that a stronger visual contrast between clickable and non-clickable content was needed along with additional audio instructions and the use of animations to highlight fields that need to be clicked if no interaction is detected.

"I think if you didn't get it immediately, because for me if I can't get something because of … things. I tend to give up and try something I can do. Because it'll make me feel better [laughs]" – Stroke survivor with pure alexia, male, 46

[On being unsure how to use an app] *"wouldn't have … confidence … to ask for help"* – Stroke survivor with aphasia, male, 65

To further simplify the app experience, a more linear flow was implemented with buttons always located in the same locations. The visual appeal of the app design was of little or no importance to the majority of participants. Participants consistently prioritised ease of use and clear navigation over the visual sophistication of the app. Participants expressed concerns about the potential learning curve for highly stylised interfaces, which could deter engagement. Alternative designs for the main menu that involved more immersive and visually-stimulating experiences were viewed as visually cluttered or difficult to interpret with concerns about learning to use a more complicated app independently. Instead, a simplified, more functional navigation to the therapy, help section and feedback graphs was largely preferred (Figure 14). High-contrast colour schemes were used to they reduced effort in search for buttons for users with visual impairments.

Figure 14 - iReadMore Main Menu displaying therapy dose information, time to next test and buttons to start therapy, access help and display statistics on therapy progress and test scores.

A further consideration is striking the right balance between providing sufficient guidance and simplicity for first-time users to navigate the app confidently and ensuring that repeated users are able to easily navigate to features such as their therapy progress and therapy performance with ease. In the iReadMore app, this was applied by ensure the therapy start button has the highest contrast with a large button centralised on the screen, with other features included on the main page with lower contrast.

5.1.2.3 Theme 3 – Motivation

Motivation underpinned many of the discussions in the co-design process. Participants thought that users of digital aphasia therapies do not need a lot of "bells and whistles" as they are highly (intrinsically) motivated by the desire to improve on their impairments

and did not respond enthusiastically to many traditional features of gamification aimed at improving extrinsic motivation. By balancing intrinsic and extrinsic motivations this could ensure that the therapy remains meaningful without relying on excessive gamification elements that do not resonate with the target demographic.

“Colours make a big difference. For using everyday, I need something a bit fun. If it's a bit simple [gestures down with hands], but colours make it [gestures upward motion with hands]” – Stroke survivor with aphasia, female, 29

Facilitator: "Would it be demotivating to get negative feedback?"

"No, no. For me personally, if I’m getting it wrong but going forward, then I'm going forward … good for my understanding." – Stroke survivor with aphasia, female, 56

Some did not understand the gamification concepts (such as points, high scores, avatars and badges) or their intended purposes, while others felt they were not of value for this demographic.

Participants thought that features to support motivation were needed later in the therapy to maintain usage over weeks to months. They proposed that the main driver of motivation long-term was the ability to track and interpret their own therapy progress using the in-app reading test which are completed after every five hours of therapy. Participants valued clear, meaningful feedback on their progress as a way to reinforce their sense of accomplishment. Many styles of presentation for this information were discussed and prototyped. The final designs are highly visual, with minimal lexical

information and multiple representations of their scores to increase accessibility (Figure 15).

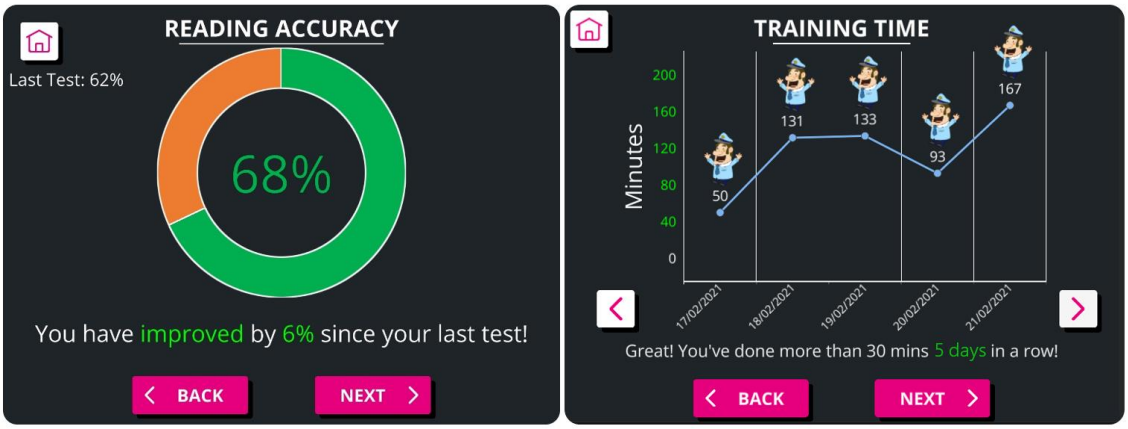

Figure 15 - iReadMore Feedback graphs and personalised messages for a) reading test accuracy and b) training time – on the graph, the stickers denote each day where 30 minutes of therapy were completed.

Adding in visual novelty was seen as another way to maintain interest and denote progression through the therapy. Therefore, a number of designs were suggested and finally, a travel-based concept with 10 destinations that users fly to around a 3-dimensional world was implemented (Figure 16). As such, when users complete 20 minutes of therapy, they visit a new destination; users are advised to use the therapy for 30 minutes a day so will visit a new location at least once a day at this rate. The destination backgrounds in the therapy are static to prevent distraction from the therapy task, instead acting as a border without text elements or animations. This balanced novelty and visual intrigue without impairing user's abilities to complete the therapy.

Figure 16 - Therapy design travel concept.

The concept of receiving negative feedback was a key subtheme in the discussions of the workshops with varied responses from participants. When asked about how they responded to the negative feedback, many believed it was acceptable and appropriate. Some thought it was key to motivating them to improve and part of the process. However, one participant reported that they would like the option to hide their test results depending on their mood. They felt that being confronted by their impairment too often would be demotivating or upsetting, making them less likely to engage with the therapy. All agreed that being able to choose was a beneficial addition to the therapy and as such, test results are viewed by clicking on the Statistics button on the main menu (Figure 14).

5.1.2.4 Theme 4 – Personal Trajectory

A clear and consistent perspective from participants was that stroke survivors with alexia are on individual journeys of rehabilitation and that gamification concepts of competition, leaderboards and other comparisons between users were viewed negatively and seen as detrimental to user motivation and likely to cause stress unnecessarily. Collaboration was also seen as pressurising due to negative feelings arising from letting others down. Instead, participants wanted to focus on their personal progress in the therapy through regular feedback and praise for consistent use of the therapy.

> *"Everyone has a different rate of improvement … So therefore, you don't want to benchmark yourself against others… I think the challenge is with you and progressing where you are and what you can do." – Stroke survivor with aphasia,* male*, 75*

A subtheme of whether being able to predict an individual's future therapeutic outcomes was of value had mixed response from the groups. There were concerns over inaccuracies as well as denial of service if it appeared it would not benefit them. Participants reported they would prefer to try it and decide whether it is not working for themselves or decide collaboratively with their clinician. However, it was also suggested that predictions could be a useful motivational tool to inspire users to continue progressing with the therapy if they were reported after the interval reading tests to

motivate users to continue with the therapy. This concept will be explored further in future work looking into the feasibility of in-app therapy prediction.

5.1.2.5 Theme 5 – Recognisable and Relatable Content

This theme relates to participants' preferences on how information is presented in the app. It was thought that a large proportion of digital therapies were designed with a young demographic in mind. There was a desire for content that feels mature and engaging reinforcing the app's focus on adult rehabilitation. However, a surprising outcome to the researchers was the pervasiveness and appeal of emoticons (emojis). Participants reported using emojis in place of words when they were having word-finding difficulties.

"Because it feels quite young, it doesn't make you feel good about doing the exercise. It makes you feel like your level of understanding is lower" – Stroke survivor with pure alexia, male*, 46*

"Yes, it suggests you're doing this at school and not as an adult. It needs to be something that we're accustomed to seeing and understanding" - Stroke survivor with pure alexia, male, 78

Some participants did not understand or engage with the gamification concepts of points or scores, seeing them as unnecessary and not desirable. Further, some

individuals had difficulty in number reading and found numerical scores distracting when incorporated into the therapy, so these were removed. Therefore, the numeric point system was replaced with styles of feedback delivery using visual and audio content delivered through an animated cartoon character (Figure 17) to provide immediate performance feedback on a therapy challenge.

Figure 17 - iReadMore character design and Challenge Phase feedback reactions.

Participants thought the language used in instructions in the app and guidance for using the therapy should be simple and unambiguous. A couple of participants referred to frustration from not receiving clear guidance on how to use a therapy effectively. The group felt that quantified, realistic goals would inspire regular use and confidence that they are using the app correctly. Ambiguous guidance such as 'use the therapy as much as you can' was seen as unhelpful. One participant described previous experiences of using therapies for long, continuous periods in the first instances led to fatigue and would not be feasible longer term. Whereas, clearer guidance such as 'Use the app for

30 minutes a day' were seen as a motivating, achievable and provide evidence-based advice; therefore, this was implemented in the app.

Exploration of implementing a virtual coach in the app received mixed feedback. Some participants thought this would distract from the therapy or overcomplicate what users would like to receive from the therapy. However, the implementation of personalised, positive feedback without embodiment of a virtual coach was unanimously supported. Examples of feedback included how often they were using the app, their performance, and overall progression in the therapy in terms of reading accuracy and speed test scores. Participants felt once or twice a week was an appropriate frequency for these types of messages and that it needed to feel sporadic and related to their personal performance.

5.1.2.6 Theme 6 – Social and Sharing

Participants wanted to be able to share their therapy progress with personal contacts and clinicians. Many participants were eager to incorporate a screenshot which they could share with their family and friends to share their therapy progression and use as a tool for fostering a sense of community and enable positive feedback outside of the app environment. One participant mentioned that it could help to act as an icebreaker and enable open discussion about their condition, something which they currently find difficult to do. Only a few participants wanted to be able to share this feedback on social media. Many wanted to share this information with close, personal contacts, either in person, via email, text message or a messaging app (such as WhatsApp).

"When I finish and go 'yay!', I want to show my family. [Picks up phone and opens WhatsApp] I love send photos!"– Stroke survivor with aphasia, female, 60

"Would be great to show to my therapist. That way she'll know that I'm actually doing the home practice! [laughs]" – Stroke survivor with pure alexia, female, 50

The other aspect of this theme was being able to share information with their clinicians, in particular, speech and language therapists (SLTs) or with facilitators and group members at their aphasia support groups. This was suggested as a feature which would be an additional benefit of using the app as it could demonstrate their therapy compliance and progression which could be used to report competence and willingness. Further, two participants mentioned this could aid discussions with their clinical team over clinical decision-making, where the SLT could advise on whether the therapy is working for them, allowing them to make more informed decisions about continuing therapy or making adjustments to their treatment plans. A feature for clinicians to directly access users' progress or provide feedback through the app could create a more integrated, continuous care model.

5.1.2.7 Theme 7 – Widening Participation

The final theme relates to accessibility barriers for digital therapies. Issues relating to usability of the app in the context of aphasia as well as prevalent co-morbidities such as

physical (hemiplegia, hemiparesis), visual (hemianopia, colour blindness, visual neglect), auditory (high frequency hearing loss) and working memory impairments were raised. Based on these, the groups developed design refinements that would make the app more accessible. Examples include how the app does not require using more than one finger to operate and does not need to be held while in use, buttons and important visual content are always located centrally on the screen, the words in the therapy are read out twice in both female and male voices, and if no response is detected, spoken instructions are repeated and in some cases, highlighted on screen through animations.

"Can't do! When you first start, you need to focus on the word… and don't want distractions. Not for me with distractions, not for me." – Stroke survivor with aphasia, female, 38

An early prototype used animations throughout the therapy trials to make it more visually stimulating, however, this prevented a number of participants from knowing where to focus on the screen and was regarded as a distraction. As a result, animations were limited to reporting feedback after the user has answered a trial as a balance between making the therapy visually stimulating while minimising distractions.

Another significant barrier to access arose from minimal prior experience with technology. Issues related to the technical difficulties of setting up and using a tablet device and downloading the therapy. In response, aphasia-friendly instructions and frequently asked questions that were generated in the co-design process have been

integrated into the app. Participants wished to be able to contact the team directly for technical support or guidance. Therefore, an anonymous ‘contact us’ button was added into the Help section of the app. This allows the research team to assist users whilst maintaining anonymity in-line with our ethical approval and data security regulations.

Finally, concerns were raised over deploying the app solely on Android tablets as initially intended due to financial constraints. Some participants were unsure of what kind of device was required to use the therapy. The majority of the group did not have a tablet at home and the minority that did were split between apple devices and Android devices. As a result, the app was developed for Apple and Android phone and tablet devices.

5.1.3 Discussion

5.1.3.1 Themes

Seven themes emerged iteratively from the framework analysis of the iReadMore co-design process. These themes evolved as the development cycle progressed, with insights from each session informing the design and refinement of individual elements. The framework of themes evolved over successive sessions, as subsequent participants provided their perspectives building upon discussions from previous sessions. Insights drawn from each focus group informed the design of future sessions, shaping questions and activities to further explore user priorities. This approach facilitated the integration of numerous design changes, ensuring that the application reflected the lived experiences and preferences of individuals with aphasia.

For iReadMore, this was exemplified by several unexpected findings, such as participants' diverse interpretations of gamified features and their desire for designs that prioritised simplicity and accessibility over traditional gaming elements. These insights illustrate how co-design can lead to innovations that are more aligned with the preferences and capabilities of the intended user group. This iterative, user-centred approach underscores the importance of involving patients early in the design of digital applications. By doing so, key decisions are informed by the lived experiences and needs of end-users, rather than being based solely on the perspectives of researchers or software developers. This is particularly critical in the context of digital therapies like iReadMore, where the target population may have limited confidence in using digital technologies.

5.1.3.2 Agency

Promoting a sense of agency in therapy was particularly significant for participants, as many reported feeling a lack of control in other areas of their lives. Participants shared that being unable to use digital therapies designed for their demographic often led to feelings of inadequacy, reduced self-confidence, and disengagement from those therapies. This aligns with recent findings on how aphasia impacts self-identity and highlights the importance of designing therapies that support autonomy and competence (Taubner, Hallén and Wengelin, 2020).

Conversely, digital therapies that could be used independently were described as having positive effects on personal empowerment and routine-building. Regaining agency through therapy may help individuals rebuild their sense of self-identity, which is frequently challenged by the effects of their impairments. This finding is consistent with broader evidence suggesting that fostering autonomy and independence significantly contributes to psychological well-being (Cardol, De Jong and Ward, 2002; Biel *et al.*, 2022).

The ability to independently engage with a digital therapy promotes regular, self-directed use, empowering participants to integrate therapy into their daily routines. Establishing such routines not only reinforces a sense of control over their rehabilitation process but also helps develop sustainable habits that support long-term recovery. These findings underscore the potential of agency-centred therapy design to enhance both clinical outcomes and overall quality of life for individuals with alexia or aphasia.

5.1.3.3 Intuitive Design

The visual appeal of the app content was not a primary concern for most participants, contrasting with findings from previous co-designed digital therapies such as EVA Park (Marshall *et al.*, 2018) and GeST (Marshall *et al.*, 2013), both of which utilise immersive virtual worlds. Instead, participants in this study expressed a preference for simpler navigation and an intuitive app flow, favouring functionality over highly gamified or visually elaborate designs. This difference may stem from the contrasting contexts of therapy delivery: EVA Park and GeST are SLT-led therapies for communication production, while the iReadMore app is designed as a self-managed therapy for reading. The context-specific nature of co-design research inherently produces findings tailored to the unique needs and preferences of the target user group.

In this study, participants appeared to prioritise ease of use, particularly during the initial stages of therapy, over immersive elements. However, visual (non-lexical) communication remained integral, as participants valued the use of graphical or symbolic elements to effectively convey feedback and progress within the app.

It is worth considering whether visual appeal may become more important with extended use of the therapy. While simplicity is crucial for onboarding and early engagement, features that enhance the visual experience and increase functionality could be gradually introduced as users become more confident and familiar with the app. Striking a balance between simplicity during the initial period and added complexity over time could address users' evolving needs without overwhelming them

at the outset. This staged approach might enhance both user satisfaction and long-term engagement with the therapy.

5.1.3.4 Motivation

Maintaining motivation was reported to be driven by intrinsic motivations and self-monitoring reading improvements through graphs or personalised messages. When participants were presented with variations of gamified therapy prototypes aimed at promoting extrinsic motivation, it was often felt these alone would have little impact on their decision to use the therapy. The subtheme on receiving negative feedback were in contrast with the concept of errorless learning, which is often applied in rehabilitation technologies, and more in line with error-reducing learning (Middleton and Schwartz, 2012). However, it may be important to consider that people with aphasia who actively take part in research may display higher intrinsic motivation than those who do not. Many of these participants had taken part in previous studies involving highly gamified digital therapies and this may have shaped their perspective. Therefore, findings may not relate to the experience of people with aphasia and lower intrinsic motivation. In order to try and gain a wider perspective in future work, all users of the therapy will be able to anonymously provide qualitative feedback through the app.

5.1.3.5 Personal Trajectory

The theme of personal trajectory underscores the importance of tailoring digital therapies to the individual journeys of stroke survivors with alexia. Participants consistently emphasised the need for a therapy to focus on their personal progress

rather than external comparisons. Traditional gamification elements such as leaderboards, competition, and collaborative performance were seen as counterproductive, potentially inducing unnecessary stress or feelings of inadequacy. This aligns with existing literature on the negative impact of external comparisons in contexts where individual differences in abilities and recovery trajectories are pronounced.

Instead, participants preferred regular, personalised feedback that highlighted their own progress and praised consistent engagement with the therapy. This approach reinforces intrinsic motivation by celebrating individual achievements and mitigating the pressures associated with direct comparisons or collaborative dynamics. Feedback strategies tailored to personal milestones could thus play a vital role in sustaining long-term therapy engagement.

Interestingly, the concept of predicting therapeutic outcomes emerged as a mixed subtheme. While some participants recognised the motivational potential of outcome predictions, others raised valid concerns regarding their accuracy and the potential for discouragement if predictions appeared unfavourable. This hesitancy highlights the need for careful implementation of predictive tools, ensuring they empower rather than discourage users. One possible approach is to integrate predictions as motivational checkpoints following interval reading tests. This allows users to see their progress contextualised within achievable goals for continuing the therapy, fostering a sense of agency and self-determination in their rehabilitation process. Alternatively, the prediction can be used to support users to decide when to alter their treatment plan.

5.1.3.6 Recognisable and Relatable Content

The importance of recognisable and relatable content in digital therapies is underscored by the perspectives of the participants, who expressed a desire for content tailored to adult rehabilitation rather than designs that might feel juvenile or simplistic. This finding challenges the conventional focus on highly gamified, youth-oriented digital platforms and highlights the need for a thoughtful approach to design that aligns with the lived experiences and preferences of adult users.

While some participants expressed reservations about elements such as emojis or animated characters, their practical utility emerged as a nuanced theme. Emojis, for instance, were valued by participants for their role in facilitating communication, particularly among individuals experiencing word-finding difficulties. This suggests that while such elements may initially seem incongruent with an adult therapy setting, their functional relevance can outweigh aesthetic concerns when thoughtfully integrated into the app's design (Foulkes, 2019).

Clear and concise language emerged as a critical requirement for app instructions and guidance. Participants expressed frustration with ambiguous or overly general directives, which can undermine confidence and lead to improper use of the therapy. In response, the implementation of specific recommendations, such as setting achievable daily usage goals, addresses this challenge by fostering a sense of structure to app usage. This aligns with broader findings in language rehabilitation research, which emphasise the importance of actionable, goal-oriented messaging to motivate sustained engagement (Hersh *et al.*, 2012; Biel *et al.*, 2022).

The discussions on integrating recognisable and relatable content have similarities with design concepts being explored in other aphasia therapies, such as Web ORLA, which utilises an embodied virtual therapist in the programme(Cherney *et al.*, 2021). Within the timeframe and financial limits available for this research, exploring implementing a virtual coach in iReadMore was deemed unfeasible and personalised feedback on therapy usage and progress was seen as an appropriate alternative to this (Figure 15). There were also concerns it may lead to accessibility issues that could preclude some users from being able to engage with the therapy due to the technical and linguistic requirements of communicating with a virtual coach. Research exploring the feasibility of applying virtual coaches in rehabilitation for older adults including people with aphasia is ongoing (Kyriazakos *et al.*, 2020); however, this study also excludes those with global aphasia.

5.1.3.7 Social and Sharing

The emphasis on integrating social opportunities into the therapy is an understudied and somewhat underutilised concept in digital therapies at present, and participants generally felt this was a key area for improvement. This relates to previous research which has found that people with aphasia tend to have a reduced social network and less frequent social interactions (Vickers, 2010; Northcott and Hilari, 2011), while also experiencing an overall reduction in quality of life compared to stroke survivors without aphasia (Hilari, 2011). It was noted by the researchers that the participants who felt they would not want to see their own progress (as highlighted in the Motivation theme), also did not want to share their progress with a clinician or friends and family. Their focus

was on making the app independently and privately usable, whereas in contrast, other participants wanted features which would enable real world connections by sharing this information to prompt conversations about their condition with friends and family. Therefore, a balance is required to appeal to these conflicting perspectives. However, there are also a number of obstacles to integrating aspects of the Social and Sharing theme into a digital therapy, including concerns of data security, regulatory affairs, content moderation and the complexity of the design required which will need to be considered.

5.1.3.8 Widening Participation

The theme of widening participation has parallels to the findings of a recent clinical review of technology use in aphasia (Cann and Bulman, 2018). This survey revealed that people with aphasia are more likely to have access to a tablet device than a mobile phone or computer. However, the population assessed were currently receiving speech and language therapy and it was more likely that the tablet was owned by the clinical service than the person with aphasia. Therefore, in order to reach people who are not currently receiving speech and language therapy, it was important to release the application on tablet and mobile devices across operating systems and in future, develop a desktop version of the app.

5.1.3.9 Co-design and Framework Analysis

Conducting a framework analysis alongside the co-design process enabled the simultaneous refinement of the app design and the development of qualitative themes

in a synergistic and efficient manner. This dual approach leveraged the iterative nature of framework analysis to provide a structured yet flexible pathway for uncovering insights. The methodology was particularly effective in distilling the diversity of participant perspectives into actionable design recommendations, a critical aspect of co-design projects.

Framework analysis offered a systematic process for managing and analysing qualitative data, ensuring that insights from focus groups were rigorously categorised, compared, and interpreted throughout the co-design process. The iterative process enabled revisiting and refining the coding framework as new themes or ideas emerged in successive focus groups. This adaptability was instrumental in shaping semi-structured questions and activities for subsequent sessions, fostering deeper exploration of evolving concepts and design recommendations.

Moreover, the structured approach enabled the meaningful organisation of diverse participant input while maintaining the flexibility to engage with unexpected themes or ideas that arose organically during discussions. This balance between structure and adaptability ensured that the process was both methodical and responsive, enhancing the quality and relevance of insights generated during the co-design process.

A significant strength of this approach was its inclusivity, particularly when collaborating with individuals with aphasia or alexia. The data collected during co-design sessions spanned multiple formats, including transcripts, video recordings, drawings, and written notes. Framework analysis proved capable of accommodating these diverse data sources, allowing the integration of non-verbal and visual data into

the analysis. This inclusivity would have been more challenging to achieve with traditional thematic analysis, which often relies primarily on textual transcripts and is less well-suited to multimodal data. The inclusion of visual and interactive elements, enabled by the co-design process, enriched the qualitative data and ensured that participants with varied communication preferences and abilities were fully represented.

5.1.3.10 Self-Determination Theory

The themes generated in this study align with the Self-Determination Theory (SDT; Deci and Ryan, 1985; Ryan and Deci, 2000), which highlights autonomy, competence, and relatedness as key psychological needs for motivation and wellbeing. These principles are particularly relevant in neurorehabilitation, where sustained engagement is essential.

In terms of autonomy, the themes of agency, intuitive design and personal trajectory emphasise self-directed therapy and individual progress, allowing users to feel in control of their rehabilitation. Customisable feedback, optional sharing, and clear, user-friendly instructions enhance autonomy, empowering users to take ownership of their recovery.

The themes of recognisable and relatable content, motivation and widening participation address competence by providing clear, constructive feedback and realistic goals, such as "30 minutes a day," which foster achievement without causing

frustration. Simplifying gamified elements supports understanding and promotes a sense of capability.

The social and sharing theme aligns with relatedness, emphasising meaningful connections with family and friends, and improve dialogues with care providers. Sharing progress privately rather than publicly highlights the importance of personalised, supportive social interactions, reinforcing the motivational role of connection in recovery.

These findings align with person-centred and life participation approaches in aphasia rehabilitation (Chapey *et al.*, 2012; Simmons-Mackie, 2012), highlighting how addressing autonomy, competence, and relatedness can enhance motivation, identity reconstruction, and engagement with therapy. Future work should build on these frameworks to further refine user-centred features.

5.1.3.11 Implementation of Design Recommendations

The inclusive co-design methodology highlighted the need for features and design elements that researchers had not previously considered. Many of these innovations emerged directly from the designs, discussions, and activities of the participants, offering novel and informative contributions to the therapy app. The iterative phases of co-design not only allowed researchers to capture participant feedback on specific aspects of the app but also provided a mechanism to verify that subsequent redesigns aligned with participants' expectations. This iterative process ensured that the app evolved in a way that was both user-centred and responsive to feedback.

Co-design also demonstrated its potential to move beyond traditional, clinician-led therapy tasks, fostering the development of innovative therapies that leverage technology to achieve outcomes unattainable through conventional methods. By engaging participants directly in the design process, the app incorporated features tailored to real user needs and preferences, rather than relying solely on preconceptions or clinical perspectives.

However, it is important to note that not all of the design ideas generated through this process were able to be incorporated into the final version of the app. Constraints such as time, funding, and resources limited the extent to which certain features could be implemented or explored. Nonetheless, the co-design methodology laid the foundation for a more inclusive, user-driven development process, providing valuable insights for future iterations of the therapy app.

5.1.3.12 Limitations

While several recommended co-design session aids and suggestions were successfully incorporated into the design of the focus groups to enhance inclusion in the co-design process, some recommendations were not feasible due to resource constraints and logistical challenges. For instance, applying gesture analysis, a method particularly valuable in aphasia research, was not integrated into the data collection. This approach offers critical insights into nonverbal communication, which is essential for individuals with severe aphasia. As highlighted by Wilson and Kim (2019), capturing nonverbal cues alongside verbal communication ensures more comprehensive participation of individuals with aphasia in qualitative research.

Additionally, other aids that could have enriched the process, such as providing participants with tablets for extended periods to trial app prototypes and employing eye-tracking technology for enhanced usability insights, were not implemented.

The unplanned shift to online beta testing following the focus groups due to the COVID-19 pandemic also posed significant challenges. However, the online format provided a realistic simulation for beta testing. It enabled the onboarding of new participants with minimal support, mimicking real-world conditions and yielding valuable insights into usability and participant experiences under less controlled circumstances.

5.1.3.13 Reflections and Future Work

This study reinforces the current literature on the ability to successfully conduct a co-design study with people with aphasia. A core component of the co-design process was establishing total communication techniques that enable participants to engage meaningfully. These techniques include incorporating drawing, writing, visual aids, and emotion scales in the co-design sessions (Neate *et al.*, 2019). It can be beneficial to know the communication profiles of participants ahead of time in order to support specific communication needs and explore how participants can be best supported to contribute (Wilson and Kim, 2019). In addition, involving carers and partners in the co-design sessions can further enable effective communication, particularly for individuals with more severe impairments (Prior *et al.*, 2020). Finally, the technique of asking participants to consider the perspectives of other individuals with aphasia that they knew personally was particularly useful in addressing issues that would form the basis of the Widening Participation theme. Participants were asked to think of other

individuals they knew with alexia or aphasia and were asked what would help make the therapy accessible and appealing to them. Additionally, participants were asked to reflect on other apps that they use for therapy purposes or general use.

The implications of the COVID-19 pandemic led us to conduct beta testing remotely with people with aphasia using the therapy at home with their own devices where possible. Testing the therapy in the same setting as it is intended to be used was highly valuable and enabled the inclusion of participants outside of our usual catchment area as an added benefit. Stratifying users by technology usage and prior participation in a digital therapy clinical trial was important for ensuring the development of an app that was accessible to first-time users while also remaining engaging after use for the substantial period of time required to achieve therapeutic gains. However, we found similar trends for both those with and without prior technology experience in wanting to prioritise the ease of use of the application over design novelty or complexity. This was in order for users to feel confident in using the app independently, as the frustration of not knowing what to do with a digital therapy was highlighted as a key reason for therapy disengagement.

A further area of interest for future work involves conducting similar co-design activities to investigate the specific needs and requirements of individuals with alexia resulting from primary progressive aphasia (PPA), a form of language-led dementia (Marshall *et al.*, 2018). While preliminary insights were gathered anecdotally from three individuals with PPA who trialled the co-designed version of iReadMore, it became evident that additional development work is necessary to tailor the therapy to this population. Their

unique challenges and evolving needs underscore the importance of targeted adaptations to ensure the app is both accessible and effective for individuals with PPA.

5.2 Study 2 - Preliminary Findings of the iReadMore Rollout Trial

5.2.1 Results

5.2.1.1 Participant demographics and usage

App Users

Between the public release of the iReadMore app on 1st March 2021 and 1st April 2023, the app was downloaded more than a thousand times from the Apple App Store and Google Play Store across 80 countries (see Figure 18). During this period, 529 users agreed to have their therapy data collected for research purposes and completed the baseline assessment. Of these, 173 users completed at least one therapy session, and only 34 users progressed to the first interval timepoint (5 hours of therapy). 14 users proceed to complete the 20 hours of therapy required to reach interval test 4 and be included in the data analysis.

A summary of user progression through therapy timepoints is provided in Table 13 and Figure 19.

Three users exceeded 100 hours of therapy, among them, one user completed 178 hours, while another reached 363 hours of therapy. Notably, these are the only users included in the data analysis that surpassed the minimum therapy dose of 60 hours that was achieved by all participants in the iReadMore randomised controlled trial.

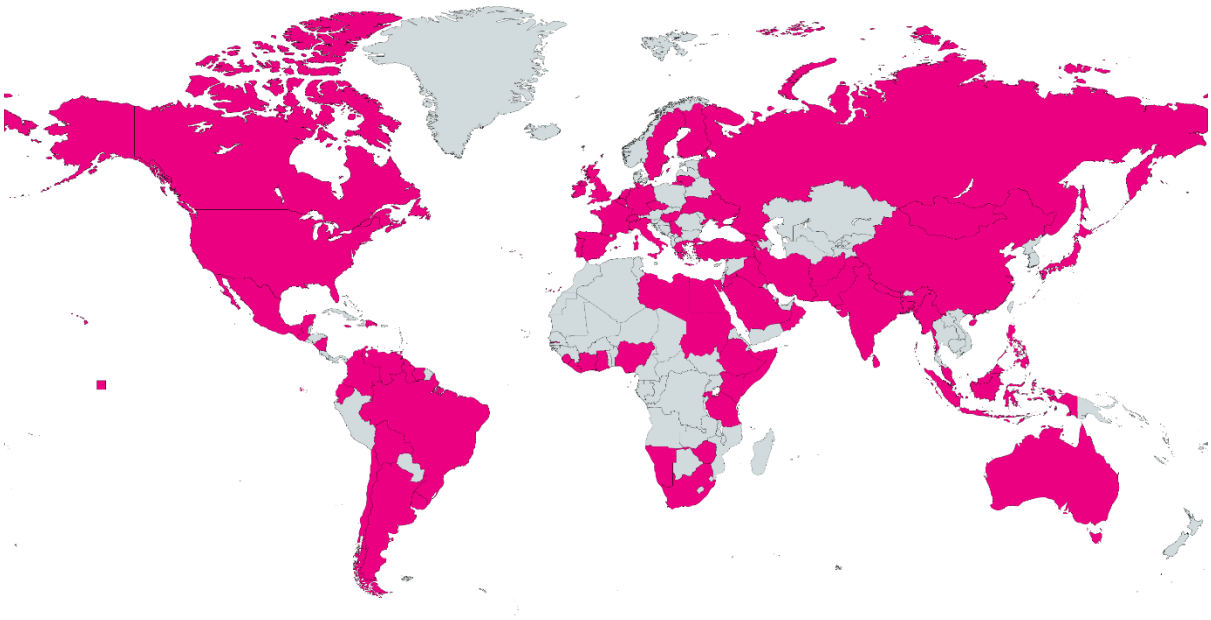

Figure 18 - Map of countries where the iReadMore app has been downloaded, highlighted in iReadMore pink (figure generated in www.mapchart.net).

Table 13 - Interval tests completed by number of users

Interval Test	Hours of Therapy Completed	Number of users
1	5	31
2	10	21
3	15	15
4	20	14
5	25	11
6	30	6
7	35	5
8	40	4
12	60	3
20	100	2

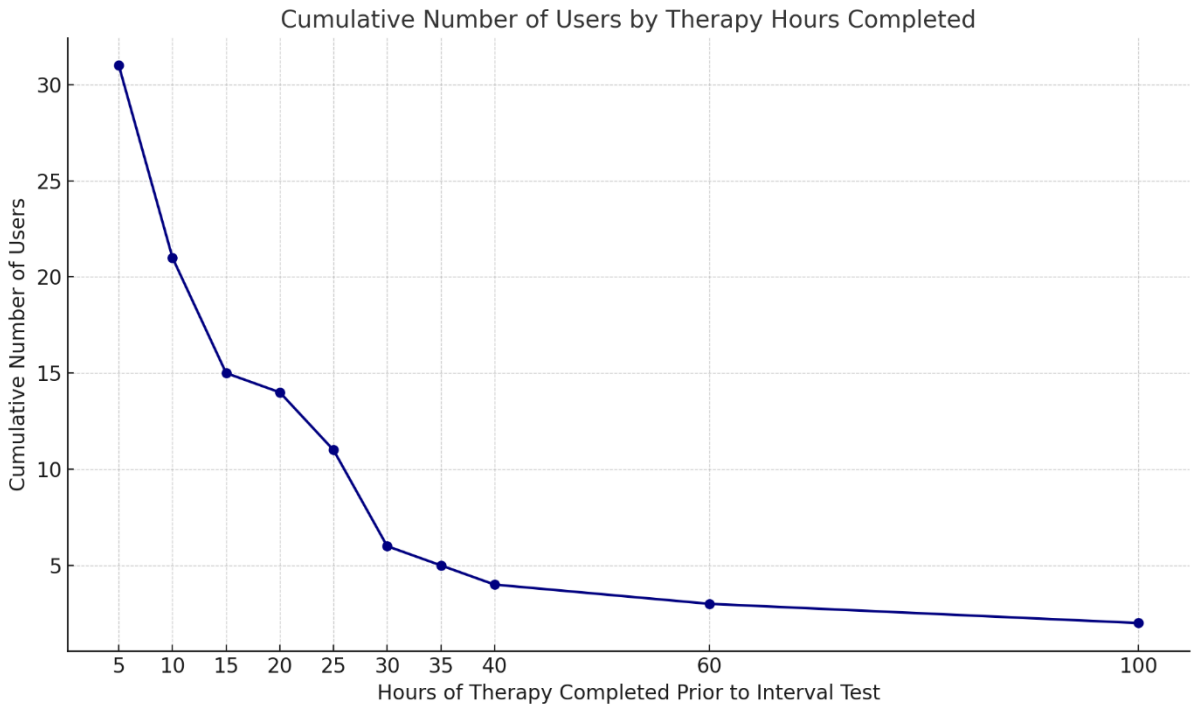

Figure 19 - Hours of therapy completed by number of users (up to 100 hour).

User Demographics

Demographics for the 14 users who completed more than 20 hours of therapy, reaching at least the fourth interval test, are presented in Table 14. Stroke was the most commonly reported cause of reading impairment (86%), followed by brain injury (14%). The majority of users identified as male (57%), with 36% identifying as female, and one user selecting another gender category. The average age of participants was 61.8 years, with a range of 28 to 78 years. On average, the reading impairment had been present for 4.8 years, with all participants in the chronic phase (ranging from 2 to 9 years). Participants demonstrated an average baseline reading accuracy of 79% on the WRT with scores ranging from 30% to 97%. Regarding therapy dose, participants achieved an average of 66.8 hours of therapy, with a range spanning 20.1 to 386 hours. VFT scores are detailed in Supplementary Table 1 of the Appendices.

User	Demographics				Baseline Tests					iReadMore Therapy
	Sex	Age (Years)	Cause of Reading Impairment	Time Since Onset of Reading Impairment (Years)	WRT - Accuracy (%)	WRT – RT (Seconds)	CAT - Spelling (%)	CAT - WPD Score	cSART - Accuracy (%)	Therapy Dose (Hours)
2	Male	55	Stroke	2.0	96	5.4	83	10	98	36.2
3	Male	71	Brain Injury	2.4	94	5.7	92	12	96	20.3
6	Female	28	Stroke	3.1	97	4.7	75	22	97	20.1
7	Male	61	Stroke	3.5	69	7.6	100	12	89	386.0
9	Male	62	Stroke	7.2	93	6.6	83	20	98	26.3
13	Male	75	Stroke	2.3	60	10.4	0	-2	88	61.5
16	Male	53	Stroke	6.2	60	7.6	0	0	96	28.5
17	Female	61	Stroke	6.3	69	8.2	42	3	92	42.2
21	Male	77	Stroke	3.0	82	14.7	58	3	97	27.2
28	Female	71	Brain Injury	9.3	30	15.9	42	3	96	33.3
30	Female	69	Stroke	4.5	90	4.8	67	9	96	27.7
31	Other	40	Stroke	7.6	91	6.8	67	8	95	21.4
32	Male	65	Stroke	4.4	91	10.2	83	6	96	26.1
34	Female	78	Stroke	4.7	97	4.8	92	4	98	178.8

Table 14 – User demographics for preliminary data analysis.

Data Handling

Duplicate Data Entries in Word Reading Tests

If a user stopped completing a test, closed the app, or experienced a prolonged interruption (e.g., due to internet connectivity issues or a system crash), the test would automatically restart from the beginning the next time the app was opened. When restarted, the test items were presented in a randomised order. Additionally, each test item included a randomised presentation of responses on the screen to minimise the impact of repeated items on test results.

In these instances, some test items were completed twice on separate occasions during the same interval test. This affected two users in the data collection:

- User 6: 44 duplicated entries across Interval Tests 3 and 4.
- User 32: 25 duplicated entries in Interval Test 6.

Duplicate data was removed from the analysis to maintain data integrity, but the affected users were not excluded.

Word Reading Tests Outliers

One user in the data collection (User 28) had irregular scores on the final interval test that they completed (Interval Test 7). This was excluded from the analysis because their reading accuracy scores were extreme outliers, measuring 4 and 6 standard deviations above the mean for trained and untrained conditions respectively. No other instances of outlier test performance were identified.

Missing Word Reading Tests

Missing interval tests were identified for two users (Users 14 and 24) who had completed more than 5 hours of therapy. Neither had achieved the minimal therapy usage to be included in the analysis and had been excluded for this reason. However, the cause of the missing data warranted investigation, which revealed that the dates coincided with a server outage caused by malware affecting our server provider. This outage occurred between 05/07/2021 and 19/07/2021. Importantly, this was the only prolonged server outage during the study, and no other users included in the analysis were affected.

Devices

Of all users who downloaded and accessed the therapy, users were more likely to be using a phone than a tablet (62% compared to 38% respectively). Users were also more likely to be using the therapy on an Apple device over an Android device or device using another operating system (57% to 43%). Three devices could not be identified.

5.2.1.2 Primary Outcomes

Word Reading Test - Accuracy

The change in word reading accuracy for all 14 participants is shown in Figure 20, illustrating improvements compared to baseline for both trained and untrained words. At every timepoint, the improvement in reading accuracy was consistently greater for

trained words than for untrained words, with differences between ranging from 2% (Intervals 3, 5, and 12) to a maximum of 14% (Interval 10).

By the 20-hour therapy mark, the average improvement across all 14 users was a 7% increase in accuracy for trained words, compared to a 3% increase for untrained words.

All participants completed Interval Test 4 (20 hours of therapy); however, participant attrition was observed beyond this timepoint. Tables 15 details the number of tests completed by users across all timepoints.

The peak change in reading performance occurred at interval Tests 8-10 (40–50 hours of therapy), with a 20% average improvement in trained word accuracy. However, this peak reflects the performance of only 3-4 participants, compared to a 6–11% improvement in untrained word accuracy observed during the same time period. These results demonstrate that while therapy outcomes for accuracy of reading trained words were consistently higher, there was notable variability in improvements across participants and timepoints.

Changes in average Word Reading Test accuracy for each user and timepoint is reported in Supplementary Table 2 of the Appendices.

Table 15 - Timepoint attrition - percentage of users at each timepoint.

Timepoint	Number of Users	Percentage of users at timepoint
Baseline	14	100
Interval 1	14	100
Interval 2	14	100
Interval 3	14	100
Interval 4	14	100
Interval 5	11	79
Interval 6	6	43
Interval 7	5	36
Interval 8	4	29
Interval 9	3	21
Interval 10	3	21
Interval 11	3	21
Interval 12	3	21
Interval 13	2	14
Interval 14	2	14
Interval 15	2	14

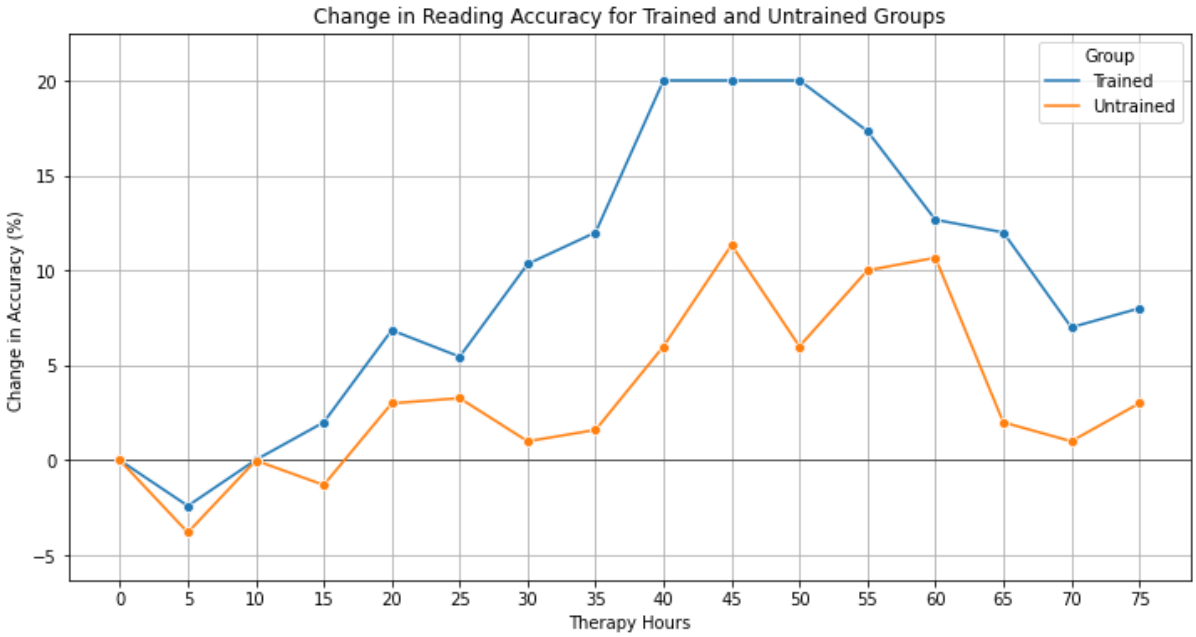

Figure 20 - Change in Reading Accuracy for Trained and Untrained Items.

The therapy effects were analysed using a mixed linear regression model which is well-suited for datasets with significant timepoint attrition. This method utilised all available accuracy data from the 14 participants to evaluate changes in word reading accuracy from Baseline to Interval Test 15.

Main Hypothesis: Testing for a Time*Item Interaction

Only at Interval Test 8 (40 hours of therapy), was a significant interaction between items (trained or untrained words) and timepoint found (p-value = 0.031), indicating significant improvements for trained items compared to untrained items at this specific timepoint with a large effect size (Cohen's d = -2.15). For Interval Test 10, the p-value for the interaction effect was insignificant but just outside the threshold for statistical

significance (p = 0.055, Cohen's d = -1.92). While not statistically significant across most timepoints, trained items generally demonstrated slightly greater improvements in accuracy. The tables of results from the mixed linear regression model are included in Table 16 on interaction effects.

*Table 16 - Interaction Effects Between Group and Timepoint on Reading Accuracy (Significant p-values are indicated with *).*

Timepoint	Unstandardised Coefficient	Standard Error	p-value	95% CI	Cohen's d
Interval 1	-0.014	0.043	0.742	(-0.099, 0.071)	-0.33
Interval 2	-0.033	0.043	0.449	(-0.118, 0.052)	-0.77
Interval 3	-0.033	0.043	0.449	(-0.118, 0.052)	-0.77
Interval 4	-0.039	0.043	0.374	(-0.124, 0.046)	-0.91
Interval 5	-0.022	0.046	0.637	(-0.112, 0.069)	-0.48
Interval 6	-0.093	0.056	0.096	(-0.203, 0.016)	-1.66
Interval 7	-0.104	0.06	0.082	(-0.221, 0.013)	-1.73
Interval 8	-0.14	0.065	0.031*	(-0.268, -0.012)	-2.15
Interval 9	-0.087	0.073	0.235	(-0.23, 0.056)	-1.19
Interval 10	-0.14	0.073	0.055	(-0.283, 0.003)	-1.92
Interval 11	-0.073	0.073	0.315	(-0.216, 0.07)	-1.00
Interval 12	-0.02	0.073	0.784	(-0.163, 0.123)	-0.27
Interval 13	-0.1	0.087	0.249	(-0.27, 0.07)	-1.15
Interval 14	-0.06	0.087	0.489	(-0.23, 0.11)	-0.69
Interval 15	-0.05	0.087	0.564	(-0.22, 0.12)	-0.57

Post-hoc Analysis: Testing for a Simple, Main Effect of Time and Item

The analysis of the main effect of item revealed no significant differences between trained and untrained items in terms of word reading accuracy (p-value = 1). These findings suggest that while significant improvements in accuracy occurred at specific intervals, only at Interval Test 8 did trained items demonstrate a statistically significant improvement over untrained items in word reading accuracy.

Considering improvements across all items, statistically significant improvements in accuracy were observed at Intervals 4, 6, 8, 9, 10, 11, and 13, with p-values ranging from 0.002 at Interval 8, to 0.031 at Interval 13. Conversely, Intervals 1–3, 5, 7, 12, 14, and 15 did not yield statistically significant findings (p-values ranged from 0.07 at Interval 7 to 0.428 at Interval 1). Among the statistically significant intervals, large effect sizes were consistently observed, with Cohen's d values ranging from 2.17 to 2.72.

The tables of results from the mixed linear regression model on timepoint effects is included in Table 17.

*Table 17 - Effects of Timepoint on Reading Accuracy (Significant p-values are indicated with *).*

Timepoint	Unstandardised Coefficient	Standard Error	p-value	95% Confidence Interval	Cohen's d
Interval 1	-0.024	0.031	0.428	(-0.084, 0.036)	-0.774
Interval 2	0.021	0.031	0.485	(-0.039, 0.082)	0.677
Interval 3	0.02	0.031	0.514	(-0.040, 0.080)	0.645
Interval 4	0.069	0.031	0.025*	(0.008, 0.129)	2.226
Interval 5	0.054	0.033	0.103	(-0.011, 0.118)	1.636
Interval 6	0.108	0.04	0.007*	(0.030, 0.187)	2.7
Interval 7	0.078	0.043	0.07	(-0.006, 0.162)	1.814
Interval 8	0.144	0.053	0.002*	(0.052, 0.236)	2.72
Interval 9	0.142	0.053	0.007*	(0.038, 0.245)	2.679
Interval 10	0.142	0.053	0.007*	(0.038, 0.245)	2.679
Interval 11	0.115	0.053	0.029*	(0.012, 0.218)	2.17
Interval 12	0.068	0.053	0.195	(-0.035, 0.172)	1.283
Interval 13	0.135	0.063	0.031*	(0.013, 0.258)	2.143
Interval 14	0.085	0.063	0.173	(-0.037, 0.208)	1.349
Interval 15	0.095	0.063	0.128	(-0.027, 0.218)	1.508

Word Reading Test – Reaction Time

In terms of reaction time for reading trained and untrained items in the word reading test, for the 14 participants, similar trends of improvements were seen for both trained and untrained items, see Figure 21. Change in average Word Reading Test reaction times for each user and timepoint are reported in Supplementary Table 3 of the Appendices.

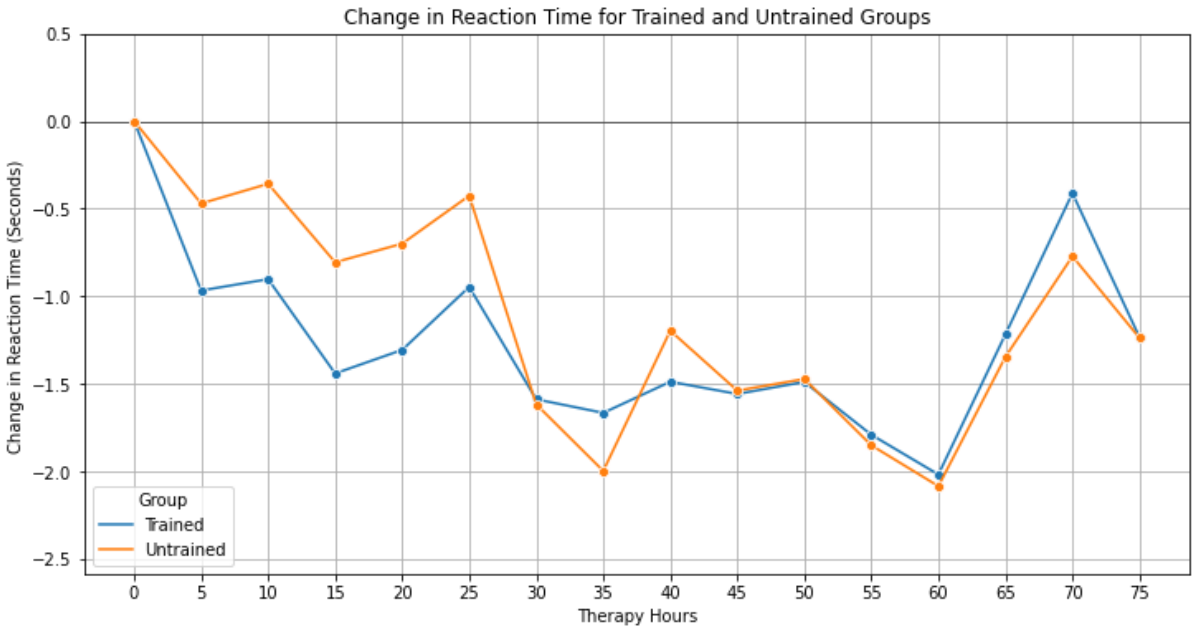

Figure 21 - Change in Reaction Time for Trained and Untrained items.

Main Hypothesis: Testing for a Time*Item Interaction

No statistically significant interaction effects were observed between timepoints and items, indicating that improvements in reaction time did not differ significantly between responses on trained and untrained words. While trained words demonstrated slightly

greater improvements at earlier intervals (Interval Tests 1–4), these differences were not statistically significant.

The standard errors for interaction effects ranged from 0.494 to 0.989, which, relative to the unstandardised coefficients, may suggest some imprecision in the effect estimates.

The table of results from the mixed linear regression model are included in Table 18 on interaction effects.

*Table 18 - Interaction Effects Between Group and Timepoint on Reaction Time (Significant p-values are indicated with *).*

Timepoint	Unstandardised Coefficient	Standard Error	p-value	95% CI	Cohen's d
Interval 1	0.497	0.494	0.315	(-0.472, 1.466)	1.006
Interval 2	0.544	0.494	0.271	(-0.425, 1.514)	1.101
Interval 3	0.635	0.494	0.199	(-0.334, 1.604)	1.285
Interval 4	0.607	0.494	0.22	(-0.362, 1.576)	1.229
Interval 5	0.539	0.527	0.307	(-0.495, 1.572)	1.023
Interval 6	-0.029	0.638	0.964	(-1.28, 1.222)	-0.045
Interval 7	-0.331	0.682	0.627	(-1.667, 1.005)	-0.485
Interval 8	0.291	0.742	0.694	(-1.162, 1.745)	0.392
Interval 9	0.019	0.832	0.982	(-1.612, 1.65)	0.023
Interval 10	0.017	0.832	0.984	(-1.614, 1.648)	0.020
Interval 11	-0.06	0.832	0.942	(-1.692, 1.571)	-0.072
Interval 12	-0.066	0.832	0.936	(-1.698, 1.565)	-0.079
Interval 13	-0.128	0.989	0.897	(-2.067, 1.81)	-0.129
Interval 14	-0.36	0.989	0.716	(-2.299, 1.578)	-0.364
Interval 15	-0.004	0.989	0.997	(-1.942, 1.934)	-0.004

Post-hoc Analysis: Testing for a Simple Main Effect of Time and Item

The analysis of the main effect of item revealed no significant differences between trained and untrained items in terms of reaction time (p-value = 1). These results suggest that reaction time performance was comparable between both sets of items.

Across all items, the mixed linear model analysis revealed significant reductions in reaction time compared to baseline across most timepoints, with the exception of Interval Test 14 (70 hours of therapy).

In terms of the standardised error, standard errors for reaction time ranged from 0.35 to 0.713 across the timepoints, with larger values observed at later intervals. The larger error values at later intervals reflecting increased variability due to smaller sample sizes and participant attrition. These variations in standard errors are consistent with wider confidence intervals observed at later intervals.

Overall, these findings indicate that reaction times improved for both trained and untrained words in the Word Reading Test, with no significant item effects. The findings are explored in detail and context in the discussion.

The table of results from the mixed linear regression model on timepoint effects is included in Table 19.

*Table 19 - Effects of Timepoint on Reaction Time (Significant p-values are indicated with *).*

Timepoint	Unstandardised Coefficient	Standard Error	p-value	95% Confidence Interval	Cohen's d
Interval 1	-0.968	0.35	0.006*	(-1.653, -0.282)	-2.77
Interval 2	-0.903	0.35	0.01*	(-1.588, -0.217)	-2.58
Interval 3	-1.441	0.35	0.0001*	(-2.126, -0.756)	-4.12
Interval 4	-1.307	0.35	0.0001*	(-1.992, -0.621)	-3.73
Interval 5	-0.983	0.375	0.009*	(-1.718, -0.249)	-2.62
Interval 6	-1.447	0.458	0.002*	(-2.344, -0.549)	-3.16
Interval 7	-1.728	0.49	0.0001*	(-2.689, -0.768)	-3.53
Interval 8	-1.64	0.534	0.002*	(-2.686, -0.593)	-3.07
Interval 9	-1.851	0.6	0.002*	(-3.028, -0.675)	-3.09
Interval 10	-1.783	0.6	0.003*	(-2.959, -0.606)	-2.97
Interval 11	-2.084	0.6	0.001*	(-3.261, -0.908)	-3.47
Interval 12	-2.311	0.6	0.0001*	(-3.488, -1.135)	-3.85
Interval 13	-1.848	0.713	0.01*	(-3.245, -0.450)	-2.59
Interval 14	-1.042	0.713	0.144	(-2.439, 0.356)	-1.46
Interval 15	-1.863	0.713	0.009*	(-3.261, -0.466)	-2.61

5.2.1.3 Secondary Outcomes

Children's Sustained Attention to Response Task (cSART)

Performance on the cSART varied inconsistently across timepoints, showing no clear trends. A mixed linear regression analysis, conducted with data from the 14 participants, revealed no significant changes in either cSART accuracy or reaction time across interval timepoints.

Frequency and Intensity of Therapy Practice

At 20 hours of therapy, no significant correlations were observed between frequency and intensity of therapy practice and the outcomes of change in accuracy or reaction time. This was analysed for the 14 users who completed more than 20 hours of therapy. The results are illustrated in Figure 22.

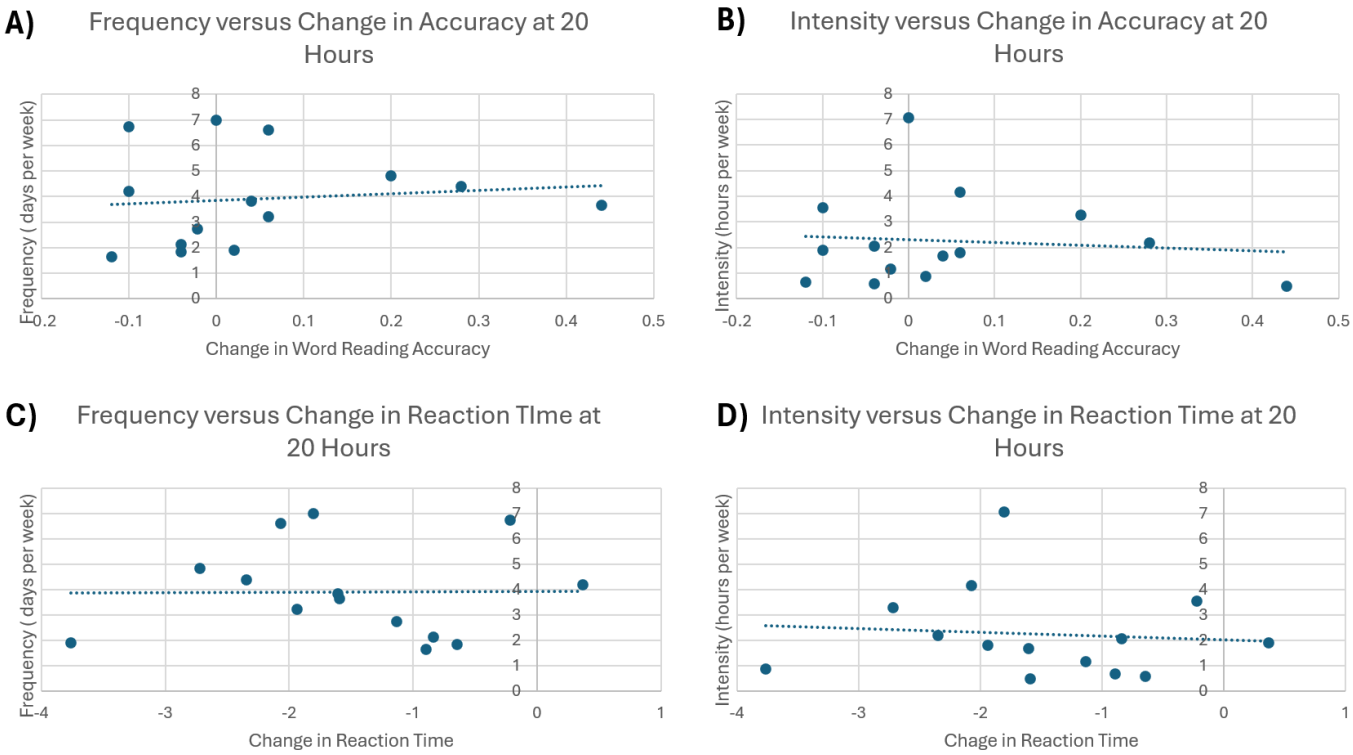

Figure 22 - Frequency and Intensity by Reading Accuracy and Reaction Time.

Quantitative Reading Self-Assessment (Patient Reported Outcome Measure)

PROMs were collected at 10-hour intervals throughout the therapy, with data reported across 7 timepoints for the 14 users.

The average user-reported PROMs, which assessed perceived abilities to read words, sentences, and text using a 7-point Likert scale, showed slight decreases compared to baseline self-assessments (see Table 20). The majority of changes were less than one point on the Likert scales, with the exception of Interval Tests 2 and 12 for reading words and reading sentences.

In mixed linear regression models, self-assessed word reading ability declines over time, and this decline is statistically significant ($p = 0.035$). However, the effect size is very small (Cohen's $d = -0.045$). There is no statistically significant change in self-assessed sentence or text reading ability over time (p-value = 0.338 and p-value = 0.242 respectively).

Table 20 - Change in PROMs by timepoint.

Timepoint	Reading Words	Reading Sentences	Reading Text
Baseline	0	0	0
Interval 2	-1.15	-1.15	-0.38
Interval 4	-0.23	-0.23	0.15
Interval 6	-0.38	-0.38	-0.23
Interval 8	-0.69	-0.69	-0.38
Interval 10	-0.82	-0.82	-0.36
Interval 12	-1	-1	-0.4
Interval 14	0	0	0

Qualitative Analysis of User Feedback

This section presents a qualitative analysis of user feedback from individuals who engaged with the iReadMore app for reading rehabilitation. A total of 55 written responses were received from the 31 users who reached the second interval timepoint when feedback is sought submitted written feedback through the app, with responses varying in length from single words to 50-word comments. The majority of responses were brief, and often consisting of only a single word.

To analyse the feedback, a thematic analysis was conducted, enabling the identification of recurring patterns and themes from these early adopters.

The purpose of this analysis was to better understand user experiences, identify common challenges, and highlight areas for potential improvement offering valuable insights into its usability, engagement, and perceived impact on reading abilities.

Therapy Difficulty and Progression

Users frequently commented on the varying levels of difficulty associated with the app's tasks and their perceived pace of progression. Users commented on the challenging nature of the therapy, calling it *"not easy"* and *"still difficult"*. Shorter words were perceived as manageable by one user, while longer words challenges:

"Longer words more difficult. I often do not have time to read the last letter."

Despite the level of difficulty, users expressed a desire to continue progressing and were hopeful about improving:

"I still hope to improve."

"I am hoping that with more therapy this will improve."

"I have to be patient with it to hopefully see large improvements in my reading."

Additionally, many users expressed that they would like the therapy difficulty to progress at a faster pace:

"I would like to progress quicker to harder words."

"I feel ready to move to harder longer words but I know I have to improve and strengthen all words."

Further, two users requested changes to therapy content training sentences:

"It would be very good if it somehow used sentences rather than just words, because I think I am just getting better at reading those words."

"Use sentences rather than just words"

Therapy Approval

The most commonly received feedback was the word "good" by 8 users. A number of short positive instances of feedback were received regarding user engagement and enjoyment:

"I like the challenge." "Its good thank you."

"It is still very good." "Great that it exist(s)"

One user mentioned a desire for the therapy to be more motivating:

"I'd like more variety or maybe be giv(e)n more feedback. Even if it's just slightly would be more motivating"

Impact and Perceived Effectiveness

User feedback on the app's impact highlighted mixed perceptions. While some users reported perceived improvements in their reading skills, others were uncertain about the extent of the app's contribution related to their therapy goals:

"It has slowly improved."

"It has slightly improved, but the words are not changing that much."

"Good. But unsure if it improves speed which is my biggest concern."

"*I think I am just getting better at reading those words*"

"Possibly a little, still early days but thi(n)k its getting slightly easier"

One user used the feedback portal to highlight they were undertaking multiple therapies simultaneously, making it difficult to attribute improvements specifically to the app:

"I am also doing multiple oral reading therapy at the same time. My reading speed has increased on my baseline reading therapy score. I am not sure what proportion each therapy is contributing to my improvement."

Accessibility Issues

One user noted access issues regarding the app's availability on different platforms, such as laptops and desktops:

"I wish it had been available on a PC as I do not have a tablet or smartphone and have to borrow an iPad every time."

Three users mentioned difficulties with using the required screen orientations needed for the therapy and testing on their devices:

"find it difficult with the elongated screen"

"I enjoy it, I do not enjoy not being able to use the vertica(l) screen"

5.2.2 Discussion

5.2.2.1 Word Reading Accuracy

Preliminary findings from the iReadMore Rollout Trial, based on a small sample of real-world users who completed more than 20 hours of therapy, provide limited indications of a treatment effect for word reading accuracy on the Word Reading Test. However, these findings are not without uncertainty.

A significant improvement in reading accuracy for trained words compared to untrained words was observed at Interval Test 8 (after 40 hours of therapy), with a large effect size (Cohen's d = 2.15). No other timepoints demonstrated significant differences in word reading accuracy for trained words compared to untrained, although further intervals showed large effect sizes that did not reach statistical significance. This preliminary finding after 40 hours of therapy aligns well with results from the iReadMore RCT (Woodhead *et al.*, 2018), which also reported a large effect size (Cohen's *d* = 1.38) after 60-70 hours of therapy split into two blocks of 30-35 hours of training on separate word lists. Potential causes for the lack of significance, including a small sample size and high attrition rate, are discussed in the Limitations section below.

In terms of the mixed linear regression that was conducted, the residual variance of the model (Scale = 0.0066) was relatively low, suggesting that the model captures much of the variability in accuracy across timepoints indicating appropriate model fit and explanation of the observed changes in accuracy.

Figure 20 in section 5.2.1.2 highlights consistent, greater average improvements in accuracy for trained words compared to untrained words. However, this graphical

representation does not account for substantial variation between individuals, with a minority of users performing at or below baseline accuracy on interval tests.

The potential for achieving significant effects with a larger sample size and higher therapy doses remains. These preliminary findings underscore the need for further investigation with a larger cohort to better understand and clarify the digital therapy's effectiveness.

5.2.2.2 Reaction Time

In terms of reaction time on the Word Reading Test, the preliminary findings suggest that reaction times improved for both trained and untrained words in the Word Reading Test, with no significant difference between the two set of items. Furthermore, no specific effect of the therapy on reaction time for trained words was identified.

In terms of the mixed linear regression that was conducted the standard error was high relative to the unstandardised coefficients and increased as timepoints continued. The increase in standard errors at later intervals reflects greater variability in the data, potentially driven by smaller sample sizes resulting from user attrition.

These findings did not align with those of the previous trial, in which a large effect size after 60-70 hours of therapy (Cohen's d = 0.98). Though, the previous trial did find similar trends in improvement in reaction times for both trained and untrained words, as seen in Figure 7 of section 1.4.3.2 (Woodhead *et al.*, 2018). Therefore, further investigation of the dataset once more users have completed significant amounts of

therapy is warranted. Based on the limit evidence provided by the preliminary findings further conclusions cannot be derived at this point. Further discussions on the reaction time findings are included in the limitation section below.

5.2.2.3 PROMs

The preliminary quantitative findings from self-reported PROMs indicated no change or a very small decline in perceived reading abilities for reading words, sentences and text. These outcomes may highlight the challenges users face in accurately self-assessing their reading abilities (Webster, Morris and Howard, 2023). The previous iReadMore RCT found only a non-significant improvement in self assessed word reading abilities. Another perspective is that the training and feedback provided during therapy may help individuals with alexia develop a deeper understanding of their reading impairments as the therapy progresses.

Interestingly, the qualitative feedback revealed that many users believed the therapy was effective, a perception that did not align with the quantitative PROM results they had provided. This discrepancy raises questions about the validity of using a Likert scale to capture nuanced user experiences and perceptions of progress, particularly for complex skills like reading.

The qualitative analysis of user feedback from early adopters of the iReadMore app provides valuable insights into their experiences, challenges and perceived benefits. While most users expressed positive sentiments about the therapy, describing it as "good" and appreciating its availability, several key themes emerged that highlight areas

for improvement. Although many users were motivated to continue with the therapy, some noted that sentence-based training could be more practical for their reading impairments.

Suggestions for improvements to the iReadMore app included the ability to progress more quickly to harder therapy difficulties. Several users expressed that the initial stages of therapy felt too easy or progressed too slowly, potentially hindering engagement and potentially lacking therapeutic effects. This point arose in the iReadMore Co-design process. However, further development to address this feedback was not feasible within the limited resources available to the project. One potential solution could involve using performance on the initial reading tests to adjust the starting therapy difficulties or responding to user performance after the first few challenges to offer users a "difficulty boost" option if desired. This approach could tailor the therapy to individual needs, enhancing user satisfaction and motivation.

Accessibility and access issues were also noted by a small number of users, including difficulties with required screen orientations as well as one expressing a desire for using the app on a computer, rather than a phone or tablet. One user noted that they were borrowing a tablet in order to use the therapy.

Addressing these refinements could improve the therapy's usability, ensuring it better meets the needs of its diverse user base. However, it is important to acknowledge that the method of collecting feedback through written responses may have excluded some users, potentially limiting the diversity and inclusivity of the feedback received.

5.2.2.4 Challenges in User Uptake

The uptake of the iReadMore app has faced several challenges, highlighting the need for continued efforts to engage more users and address barriers. While user numbers are growing, the current dataset is limited, and more users are needed to strengthen the analysis and validate the findings.

Despite the app being CE-marked and listed in reputable directories such as the NHS App Library, ORCHA, and the Aphasia Software Finder, achieving broader adoption continues to remain a challenge.

Various efforts have been made to increase user engagement and uptake, including demos at conferences and patient awareness events, maintaining an up-to-date webpages on the UCL website, and creating YouTube videos including adverts, demos, and aphasia-friendly instructional content. Additional initiatives have included targeted online and print advertisements in the *Royal College of Speech and Language Therapists Bulletin*, features in subject-specific publications, and coverage in a national newspaper both in print and online (Bawden, 2021). While some of these activities generated temporary spikes in interest, the conversion of this interest into sustained user registrations and long-term therapy engagement has remained slow, with increases proving short-lived and insufficient to achieve widespread uptake.

One notable barrier to adoption may be the subscription model, deemed essential to cover ongoing costs of maintaining the app across two app stores. Securing funding from research funding providers to support these operational costs and mitigate costs to the users has proven challenging. This appears to impact the conversion of

downloads into active users beyond the free trial period of one week. Additionally, access to suitable devices, such as tablets or smartphones, has been raised as a challenge. Some users reported borrowing devices to participate in therapy, suggesting that a lack of personal access to compatible technology is a significant obstacle (Cuperus *et al.*, 2023; Kearns and Kelly, 2023). Introducing a version of the app accessible on a computer could help mitigate this issue and expand accessibility.

Despite the previous efforts, a paradox is apparent between the positive qualitative feedback received from users and the high therapy non-adherence observed from real world users of the iReadMore therapy app. While users reported positive experiences, the high attrition rates suggest that long-term engagement may not align with the initial satisfaction reported.

Several biases may explain this paradox. Social desirability bias may lead users to give overly positive feedback to please researchers and limit negative feedback, even if the engagement of the researchers was limit (Paulhus, 1984). Contextual motivation may also play a role, as users feel motivated as participants in the structured environment of research, but struggle to maintain intrinsic motivation when the therapy engagement is self-directed. Sampling bias is another factor, as qualitative feedback may come from users with positive experiences, leaving those who dropped out underrepresented. Lastly, sustained engagement versus initial interest shows that initial enthusiasm may fade over time as users encounter barriers, such as frustration with their perceived progress, the app design or the repetitive nature of the therapy mechanism.

5.2.2.5 Limitations

A number of limitations were identified which should be considered when interpreting the findings.

The preliminary analyses were conducted on a small sample (14 users) with substantial attrition beyond Interval Test 4 (20 hours of therapy), limiting available data for analysis at later timepoints. Since the collation of this preliminary dataset, over 300 additional users have completed the baseline assessment. It is hoped that, over time, more users will complete sufficient therapy for inclusion in the study.

Although mixed linear models effectively manage participant attrition, this results in elevated standard errors and reduced statistical power as sample sizes diminish at later timepoints. This is evident in the higher standard errors and reduced significances observed at later interval tests.

Unexpected improvements in untrained items were observed, which do not align with prior research on iReadMore (Woodhead *et al.*, 2013, 2018). Unlike the 2018 RCT, which utilised multiple baseline assessments to account for test-retest effects. In contrast, the current study only includes a single baseline assessment, as it was anticipated that app users, whose primary motivation is to access the therapy to improve their reading speed and accuracy, rather than contribute to research and would not tolerate multiple baseline assessments. The preliminary data shows that nine out of fourteen participants improved their reaction time for untrained words between Baseline and Interval Test 1 indicating a plausible test-retest effect (Scharfen, Peters and Holling, 2018). However, only four showed improvements in accuracy for untrained words,

indicating that other factors may better explain the lack of significant differences in word reading accuracy.

Of the 14 users included, 8 had baseline word reading accuracies above 90%, and 3 of these were above 95%, leading to potential ceiling effects that may have limited the observable improvements in reading accuracy. In future analyses, as the dataset expands, it may be beneficial to segregate users based on the type of alexia (pure vs. central), as people with pure alexia are likely closer to ceiling and may mask therapy effects on reading accuracy in mixed analyses.

Another limitation was the presence of unusually long reaction times in the dataset, which may not accurately reflect reading speed as intended. Several factors likely contributed to these extended reaction times, including the absence of controlled testing environments, where distractions or differing user priorities, such as focusing on accuracy over speed and may have influenced performance. Furthermore, technical issues, such as reaction times being recorded from the moment a page was opened or delays caused by slow internet connections, were observed during testing and are likely to have artificially inflated the recorded reaction times.

Additional design differences between the previous studies and the present may have influenced the findings. The differences in external support provided by the research team in the structured RCT environment, compared to the self-managed nature of the real-world deployment, likely had a significant impact on therapy attrition and therapeutic outcomes. In the RCT, consistent, personalised support was available from the research team, through in-person appointments and phone calls to ensure

participants accessed the therapy daily helped maintain engagement through extrinsic motivations. In contrast, the real-world deployment, which relied on users to self-manage their therapy, likely contributed to inconsistent engagement and variability in the data. Additionally, some users may have encountered challenges related to motivation, technical difficulties, or limited access to support, all of which could have negatively impacted their therapy progression.

In the RCT, there were two intervals spaced 30–35 hours of therapy apart, whereas the current analysis included 15 intervals spaced 5 hours apart. This increased frequency of testing may have amplified test-retest effects, potentially contributing to the observed improvements in untrained items. Indeed, significant improvements in reading accuracy were observed for both trained and untrained words at seven intervals beyond Interval Test 3. The decision to conduct assessments every 5 hours aimed to address unpredictable attrition and provide personalised feedback to enhance user motivation and engagement. However, this frequent testing may have influenced test performance.

Another key difference lies in the assessment environment. In the previous trials, tests were conducted in-person under controlled research settings, whereas the present study relied on remote, self-managed assessments. This change in methodology may have introduced variability and uncontrollable extrinsic factors, contributing to inconsistencies in the findings. Examples of extrinsic factors observed in the preliminary analysis included a 60% increase in reading accuracy from Interval Test 5 to Interval Test 6 (which was excluded from the analysis). Secondly, another participant

reported in the qualitative feedback that they were undertaking multiple oral re-reading therapy (Moyer, 1979) at the same time as using iReadMore therapy.

Overall, these limitations emphasise the challenges of conducting remote, self-managed assessments and highlight the importance of considering external factors when interpreting findings.

5.2.2.6 Future Work

Several areas for future development and research have been identified for iReadMore therapy. One consideration involves potential design changes to the app, such as introducing a "difficulty boost" feature to better tailor therapy difficulty to individual users. This adjustment could improve user engagement and ensure the therapy remains appropriately challenging. Additionally, developing a computer-based version of the app could significantly increase access, addressing barriers for users who lack access to tablets or smartphones.

Ongoing efforts to promote the app will remain a priority to expand its user base and facilitate further data collection. These activities will build on existing promotional strategies to sustain interest and engagement among potential users and healthcare professionals.

Follow this, the main focus will be the completion of the iReadMore rollout trial, which will provide valuable evidence of the app's effectiveness in real-world users. Further evaluation of iReadMore as a component contributing to an Intensive Comprehensive

Aphasia Program (ICAP) is also planned, examining its contribution within a broader therapeutic framework.

5.3 Study 3 - Prediction of Therapy Outcomes Using Therapy Data

5.3.1 iReadMore Results

5.3.1.1 Predicting Reading Accuracy

Training Models

Models were trained to predict reading accuracy post-therapy. Training performance of the models are included below in Table 21.

Table 21 - Training model error summary for predicting word reading accuracy.

Model	RMSE	SD
Linear Regression	14.4	3.22
Decision Tree Regression	12.8	3.91
Random Forest Regression (Untuned)	13.7	2.60
Random Forest Regression (Tuned)	5.59	2.2
Support Vector Regression (Untuned)	10.0	1.80
Support Vector Regression (Tuned)	7.18	3.41

The models tested varied in accuracy and consistency across cross-validations for predicting reading accuracy post-therapy, with the Linear Regression model exhibiting the highest mean error and the Decision Tree Regression displaying the most fluctuating performance across the leave-one-out cross validations.

Prior to hyperparameter tuning, both Random Forest Regression and Support Vector Regression models demonstrated improved stability. Following hyperparameter tuning, the tuned Random Forest Regression model displayed a significant improvement, with a mean RMSE of 5.59 and standard deviation of 2.2, showing both more accurate predictions and low error variability comparative to other models.

Final Model Performance

Table 22 - Final model performance for predicting word reading accuracy.

Final Models	RMSE	95% Confidence Interval	R^2
Random Forest Regression	2.93	(0, 6.49)	0.74

The final model performance results are reported in Table 22. The tuned Random Forest Regression model produced a final RMSE of 2.93, meaning that model predictions of reading accuracy therapy gains are, on average, within ±3% of the actual improvement observed. Further, a 95% confidence interval ranging from 0 to 6.49, indicating high accuracy, however with some variation in error across different samples. The model

demonstrated strong predictive performance, explaining 74% of the variance in the data (R^2 = 0.74), indicating a good fit to the data.

Figure 23 provides a visualisation of the predicted values compared to the true observed values. Perfect predictions would align with the solid diagonal line, while an acceptable margin of error is illustrated by two dashed diagonal lines representing a 10% deviation from the true values. Of the 21 predictions, 8 fall outside this acceptable range.

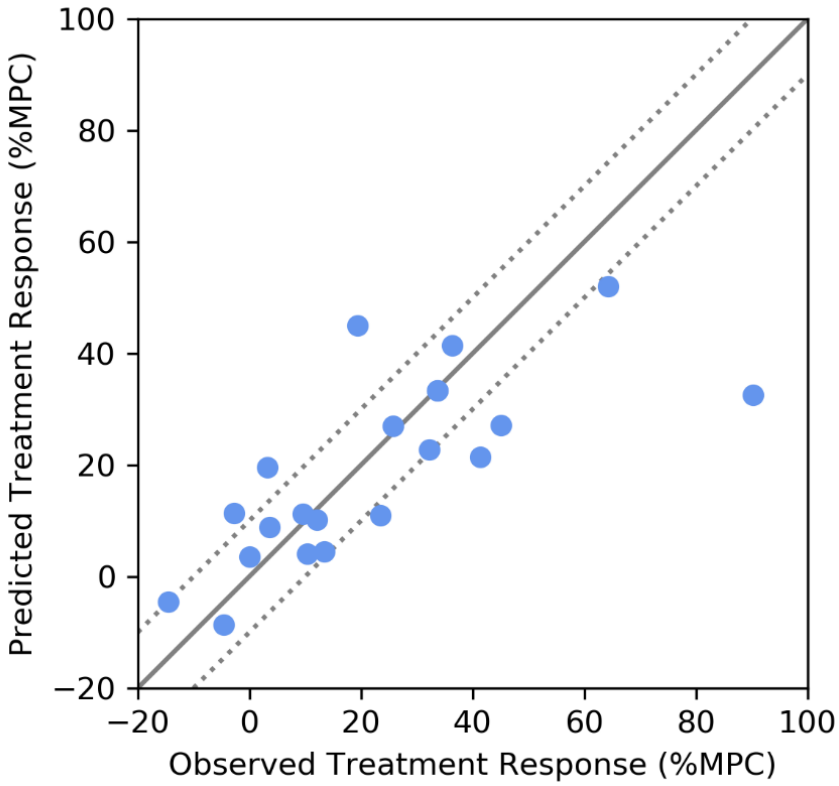

Figure 23 - Predicted treatment response against observed treatment response. A perfect prediction will fall on the diagonal solid line (y=x). The dotted lines denote 10% away from the line y=x.

Sensitivity and Specificity

To calculate the sensitivity and specificity of the model for reading accuracy, a binary classification of Therapy Responders and Non-Responders was applied to the continuous predictions. Therapy Responders are defined as participants achieving ≥ 20% improvement, and Non Responders below this this threshold. The prediction and observed classifications are presented in Table 23, which presents the confusion matrix.

Table 23 - Confusion Matrix for Predicted and Observed Therapy Responses

	Observed Therapy Responder	**Observed Therapy Non-Responder**
Predicted Therapy Responder	**True Positives = 9**	**False Positives = 1**
Predicted Therapy Non-Responder	**False Negatives = 1**	**True Negatives = 10**

From this, the accuracy of the model in classifying therapy responders and non-responders was found to be 90.5%, with the sensitivity (True Positive Rate) being 90% and the specificity (True Negative Rate) being 90.9%.

5.3.1.2 Predicting Reading Reaction Time

Training Models

Table 24 - Training model error summary for predicting reaction time.

Model	RMSE	SD
Linear Regression	188.8	56.25
Decision Tree Regression	368.4	204.0
Random Forest Regression (Untuned)	248.3	155.0
Random Forest Regression (Tuned)	239.0	87.17
Support Vector Regression (Untuned)	290.9	189.5
Support Vector Regression (Tuned)	167.2	42.54

Similarly, models varied for predicting reading reaction time post-intervention (see Table 24). The Decision Tree Regression model had the highest error and variability. The Tuned Support Vector Regression model performed the best with an RMSE of 167.2 and a low SD of 42.54, suggesting significantly improved prediction accuracy and consistency. This was closely followed by the linear regression model, with an RMSE of 188.8 and SD of 56.25.

Final Model Performance

Table 25 - Final model performance for predicting reaction time.

Final Models	RMSE	95% Confidence Interval	R^2
Linear Regression	236.6	(0, 851)	0.20
Support Vector Regression	250.6	(156.7, 317.9)	0.0089

The final results for the two models show substantial differences in their performance and uncertainty for prediction of reading reaction time (Table 25). The Linear Regression model produced an RMSE of 236.6 with a very wide confidence interval ranging from 0 to 851. The wide spread of the confidence interval reflects the model's inconsistency or sensitivity to different subsets of data.

On the other hand, the Support Vector Regression model displayed an RMSE of 250.6, with a narrower confidence interval. Although the RMSE is slightly higher than that of Linear Regression, the confidence interval is more defined, indicating that the SVR model produces more consistent predictions with less variance.

While the SVR model seems to offer greater consistency, both models show high error rates suggesting that neither provides strong predictive accuracy. The low R^2 values reinforce this, indicating that a significant portion of variance remains unexplained. Among the two models, Linear Regression demonstrated a better overall fit ($R^2 = 0.20$), while SVR performed poorly ($R^2 = 0.0089$), meaning SVR is nearly ineffective at capturing meaningful relationships in the data. Notably, when SVR was restricted to a linear kernel, it replicated the results of the Linear Regression model, as expected.

Feature Importance

Feature importance was investigated for the top-performing model for predicting reading accuracy. This analysis aimed to identify the variables most influential in driving predictions. The CAT Word Reading score emerged as the most critical variable. Other factors that were influential to a lesser extent were the cSART Correct Rejections, CAT Picture Description Score, time post-stroke and Word Reading Test Accuracy. A visual representation of these feature contributions is provided in Figure 24.

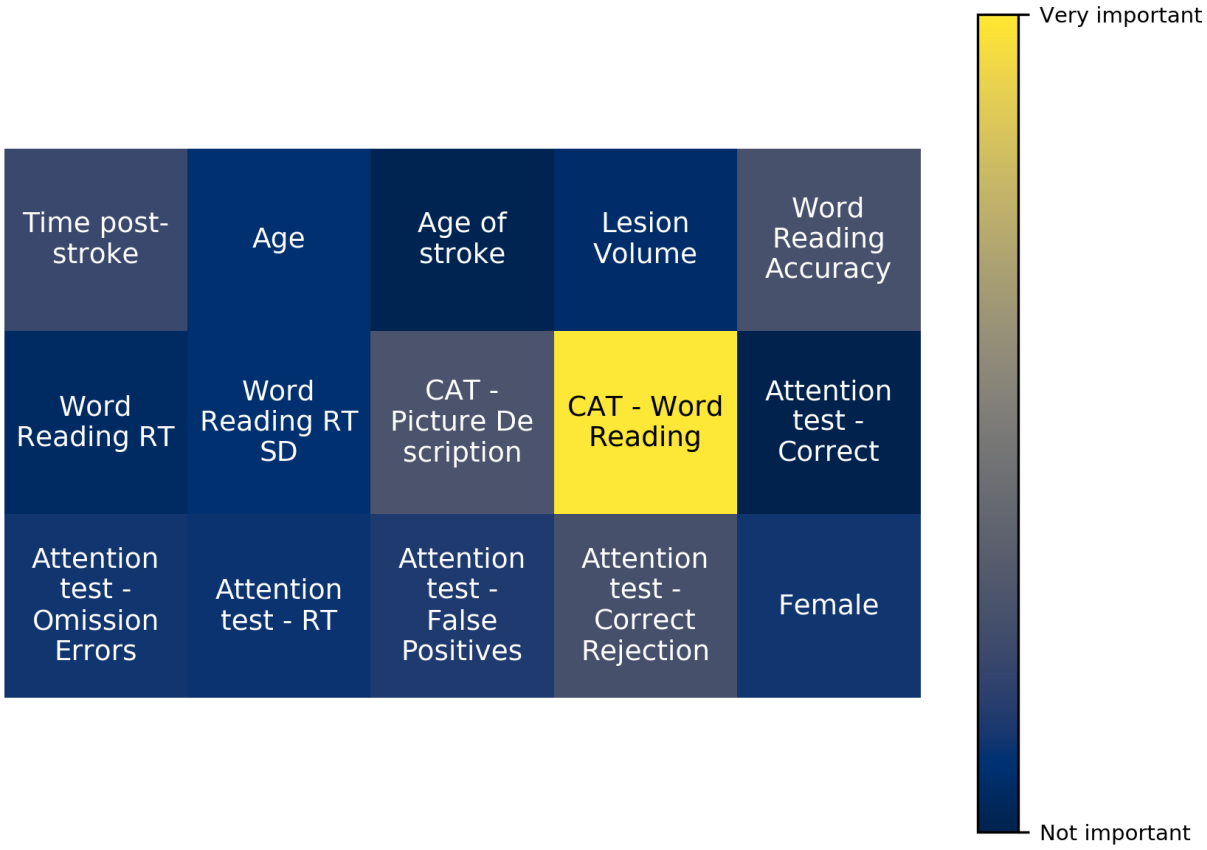

Figure 24 - Feature importance heatmap for predicting word reading accuracy. The colour scale represents the relative importance of different variables, with yellow indicating the most influential factors and dark blue indicating the least important.

5.3.2 iReadMore Discussion

5.3.2.1 Reading Accuracy

The winning model for predicting reading accuracy utilised a random forest regression, which was heavily reliant on the CAT word reading score variable as a predictor. This indicates that the model found the baseline reading score to be a strong determinant of future reading outcomes. Interestingly, this highlights the point that neuroimaging data in the context of predicting treatment outcomes for aphasic patients may not always be necessary and there is a need for finding alternative variables to drive prediction algorithms (Price, Seghier and Leff, 2010; Harvey, 2015). The data from iReadMore demonstrated that non-imaging features, like baseline reading accuracy, can provide a useful framework for predicting treatment outcomes, supporting the growing body of work suggesting that simple, accessible data can be powerful for prediction (Iorga *et al.*, 2021).

By converting the outcomes prediction into a classification of therapy responders and non-responders, the practical utility of the model was explored. Classification accuracy of approximately 90.5 % and sensitivity of 90 % indicate reliable identification of true therapy responders. Specificity of approximately 90.9 % demonstrates robust exclusion of non-responders, despite the limited sample sizes.

In comparison to a previous regression model developed using the iReadMore randomised trial data (Aguilar *et al.*, 2018), which incorporated neuroimaging data, the current model demonstrated superior predictive accuracy. The earlier model yielded a relatively lower R^2 value of 0.48 (95% CI: 0.08–0.75, p=0.02), indicating substantial error,

whereas the current model achieved a higher R^2 of 0.7 (95% CI: 0.48–1.00, p=0.0001) indicates that 70% of the variance in reading gains is explained by the model. This suggests that the inclusion of neuroimaging data may not always improve model performance (Chang *et al.*, 2021).

However, while this model showed promise, it did not reach the precision necessary for clinical application, particularly in the context of aphasia treatment. Further, the utility of therapy data itself warrants exploration. In this study, there was considerable variability in baseline reading accuracy, with scores ranging from as low as 3% to as high as 97%, reflecting a wide variety of starting points among participants. This variability, along with significant missing data (5.5% across all variables, and as high as 23.8% for some variables), adds complexity to the prediction task. While the missing data was largely due to manual data collection methods and the challenges of measuring certain variables in severely impaired participants, this issue could be mitigated in future studies with automated data collection systems (Bishop, 2019). This would reduce the risk of missing data and improve the model's accuracy over time.

Imputing missing data also remains a challenge, particularly when the cause of missingness is unknown, as this can lead to biased estimates. For instance, when missing data was due to low or average scores, it was difficult to impute accurately, potentially skewing the results. Despite these limitations, the model's reliance on baseline reading accuracy and the exploration of therapy data demonstrates the value of these features for predicting treatment outcomes, while also highlighting the need for more robust data handling strategies in future work.

5.3.2.2 Reading Speed

The prediction of reading reaction time in the iReadMore trial was less successful compared to the prediction of reading accuracy. Linear Regression achieved an R^2 of 0.20, indicating a weak but non-random relationship between features and reading speed, while Support Vector Regression (SVR) performed notably worse, with an R^2 of just 0.0089, suggesting almost no predictive value.

The Support Vector Regression (SVR) model, although widely used in machine learning for its robustness in high-dimensional data, did not show significant improvement in this case. The model's performance was constrained by factors such as the variability in individual reading speeds and the challenge of capturing complex, nonlinear relationships with limited data. The low R^2 values suggest that the current feature set does not adequately explain variance in reaction time, highlighting the need for additional predictive factors.

It is possible that the model's ability to predict reading speed is hindered by insufficient features or data related to underlying cognitive and language processes. For example, the inclusion of more nuanced in-therapy data or more targeted behavioural metrics, such as attentional focus or response time to specific task types, might improve predictions. Additionally, like other models in this space (Chen *et al.*, 2020), SVR can struggle when the data does not exhibit clear, linear patterns or when the number of samples is too small to effectively capture variability across individuals.

Further refinement of the model, including exploration of additional features (e.g., fine-grained therapy progress, patient-specific factors), could help enhance prediction accuracy for reading speed. Additionally, combining SVR with other machine learning approaches, such as ensemble methods or deep learning, might improve its ability to handle the complex relationships that affect reading performance.

5.3.2.3 Limitations

The sample size of 21 participants used to train the prediction model is very small. This is limiting the statistical power and affects the model's ability to learn and generalise effectively. Additionally, the dataset is drawn from a single-centre cohort with a relatively homogeneous demographic, which may not accurately reflect the broader real-world clinical population who will ultimately use the therapy. To enhance the model's robustness and validity, a larger and more diverse sample is needed for retraining, ensuring it is better suited to generalise across different patient populations and clinical settings.

Finally, therapy dosage and participant engagement metrics were collected under controlled study conditions; real-world implementation in diverse clinical settings may introduce variability not captured here. Therefore, these findings should be interpreted cautiously, and further development and validation is required before clinical implementation.

5.3.2.4 Next Steps

Given the limited dataset currently available for the iReadMore app, and the interest in exploring in-therapy data as a potential additional variable for the prediction of treatment outcomes, the next phase of this study utilised data from the randomised trial of the Listen-In speech comprehension app. This trial provided a more robust dataset including in-therapy data to further investigate the relationship between in-therapy data and treatment efficacy.

In the future, once sufficient real-world data becomes available for iReadMore, it would be valuable to extend this work by leveraging this data to train models for predicting reading accuracy and speed. The integration of such data could provide deeper insights into predicting outcomes and refining treatment strategies for users outside of controlled research settings.

5.3.3 Listen-In Results

5.3.3.1 Model Performance

Four models were trained on varying amounts of in-therapy data from no data (baseline variables only) up to all therapy data completed for each user. The model performances are includes in Figure 25.

With no in-therapy data, a tuned random forest model was the best performing. With the addition of in-therapy data relating to the first 10 therapy blocks completed, an elastic net regression model achieved the lowest RMSE at this level of data provided (RMSE = 7.1). This continued with the Elastic Net models outcompeting with the addition of further in-therapy data. The lowest RMSE was achieved at 50 blocks for the Elastic Net model (RMSE = 6.1); which was lower than with all data added (RMSE = 6.5)

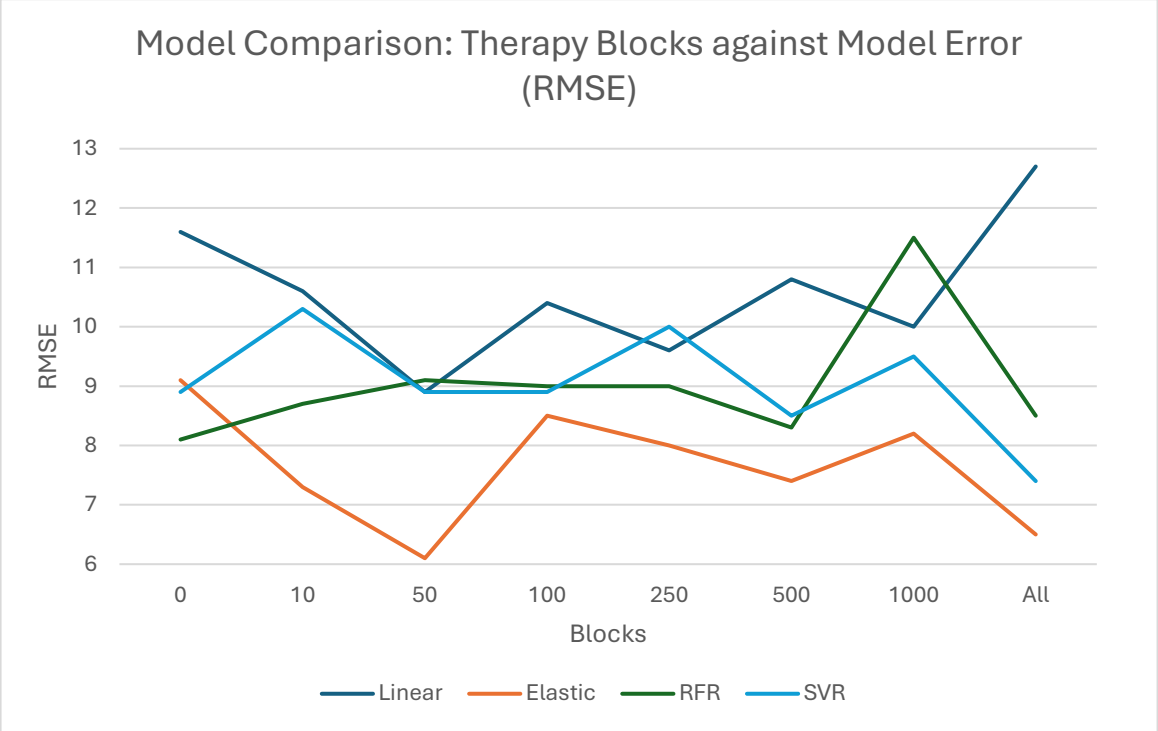

Figure 25 - RMSE model comparison for training models.

For Elastic Net models only, the RMSE, RMSE SD and R^2 values by varying therapy blocks are presented in Table 26 below. Across measures, the Elastic Net models trained with 10 and 50 blocks of data outperformed the other models, not including all therapy data. These blocks translate to completing approximately 1 and 5 hours of therapy respectively in order to generate a prediction.

Table 26 - Elastic Net model error by number of therapy blocks used in training data.

Number of therapy blocks used in training dataset	RMSE	SD	R^2
0	9.1	2.9	0.61
10	7.3	2.6	0.86
50	6.1	2.7	0.85
100	8.5	2.9	0.8
250	8	3.4	0.84
500	7.4	2.8	0.83
1000	8.2	2.8	0.93
All (1865)	6.5	3.1	0.9

5.3.3.2 Final model performance

The scores for the winning model are included in Table 27.

Table 27 - Winning model summary.

Model	RMSE	SD	R^2
Elastic Net Regression using in-therapy data from 50 Therapy Blocks (approximately 5 hours of therapy)	6.1	2.7	0.85

5.3.3.3 Feature Importance

A correlation matrix presents a comparison of all 46 features in Figure 26. High correlation was expected between the linguistic variables and variables derived from the same behavioural assessments.

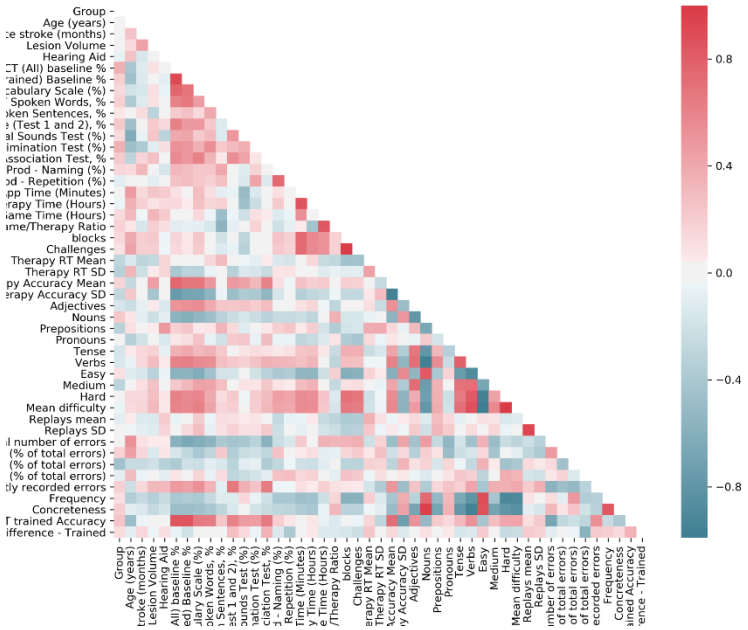

Figure 26 - Correlation matrix of all features used in model development.

Due to the high correlations, an association dendrogram was constructed using hierarchal clustering (Figure 27) to visualise the relatedness of the features utilised in the final model. The groupings by colour denote three distinct groups of related variables (green, blue and red groupings). This found that the predictor (ACT trained accuracy) is different to the dependent variables as it is the only blue variable. The variables names have been highlighted to demonstrate those which were driving the prediction in the final model. The 7 variables were: sex, percentage of challenges presented that had an 'easy' difficulty, performance on trained words in the ACT baseline test, the number of unrelated errors generated by the user, their mean therapy accuracy and SD, and the total number of errors generated.

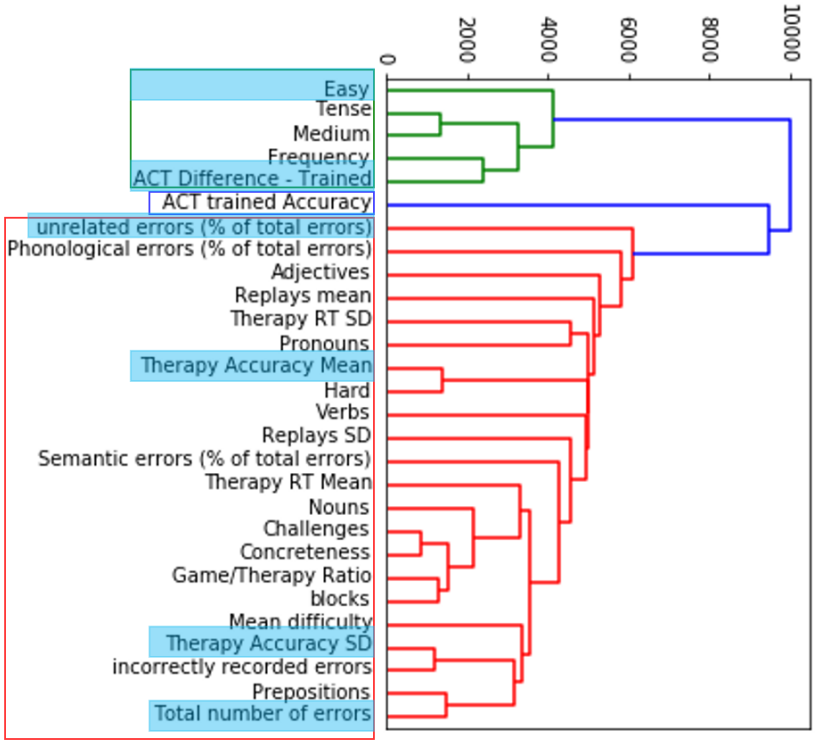

Figure 27 - Association dendogram for all features in the dataset. Three groups of features were identified (highlighted in green, blue and red). Features highlighted in a blue box were used in the final model.

5.3.4 Listen-In Discussion

5.3.4.1 Appropriateness of Winning Model

The inclusion of in-therapy data substantially enhanced the predictive accuracy of the winning model. Among all models and input combinations, the highest performance was achieved using elastic net regression and data from the first 50 therapy blocks (approximately 5 hours of therapy), yielding an r^2 of 0.85 and RMSE of 6.1; meaning that 85% of the variance in speech comprehension gains are explained by the model and predicted improvements in speech comprehension deviate by around ±6% from the actual observed values. Notably, predictions based on just the first hour of therapy data (r^2 = 0.86, RMSE = 7.3) were only marginally less accurate and remained significantly more effective than predictions based solely on baseline data without therapy input (r^2 = 0.61, RMSE = 9.1). These findings underscore the value of early in-therapy data for enhancing prediction accuracy, even with limited initial therapy exposure.

The elastic net regression model's suitability for this dataset is supported in the literature. Elastic net regression is particularly effective for applications with small sample sizes, a common limitation in aphasia therapy research, as it balances the strengths of Lasso and Ridge regression. This hybrid approach minimises overfitting while effectively removing non-contributory variables, ensuring robust, interpretable models even in the presence of high dimensionality or multicollinearity (Zou and Hastie, 2005). Moreover, the application of elastic net regression in previous studies on aphasia recovery and language therapy outcomes further validates its utility in this domain (Chang *et al.*, 2021; Iorga *et al.*, 2021). These studies demonstrate elastic net regression's capacity to identify and model complex relationships within heterogeneous

datasets, a critical feature for accurately predicting therapy outcomes. By integrating diverse predictors and maintaining high predictive accuracy, elastic net regression emerges as a valuable tool for exploring treatment effects and facilitating personalised therapy approaches in digital interventions.

5.3.4.2 Model Drivers

The optimal predictive model incorporated a mix of baseline and in-therapy features, highlighting the nuanced interplay between pre-therapy performance and real-time engagement metrics. Key features included accuracy on trained words in the ACT baseline test, which may have served as a foundational measure of participants' initial proficiency with therapy-specific items. Within the therapy, the percentage of easy therapy challenges may capture performance on simpler tasks and rate of progression to more difficult challenges, reflecting learning potential and engagement during therapy. The inclusion of the total number of errors generated in the therapy and the percentage that were unrelated errors may offered insights into error patterns associated with the users' impairment or lack of attention on the task. While mean therapy accuracy (%) and standard deviation quantified not only the overall success rate but also the consistency and variability in therapy performance. Finally, Sex was incorporated potentially to capture a demographic variable to account for potential differences in therapy outcomes. Sex has been identified as a predictor for other aphasia therapies outcomes (Kristinsson *et al.*, 2022), however it could also be due to overfitting in a small and imbalanced dataset.

5.3.4.3 Prediction of Outcomes for Therapy Non-responders

Despite the promising evaluation scores of the winning model, a limited dataset contributed to its underperformance, particularly in predicting outcomes for participants whose comprehension decreased post-therapy (n=4). This is illustrated in Figure 28 showing the predictions for the 4 participants in the test dataset, where one participant's prediction falls outside the acceptable margin of error. Efforts to address this, including trialling down-sampling, did not yield improvements. The model's performance would benefit from a larger pool of data. Other regression algorithms have employed the development of a responders and non-responders model (similar to Billot *et al.*, 2022) as a consideration for future work in this area.

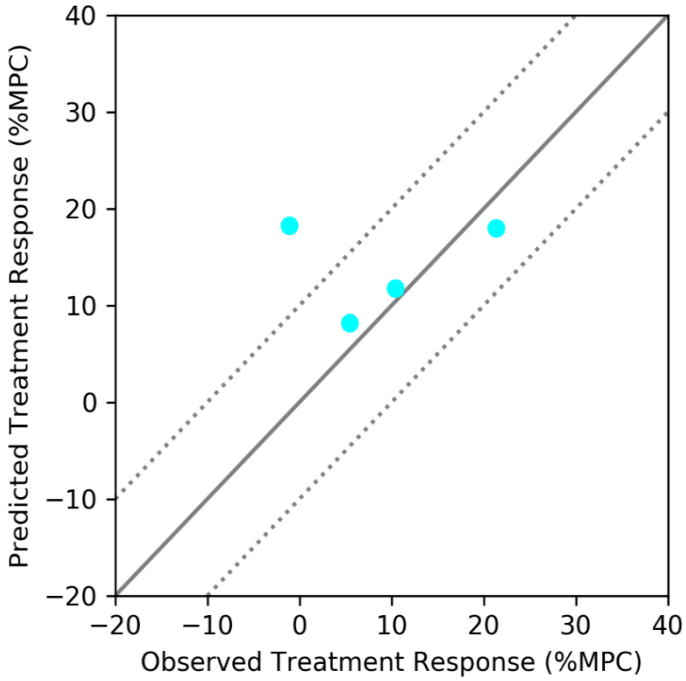

Figure 28 - Prediction of treatment response in the test dataset (n=4). Predictions are generally accurate, except for one participant who did not improve in observed treatment response, the prediction expected around a 20% increase.

5.3.4.4 Challenges with Alternative Models

Interestingly, the addition of in-therapy data did not enhance the performance of Random Forest Regression (RFR) and Support Vector Regression (SVR) models. This outcome can be attributed to several factors. RFR, while effective for handling high-dimensional data by averaging predictions from multiple decision trees, may struggle to capitalise on incremental information provided within otherwise noisy and redundant in-therapy data. If the new data adds only marginal value or introduces noise, RFR's ability to extract meaningful patterns diminishes. Similarly, SVR relies heavily on parameter tuning and data structure to perform effectively. Furthermore, the smaller sample size and added complexity of the dataset could further exacerbate these issues, particularly for models sensitive to dimensionality.

5.3.4.5 Implementation

Integrating a predictive algorithm into the Listen-In digital therapy warrants further evaluation, particularly because the model relies on baseline assessments, some of which, like the ACT baseline, were not delivered within the app as in its publicly available version. However, the ability to generate predictions after just 1 to 5 hours of therapy makes this approach practical for prospective users, with predictions potentially available within a day to a week, depending on therapy usage patterns, as observed in the randomised controlled trial. Despite this, data from the iReadMore rollout trial highlight the significant variability between research settings and real-world usage, especially in terms of therapy dose and intensity. This discrepancy poses challenges for broader implementation and emphasises the need for further exploration when real-world data becomes more abundant.

With the rising global incidence of stroke and the increasing number of stroke survivors each year (Seminog *et al.*, 2019), the demand for neurorehabilitation services is growing, and addressing this clinical unmet need is apparent. Innovative methods to deliver effective, evidence-based, high-dose therapies to the right candidates (or high responders) are necessary. The combination of digital health tools and treatment outcome prediction models, particularly within a feedback loop, presents a significant opportunity to scale treatment planning and address the needs of a larger patient population, ensuring timely and personalised care (Sutton *et al.*, 2020; Zhong, 2024).

5.3.4.6 Limitations

The number of appropriate regression models was limited by the small sample size, which may have affected model performance. Some models, like random forest regression and support vector regression, may not have been suitable for a small dataset. Elastic Net Regression is well-suited for smaller sample sizes due to its ability to prevent overfitting. However, it cannot model non-linear relationships, potentially missing valuable insights. Additionally, the model was trained on data from a randomised controlled trial, which involved high therapy doses (70-100 hours) delivered in a research context, and may not be reflective of real-world usage in a self-led digital therapy. Further, regarding data quality, although the amount of missing data was minimal, missing therapy data could have impacted the accuracy of predictions.

The Listen-In model was trained on data from only 17 participants, and the independent test sample was even more limited (n = 4), both of which severely restrict statistical power and undermine the reliability of performance estimates. With so few individuals contributing to model fitting, even minor fluctuations in individual outcomes can disproportionately influence the fitted parameters and inflate apparent performance.

The representativeness of the sample is also a consideration. The trial's inclusion and exclusion criteria, such as participant severity, comorbidities, and the male/female ratio, may limit the generalisability of the model to a real-world population. The training cohort was drawn from a controlled trial with homogeneous demographics and consistent protocol adherence, failing to capture variability typical of real-world clinical populations. The absence of an external validation cohort precludes any assessment of

overfitting, meaning that the strong reported R^2 values (0.85–0.86) may not translate to independent samples.

5.3.4.7 Future Work

This study represents a preliminary exploration of using in-therapy data to enhance predictive models for aphasia treatment response. However, further research and data collection are necessary before implementation is possible.

As digital health technologies have advanced, routine collection of in-therapy data is now highly feasible. The Listen-In app, now available on the Google Play Store for Android devices, presents an opportunity for large-scale real-world data collection, which will be essential for refining the predictive model. To improve the generalisability of the model, future work should involve retraining the model using a much larger dataset from app users using the therapy in a self-led manner outside of a research context. Understanding how smaller therapy doses may impact predictions will be key to improving the model's ability to predict outcomes for a broad range of users outside of a research context.

Another next step to explore is integrating in-therapy data alongside neuroimaging data to train models; this was beyond the scope of the current study, which prioritised ecological validity and the feasibility of real-world implementation in a digital therapy.

Finally, to transition these predictive technologies to clinical practice, additional challenges, not previously mentioned herein will need to be addressed. These include ensuring that the model is equitable and fair, overcoming implementation barriers, and meeting regulatory standards for medical applications.

6 Discussion

6.1 Summary of Findings

This thesis consists of the co-design of the iReadMore app, preliminary findings of the iReadMore rollout trial investigating clinical effectiveness, and explorations in treatment outcome prediction for digital aphasia therapies.

The three overarching aims of this thesis were:

1. **Chapter 1** – To co-design the iReadMore app to enhance accessibility and therapy engagement.
2. **Chapter 2** – To assess the effectiveness of iReadMore in improving reading accuracy and reaction time.
3. **Chapter 3** – To explore the potential for predicting treatment outcomes using only variables collected via digital therapy.

In **Chapter 1**, I co-designed the iReadMore app with 50 participants, including 42 individuals with aphasia or alexia and 8 family members or carers. The co-design process occurred in two phases: focus groups and beta testing. This process led to a publication on design recommendations for digital alexia therapies, utilising a qualitative framework analysis methodology. Key themes identified from the framework analysis included motivation, accessibility, social sharing, widening participation, relatable content, personal trajectory, agency, and intuitive design. Further development of the co-designed version, supported by additional funding from

Research England, resulted in its release on both the Apple App Store and Google Play Store. Additionally, the iReadMore app received CE marking as a medical device.

Future developments, which were not possible due to resource constraints, include creating a laptop version, offering more personalised therapy difficulty starting points, providing more tailored feedback, and exploring how digital therapies can better promote agency and social participation.

In **Chapter 2**, I evaluated preliminary findings from a small cohort of 14 real-world users. The limited results revealed a significant effect of the therapy at one time point (Interval Test 8) for reading accuracy on trained words compared to untrained words (large effect size, Cohen's d = 2.15). However, no significant effects were found for trained words with regards to reaction time. The trial is ongoing, and further data will determine whether additional significant findings emerge, consistent with previous research on the iReadMore therapy mechanism.

Interestingly, the real-world usage data suggested that therapy engagement, including dose, frequency, and intensity, may differ from research trial settings. This may provide a valuable dataset for investigating self-led therapy usage, where higher doses and intensity could be crucial for enhancing therapeutic effects. During the co-design phase, it was believed that extrinsic motivators, such as push notifications, were unnecessary, as users would be intrinsically motivated to use the therapy. However, further exploration of push notifications, as one example, could help boost engagement and therapy dosage.

In **Chapter 3**, I investigated whether data commonly collected in digital aphasia therapies could be used to predict treatment outcomes. An initial study using data from the iReadMore RCT provided limited therapy data for model training. However, a richer dataset from the Listen-In RCT offered more robust data for training models based on therapy progression. This suggests that treatment outcome prediction for reading and speech comprehension accuracies may be feasible and scalable in digital therapies. Both iReadMore and Listen-In are currently available (though Listen-In is only on the Google Play Store). As user data increases, revisiting this concept could lead to the development of implementable prediction algorithms.

Such predictive algorithms could empower users to determine whether a therapy is suitable for them and enhance motivation by providing feedback that reinforces their chances of improvement. However, the framework analysis of the iReadMore co-design process revealed mixed findings regarding participant perspectives on outcome prediction in digital aphasia therapies. Concerns were raised about the accuracy of predictions and the potential denial of therapy access due to negative outcomes. Conversely, positive predictions were seen as motivational tools that encouraged continued therapy engagement. These findings underscore the importance of addressing user concerns with careful implementation of predictive algorithms to ensure the transparent communication of predictive results, without discouraging user motivation.

6.2 Future Direction

Looking ahead, the primary focus of the iReadMore project will be the completion of the clinical effectiveness rollout trial. This will involve recruiting additional users to complete sufficient therapy sessions for inclusion in the trial. Strategies to enhance user engagement and improve therapy adherence will be essential for continued data collection. One potential approach to support this is the development of a laptop-based version of the therapy.

As more data becomes available, the additional baseline assessment datapoints can be leveraged to gain deeper insights into reading rehabilitation in individuals with aphasia and alexia.

Additionally, once substantial data becomes available for iReadMore or Listen-In, further exploration into treatment outcome prediction will be pursued. One promising avenue to support both the completion of the trial and the prediction efforts is the development of a dashboard for real-time monitoring of user therapy progress. This could provide valuable insights into user engagement and therapy effectiveness, helping to drive both participation and outcomes.

7 References

Aguilar, O. M. *et al.* (2018) 'Lesion-site-dependent responses to therapy after aphasic stroke', *Journal of Neurology, Neurosurgery and Psychiatry*. doi: 10.1136/jnnp-2017-317446.

Altmeyer, M., Lessel, P. and Krüger, A. (2018) 'Investigating gamification for seniors aged 75+', in *DIS 2018 - Proceedings of the 2018 Designing Interactive Systems Conference*. doi: 10.1145/3196709.3196799.

Amaya, A. *et al.* (2018) 'Receiving aphasia intervention in a virtual environment: the participants' perspective', *Aphasiology*. doi: 10.1080/02687038.2018.1431831.

Barbur, V. A., Montgomery, D. C. and Peck, E. A. (1994) 'Introduction to Linear Regression Analysis.', *The Statistician*. doi: 10.2307/2348362.

Bawden, T. (2021) 'New app can boost the reading ability of stroke survivors with language impairment', *The i Paper*. Available at: https://inews.co.uk/news/science/new-app-boost-reading-ability-stroke-survivors-language-impairment-1089400?srsltid=AfmBOoprJ998bmJ1xnOqsJQw90EI5ratdbHgRu26ZTqh_H1muQrwt19e.

Beber, B. C. *et al.* (2014) 'Logopenic aphasia or Alzheimer's disease: Different phases of the same disease?', *Dementia & Neuropsychologia*. doi: 10.1590/s1980-57642014dn83000016.

Beeson, P. M. (1999) 'Treating acquired writing impairment: Strengthening graphemic representations', in *Aphasiology*. doi: 10.1080/026870399401867.

Beeson, P. M. *et al.* (2010) 'A treatment sequence for phonological alexia/agraphia', *Journal of Speech, Language, and Hearing Research*, 53(2), pp. 450–468. doi: 10.1044/1092-4388(2009/08-0229).

Ben-Shachar, M. *et al.* (2011) 'The development of cortical sensitivity to visual word forms', *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn.2011.21615.

Berthier, M. L. (2005) 'Poststroke aphasia: Epidemiology, pathophysiology and treatment', *Drugs and Aging*. doi: 10.2165/00002512-200522020-00006.

Bhogal, S. K., Teasell, R. and Speechley, M. (2003) 'Intensity of aphasia therapy, impact on recovery', *Stroke*. doi: 10.1161/01.STR.0000062343.64383.D0.

Biel, M. *et al.* (2022) 'Motivation Theory and Practice in Aphasia Rehabilitation: A Scoping Review', *American Journal of Speech-Language Pathology*. doi: 10.1044/2022_AJSLP-22-00064.

Billot, A. *et al.* (2022) 'Multimodal Neural and Behavioral Data Predict Response to Rehabilitation in Chronic Poststroke Aphasia', *Stroke*, 53(5), pp. 1606–1614. doi: 10.1161/STROKEAHA.121.036749.

Binder, J. R. *et al.* (2006) 'Tuning of the human left fusiform gyrus to sublexical orthographic structure', *NeuroImage*. doi: 10.1016/j.neuroimage.2006.06.053.

Binney, R. J. *et al.* (2010) 'The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: Evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia', *Cerebral Cortex*. doi: 10.1093/cercor/bhq019.

Bishop, D. (2019) 'Rein in the four horsemen of irreproducibility', *Nature*. doi: 10.1038/d41586-019-01307-2.

Brady, M. C. *et al.* (2012) 'Speech and language therapy for aphasia following stroke', *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD000425.pub3.Copyright.

Brady, M. C. *et al.* (2016) 'Speech and language therapy for aphasia following stroke', *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD000425.pub4.

Brady, M. C., Ali, M., Berg, K. Vanden, *et al.* (2022) 'Complex speech-language therapy interventions for stroke-related aphasia: The RELEASE study incorporating a systematic review and individual participant data network meta-analysis', *Health and Social Care Delivery Research*. doi: 10.3310/RTLH7522.

Brady, M. C., Ali, M., VandenBerg, K., *et al.* (2022) 'Dosage, Intensity, and Frequency of Language Therapy for Aphasia: A Systematic Review-Based, Individual Participant Data Network Meta-Analysis', *Stroke*, 29(2), pp. 956–967. doi: 10.1161/STROKEAHA.121.035216.

Brambati, S. M. *et al.* (2009) 'Reading disorders in primary progressive aphasia: A behavioral and neuroimaging study', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2009.02.033.

Breiman, L. (2001) 'randomForest: Breiman and Cutler's Random Forests for Classification and Regression', *Machine Learning*.

Breiman, L. *et al.* (2017) *Classification and regression trees*, *Classification and Regression Trees*. doi: 10.1201/9781315139470.

Brookshire, C. E. *et al.* (2014) 'Frequency, nature, and predictors of alexia in a convenience sample of individuals with chronic aphasia', *Aphasiology*. doi: 10.1080/02687038.2014.945389.

Brysbaert, M. and New, B. (2009) 'Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and

improved word frequency measure for American English', *Behavior Research Methods*. doi: 10.3758/BRM.41.4.977.

Butts, A. M. *et al.* (2015) 'Neuropsychological Profiles Differ among the Three Variants of Primary Progressive Aphasia', *Journal of the International Neuropsychological Society*. doi: 10.1017/S1355617715000399.

Cahana-Amitay, D. *et al.* (2011) 'Language as a stressor in aphasia', *Aphasiology*. doi: 10.1080/02687038.2010.541469.

Cann, K. and Bulman, L. (2018) *Technology and Aphasia Review*. Available at: http://www.aphasiafriendly.co/aphasia-and-technology.html.

Cardol, M., De Jong, B. A. and Ward, C. D. (2002) 'On autonomy and participation in rehabilitation', *Disability and Rehabilitation*. doi: 10.1080/09638280210151996.

Carlson, T. A. *et al.* (2014) 'The emergence of semantic meaning in the ventral temporal pathway', *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn_a_00458.

Carreiras, M. *et al.* (2014) 'The what, when, where, and how of visual word recognition', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2013.11.005.

Cattell, R. and Cattell, A. (1960) 'Handbook for the individual or group culture fair intelligence test', *Savoy, IL: Institute for Personality and Ability Testing*.

Caute, A. *et al.* (2018) 'Assistive technology approaches to reading therapy for people with acquired dyslexia', *Aphasiology*. doi: 10.1080/02687038.2018.1489119.

Chang, A. J. *et al.* (2021) 'Cortical microstructural changes associated with treated aphasia recovery', *Annals of Clinical and Translational Neurology*. doi: 10.1002/acn3.51445.

Chapey, R. *et al.* (2012) 'Life-participation approach to aphasia: A statement of values for the future', in *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders: Fifth Edition*. doi: 10.1044/leader.ftr.05032000.4.

Chen, L. *et al.* (2020) 'Improving the Assessment of Mild Cognitive Impairment in Advanced Age With a Novel Multi-Feature Automated Speech and Language Analysis of Verbal Fluency', *Frontiers in Psychology*. doi: 10.3389/fpsyg.2020.00535.

Cherney, L. (2010) 'Oral Reading for Language in Aphasia (ORLA): Evaluating the efficacy of computer-delivered therapy in chronic nonfluent aphasia', *Topics in Stroke Rehabilitation*, 17(6), pp. 423–431. doi: 10.1310/tsr1706-423.

Cherney, L. R. (2012) 'Aphasia treatment: Intensity, dose parameters, and script training', *International Journal of Speech-Language Pathology*. doi: 10.3109/17549507.2012.686629.

Cherney, L. R. *et al.* (2021) 'Web-based Oral Reading for Language in Aphasia (Web ORLA®): A pilot randomized control trial', *Clinical Rehabilitation*. doi: 10.1177/0269215520988475.

Clarke, D. J. *et al.* (2018) 'Why do stroke survivors not receive recommended amounts of active therapy? Findings from the ReAcT study, a mixed-methods case-study evaluation in eight stroke units', *Clinical Rehabilitation*. doi: 10.1177/0269215518765329.

Cohen, L. *et al.* (2004) 'The pathophysiology of letter-by-letter reading', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2004.04.018.

Cohen, L. *et al.* (2008) 'Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways', *NeuroImage*. doi: 10.1016/j.neuroimage.2007.11.036.

Coltheart, M. *et al.* (1993) 'Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches', *Psychological Review*. doi: 10.1037/0033-295X.100.4.589.

Coltheart, M. (1998) 'SEVEN QUESTIONS ABOUT PURE ALEXIA (LETTER-BY-LETTER READING)', *Cognitive Neuropsychology*. doi: 10.1080/026432998381203.

Coltheart, M. *et al.* (2001) 'DRC: A dual route cascaded model of visual word recognition and reading aloud', *Psychological Review*. doi: 10.1037/0033-295X.108.1.204.

Coltheart, M. and Rastle, K. (1994) 'Serial Processing in Reading Aloud: Evidence for Dual-Route Models of Reading', *Journal of Experimental Psychology: Human Perception and Performance*. doi: 10.1037/0096-1523.20.6.1197.

Conroy, P. *et al.* (2018) 'Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia', *Brain*. doi: 10.1093/brain/awy087.

Coppens, P. and Patterson, J. P. (eds) (2018) *Aphasia Rehabilitation : Clinical Challenges*. Jones & Bartlett Publishing.

Cornelissen, P. L. *et al.* (2009) 'Activation of the left inferior frontal gyrus in the first 200 ms of reading: Evidence from Magnetoencephalography (MEG)', *PLoS ONE*. doi: 10.1371/journal.pone.0005359.

Coyle-Gilchrist, I. T. S. *et al.* (2016) 'Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes', *Neurology*. doi: 10.1212/WNL.0000000000002638.

Crisp, J. and Lambon Ralph, M. A. (2006) 'Unlocking the nature of the phonological-deep dyslexia continuum: The keys to reading aloud are in phonology and semantics', *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn.2006.18.3.348.

Cuperus, P. *et al.* (2023) 'Understanding User Needs for Digital Aphasia Therapy: Experiences and Preferences of Speech and Language Therapists', *Aphasiology*. doi: 10.1080/02687038.2022.2066622.

Deci, E. L. and Ryan, R. M. (1985) *Intrinsic Motivation and Self-Determination in Human Behavior*, *Intrinsic Motivation and Self-Determination in Human Behavior*. doi: 10.1007/978-1-4899-2271-7.

Dehaene, S. *et al.* (2005) 'The neural code for written words: A proposal', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2005.05.004.

Dejerine, M. (1892) 'Contribution à l'étude anatomopathologique et clinique des différents variétés de cécité verbale', *Mémoires de la Société de Biologie*, 9(IV), pp. 61–90.

Deterding, S. (2013) 'Skill Atoms as Design Lenses for User-Centred Gameful Design', in *CHI 2013 Workshop 'Designing Gamification'*.

Dignam, J. *et al.* (2015) 'Intensive Versus Distributed Aphasia Therapy: A Nonrandomized, Parallel-Group, Dosage-Controlled Study', *Stroke; a journal of cerebral circulation*. doi: 10.1161/STROKEAHA.115.009522.

Doogan, C. *et al.* (2018) 'Aphasia Recovery : When , How and Who to Treat ?' Current Neurology and Neuroscience Reports, pp. 1–7. doi: 10.1007/s11910-018-0891-x.

Drucker, H. *et al.* (1997) 'Support vector regression machines', in *Advances in Neural Information Processing Systems*.

Dunn, Lloyd M. *et al.* (1997) *British Picture Vocabulary Scale (BPVS-II) 2nd ed.*, *NFER-Nelson Publishing Company*.

Feigin, V. L., Norrving, B. and Mensah, G. A. (2017) 'Global Burden of Stroke', *Circulation Research*. doi: 10.1161/CIRCRESAHA.116.308413.

Fleming, V. *et al.* (2021) 'Efficacy of spoken word comprehension therapy in patients with chronic aphasia: A cross-over randomised controlled trial with structural imaging', *Journal of Neurology, Neurosurgery and Psychiatry*. doi: 10.1136/jnnp-2020-324256.

Foulkes, J. (2019) *Is there a place for emojis in aphasia intervention?* Available at: https://pbs.twimg.com/media/EAVzhviWsAE8ysz?format=jpg&name=4096x4096.

Fridriksson, J. and Hillis, A. E. (2021) 'Current approaches to the treatment of post-stroke aphasia', *Journal of Stroke*. doi: 10.5853/jos.2020.05015.

Friston, K. (2005) 'A theory of cortical responses', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2005.1622.

Gerling, K. M., Schild, J. and Masuch, M. (2010) 'Exergame design for elderly users: The

case study of SilverBalance', in *ACM International Conference Proceeding Series*. doi: 10.1145/1971630.1971650.

Glezer, L. S., Jiang, X. and Riesenhuber, M. (2009) 'Evidence for Highly Selective Neuronal Tuning to Whole Words in the "Visual Word Form Area"', *Neuron*. doi: 10.1016/j.neuron.2009.03.017.

Gorno-Tempini, M. L. *et al.* (2004) 'Cognition and Amatomy in Three Variants of Primany Progressive Aphasia', *Annals of Neurology*. doi: 10.1002/ana.10825.

Gorno-Tempini, M. L. *et al.* (2011) 'Classification of primary progressive aphasia and its variants', *Neurology*. doi: 10.1212/WNL.0b013e31821103e6.

Hallams, S. and Baker, K. (2009) 'The development of a questionnaire to assess motivation in strokesurvivors: a pilot study', *New Zealand Journal of Physiotherapy*.

Harm, M. W. and Seidenberg, M. S. (1999) 'Phonology, reading acquisition, and dyslexia: Insights from connectionist models', *Psychological Review*. doi: 10.1037/0033-295X.106.3.491.

Harm, M. W. and Seidenberg, M. S. (2004) 'Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes', *Psychological Review*. doi: 10.1037/0033-295X.111.3.662.

Harris, J. M. *et al.* (2013) 'Classification and pathology of primary progressive aphasia', *Neurology*. doi: 10.1212/01.wnl.0000436070.28137.7b.

Harrison, M., Palmer, R. and Cooper, C. (2020) 'Factors Associated With Adherence to Self-Managed Aphasia Therapy Practice on a Computer—A Mixed Methods Study Alongside a Randomized Controlled Trial', *Frontiers in Neurology*. doi: 10.3389/fneur.2020.582328.

Harvey, R. L. (2015) 'Predictors of Functional Outcome Following Stroke', *Physical Medicine and Rehabilitation Clinics of North America*. doi: 10.1016/j.pmr.2015.07.002.

Hersh, D. *et al.* (2012) 'An analysis of the "goal" in aphasia rehabilitation', *Aphasiology*. doi: 10.1080/02687038.2012.684339.

Hickok, G. and Poeppel, D. (2007) 'The cortical organization of speech processing', *Nature Reviews Neuroscience*. doi: 10.1038/nrn2113.

Hilari, K. (2011) 'The impact of stroke: Are people with aphasia different to those without?', *Disability and Rehabilitation*. doi: 10.3109/09638288.2010.508829.

Holcomb, P. J. and Anderson, J. E. (1993) 'Cross-Modal semantic priming: A time-course analysis using event-related brain potentials', *Language and Cognitive Processes*. doi: 10.1080/01690969308407583.

Hope, T. M. H., Leff, A. P. and Price, C. J. (2018) 'Predicting language outcomes after stroke: Is structural disconnection a useful predictor?', *NeuroImage: Clinical*. doi: 10.1016/j.nicl.2018.03.037.

Howard, D. *et al.* (2006) 'Re-visiting "semantic facilitation" of word retrieval for people with aphasia: Facilitation yes but semantic no', *Cortex*. doi: 10.1016/S0010-9452(08)70439-8.

Iorga, M. *et al.* (2021) 'Predicting language recovery in post-stroke aphasia using behavior and functional MRI', *Scientific Reports*. doi: 10.1038/s41598-021-88022-z.

Jobard, G., Crivello, F. and Tzourio-Mazoyer, N. (2003) 'Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies', *NeuroImage*. doi: 10.1016/S1053-8119(03)00343-4.

Katz, R. C. (2010) 'Computers in the treatment of chronic aphasia', *Seminars in Speech and Language*. doi: 10.1055/s-0029-1244951.

Katz, R. C. and Wertz, R. T. (1997) 'The Efficacy of Computer-Provided Reading Treatment for Chronic Aphasic Adults', *J Speech Lang Hear Res*. doi: 10.1044/jslhr.4003.493.

Ke, G. *et al.* (2017) 'LightGBM: A highly efficient gradient boosting decision tree', in *Advances in Neural Information Processing Systems*.

Kearns, Á. and Kelly, H. (2023) 'ICT usage in aphasia rehabilitation–beliefs, biases, and influencing factors from the perspectives of speech and language therapists', *Aphasiology*. doi: 10.1080/02687038.2022.2030462.

Kerry, S. J. *et al.* (2019) 'How does iReadMore therapy change the reading network of patients with central alexia?', *The Journal of Neuroscience*. doi: 10.1523/jneurosci.1426-18.2019.

Kim, M. and Russo, S. (2010) 'Multiple Oral Rereading (MOR) Treatment: Who Is It for?', *Contemporary Issues in Communication Science and Disorders*. doi: 10.1044/cicsd_36_s_58.

Koiava, N. *et al.* (2012) 'A "web app" for diagnosing hemianopia', *Journal of Neurology, Neurosurgery and Psychiatry*. doi: 10.1136/jnnp-2012-302270.

Kristinsson, S. *et al.* (2021) 'Individualized response to semantic versus phonological aphasia therapies in stroke', *Brain Communications*. doi: 10.1093/braincomms/fcab174.

Kristinsson, S. *et al.* (2022) 'Predictors of Therapy Response in Chronic Aphasia: Building a Foundation for Personalized Aphasia Therapy', *Journal of Stroke*, 24(2), pp. 189–206. doi: 10.5853/jos.2022.01102.

Kristinsson, S. *et al.* (2023) 'Predicting Outcomes of Language Rehabilitation: Prognostic Factors for Immediate and Long-Term Outcomes After Aphasia Therapy', *Journal of Speech, Language, and Hearing Research*. doi: 10.1044/2022_JSLHR-22-00347.

Kurland, J. *et al.* (2008) 'Neural mechanisms underlying learning following semantic mediation treatment in a case of phonologic alexia', *Brain Imaging and Behavior*. doi: 10.1007/s11682-008-9027-2.

Kurland, J., Wilkins, A. R. and Stokes, P. (2014) 'IPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatment', *Seminars in Speech and Language*. doi: 10.1055/s-0033-1362991.

Kyriazakos, S. *et al.* (2020) 'A Novel Virtual Coaching System Based on Personalized Clinical Pathways for Rehabilitation of Older Adults—Requirements and Implementation Plan of the vCare Project', *Frontiers in Digital Health*. doi: 10.3389/fdgth.2020.546562.

Lam, J. M. C. and Wodchis, W. P. (2010) 'The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in ontario hospital-based long-term care residents', *Medical Care*. doi: 10.1097/MLR.0b013e3181ca2647.

Lambon Ralph, M. A. *et al.* (2010) 'Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors.', *Neuropsychological rehabilitation*, 20(2), pp. 289–305.

Langford, T. *et al.* (2022) 'Design Innovation for Engaging and Accessible Digital Aphasia Therapies: Framework Analysis of the iReadMore App Co-Design Process', *JMIR Neurotechnology*. doi: 10.2196/39855.

Langford, T., Leff, A. and Romano, D. (2021) 'IReadMore: A Reading Therapy App Co-Designed by People with Aphasia and Alexia', in *ASSETS 2021 - 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. doi: 10.1145/3441852.3476518.

Lee, J. B. *et al.* (2018) 'Attention Process Training-3 to improve reading comprehension in mild aphasia: A single-case experimental design study', *Neuropsychological Rehabilitation*. doi: 10.1080/09602011.2018.1477683.

Leech, R. *et al.* (2009) 'Informational factors in identifying environmental sounds in natural auditory scenes', *The Journal of the Acoustical Society of America*. doi: 10.1121/1.3238160.

Leff, A. P. *et al.* (2001) 'The functional anatomy of single-word reading in patients with hemianopic and pure alexia', *Brain*. doi: 10.1093/brain/124.3.510.

Leff, A. P. *et al.* (2006) 'Structural anatomy of pure and hemianopic alexia', *Journal of Neurology, Neurosurgery and Psychiatry*. doi: 10.1136/jnnp.2005.086983.

Leff, A. and Starrfelt, R. (2014) *Alexia: Diagnosis, Treatment and Theory*, *Alexia: Diagnosis, Treatment and Theory*. doi: 10.1007/978-1-4471-5529-4.

Little, R. J. A. and Rubin, D. B. (2014) *Statistical analysis with missing data*, *Statistical Analysis with Missing Data*. doi: 10.1002/9781119013563.

Lott, S. N. *et al.* (2010) 'Overt use of a tactile/kinaesthetic strategy shifts to covert processing in rehabilitation of letter-by-letter reading', *Aphasiology*. doi: 10.1080/02687030903580333.

Magnin, E. *et al.* (2016) 'Primary Progressive Aphasia in the Network of French Alzheimer Plan Memory Centers', *Journal of Alzheimer's Disease*. doi: 10.3233/JAD-160536.

Manly, T. *et al.* (2000) 'Not enough time or not enough attention ? Speed , error and self-maintained control in the Sustained Attention to Response Test (SART)', *Clinical Neuropsychological ASsessment*, 3(December 2013), pp. 167–177.

Marshall *et al.* (2018) 'Primary progressive aphasia: a clinical approach', *Journal of Neurology*. doi: 10.1007/s00415-018-8762-6.

Marshall, J. *et al.* (2013) 'Computer delivery of gesture therapy for people with severe aphasia', *Aphasiology*. doi: 10.1080/02687038.2013.786803.

Marshall, J. *et al.* (2016) 'Evaluating the benefits of aphasia intervention delivered in virtual reality: Results of a quasi-randomised study', *PLoS ONE*. doi: 10.1371/journal.pone.0160381.

Marshall, J. *et al.* (2018) 'Delivering word retrieval therapies for people with aphasia in a virtual communication environment', *Aphasiology*. doi: 10.1080/02687038.2018.1488237.

Matías-Guiu, J. A. *et al.* (2017) 'Reading difficulties in primary progressive aphasia in a regular language-speaking cohort of patients', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2017.05.018.

Mayer, J. F. and Murray, L. L. (2002) 'Approaches to the treatment of alexia in chronic aphasia', *Aphasiology*. doi: 10.1080/02687030143000870.

McClelland, J. L. and Rumelhart, D. E. (1981) 'An interactive activation model of context effects in letter perception: I. An account of basic findings', *Psychological Review*. doi: 10.1037/0033-295X.88.5.375.

Menger, F., Morris, J. and Salis, C. (2016) 'Aphasia in an Internet age: wider perspectives on digital inclusion', *Aphasiology*. doi: 10.1080/02687038.2015.1109050.

Messamer, P., Ramsberger, G. and Atkins, A. (2016) 'BangaSpeak: an example of app design for aphasia clients and SLP users', *Aphasiology*. doi:

10.1080/02687038.2015.1102856.

Mesulam, M. -M (1987) 'Primary progressive aphasia—differentiation from Alzheimer's disease', *Annals of Neurology*. doi: 10.1002/ana.410220414.

Middleton, E. L. and Schwartz, M. F. (2012) 'Errorless learning in cognitive rehabilitation: A critical review', *Neuropsychological Rehabilitation*. doi: 10.1080/09602011.2011.639619.

Mortley, J. *et al.* (2004) 'Effectiveness of computerised rehabilitation for long-term aphasia: A case series study', *British Journal of General Practice*.

Moyer, S. B. (1979) 'Rehabilitation of Alexia: A Case Study', *Cortex*. doi: 10.1016/S0010-9452(79)80015-5.

Mumford, D. (1992) 'On the computational architecture of the neocortex - II The role of cortico-cortical loops', *Biological Cybernetics*. doi: 10.1007/BF00198477.

Munsell, M. *et al.* (2020) 'Closing the digital divide in speech, language, and cognitive therapy: Cohort study of the factors associated with technology usage for rehabilitation', *Journal of Medical Internet Research*. doi: 10.2196/16286.

Neate, T. *et al.* (2019) 'Co-created personas: Engaging and empowering users with diverse needs within the design process', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/3290605.3300880.

Northcott, S. and Hilari, K. (2011) 'Why do people lose their friends after a stroke?', *International Journal of Language and Communication Disorders*. doi: 10.1111/j.1460-6984.2011.00079.x.

Nouwens, F. *et al.* (2018) 'Validation of a prediction model for long-term outcome of aphasia after stroke', *BMC Neurology*. doi: 10.1186/s12883-018-1174-5.

Orjada, S. A. and Beeson, P. M. (2005) 'Concurrent treatment for reading and spelling in aphasia', *Aphasiology*. doi: 10.1080/02687030444000796.

Palmer, R. (2015) 'Innovations in aphasia treatment after stroke: technology to the rescue', *British Journal of Neuroscience Nursing*. doi: 10.12968/bjnn.2015.11.sup2.38.

Palmer, R., Witts, H. and Chater, T. (2018) 'What speech and language therapy do community dwelling stroke survivors with aphasia receive in the UK?', *PLoS ONE*. doi: 10.1371/journal.pone.0200096.

Paulhus, D. L. (1984) 'Two-component models of socially desirable responding', *Journal of Personality and Social Psychology*. doi: 10.1037/0022-3514.46.3.598.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*.

Peelle, J. E. *et al.* (2008) 'Sentence comprehension and voxel-based morphometry in progressive nonfluent aphasia, semantic dementia, and nonaphasic frontotemporal dementia', *Journal of Neurolinguistics*. doi: 10.1016/j.jneuroling.2008.01.004.

Perrone-Bertolotti, M. *et al.* (2017) 'Effective connectivity between ventral occipito-temporal and ventral inferior frontal cortex during lexico-semantic processing. A dynamic causal modeling study', *Frontiers in Human Neuroscience*. doi: 10.3389/fnhum.2017.00325.

Petheram, B. (2004) 'Computers and aphasia: A means of delivery and a delivery of means', *Aphasiology*. doi: 10.1080/02687030444000020.

Piguet, O. *et al.* (2015) 'Memory and emotion processing performance contributes to the diagnosis of non-semantic primary progressive aphasia syndromes', *Journal of Alzheimer's Disease*. doi: 10.3233/JAD-141854.

Plaut, D. C. *et al.* (1996) 'Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains', *Psychological Review*. doi: 10.1037/0033-295X.103.1.56.

Plowman, E., Hentz, B. and Ellis, C. (2012) 'Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors', *Journal of Evaluation in Clinical Practice*. doi: 10.1111/j.1365-2753.2011.01650.x.

Price, C. J. (2012) 'A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading', *NeuroImage*. doi: 10.1016/j.neuroimage.2012.04.062.

Price, C. J. and Devlin, J. T. (2011) 'The Interactive Account of ventral occipitotemporal contributions to reading', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2011.04.001.

Price, C. J., Seghier, M. L. and Leff, A. P. (2010) 'Predicting language outcome and recovery after stroke: The PLORAS system', *Nature Reviews Neurology*. doi: 10.1038/nrneurol.2010.15.

Prior, S. *et al.* (2020) 'The challenges of including patients with aphasia in qualitative research for health service redesign: Qualitative interview study', *Journal of Participatory Medicine*. doi: 10.2196/12336.

Purdy, M. *et al.* (2019) 'Reading comprehension treatment in aphasia: a systematic review', *Aphasiology*. doi: 10.1080/02687038.2018.1482405.

Richie, J. and Spencer, L. (1994) 'Qualitative data analysis for applied policy research', in *Analyzing qualitative data*.

Riley, E., Brookshire, C. and Kendall, D. (2017) 'The Acquired Disorders of Reading.', in Papathanasiou, I. and Coppens, P. (eds) *Aphasia and Related Neurogenic Communication Disorders*. Second Edi. Burlington, MA: Jones & Bartlett Learning.

Robson, H. *et al.* (2012) 'Revealing and quantifying the impaired phonological analysis underpinning impaired comprehension in Wernicke's aphasia', *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2011.11.022.

Des Roches, C. A. *et al.* (2015) 'Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform', *Frontiers in Human Neuroscience*. doi: 10.3389/fnhum.2014.01015.

Rogalski, Y. and Edmonds, L. (2008) 'Attentive reading and constrained summarisation (ARCS) treatment in primary progressive aphasia: A case study', in *Aphasiology*. doi: 10.1080/02687030701803796.

Rohrer, J. D. *et al.* (2013) 'Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia', *Brain and Language*. doi: 10.1016/j.bandl.2012.12.008.

Rohrer, J. D. and Warren, J. D. (2016) 'Oxford Textbook of Cognitive Neurology and Dementia', in *Oxford Textbook of Cognitive Neurology and Dementia*, pp. 381–389. doi: 10.1093/med/9780199655946.001.0001.

Roper, A. *et al.* (2018) 'Usability Testing - An Aphasia Perspective', in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*. New York, New York, USA: ACM Press, pp. 102–106. doi: 10.1145/3234695.3241481.

Ryan, R. M. and Deci, E. L. (2000) 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being', *American Psychologist*. doi: 10.1037/0003-066X.55.1.68.

Sajjadi, S. A. *et al.* (2012) 'Primary progressive aphasia: A tale of two syndromes and the rest', *Neurology*. doi: 10.1212/WNL.0b013e3182574f79.

Sarzynski, E. *et al.* (2017) 'Beta Testing a Novel Smartphone Application to Improve Medication Adherence', *Telemedicine and e-Health*. doi: 10.1089/tmj.2016.0100.

Saur, D. *et al.* (2006) 'Dynamics of language reorganization after stroke', *Brain*. doi: 10.1093/brain/awl090.

Saur, D. *et al.* (2008) 'Ventral and dorsal pathways for language', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0805234105.

Scharfen, J., Peters, J. M. and Holling, H. (2018) 'Retest effects in cognitive ability tests: A meta-analysis', *Intelligence*. doi: 10.1016/j.intell.2018.01.003.

Seghier, M. L. *et al.* (2016) 'The PLORAS Database: A data repository for Predicting Language Outcome and Recovery After Stroke', *NeuroImage*. doi: 10.1016/j.neuroimage.2015.03.083.

Seidenberg, M. S. (2005) ‘Connectionist models of word reading’, *Current Directions in Psychological Science*. doi: 10.1111/j.0963-7214.2005.00372.x.

Seidenberg, M. S. and McClelland, J. L. (1989) ‘A Distributed, Developmental Model of Word Recognition and Naming’, *Psychological Review*. doi: 10.1037/0033-295X.96.4.523.

Seminog, O. O. *et al.* (2019) ‘Determinants of the decline in mortality from acute stroke in England: Linked national database study of 795 869 adults’, *The BMJ*. doi: 10.1136/bmj.l1778.

Simmons-Mackie, N. (2012) ‘Social approaches to aphasia intervention’, in *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders: Fifth Edition*.

Stark, B. C. and Warburton, E. A. (2018) ‘Improved language in chronic aphasia after self-delivered iPad speech therapy’, *Neuropsychological Rehabilitation*. doi: 10.1080/09602011.2016.1146150.

Starrfelt, Randi *et al.* (2013) ‘How low can you go: spatial frequency sensitivity in a patient with pure alexia.’, *Brain and language*, 126(2), pp. 188–92. doi: 10.1016/j.bandl.2013.05.006.

Starrfelt, R. *et al.* (2013) ‘Word-superiority in pure alexia’, in *Behavioural Neurology*. doi: 10.3233/BEN-2012-129002.

Stroke Association (2016) *A New Era for Stroke*. Available at: https://www.stroke.org.uk/get-involved/campaigning/new-era-for-stroke.

Stroke Association (2018) *State of the Nation: stroke statistics 2018*. Available at: https://www.stroke.org.uk/resources/state-nation-stroke-statistics.

Sutton, R. T. *et al.* (2020) ‘An overview of clinical decision support systems: benefits, risks, and strategies for success’, *npj Digital Medicine*. doi: 10.1038/s41746-020-0221-y.

Swinburn, K., Howard, D. and Porter, G. (2004) ‘CAT : comprehensive aphasia test’.

Taubner, H., Hallén, M. and Wengelin, Å. (2020) ‘Still the same?–Self-identity dilemmas when living with post-stroke aphasia in a digitalised society’, *Aphasiology*. doi: 10.1080/02687038.2019.1594151.

Taylor, J. S. H., Rastle, K. and Davis, M. H. (2013) ‘Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies’, *Psychological Bulletin*. doi: 10.1037/a0030266.

Taylor, J. S. H., Rastle, K. and Davis, M. H. (2014) ‘Distinct Neural Specializations for Learning to Read Words and Name Objects’, *Journal of Cognitive Neuroscience*. doi:

10.1162/jocn_a_00614.

Tishbirani, R. (1996) ‘Regression shrinkage and selection via the Lasso’, *Journal of the Royal Statistical Society. Series B (Methodological)*.

Vansteenkiste, M., Niemiec, C. P. and Soenens, B. (2010) ‘The development of the five mini-theories of self-determination theory: An historical overview, emerging trends, and future directions’, *Advances in Motivation and Achievement*. doi: 10.1108/S0749-7423(2010)000016A007.

Vickers, C. P. (2010) ‘Social networks after the onset of aphasia: The impact of aphasia group attendance’, *Aphasiology*. doi: 10.1080/02687030903438532.

Visch-Brink, E. G., Stronks, D., & Denes, G. (2005) *De Semantische Associatie Test [Semantic Association Test]*. Lisse, the Netherlands: Swets & Zeitlinger.

Visser, M. *et al.* (2012) ‘Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: Distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes’, *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn_a_00244.

Volkmer, A. *et al.* (2018) ‘Speech and language therapy for primary progressive aphasia: Referral patterns and barriers to service provision across the UK’, *Dementia*. doi: 10.1177/1471301218797240.

Volkmer, A. *et al.* (2019) ‘Speech and language therapy approaches to managing primary progressive aphasia’, *Practical Neurology*. doi: 10.1136/practneurol-2018-001921.

Wafa, H. A. *et al.* (2020) ‘Burden of Stroke in Europe: Thirty-Year Projections of Incidence, Prevalence, Deaths, and Disability-Adjusted Life Years’, *Stroke*, 51(8), pp. 2418–2427. doi: 10.1161/STROKEAHA.120.029606.

Wang, J. *et al.* (2013) ‘Predicting speech fluency and naming abilities in aphasic patients’, *Frontiers in Human Neuroscience*. doi: 10.3389/fnhum.2013.00831.

Watila, M. M. and Balarabe, B. (2015) ‘Factors predicting post-stroke aphasia recovery’, *Journal of the Neurological Sciences*. doi: 10.1016/j.jns.2015.03.020.

Webster, J., Morris, J. and Howard, D. (2023) ‘Reading comprehension in aphasia: the relationship between linguistic performance, personal perspective, and preferences’, *Aphasiology*. doi: 10.1080/02687038.2022.2039999.

Wheat, K. L. *et al.* (2010) ‘During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: Evidence from magnetoencephalography’, *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.4448-09.2010.

Wilson, C. and Kim, E. S. (2019) 'Qualitative data collection: considerations for people with Aphasia', *Aphasiology*. doi: 10.1080/02687038.2019.1693027.

Wilson, S. *et al.* (2015) 'Codesign for people with aphasia through tangible design languages', *CoDesign*. doi: 10.1080/15710882.2014.997744.

Wilson, S. M. *et al.* (2009) 'The neural basis of surface dyslexia in semantic dementia', *Brain*. doi: 10.1093/brain/awn300.

Wilson, S. M. *et al.* (2010) 'Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.2547-10.2010.

Woodhead, Z. V. J. *et al.* (2013) 'Reading therapy strengthens top-down connectivity in patients with pure alexia', *Brain*. doi: 10.1093/brain/awt186.

Woodhead, Z. V. J. *et al.* (2018) 'Randomized trial of iReadMore word reading training and brain stimulation in central alexia', *Brain*. doi: 10.1093/brain/awy138.

Woollams, A. M. *et al.* (2007) 'SD-squared: On the association between semantic dementia and surface dyslexia', *Psychological Review*. doi: 10.1037/0033-295X.114.2.316.

Zhong, X. (2024) 'AI-assisted assessment and treatment of aphasia: a review', *Frontiers in Public Health*, 12(August), pp. 1–11. doi: 10.3389/fpubh.2024.1401240.

Zhou, W. and Shu, H. (2017) 'A meta-analysis of functional magnetic resonance imaging studies of eye movements and visual word reading', *Brain and Behavior*. doi: 10.1002/brb3.683.

Zou, H. and Hastie, T. (2005) 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. doi: 10.1111/j.1467-9868.2005.00503.x.

8 Appendices

Appendix 1 - Visual Fields Test scores

User	T1	T2.5	H2.5	I2.5	T5	T10	S1	S2.5	G2.5	F2.5	S5	S10
2	2	2	2	4	2	1	2	2	4	4	2	2
3	0	2	4	4	2	2	2	2	4	4	2	2
6	2	2	4	4	2	2	2	2	4	4	2	2
7	2	2	4	4	2	2	2	2	4	4	2	2
9	2	2	3	4	2	0	2	2	4	4	2	2
13	2	2	2	3	2	0	1	2	3	3	2	2
16	2	2	4	4	2	2	2	2	3	4	2	2
17	2	2	4	4	2	2	2	2	4	4	2	2
21	2	2	4	4	2	2	2	2	4	4	2	2
28	1	0	2	2	1	1	1	1	3	2	1	2
30	1	2	3	2	1	2	1	1	4	3	2	2
31	2	1	1	1	1	2	2	2	4	4	2	2
32	1	2	3	1	0	1	2	2	4	4	2	2
34	2	1	3	3	0	2	1	2	4	4	2	2

Appendix 2 - Change in Word Reading Accuracy on the Word Reading Test for all users and timepoints

	User	2	3	6	7	9	13	16	17	21	28	30	31	32	34
Baseline	**Trained**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Untrained	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interval 1	**Trained**	0	-0.04	-0.06	0	0.02	0.12	0.04	0.12	-0.34	-0.06	-0.02	0.02	-0.1	-0.04
	Untrained	-0.04	0.02	0	-0.02	0	-0.04	0.1	-0.04	-0.36	0.02	0.02	-0.04	-0.08	-0.08
Interval 2	**Trained**	0	-0.1	0.02	0.26	0.02	0.28	0.16	0.16	-0.26	-0.06	-0.04	0.04	-0.08	-0.1
	Untrained	0.02	0	0.02	0.02	0	0.12	0.12	-0.08	-0.46	0.12	0.02	0.04	-0.08	-0.02
Interval 3	**Trained**	0	-0.02	-0.06	0.22	0.02	0.46	0.12	0.16	-0.26	-0.08	-0.06	0	-0.1	-0.12
	Untrained	-0.02	-0.04	0.02	0.1	0.04	0.2	0.26	-0.06	-0.56	-0.04	0.06	0.04	-0.06	-0.12
Interval 4	**Trained**	0.06	0.04	-0.0212	0.18	0	0.44	0.28	0.2	-0.12	0.02	-0.04	0.06	-0.04	-0.1
	Untrained	-0.06	0	0.02	0	0.06	0.28	0.26	0.1	-0.36	0.08	0.06	0.08	-0.08	-0.02
Interval 5	**Trained**	0.06			0.18	-0.02	0.38	0.14	0.14	-0.1	0	-0.02		-0.08	-0.08
	Untrained	-0.02			0.08	0.04	0.22	0.12	0.04	-0.22	0.08	0.08		0	-0.06
Interval 6	**Trained**	0.04			0.22		0.28		0.24	-0.08					-0.08
	Untrained	-0.02			0.1		0.12		0.08	-0.14					-0.08
Interval 7	**Trained**	0.06			0.14		0.38		0.14						-0.12
	Untrained	0.02			0.06		0.12		-0.02						-0.1
Interval 8	**Trained**				0.22		0.46		0.22						-0.1

	Untrained				0.12		0.2		-0.04					-0.04
Interval 9	**Trained**				0.18		0.46							-0.04
	Untrained				0.14		0.26							-0.06
Interval 10	**Trained**				0.24		0.4							-0.04
	Untrained				0.12		0.2							-0.14
Interval 11	**Trained**				0.12		0.48							-0.08
	Untrained				0.1		0.24							-0.04
Interval 12	**Trained**				0.14		0.28							-0.04
	Untrained				0.14		0.2							-0.02
Interval 13	**Trained**				0.22									0.02
	Untrained				0.12									-0.08
Interval 14	**Trained**				0.18									-0.04
	Untrained				0.06									-0.04
Interval 15	**Trained**				0.24									-0.08
	Untrained				0.1									-0.04

Appendix 3 - Change in Reaction Time on the Word Reading Test for all users and timepoints

Participant		2	3	6	7	9	13	16	17	21	28	30	31	32	34
Baseline	**Trained**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Untrained	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interval 1	**Trained**	-1.2	-0.6	-1.0	-1.1	-1.2	-1.2	-0.8	0.9	-3.4	-1.0	-0.1	-0.7	-2.2	0.2
	Untrained	-1.1	-0.2	-0.4	-2.1	-1.2	-1.0	0.6	0.4	-2.7	1.1	-0.1	-0.4	0.1	0.4
Interval 2	**Trained**	-0.9	-0.8	-1.0	-0.8	-0.6	1.2	-0.7	-1.3	-1.7	-2.4	-0.5	-1.3	-2.8	0.8
	Untrained	-0.6	-0.7	-0.4	-1.5	-1.2	0.6	0.3	-1.4	-1.8	1.2	0.0	-0.8	0.0	1.2
Interval 3	**Trained**	-1.2	-1.2	-1.1	-1.5	-1.4	0.0	-1.4	-1.5	-5.1	-1.3	-0.8	-1.8	-2.2	0.4
	Untrained	-1.6	-1.4	-0.5	-2.1	-1.6	0.6	0.0	-1.0	-4.8	2.3	-0.3	-1.5	-0.2	0.9
Interval 4	**Trained**	-1.6	-1.3	-1.1	-0.6	-1.8	-1.6	-0.9	-1.7	-1.1	-2.6	-1.0	-1.9	-1.4	0.1
	Untrained	-1.6	-1.2	-0.4	-1.2	-1.9	-0.7	0.8	-1.4	-0.3	-0.7	-0.4	-1.7	0.5	0.4
Interval 5	**Trained**	-1.8			-0.6	-1.7	-1.1	3.0	-1.7	-2.7	-1.7	-0.8		-1.4	-0.1
	Untrained	-1.4			-1.2	-1.6	0.0	4.3	-2.6	-2.9	1.6	-0.6		-0.1	-0.1
Interval 6	**Trained**	-2.2			-1.0		-2.8		-1.6	-1.9					-0.1
	Untrained	-1.5			-1.1		-2.8		-2.4	-1.6					-0.2
Interval 7	**Trained**	-1.9			-0.4		-3.8		-1.5						-0.6
	Untrained	-1.8			-1.6		-3.9		-2.5						-0.2
Interval 8	**Trained**				-1.1		-2.6		-2.2						-0.1
	Untrained				-1.6		-0.8		-2.3						-0.1
Interval 9	**Trained**				-1.3		-3.2								-0.1
	Untrained				-1.8		-2.5								-0.3
Interval 10	**Trained**				-0.7		-2.9								-0.8
	Untrained				-1.0		-3.1								-0.3
Interval 11	**Trained**				-1.1		-3.7								-0.6
	Untrained				-1.9		-3.0								-0.6

Interval 12	**Trained**				-1.9		-3.4								-0.8
	Untrained				-2.2		-3.2								-0.9
Interval 13	**Trained**				-1.7										-0.7
	Untrained				-2.2										-0.5
Interval 14	**Trained**				0.2										-1.0
	Untrained				-0.8										-0.8
Interval 15	**Trained**				-1.5										-1.0
	Untrained				-2.1										-0.4